DISSERTATION

ORTHOGONAL PAIR-DIRECTED CODON REASSIGNMENT AS A TOOL FOR EVALUATING THE FACTORS AFFECTING TRANSLATION IN *E. COLI*

Submitted by

David Schwark

Graduate Degree Program in Bioengineering

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2018

Doctoral Committee:

Advisor: John (Nick) Fisk

Chris Ackerson
Alan Kennan
Christie Peebles
Chris Snow

ABSTRACT

ORTHOGONAL PAIR-DIRECTED CODON REASSIGNMENT AS A TOOL FOR EVALUATING THE FACTORS AFFECTING TRANSLATION IN *E. COLI*

Proteins are polymers of amino acids that are essential for life, central to cellular function, and have applications in fields ranging from materials science to biomedicine. Proteins in nature are composed of 20 amino acids with limited variability in size and chemical properties. Expanding the genetic code to contain non-canonical amino acids (ncAAs) that contain functionalities not contained in nature is a powerful strategy for probing and extending the properties of proteins. Current *in vivo* systems for expanding the genetic code have focused on using an engineered orthogonal aminoacyl-tRNA and aminoacyl tRNA-synthetase pairs (tRNA/aaRS) to direct incorporation of ncAAs at amber stop codons. In order to further expand the genetic code to 22 or more amino acids, additional codons must be targeted for reassignment to ncAAs. The genetic code is degenerate; 18 of the 20 canonical amino acids are encoded by more than one codon. Breaking the degeneracy of the genetic code by orthogonal pair directed sense codon reassignment is one pathway to genetic codes of 22 or more amino acids.

However, orthogonal pair directed sense codon reassignment is hampered by a limited understanding of the relative importance of the factors that affect the translation of proteins. Here, we describe the repurposing of two commonly used orthogonal pairs from *Methanocaldococcus jannaschii* (*M. jannaschiil)* and *Methanosarcina barkeri* (*M.*

barkeri) to measure the *in vivo* reassignment efficiency of 30 different sense codons to tyrosine in *E. coli* with a simple fluorescence-based screen.

The suite of sense codon reassignment efficiencies identified multiple promising codons for reassignment to ncAAs that have not been previously identified. Importantly, every sense codon was partially reassigned to tyrosine when either orthogonal tRNA/aaRS pair was used. Sense codons reassigned to tyrosine with high efficiency may be used directly for reassignment to ncAAs, and any sense codon with measurable reassignment to tyrosine may be improved through directed evolution. The sets of *in vivo* sense codon reassignment also revealed that *E. coli* are broadly tolerable to a large number of amino acid substitutions to tyrosine throughout the proteome.

The codon reassignment efficiency measurements also enabled an analysis of the *in vivo* importance of local codon context effects, tRNA abundance, aminoacylation level, tRNA modifications, and codon-anticodon binding energy in determining translational fidelity. Quantitative sense codon reassignment efficiency measurements showed that the process of translation is highly balanced and both tRNA abundance and aminoacylation efficiency do not appear to be dominant factors in determining translational fidelity. Furthermore, quantitative measurements of amber stop codon reassignment efficiencies to tyrosine with the orthogonal *M. jannaschii* pair revealed that local codon context is an important factor for orthogonal pair directed amber stop codon reassignment.

ACKNOWLEDGEMENTS

I have a lot of people to acknowledge for helping me throughout my graduate career. First, I would like to thank my advisor, Dr. Nick Fisk. I first met Nick when he was the instructor for my molecular concepts and applications course during my undergraduate studies in chemical and biological engineering here at CSU. Nick's passion for science and engineering and his dedication to teaching were evident, and make him an ideal advisor. Nick gave me my first opportunity to pursue research as an undergraduate assistant in his lab, and I am grateful for that opportunity and to continue on as a graduate student in his lab. I cannot thank Nick enough for his patience, guidance, support, and encouragement during my time pursuing a Ph.D. in his lab. Science and engineering are difficult – and experiments fail; Nick taught me how to think critically about experimental failures, and how to overcome them.

I would also like to thank Dr. Meg Schmitt. Meg is a staff scientist in Nick's lab, was an unofficial co-advisor for me, and has become a great friend. She was always willing to discuss my projects with me, how they fit into a bigger scientific picture, and how to troubleshoot any issues I had in the lab. Meg always supported and encouraged me, and specifically helped me become a better writer and speaker. I cannot thank her enough for all of her help throughout my time in the Fisk lab.

I would like to thank my committee members, Dr. Alan Kennan, Dr. Chris Snow, Dr. Christie Peebles, and Dr. Chris Ackerson. I really appreciate them for dedicating their time to be a part of my committee. I would also like to thank my fellow graduate students in the Fisk lab, including Steven Smeal, Dr. Wil Biddle, Dr. Ning Zhao, Zach

Butz and Jenn Bjerke. In addition to discussing and working on projects in the lab with me, you have become my friends.

I would also like to thank my family. They have supported me from day 1, when I decided I wanted to pursue a career in science and engineering. Without your love and support, I would not be where I am today.

Finally, I want to thank my wife. Trish and I met back in the fall of 2009, before I started as an undergraduate assistant in Nick's lab. She has stuck with me through the ups and downs of graduate school, through almost three years of a long distance relationship, and through many trying times. Thank you for always supporting and loving me, no matter how many late nights or weekends I worked.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 OVERVIEW

This dissertation contains 6 chapters that describe the use of orthogonal translation systems, directed evolution, and quantitative *in vivo* fluorescence measurements to investigate the plasticity of the *Escherichia coli* (*E. coli*) genetic code and address questions about relative quantitative importance of multiple factors contributing to protein translation. The ultimate goals of this work are a deeper understanding of the factors controlling the fidelity of translation and the development of tools to allow biosynthetic incorporation of 22 or more amino acids by *E. coli* translational machinery.

The first chapter provides background on the protein translation machinery as a framework for understanding the work presented in the remainder of the dissertation. The introduction includes a brief description of the relevance of proteins in biomedical engineering applications, an overview of the central dogma of molecular biology, current limitations in the understanding of *in vivo* protein synthesis, a description of both *in vivo* and *in vitro* methods used for incorporating non-canonical amino acids (ncAAs) into proteins, and the fluorescence based screen developed in the Fisk lab to measure the *in vivo* incorporation of amino acids in response sense codons using an engineered orthogonal tRNA / aminoacyl-tRNA synthetase (aaRS) pairs.

Chapter 2 describes the measurement of sense codon reassignment efficiencies for rarely used sense codons in *E. coli* using the orthogonal *Methanocaldococcus jannaschii* (*M. jannaschii*) tRNA/aaRS pair. The quantitative measurements of sense codon reassignment efficiencies at the 10 additional rare codons augments the previous 20 measurements at *E. coli* wobble codons. Significantly, every codon was partially reassigned to tyrosine with efficiencies ranging from 3.5% to 70%. Furthermore, in general sense codon reassignment efficiencies were higher for rarely used codons than reassignment efficiencies at wobble codons. In addition, substituting the encoded amino acid for tyrosine at codons throughout the proteome was well-tolerated; only two reassignment systems, Ile AUA and Leu CUA, had doubling times increased more than 2 standard deviations above a non-reassigning wild-type GFP expressing control despite estimated 100,000s to 1,000,000s of substitutions. Analysis of wobble and rare codon data sets together show a weak correlation between sense codon reassignment efficiency to tyrosine and either competing endogenous *E. coli* tRNA abundance or predicted-fold reduction in aminoacylation efficiency. In contrast to the previous data analysis of the wobble codon only data set, analysis of the wobble codon and rare codon data sets together suggests that endogenous *E. coli* tRNA abundance may be more influential than aminoacylation efficiency on sense codon reassignment. Additional sense codon reassignment efficiency data from another orthogonal pair may further assist in determining the relative importance of factors involved in protein translation.

Chapter 3 describes the engineering of a *Methanosarcina barkeri* (*M. barkeri*) pyrrolysyl aaRS to aminoacylate tyrosine onto its corresponding tRNA, making it compatible with the previously developed GFP screen for sense codon reassignment.

The *M. barkeri* and *M. jannaschii* orthogonal pair systems are the most commonly used systems for ncAA incorporation. The amber stop codon suppression efficiency of the engineered *M. barkeri* tRNA/aaRS pair was 8.2% of wild-type GFP. The substrate specificity of the engineered pair was characterized by electrospray ionization (ESI) mass spectrometry of a target protein produced by suppressing an amber stop codon with the engineered *M. barkeri* pair. The ESI-MS data suggested that a small amount of phenylalanine was also aminoacylated by the engineered *M. barkeri* aaRS. The engineered pair was also used to measure sense codon reassignment at eight different codons with the GFP screen. Significantly, all sense codons showed measurable levels of sense codon reassignment, although the dynamic range of the GFP screen for reassignment efficiency was reduced.

Chapter 4 describes the improvement of the tyrosine-incorporating pyrrolysyl *M. barkeri* tRNA/aaRS pair through error-prone PCR and fluorescence-activated cell sorting (FACS). After 2 rounds of library generation and screening, a tyrosine-incorporating Pyl *M. barkeri* tRNA/aaRS pair capable of suppressing an amber stop codon at position 66 in GFP with 92.5% efficiency of wild-type GFP in *E. coli* DH10B was identified. When the identified *M. barkeri* pair was used in *E. coli* SB3930, the strain used for all sense codon reassignment efficiency experiments with the orthogonal *M. jannaschii* pair, 65.2% efficiency of wild-type GFP was observed. The difference in amber stop codon suppression efficiencies with different *E. coli* strains may be due to differences in expression of orthogonal translation machinery or the reporter protein, or differences in endogenous translational machinery components of the strains. The starting tyrosine-incorporating Pyl *M. barkeri* tRNA/aaRS pair's amber stop codon

suppression efficiency was 12.6% in *E. coli* SB3930. The improvements of the aaRS resulted in a 5.2-fold increase in amber stop codon suppression. In addition, the substrate specificity of the engineered tRNA/aaRS pair was evaluated through ESI-mass spectrometry of the same target protein previously used, and showed that the substrate specificity of the aaRS was also improved. The improved orthogonal tyrosine-incorporating *M. barkeri* pair was then used to measure the sense codon reassignment efficiencies at the same 8 sense codons tested with the original engineered tyrosine-incorporating *M. barkeri* pair. Reassignment efficiencies to tyrosine at the 8 sense codons ranged from 0.8% to 65.0% of wild-type GFP, corresponding to 3.1-10.6 fold increases over the sense codon reassignment efficiencies measured with the starting orthogonal *M. barkeri* pair. Furthermore, all sense codon reassignment efficiencies measured were above the limit of quantification.

Chapter 5 describes the measurement of the full suite of sense 30 sense codons using the improved orthogonal tyrosine-incorporating *M. barkeri* tRNA/aaRS and a comparison of those measurements to the sense codon reassignment efficiencies with the orthogonal *M. jannaschii* pair. Significantly, every sense codon measured had quantifiable reassignment to tyrosine with the orthogonal *M. barkeri* pair, ranging from 0.9% to 70.3%. Furthermore, sense codon reassignment efficiencies with the orthogonal *M. barkeri* pair correlated strongly with the sense codon reassignment measurements made with the orthogonal *M. jannaschii* pair. The abilities of the orthogonal *M. barkeri* tRNAs to discriminate between the codon targeted for reassignment and closely related codons were very similar to those observed with the orthogonal *M. jannaschii* pair. In addition, substituting the canonically encoded amino

acids for tyrosine by the orthogonal *M. barkeri* pair at codons throughout the proteome was well-tolerated; despite 100,000s to 1,000,000s of estimated amino acid substitutions in the proteome of the host *E. coli* cells during reassignment, the instantaneous doubling times of the cells during the exponential phase were only increased by slightly more than 2 standard deviations of a non-reassigning, wild-type GFP expressing control. Analysis of the sense codon reassignment data sets from both orthogonal pairs together show a weak correlation between sense codon reassignment efficiency to tyrosine and either competing endogenous *E. coli* tRNA abundance or predicted-fold reduction in aminoacylation efficiency when all 30 sense codons are considered. Both competing endogenous tRNA concentration and aminoacylation efficiency do not appear to be dominant factors in orthogonal pair directed sense codon reassignment.

Chapter 6 describes the application of a GFP fluorescence-based screen to quantitatively evaluate the effect of codon context on amber stop codon reassignment tyrosine by the orthogonal *M. jannaschii* pair. Amber stop codon reassignment efficiencies ranged from 51%-116% in *E. coli* DH10B and 75%-103% in a genomically recoded strain of *E. coli* that does not contain ribosomal release factor 1 for eight single amber stop codon (UAG) containing GFP variants. The only difference between each of the reporter GFP variants was the location of the UAG codon, demonstrating that local codon context is an important factor for UAG reassignment with the orthogonal *M. jannaschii* pair. Measurement of suppression efficiencies for 14 multiple UAG-containing GFP variants showed that the codon context effects of individual UAG codon suppression are not strictly independent and additive. Some factors affecting codon

context are non-local and negatively synergistic. The comparison of context effects between a normal cell line and a strain where the competing release factor RF1 for amber stop codon reading has been deleted supports the conclusion that local codon context effects are in part related to the context preferences of the release factors, but clearly involve other contributions. The study presented in this chapter is a first small step towards the application of orthogonal tRNA/aaRS pairs to understanding local codon context effects on codon reassignment.

1.2 PROTEINS AS MOLECULAR TOOLS IN BIOMEDICAL ENGINEERING

Proteins are polymers of amino acids that are essential for life and central to cellular function. Proteins in nature are typically composed of 20 amino acids [1]. The side chains of the 20 genetically encoded amino acids vary in size and chemical properties (Figure 1.1). The well-defined 3-dimenisonal structures of proteins, essential for function, are determined by the protein sequence. Interactions between side chains of the amino acids, particularly the partitioning of hydrophobic side chains in protein interiors, drive the transition from an unstructured sequence to a folded form [1, 2]. An enormous number of proteins exist in nature and are responsible for many different functions in a cell, including but not limited to acting as catalysts in cellular reactions, as regulators in metabolic processes, as structural supports for cells, and as transport vehicles for small molecules [3-6].

Proteins are useful in many biomedical engineering applications, including diagnostics and therapeutics [7-10]. Differences in cellular protein concentrations, modifications, and folded states are hallmarks of a variety of disease states [11]. The ability to identify and distinguish these differences enables the characterization and

diagnose of disease. Proteins also have varied applications in therapeutics, including as replacements for abnormal or deficient proteins or as vehicles for delivering a small molecule drug to a desired location. One of the most common classes of protein

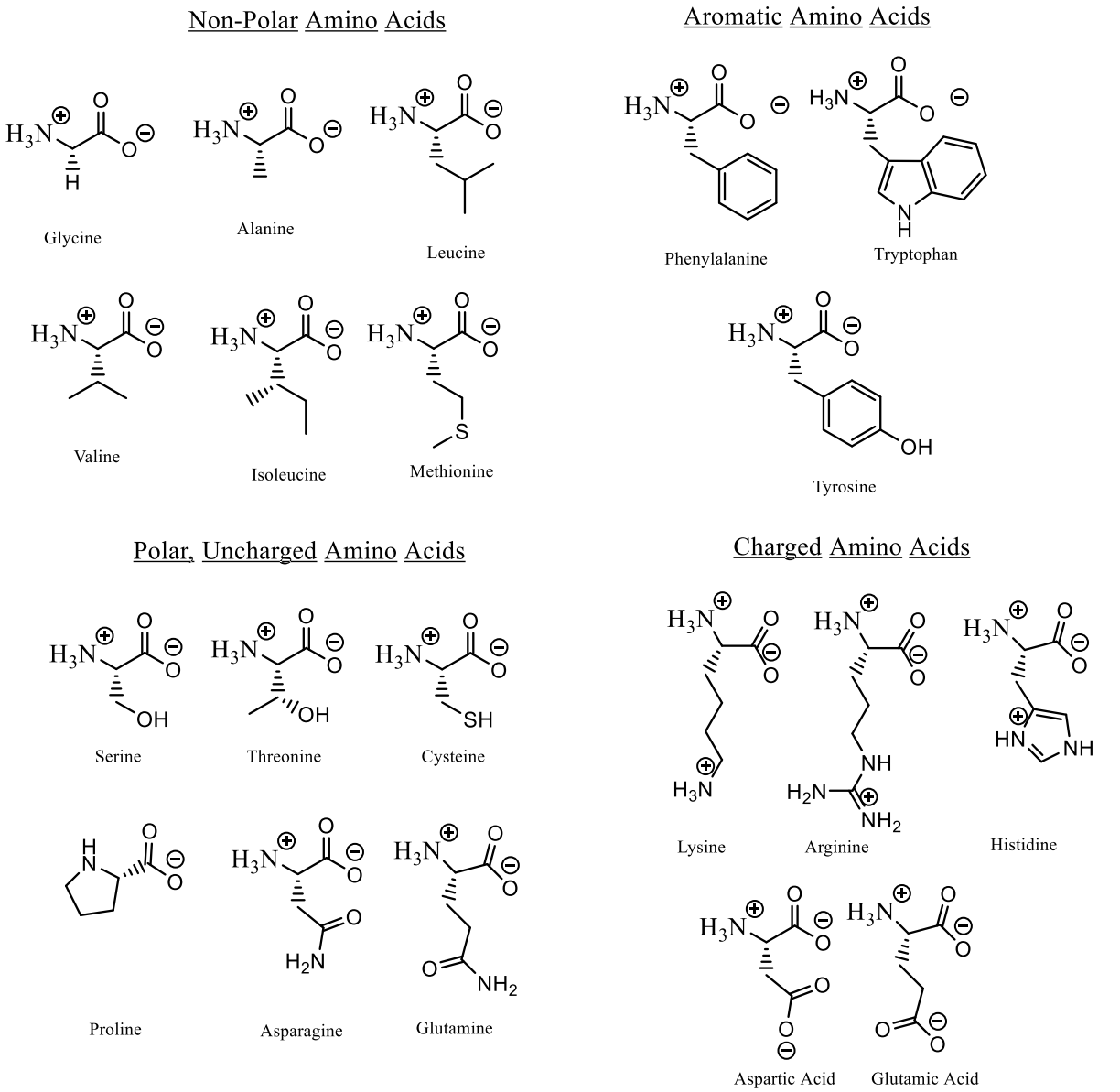

Figure 1.1. The 20 natural amino acids grouped by side chain chemistries. Figure adapted from [1].

therapeutics is based on the protein scaffolds of native immunity: antibodies [12]. The immune system is able to generate tight, specific antibodies for a variety of antigens.

Biomedical science has co-opted the antibody scaffold to engineer antibodies to a large number of additional antigens [12]. Engineered antibodies are becoming increasingly prevalent in the biomedicine industry. In 2013, each of the top six selling monoclonal antibodies on the market reported sales of $6.5 billion or more [13]. In addition, the last seven years has seen the number of monoclonal antibodies in phase III clinical trials increase by 100%, from 26 to 52 [14].

1.3 THE CENTRAL DOGMA OF MOLECULAR BIOLOGY

The central dogma of molecular biology describes the set of steps by which the information in an organism's genome stored as a polymer of nucleic acids is faithfully converted into the polyamide proteins that provide cellular structure and function (Figure 1.2) [15]. Genomic information stored as the more stable, double-stranded nucleic acid form DNA is first transcribed into messenger RNA (mRNA). Single-stranded mRNA serves as the template from which specific proteins are made in the process of translation.

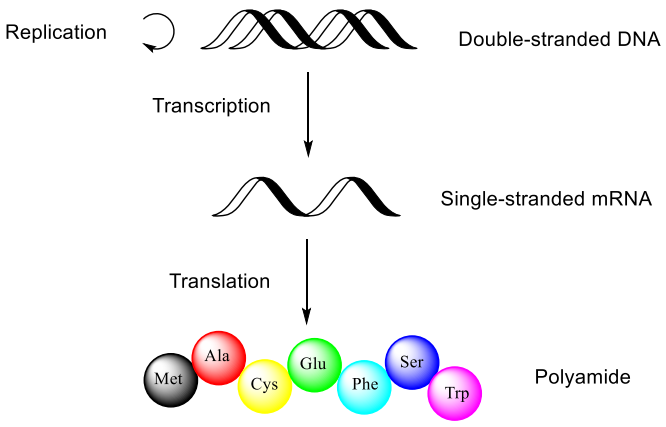

Figure 1.2. The central dogma of molecular biology [15]. DNA is transcribed into mRNA, which is then translated into a polyamide.

DNA is a stable double-stranded molecule composed of a sequence of four different nucleotides: Adenine (A), Guanine (G), Thymine (T), and Cytosine (C) [16]. A nucleotide is composed of a nitrogenous base, deoxyribose, and a triphosphate (Figure 1.2). Nucleotides in a strand of DNA are covalently linked by a phosphodiester bond at the 5' and 3' positions of the deoxyriboses. The two strands of DNA interact through hydrogen bonding and stacking of the nitrogenous bases [16-18]. Watson-Crick base-pairing typically occurs in DNA involving adenine-thymine, and cytosine-guanine pairs (Figure 1.3) [15, 16]. The A-T base pair has two hydrogen bonds and is weaker than the G-C base pair that has three hydrogen bonds [19].

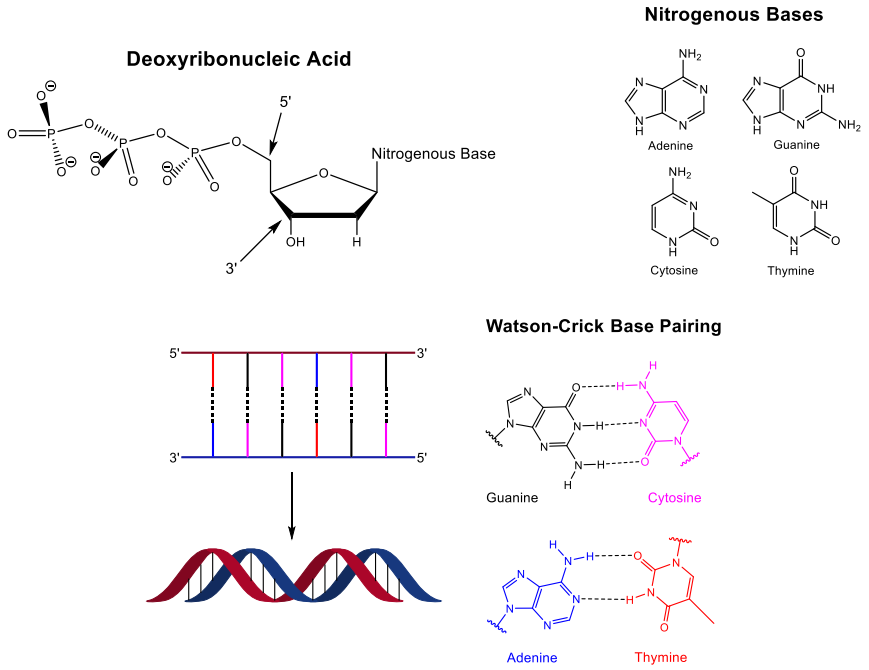

Figure 1.3. Top Left: Structure of a nucleotide [16]. Top right: Structures of the four nitrogenous bases in DNA [1]. Bottom Left: Cartoon representation of double-stranded DNA. The two strands of DNA interact through hydrogen bonding of base-pairs [16]. Bottom right: Watson-Crick base-pairs typically observed in DNA [1].

DNA sequences of nucleotide triplets called codons make up the blueprint for a protein. The ordered combination of the four nucleotides grouped three at a time results in a genetic code with 64 different codons, each of which has an assigned function [20]. These 64 distinct codons encode the 20 canonical amino acids as well as initiation and termination signals of protein translation (Table 1.1) [1]. DNA codons are transcribed into complementary single-stranded mRNA before translation. RNA is also composed of a sequence four nucleotides: adenine, uracil (instead of thymine), cytosine, and guanine.

Table 1.1. The genetic code [20].

	T	C	A	G	
T	TTT Phe TTC Phe TTA Leu TTG Leu	TCT Ser TCC Ser TCA Ser TCG Ser	TAT Tyr TAC Tyr TAA Stop TAG Stop	TGT Cys TGC Cys TGA Stop TGG Trp	T C A G
C	CTT Leu CTC Leu CTA Leu CTG Leu	CCT Pro CCC Pro CCA Pro CCG Pro	CAT His CAC His CAA Gln CAG Gln	CGT Arg CGC Arg CGA Arg CGG Arg	T C A G
A	ATT Ile ATC Ile ATA Ile ATG Met	ACT Thr ACC Thr ACA Thr ACG Thr	AAT Asn AAC Asn AAA Lys AAG Lys	AGT Ser AGC Ser AGA Arg AGG Arg	T C A G
G	GTT Val GTC Val GTA Val GTG Val	GCT Ala GCC Ala GCA Ala GCG Ala	GAT Asp GAC Asp GAA Glu GAG Glu	GGT Gly GGC Gly GGA Gly GGG Gly	T C A G

1.4 TRANSLATION

Translation is the process by which the codons in a mRNA sequence of a gene are converted into a polypeptide of amino acids (Figures 1.4-1.5). The process of translation is catatlyzed by ribososmes and employs amino acyl tRNA molecules as intermediaries. The process of translation initiation involves the small subunit of the

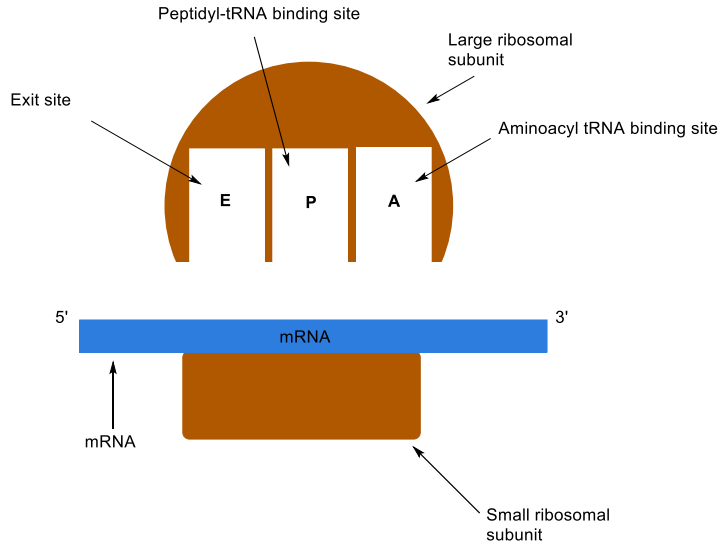

Figure 1.4. Cartoon diagram of a ribosome with the large and small subunits, a strand of mRNA, and the aminoacyl tRNA binding site, peptidyl-tRNA binding site, and exit site [1, 21].

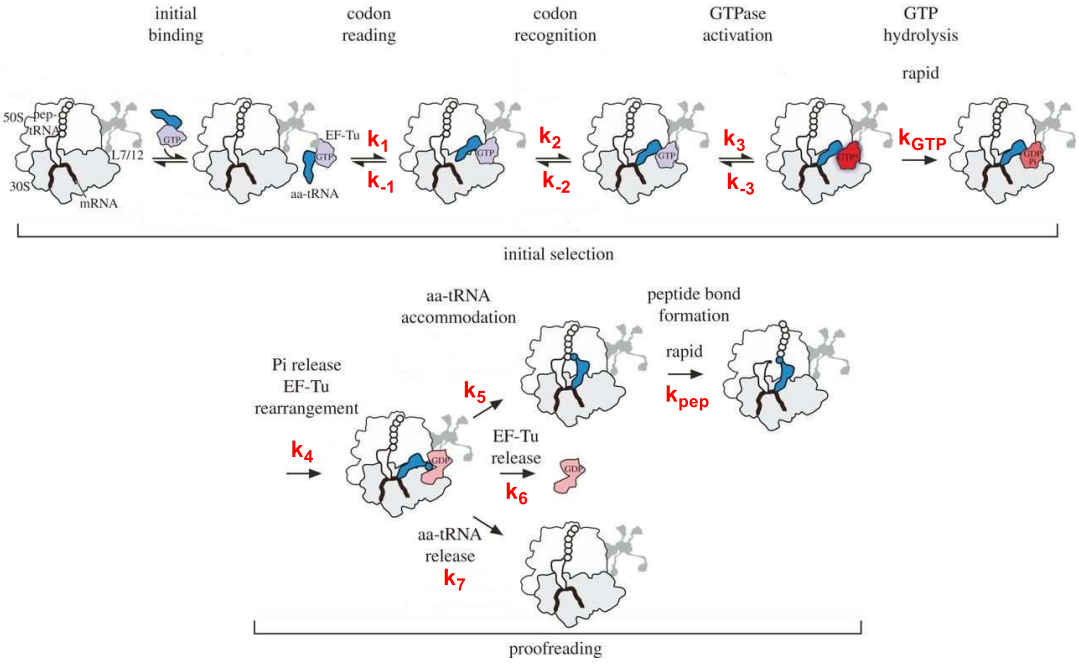

Figure 1.5. Representation of translation of a mRNA sequence into a growing polypeptide chain. Figure adapted from [22].

ribosome, several protein translation initiation factors, and the initiating tRNA binding to a molecule of mRNA [1]. Translation is usually initiated at methionine AUG, and in some less common cases valine GUG, and employs a special n-formylated methionine tRNA species [23]. The full, translation capable ribosome is assembled when the large ribosomal subunit displaces the initiating proteins and places the amino acid end of the initiating aminoacyl-tRNA sitting on the first codon of the mRNA into the peptidyl transferase center of the ribosome [1, 24]. The ribosome contains three sites for bound tRNA molecules and the process of translation involves coordinated selection and reaction of amino acids attached to tRNAs and subsequent ratcheting movement of the ribosome along the mRNA (Figure 1.4).

Once the mRNA is bound to the ribosome and translation has been initiated, the ribosome moves along the mRNA strand in a 5' to 3' direction. The translation of each codon, occurs at the acceptor, A, site in the ribosome (Figure 1.4). Complexes containing a charged, aminoacylated, tRNA, elongation factor Tu (EF-Tu) and a molecule of guanosine triphosphate (GTP) attempt to interact with the mRNA codon in the A site through codon-anticodon base-pairing [25]. If the codon-anticodon interaction is strong enough, translation moves forward (Figure 1.5). If the interaction is weak, another tRNA/EF-Tu/GTP complex attempts to bind the mRNA codon. The stochastic sampling of ternary complexes of tRNA/EF-Tu/GTP requires 10-100's of attempts before an appropriate tRNA is selected [26, 27].

When a proper tRNA has base-paired with the mRNA codon, a peptide bond is then formed between the amino acid at the A site and the polypeptide chain at the P site [28]. The formation of the peptide bond is catalyzed by the ribosome. After peptide bond

formation an additional proteins factor, EF-G ratchets the tRNAs and the bound mRNA (Figure 1.5). The tRNA (and codon) at the A site (now with the growing peptide attached) move to the P-site, the P-site tRNA (and codon) moves to the E site, and the E-site tRNA is released [28]. The ratcheting process exposes a new codon at the A site and the process is repeated. The released tRNAs, no longer charged with an amino acids, are recycled through interactions with amino acyl tRNA synthetase enzymes.

The process of tRNA selection and peptide bond formation are repeated until a termination codon UAG, UAA, or UGA in the mRNA is reached. When a termination codon reaches the A site in the ribosome, a ribosomal release factor binds instead of a charged tRNA. The peptide is hydrolyzed from the tRNA by the ribosome, and the different components dissociate, ending translation [1].

Ribosomes are large protein-RNA complexes that facilitate protein translation. The ribosome is enormously complex and is responsible for multiple functions in the translation process, including binding mRNA and catalyzing the peptide bond between individual amino acids in a protein. The ribosome is composed of a small and large subunit (Figure 1.4). The small subunit of the ribosome is a complex of a ribosomal RNA (rRNA) molecule and 19 proteins [21]. During translation initiation, nucleotides in the rRNA of the small subunit bind a molecule of mRNA at a sequence of nucleotides upstream of the translation start codon called a ribosomal binding site (RBS). Binding at the RBS by the small ribosomal subunit is critical for translation to start at the correct location. The small ribosomal subunit also binds three translation initiation factors and the special fMet-tRNA corresponding to the start codon. After the small ribosomal subunit binds the mRNA, translation initiation factors and the fMet-tRNA, the large

ribosomal subunit binds to the small subunit and translation begins. The large ribosomal subunit is a complex of two rRNA molecules and 31 proteins [29]. The large ribosomal subunit has multiple functions, the most critical of which is catalyzing the formation of a peptide bond between amino acids on tRNAs located at the A and P sites during translation. Once a peptide bond is formed between amino acids, and the tRNA without an amino acid leaves the P site, the ribosome moves the mRNA to the next codon in a 5'-3' direction.

Elongation factor Tu (EF-Tu) and GTP assist in bringing aminoacylated tRNAs to ribosome for sampling and acceptance at the A site [30, 31]. EF-Tu and a molecule of GTP complex with an aminoacylated tRNA and the tRNA anticodon attempts to base-pair with a mRNA codon in the A site of the ribosome. Sampling of different ternary complexes is stochastic. Acceptance of a complex at the A site in the ribosome is related to the strength of the hydrogen bonding between the anticodon of the aminoacyl-tRNA and the mRNA codon evaluated through specific ribosome contacts related to the geometry of the codon-anticodon RNA duplex. Correct pairing of the tRNA anticodon with the mRNA codon and recognition by the ribosome induces a conformational shift in EF-Tu that activates the GTPase domain of EF-Tu [32]. The GTPase hydrolyzes GTP into guanosine diphosphate (GDP) and an inorganic phosphate. Hydrolysis of GTP causes the ternary complex to dissociate, leaving the aminoacyl-tRNA bound in the ribosome at the A site. The aminoacyl-tRNA at the A-site of the ribosome may either dissociate from the ribosome or proceed with peptide bond formation. The strength of the codon-anticodon interaction biases the time the ternary complex sits in the A site and is interrogated by the ribosome before GTP hydrolysis,

and the probability of dissociation from the ribosome after release into the A site. The GTP hydrolysis and release steps serve to allow the codon – anticodon interaction to be evaluated twice, which contributes to the ability of the ribosome to discriminate between interactions that differ only slightly in energy [33, 34].

The tRNA is the physical link between the ordered codons of the mRNA and the ordered primary sequence of the protein. tRNAs are between 77 and 90 nucleotides, contain highly conserved secondary and tertiary structure, and are frequently highly modified to contain nucleotides other than U,G,C, and A [35, 36]. Regions of self-complementarity within a tRNA form four stem loops, commonly referred to as a cloverleaf structure; the tRNA then adopts an L-shaped tertiary structure (Figure 1.6). The 3-nucleotide anticodon sequence at the bottom end of the anticodon stem loop recognizes a specific mRNA codon through base-pairing interactions. The acceptor stem contains the 3' CCA tail is where the amino acid is covalently linked [36].

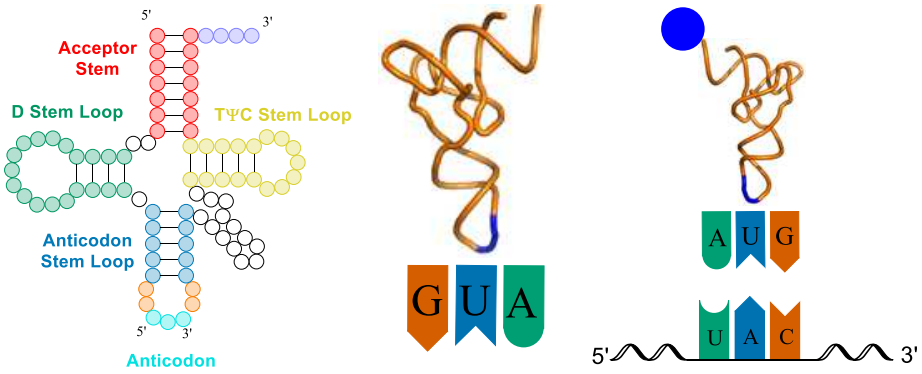

Figure 1.6. Left: Cloverleaf structure of a tRNA. Middle: Cartoon representation of the 3-dimensional structure of a tRNA, with the anticodon highlighted in blue. Right: A tRNA with an amino acid covalently attached at the 3' end of the CCA tail and base-pairing with a tyrosine UAC codon in mRNA. The 3-dimensional rendering of the tRNA is from PDB 1J1U [37].

In *E. coli*, 43 different tRNA species decode the 61 sense codons, meaning some tRNAs must decode more than one codon [38]. Decoding of multiple codons by a tRNA is usually achieved by a non-traditional “wobble” base-pairing between the first nucleotide of the tRNA anticodon and the third nucleotide of the mRNA codon. Wobble base-pairs are generally weaker than transitional Watson-Crick base-pairs (Table 1.2). The most commonly observed wobble base-pair is guanine-uridine (Figure 1.7); 14 codons in *E. coli* that do not have a Watson-Crick base-pairing tRNA are read by a G-U wobble (Table 1.1) [19].

Table 1.2. Free Energies of Watson-Crick and Wobble base-pairs [19].

Base-Pair	Type	Number of Hydrogen Bonds	ΔG^{O}_{37} (kcal/mol)
G-C	Watson-Crick	3	-6.5
A-U	Watson-Crick	2	-6.3
G-U	Wobble	2	-4.1

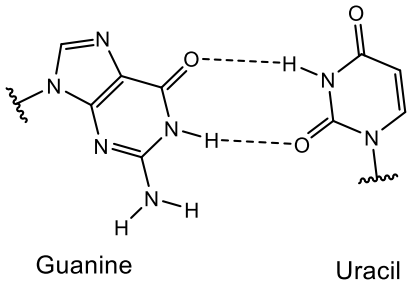

Figure 1.7. Wobble base-pairing between guanine and uracil [39, 40].

Other wobble base-pairings between a modified first nucleotide of the tRNA anticodon (position 34) and third position of the mRNA codon also occur in *E. coli*. For example, codons with the identity NAC (N represents any of the four nucleotides) are read by tRNAs that contain the nucleotide queosine (Q) at the first anticodon position [41, 42]. tRNA anticodons that contain queosine at the first position base-pair with mRNA codons that have U or C at the third position (Table 1.3). A second example of a tRNA in *E. coli* containing a modified nucleotide in the anticodon is $tRNA^{Arg}_2$, which is responsible for decoding codons CGU, CGC, and CGA as arginine [39, 43]. The first nucleotide of the anticodon in $tRNA^{Arg}_2$ is post-transcriptionally modified to inosine (I), which base-pairs with uracil, cytosine, and adenosine (Table 1.3).

Table 1.3. Expanded *E. coli* genetic code table. Shown are codons, corresponding tRNA anticodons (if present), and amino acids. Codons highlighted in green are read through a G-U wobble. Codons highlighted in blue are read through wobble base-pairing with queosine [41, 42]. Codons highlighted in orange are read through non Watson-Crick base-pairing with inosine [39, 43].

	codon	anti codon	AA	codon	anti codon	AA	codon	anti codon	AA	codon	anti codon	AA	
U	UUU		Phe	UCU		Ser	UAU		Tyr	UGU		Cys	U
	UUC	GAA		UCC	GGA		UAC	QUA		UGC	GCA		C
	UUA	UAA	Leu	UCA	UGA		UAA	------	oc	UGA	-----	op	A
	UUG	CAA		UCG	CGA		UAG	------	am	UGG	CCA	Trp	G
C	CUU		Leu	CCU		Pro	CAU		His	CGU	ICG	Arg	U
	CUC	GAG		CCC	GGG		CAC	QUG		CGC			C
	CUA	UAG		CCA	UGG		CAA	UUG	Gln	CGA			A
	CUG	CAG		CCG	CGG		CAG	CUG		CGG	CCG		G
A	AUU		Ile	ACU		Thr	AAU		Asn	AGU		Ser	U
	AUC	GAU		ACC	GGU		AAC	QUU		AGC	GCU		C
	AUA	LAU		ACA	UGU		AAA	UUU	Lys	AGA	UCU	Arg	A
	AUG	CAU	Met	ACG	CGU		AAG			AGG	CCU		G
G	GUU		Val	GCU		Ala	GAU		Asp	GGU		Gly	U
	GUC	GAC		GCC	GGC		GAC	QUC		GGC	GCC		C
	GUA	UAC		GCA	UGC		GAA	UUC	Glu	GGA	UCC		A
	GUG			GCG			GAG			GGG	CCC		G

The aminoacyl tRNAs that participate in translation on the ribosome are produced by a class of enzymes, the aminoacyl tRNA synthetases (aaRS), that are as responsible for the genetic code as the pairing of codon and anticodon during translation [1, 44]. The set of protein-tRNA interactions that lead to the attachment of the appropriate amino acid to the appropriate tRNA by the aaRSs constitutes a second genetic code [45]. Most organisms have 20 aminoacyl tRNA synthetases: one for each amino acid [44]. In *E. coli*, there are more tRNAs than aminoacyl tRNA synthetases; many aminoacyl tRNA synthetases accept more than one tRNA as a substrate. tRNA aminoacylation is a multi-step process that begins with the aaRS binding the amino acid and a molecule of ATP in the active site, resulting in the formation of an aminoacyl adenosine monophosphate (Figure 1.8).

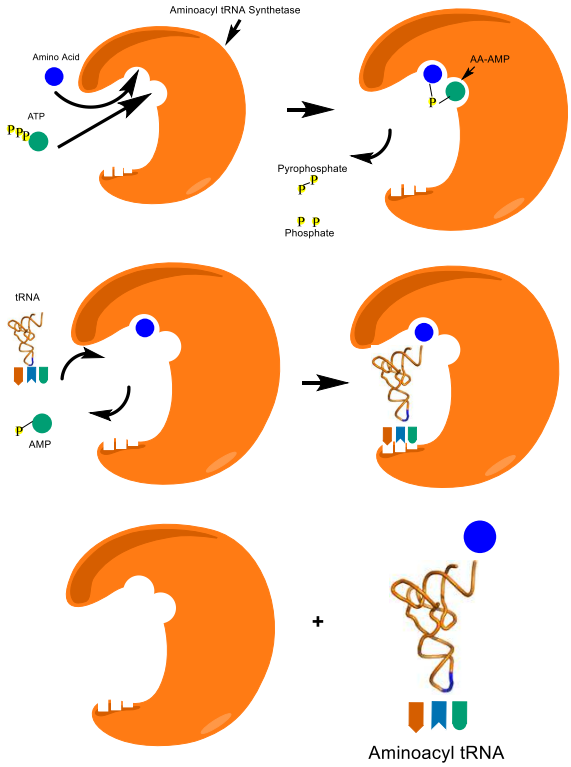

Figure 1.8. Cartoon schematic of the charging of a tRNA by an aaRS. The tRNA is from PDB 1J1U [37, 44].

Synthetases and their corresponding tRNAs are largely orthogonal, meaning cross-talk between different tRNA and synthetase pairs rarely occurs. Recognition of a tRNA by its corresponding aaRS is driven by identity elements contained in the tRNA [46, 47]. Identity elements of tRNAs can be both sequence and structure. For the majority of tRNAs, the anticodon stem loop (and in particular the anticodon sequence) and the acceptor stem are important identity elements [47, 48]. Modification of the nucleotides in a tRNA that act as identity elements has the potential to break orthogonality and cause the tRNA to be charged with an incorrect amino acid by a non-cognate aaRS. In addition, modifying the identity elements of a tRNA may reduce the efficiency of aminoacylation by the appropriate aaRS.

The process of translation as described up to this point is really the first approximation of the actual somewhat messy stochastic process. The simple model of translation connecting mRNA sequence to protein sequence via the action of ribosomes and aminoacyl tRNAs is useful in conveying the gross information transfer that occurs. The simple description does not contain sufficient detail to evaluate the rate of translation, the error rates that occur, and how composition of the translational apparatus and various modifications to tRNAs affect these quantities. A more detailed understanding of translation is required if one desires to manipulate the system to expand the genetic code. Translation is an immensely complex process, and the speed and efficiency of translation is affected by a number of factors, including the abundance of the different aminoacyl-tRNAs, energetics of codon-anticodon pairings, aminoacylation efficiency of the amino acid onto the tRNA by its aaRS, and tRNA modifications [22, 49-51].

The different tRNA species in *E. coli* are present varying concentrations, typically ranging from hundreds to tens of thousands in a cell [38]. If all other contributing factors than endogenous tRNA abundances were held constant (for example, tRNA binding efficiency to EF-Tu and the ribosome) during the process of translation, one may expect that amino acid incorporations would directly correlate to the endogenous tRNA concentration [52]. In reality, binding of different tRNAs to EF-Tu and to the ribosome differ slightly, and how these differences contribute to the process of translation is relatively unknown [30, 50, 53-55].

Relative energetics of tRNA anticodon-codon pairings in the ribosome are also a significant factor during translation. When an aminoacylated tRNA/EF-Tu/GTP ternary complex enters the A site of the ribosome during translation, the anticodon of the tRNA attempts to base pair with the exposed mRNA codon. In addition to the base pairing interactions between the anticodon nucleotides and the nucleotides of the mRNA codon, other interactions between the ribosome and the ternary complex also occur, including between the ribosomal RNA residues and the tRNA anticodon-codon complex [56]. These interactions may have a significant impact on the geometry of the anticodon-codon base pairing [22]. Furthermore, modifications of nucleotides in the anticodon of the tRNA may impact the geometry and strength of the tRNA anticodon-codon complex at the A site in the ribosome during translation. The sum of the energetic contributions of many different interactions determines which tRNAs decode which codons [53, 57].

The aminoacylation efficiencies of tRNAs by their corresponding aaRSs also indirectly affect the process of translation. Aminoacylation efficiency of a tRNA by its

cognate aaRS depends on multiple variables, including the concentrations of available tRNA, aaRS, and amino acid, all of which are present at different concentrations in a cell [44]. Variations in the concentrations of any of the three variables affects the aminoacylation efficiency, and as a result the aminoacylated tRNA concentration available to the cell for use in translation.

The relative quantitative importance of the contribution of each of the factors has not been well-established. The work in this dissertation describes progress towards quantifying the relative importance of the factors described above to the further understand of natural translation systems and to guide the construction of more efficient genetic code expansion systems. Quantifying the relative importance of the factors involved in protein translation may be achieved by precisely measuring the *in vivo* reassignment of sense codons using an engineered orthogonal tRNA/aaRS pair.

1.5 METHODS FOR EXPANDING THE GENETIC CODE

Nature uses a nearly universally conserved set of 20 amino acids as building blocks for proteins. The natural amino acids have only a very limited chemical properties, being made up of aliphatic, aromatic polar and charged amino acids (Figure 1.1) [58, 59]. Expansion of the genetic code to include non-canonical amino acids (ncAAs) with chemical functional groups, common in organic chemistry, but not found in nature is a powerful strategy for probing and extending the properties of proteins.

Non-canonical amino acids may be used to study a particular protein's function, investigate cellular processes, or generate novel proteins with entirely new functions [58-62]. Over 150 different ncAAs have been incorporated into proteins and include side chains with a wider range of size and chemistries than the natural set, some of which

are shown below (Figure 1.9). Non-canonical amino acids that act as spectroscopic probes, such as *p*-cyanophenylalanine, an IR probe, may be used to extract information about a protein's conformation or investigate interactions with other proteins in the cell's

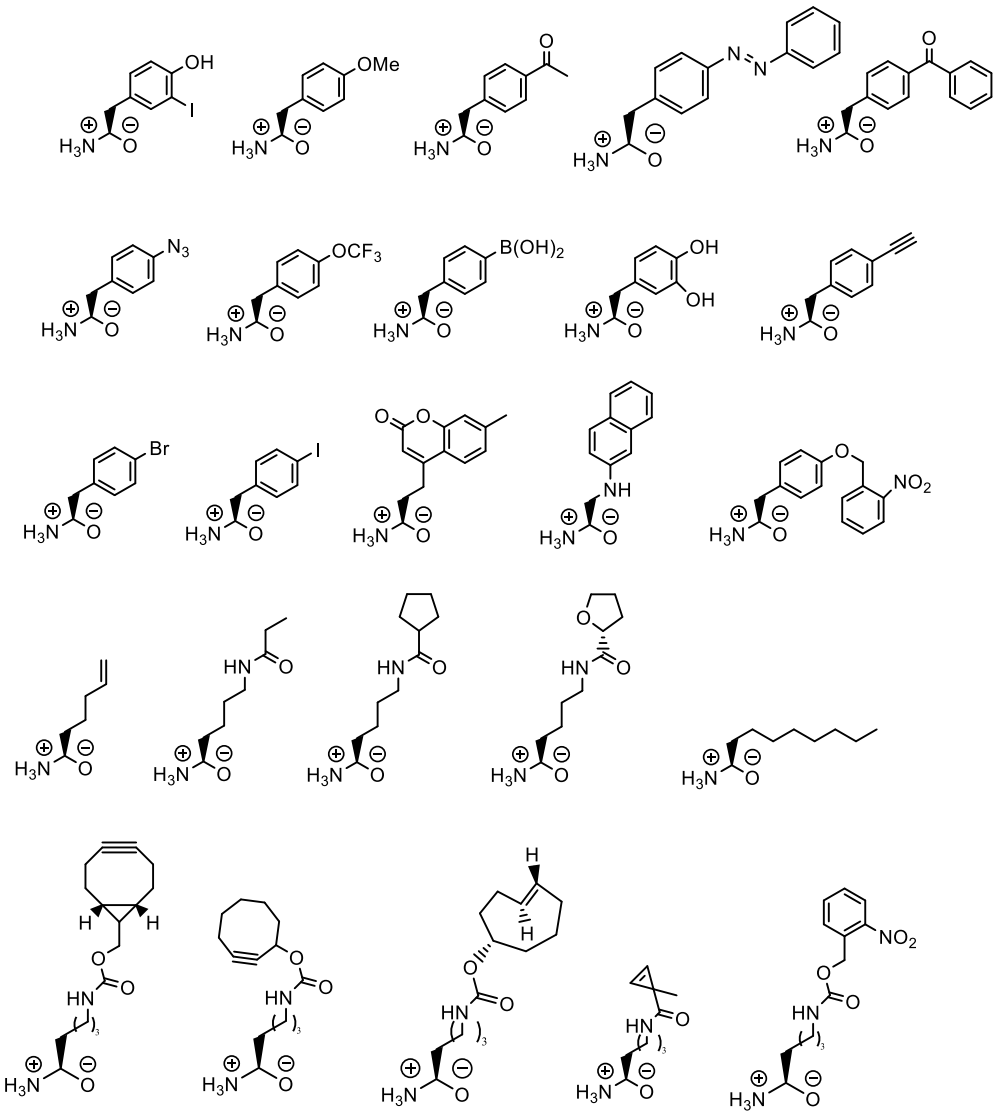

Figure 1.9. Examples of ncAAs that have been incorporated into proteins [58, 59].

environment. Residues like *p-bromo* or *p*-Iodo phenylalanine or selenomethionine can be used to aid in crystal structure determination through MAD phasing [63-66]. Incorporation of ncAAs that are analogues of post-translational modifications (PTMs) allows for investigation of the role a particular PTM (such as phosphorylation or

sulfation) may have in a biological process [67, 68]. Photocaged amino acids can act as switches for protein function by blocking activity of the protein until irradiated with light [69]. Some ncAAs have selectively reactive side-chains, such as azids or alkenes that may be used as handles for attaching chemical labels or tags at a specific site in a protein through a bio-orthogonal reaction [70-72].

Multiple *in vitro* and *in vivo* methods for the incorporation of ncAAs exist. The two most commonly employed *in vivo* strategies for biological incorporation of ncAAs into proteins are global amino acid reassignment amber stop codon suppression [63, 73-75]. Significant advantages and disadvantages of each method for biological incorporation of ncAAs into proteins are discussed below. Common *in vitro* methods for incorporation of ncAAs into proteins involving protein semi-synthesis and in vitro translation systems are also discussed briefly for comparison.

The first method developed for introduction of ncAAs was global amino acid reassignment where every occurrence of one of the 20 canonical amino acids is replaced with an ncAA that is a close structural analogue [63, 73, 76, 77]. In order to achieve high ncAA incorporation efficiency the canonical amino acid to be replaced and its corresponding tRNA/aaRS pair must be selected carefully, and the host organism and environment must be engineered such that incorporation is heavily biased towards the ncAA. Global amino acid reassignment uses an endogenous aaRS to charge the ncAA onto its corresponding tRNA, which then ribosomally incorporates the ncAA throughout the proteome [78]. An aaRS that either naturally has or that has been engineered to have relaxed substrate specificity must be used to enable efficient charging of the ncAA onto the corresponding tRNA. If an unmodified endogenous aaRS

is employed, only close structural analogues of the natural amino acid to be replaced can typically be activated by the endogenous aaRS.

Competition between the ncAA and the canonical amino acid as a substrate for the aaRS also needs to be eliminated for efficient amino acid replacement. Removal of the canonical amino acid from the environment eliminates competition, and is achieved by using a host organism that is unable to biologically produce the canonical amino acid to be replaced (auxotroph), and by removing the canonical amino acid from the host's growth medium [78, 79]. Minimizing the presence of the canonical amino acid biases incorporation strongly towards the ncAA.

Although ncAA incorporation is achieved at multiple sites with high efficiency in the target protein, global amino acid reassignment is limited in multiple ways. The ncAA must be a close structural analogue of a natural amino acid, eliminating some ncAAs with structures and chemistries significantly different from the canonical amino acid from consideration [78]. In addition, the sites that the ncAA is incorporated at are not specific; all instances of the canonical amino acid are replaced, and as a result the structure and function of the target protein may be compromised. Proteome-wide replacement of a canonical amino acid may also critically affect cell health leading to lower yields of the target protein. An auxotrophic host organism may also be difficult to engineer and maintain [79]. Finally, global amino acid replacement generates an alternate 20 amino acid genetic code instead of an expanded 21 amino acid genetic code.

In contrast the second more commonly employed strategy, amber stop codon suppression, expands the genetic code to include a 21st amino acid. An engineered, imported orthogonal tRNA/aaRS pair is used to direct incorporation of the ncAA at the

amber (UAG) stop codon. *In vivo* amber stop codon suppression employing an expressed orthogonal pair was first described nearly 20 years ago by Furter and has been extensively developed by the Schultz lab [59, 74, 75, 80]. The amber UAG codon was targeted for three main reasons. First, natural UAG suppressors already existed in nature, suggesting that the amber codon may also be exploited for ncAA incorporation [81, 82]. Second, the UAG codon is a termination codon in *E. coli*; incorporation of a ncAA in response to the UAG codon results in expansion of the genetic code from 20 to 21 amino acids. Third, only 321 instances of the amber stop codon occur in the genome of *E. coli* (out of ~1.5 million codons), minimizing the potential effects of ncAA incorporation on the organism's proteome [83, 84].

Amber stop codon suppression requires introduction of an additional tRNA/aaRS pair that specifically targets the UAG codon [74, 75]. Multiple tRNA/aaRS pairs have been engineered for ncAA incorporation in response to the UAG codon [58, 85-88]. The two most commonly used pairs are the tyrosyl tRNA/aaRS from *Methanocaldococcus jannaschii* (*M. jannaschii*), and the pyrrolysyl tRNA/aaRS from the *Methanosarcina* species [85, 86, 89]. Both the *M. jannaschii* tyrosyl tRNA/aaRS and the *Methanosarcina* pyrrolysyl tRNA/aaRS have been shown to be orthogonal to each of the *E. coli* tRNA/aaRS pairs [85, 86, 89]. The imported tRNAs are not charged with a canonical amino acid by any of the *E. coli* synthetases, and the imported synthetases do not charge any of the *E. coli* tRNAs with a ncAA. The result of orthogonality is site-specific incorporation of the ncAA at the UAG codon, without disruption of canonical amino acid incorporation at the 61 sense codons.

The *M. jannaschii* tRNA was modified to have a CUA anticodon; the pyrrolysyl tRNA from the *Methanosarcina* species did not require modification, as the wild type (WT) tRNA already had a CUA anticodon [85, 86]. The *M. jannaschii* and *Methanosarcina* aaRSs are used widely for amber stop codon suppression, and have been repeatedly engineered for incorporation of over 150 different ncAAs [59, 88].

In order to produce proteins with an ncAA incorporated in response to the amber stop codon, the gene coding for the protein of interest is modified to have a UAG codon at the desired position. The protein of interest and the engineered tRNA/aaRS are expressed in *E. coli* and the ncAA is added to the growth medium. The mRNA transcript of the target protein is then read in one of two ways (Figure 1.10) [58, 59, 70].

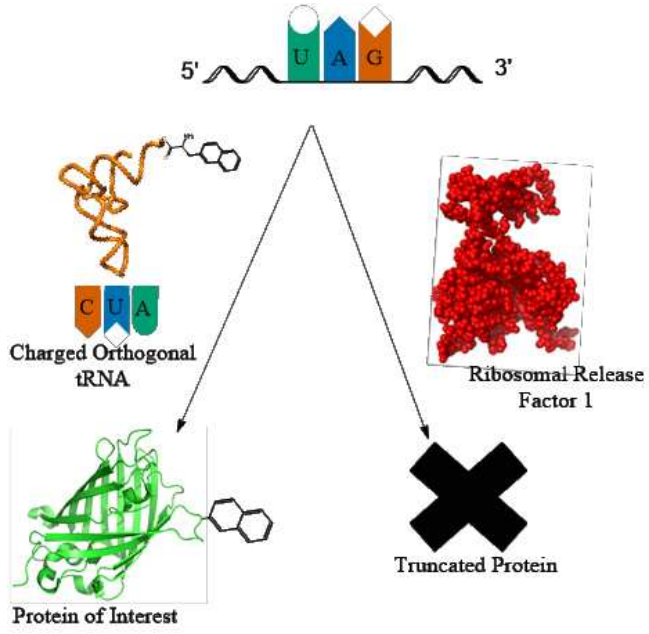

Figure 1.10. Introduction of an engineered orthogonal tRNA/aaRS pair for amber suppression generates two options for translation of a single mRNA transcript. PDB 2B3P,1J1U, 2B3T [37, 90-92].

If the orthogonal tRNA charged with ncAA outcompetes the endogenous translational termination release factor 1 (RF1) and reads the amber stop codon, the ncAA is incorporated and full length protein is produced. If the release factor reads the UAG codon instead of the charged orthogonal tRNA, translation is terminated and full length protein is not produced [87]. Depending on the location of the amber stop codon in the target protein, two products of very different sizes may be produced and easily separated, especially if a purification tag is downstream of the UAG codon.

Amber stop codon suppression has been limited to ~1 position in a protein until recently because of competition with ribosomal release factor 1 (RF1) [84, 93, 94]. The overall yield of a target protein is dependent on the suppression ability of the charged orthogonal tRNA. Typical ncAA incorporation at 1 UAG codon in a target protein yields on average 25% of the wild-type protein [95]. Adding additional UAG codons in the target protein is additive: suppressing 2 UAG codons in a target protein has a yield of ~6% (.25 x .25) of the wild-type protein, and suppressing 3 stop codons has a yield of ~1.6% (.25 x .25 x .25). Amber stop codon suppression becomes cost prohibitive and inefficient when more than 1 UAG codon is present.

Recent advances in amber stop codon incorporation have focused on achieving efficient multi-site ncAA incorporation at UAG codons through engineering the genome of *E. coli* to eliminate competition of the charged orthogonal tRNA with RF1 [93, 94]. The Sakamoto, Wang, and Church groups have each demonstrated different approaches to engineering the genome that allowed the removal of the normally essential release factor RF1. Multi-site ncAA incorporation in response to the UAG codon was then demonstrated with high efficiencies. The Sakamoto group hypothesized

that the gene for RF1 could be knocked out from *E. coli* if UAG codons that terminated essential genes were replaced with alternate stop codons [94]. The Wang group hypothesized that the gene for RF1 could be removed by improving expression of ribosomal release factor 2 (RF2) which recognizes UAA and UGA stop codons by fixing a programed frame shift in the RF2 gene [93, 96]. Each method allowed for knockout of RF1 and resulted in multi-site UAG ncAA incorporation with increased yields.

Despite the successful knockout of RF1 the amber stop codon was not truly a free codon; in both cases, UAG codons remained in the genome of *E. coli* at terminating positions. In order to completely reassign the amber stop codon to a free codon, the Church group used multiplex automated genome engineering (MAGE) to replace all 321 instances of the amber stop codon with a an alternate termination codon (UAA) (Figure 1.11) [84].

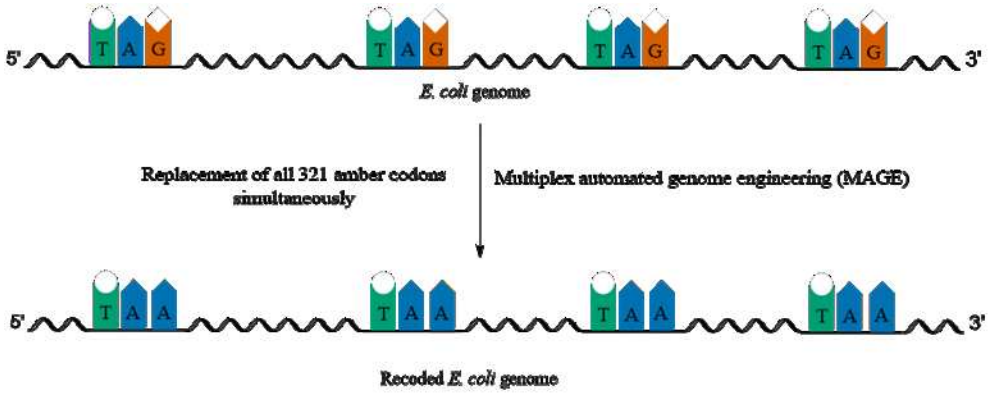

Figure 1.11. Generation of a genomically recoded strain of *E. coli* in which all 321 instances of the amber stop codon were replaced with the alternate stop codon UAA. RF1 was also removed from the genome. The result is a UAG-free strain of *E. coli*. [84].

Furthermore, RF1 was removed from genome of the recoded *E. coli* strain. The genomically recoded organism (GRO, C321.ΔRF1) was the first demonstration of

genomic engineering to completely recode one of the 64 codons. The GRO *E. coli* strain displays significantly improved efficiency of multi-site incorporation of ncAAs in response to UAG codons [84].

The brute-force method of freeing codons through MAGE based mutagenesis is not easily extended to other codons. The least frequent *E. coli* sense codon (AGG) is used ~1400 times in the genome,~5x more frequently than the UAG stop codon [97]. Replacing such a large number of codons presents a significant challenge; the Church lab is currently constructing strains with additional free codons. However, the new approach involves total de novo genome synthesis. The project of completely synthesizing 5 X 10^6 bases is expected to take a significant amount of time and money, and is currently the largest genome synthesis project [98]. To “free” sense codons, additional modifications of the translational machinery beyond genome recoding are required. The tRNAs that read the codons targeted for replacement need to be removed, which may require significant modification of the set of tRNAs used in *E coli*.

A promising, alternate method for multi-site ncAA incorporation is sense codon reassignment [99, 100]. Of the 20 canonical amino acids, 18 are encoded by multiple codons (methionine and tryptophan each have a single codon). Sense codon reassignment takes advantage of the degeneracy of the genetic code by reassigning one of multiple codons that code for the same amino acid. The remaining redundant codons still code for the canonical amino acid – resulting in an expanded genetic code. Furthermore, multiple sense codons may be targeted for reassignment at the same time, allowing 22 or more amino acid genetic codes. Sense codon reassignment uses an engineered orthogonal tRNA/aaRS pair to incorporate a ncAA in response to one of

the 61 sense codons [97, 99-101]. The extent to which the genetic code can be modified, the best places to infiltrate the code in different organisms (specifically *E. coli*), and how the already evolved ncAA incorporating orthogonal pairs can be made to function in sense codon reassignment are largely unexplored questions addressed through studies in this dissertation.

Table 1.4. Expanded genetic code table, including codon usage and tRNAs. Columns from left to right: codon, anticodon of tRNA (blank means no tRNA with a Watson-Crick base-pairing anticodon exists), relative codon usage, number of *E. coli* codons used per 100 codons, and amino acid [102, 103].

	codon	anti codon	Rel codon usage	E. coli codon usage	AA	codon	anti codon	Rel codon usage	E. coli codon usage	AA	codon	anti codon	Rel codon usage	E. coli codon usage	AA	codon	anti codon	Rel codon usage	E. coli codon usage	AA	
T	TTT		0.57	18.4	F	TCT		0.11	7	S	TAT		0.53	13.4	Y	TGT		0.42	4.3	C	T
	TTC	GAA	0.43	13.7		TCC	GGA	0.11	7.1		TAC	QUA	0.47	10.1		TGC	GCA	0.58	5.3		C
	TTA	LJAA	0.15	11.3	L	TCA	LJGA	0.15	5.9		TAA	—	0.64	1.7	oc	TGA	—	0.36	0.7	op	A
	TTG	CAA	0.12	11.4		TCG	CGA	0.16	7.3		TAG	—	0	0.2	am	TGG	CCA	1	12.6	W	G
C	CTT		0.12	9.1	L	CCT		0.17	5.8	P	CAT		0.55	10.6	H	CGT	ICG	0.36	17.3	R	T
	CTC	GAG	0.1	9.1		CCC	GGG	0.13	4.5		CAC	QUG	0.45	8		CGC		0.44	18.1		C
	CTA	LJAG	0.05	3.2		CCA	LJGG	0.14	7		CAA	LJUG	0.3	12.7	Q	CGA		0.07	2.9		A
	CTG	CAG	0.46	43.5		CCG	CGG	0.55	19.1		CAG	CUG	0.7	23.7		CGG	CCG	0.07	4.4		G
A	ATT		0.58	25	I	ACT		0.16	7.3	T	AAT		0.47	14.5	N	AGT		0.14	7.2	S	T
	ATC	GAU	0.35	20.6		ACC	GGU	0.47	19.2		AAC	QUU	0.53	17.8		AGC	GCU	0.33	13.2		C
	ATA	LAU	0.07	3.5		ACA	LJGU	0.13	5.8		AAA	LJUU	0.73	27.7	K	AGA	LJCU	0.02	1.7	R	A
	ATG	CAU	1	22.9	M	ACG	CGU	0.24	11.8		AAG		0.27	8.4		AGG	CCU	0.03	1		G
G	GTT		0.25	15.1	V	GCT		0.11	12.6	A	GAT		0.65	26.5	D	GGT		0.29	20.5	G	T
	GTC	GAC	0.18	12.5		GCC	GGC	0.31	21		GAC	QUC	0.35	15.7		GGC	GCC	0.46	24.2		C
	GTA	LJAC	0.17	8.9		GCA	LJGC	0.21	16.7		GAA	LJUC	0.7	32.6	E	GGA	LJCC	0.13	6.5		A
	GTG		0.4	21.6		GCG		0.38	27.7		GAG		0.3	14.6		GGG	CCC	0.12	9		G

A number of different codon groupings present opportunities for genetic code expansion by sense codon reassignment (Table 1.4) [102, 103]. The first set of codons is those not read by traditional Watson-Crick base-pairing. The compliment of tRNA anticodon sequences in most organisms is smaller than the number of sense codons. In *E. coli* 43 tRNA species are used to read 61 sense codons. The subset of codons that do not have an endogenous tRNA is read through a non-Watson-Crick base-pairing (wobble) interaction [39, 99]. Wobble base-pairing (for example, the G-U wobble)

between a codon and tRNA anticodon occurs for 21 of the sense codons (Figure 1.7) [39, 104]. Wobble pairs are less stable than traditional Watson-Crick G-C and A-U pairings. The difference in free energies between wobble and Watson-Crick base-pairing could potentially be harnessed for sense codon reassignment by introducing an orthogonal tRNA that is capable of Watson-Crick base-pairing with a desired codon [52, 99]. The more energetically favorable Watson-Crick base-pairing may more effectively out-compete the wobble base-pairing of the endogenous tRNA, allowing the codon to be reassigned at a high level [99]. The extent to which differences in energetics of codon–anticodon interaction, tRNA concentration, tRNA modifications, sequence context effects, EF-Tu binding and other variables in the process affect the fidelity and the ability to reprogram translation have not been thoroughly investigated. The measurements described in this dissertation are also directed at beginning to quantify the relative importance of the various system variables on the process of translation.

The use of energetic differences between codon-anticodon base pairing has been applied to reassign a sense codon in the original demonstration of breaking the degeneracy of the genetic code. Tirrell and co-workers used the difference in base-pairing energies between wobble and Watson-Crick base-pairing to reassign one of the phenylalanine (Phe) codons to a ncAA [99]. In *E. coli* phenylalanine is encoded by two codons: UUC and UUU. A single tRNA species with a GAA anticodon ($tRNA^{Phe}_{GAA}$) decodes both codons [99]. The first position of the *E. coli* $tRNA^{Phe}_{GAA}$ is the wobble position; the guanosine is capable of decoding the third position of the UUC codon via a Watson-Crick G-C base-pairing or the third position of the UUU codon through a G-U wobble base-pairing [39]. In order to reassign the UUU codon, an orthogonal yeast

phenylalanine $tRNA^{Phe}_{AAA}$ that was capable of Watson-Crick base-pairing with the UUU codon was expressed in *E. coli*; the yeast $tRNA^{Phe}_{AAA}$ was charged with the ncAA L-3-(2-naphthyl)alanine (NapA) by an engineered yeast phenylalanine aaRS. The charged orthogonal yeast $tRNA^{Phe}_{AAA}$ was able to out-compete the endogenous *E. coli* $tRNA^{Phe}_{GAA}$ because of the favorable codon-anticodon energetics, and near quantitative reassignment of the UUU codon to NapA was reported [19, 99].

Another codon group of codons that have been targeted for sense codon reassignment are codons that are used rarely in *E. coli*. Rare codons typically have endogenous tRNA/aaRS pairs that are present at lower concentrations in the cell [105]. In *E. coli*, the various tRNA species are present across an ~40 fold concentration range [106]. Because the ribosome requires sampling of the tRNA pool before decoding any mRNA codon, a highly expressed orthogonal tRNA may be able to out-compete the less abundant endogenous tRNA for the codon to be reassigned. Reassigning a rare codon to a ncAA may also minimize the impact on the host's proteome because of low usage.

Multiple groups have recently demonstrated that the least used sense codon in *E. coli*, the AGG codon, can be reassigned to a ncAA at near quantitative levels. Incorporation of ncAAs in response to the AGG codon was recently demonstrated by the Liu, Sakamoto and Yoo labs [97, 101, 107]. The Liu lab showed high level reassignment of the AGG codon by over expressing orthogonal tRNA and aaRS that was not further improved by attempts to remove the competing tRNA. Both the Sakamoto and Yoo labs demonstrated that the gene coding for the *E. coli* $tRNA^{Arg}_{CCU}$ could be knocked out, removing competition between an orthogonal $tRNA_{CCU}$ and the

endogenous $tRNA^{Arg}_{CCU}$. Reassignment of the arginine AGG codon to a ncAA was observed at high levels in both studies.

The Yoo and Sakamoto labs hypothesized that removing the endogenous *E. coli* $tRNA^{Arg}_{CCU}$ was critical for efficient reassignment of the arginine AGG codon to a ncAA. However, a recent study by Fisk and co-workers suggests that knocking out the *E. coli* $tRNA^{Arg}_{CCU}$ is not critical for efficient reassignment of the arginine AGG codon to a ncAA [108]. Fisk and co-workers demonstrated that the reassignment efficiency of the Arg AGG codon to tyrosine could be improved to near quantitative levels through directed evolution of the orthogonal *M. jannaschii* tyrosyl tRNA/aaRS pair; the near quantitative level of arginine AGG reassignment to tyrosine with the improved orthogonal *M. jannaschii* tyrosyl tRNA/aaRS pair was observed in the presence and absence of the *E. coli* $tRNA^{Arg}_{CCU}$, showing that the knockout of the endogenous *E. coli* $tRNA^{Arg}_{CCU}$ was not critical for efficient reassignment of the arginine AGG codon to tyrosine. The contrasting results of studies involving the reassignment of the rare Arg AGG sense codon show that codon selection for reassignment to a ncAA using an orthogonal tRNA/aarS pair and predicting the relative importance of the factors involved (such as competing endogenous tRNA concentration) needs to be further investigated.

Although not the focus of this dissertation, multiple *in vitro* methods exist for incorporating ncAAs into proteins and two of them are briefly described here. Cell-free translation systems can be used to incorporate ncAAs into proteins [109]. The most tunable cell-free translation systems are built by extracting and purifying the components necessary for translation separately, and then combining them together in a reaction tube. Specific components (such as tRNAs and aaRSs) can be added or

removed from the reaction, allowing for a highly tunable system capable of incorporating multiple ncAAs at multiple positions in a protein simultaneously [109]. However, the time and cost required to set up and use a cell-free translation system can be prohibitive [110]. Furthermore, cell-free translation systems cannot produce proteins or protein assemblies that require additional cellular components for proper assembly or function. For example, M13 bacteriophage (composed of multiple copies of 5 different coat proteins and a single-stranded DNA genome) is assembled from into a functional particle as it exits the host *E. coli* cell [111, 112]. M13 bacteriophages containing ncAAs in the coat proteins could not be readily prepared using a cell-free translation system.

Total protein synthesis or semi synthesis through solid phase peptide synthesis can be used to generate non-natural peptides and proteins [113]. In solid phase peptide synthesis, the first amino acid in the protein sequence is covalently linked to a solid support at the C-terminus of the amino acid [113]. The side-chain and N-terminus of the amino acid are protected during the addition to the solid phase support. After addition, the surface is washed to remove any excess reagents, and the N-terminus of the amino acid is de-protected. The next amino acid is then added, and a peptide bond is formed between the N-terminus and C-terminus of the two amino acids. This cycle is repeated as many times as necessary to achieve the desired protein [113]. The advantage of solid-phase peptide synthesis is that any amino acid sequence can be synthesized. A major drawback of solid-phase peptide synthesis is less than a 100% yield after each step. If the synthesis was 99% efficient at each step, after the addition of 100 amino acids the theoretical yield of the full-length product would be 36%. However, separation of the full-length product from side products that are very similar in chemical properties

(for example, separating 99-mers, 98-mers, etc. from a full-length 100-mer) is extremely difficult [114]. As a result, the realistic yield of the full-length product is usually much, much lower than the theoretical yield, and the cost of synthesis and purification may be prohibitive. Solid phase peptide synthesis is typically limited to peptides in the 50-70 amino acid range [113, 114]. Typical bacterial proteins contain ~300 amino acids. Solid phase synthesized peptides can also be linked together to form proteins through native chemical ligation.

Native chemical ligation covalently links two peptides together, and requires the C-terminus of the first peptide to be a thioester. The second peptide must have a cysteine residue at the N-terminus [115]. An initially reversible thioester exchange reaction occurs to bring the two peptides together, resulting in a peptide with a non-natural backbone. The non-natural thioester backbone then undergoes an irreversible arrangement to produce the natural peptide bond at the backbone [115]. One advantage of native chemical ligation is the reaction is done under mild conditions, neutral pH and in water [115]. Furthermore, the reaction can be repeated to link multiple peptides together to give a large protein. However, native chemical ligation may be difficult to use with many proteins of interest, because cysteines may not be present at the necessary location in the protein for ligation.

1.6 QUANTIFICATION OF ORTHOGONAL tRNA/aaRS PAIR DIRECTED CODON REASSIGNMENT EFFICIENCY *IN VIVO* USING A GFP REPORTER SYSTEM

Unlike stop codon suppression, where the orthogonal tRNA competes with translation termination machinery and missed incorporations lead to truncated products, in a sense codon reassignment systems missed incorporations produce full-length

polyamides [100]. Sense codon reassignment efficiency cannot be measured simply as the total amount of full-length product produced.

In order to further understand protein translation and improve sense codon targeting genetic code expansion systems, the Fisk lab recently reported a fast, quantitative method for measuring the efficiency of sense codon reassignment to tyrosine [100]. The system makes use of the fact that tyrosine is essential for the fluorescence of GFP, and measures the reassignment of a sense codon replacing a critical tyrosine in GFP [116, 117]. By expressing an orthogonal *M. jannaschii* aaRS that directs incorporation of tyrosine and an orthogonal *M. jannaschii* tRNA with a modified anticodon sequence that can decode the codon specifying the critical tyrosine position in the engineered GFP reporter protein, the system allows quantification of the efficiency of sense codon reassignment as the restoration of fluorescence of GFP [100].

GFP fluorescence was used as an indicator of whether a targeted codon was read by an endogenous *E. coli* tRNA or the orthogonal *M. jannaschii* tRNA (Figure 1.12). If the endogenous *E. coli* tRNA translates the sense codon in the fluorophore, an amino acid other than tyrosine is incorporated and a non-fluorescent GFP results. If the orthogonal *M. jannaschii* tRNA translates the sense codon and tyrosine is incorporated, a functional fluorophore forms and fluorescent GFP is produced [100]. Fisk and co-workers demonstrated that sense codon reassignment efficiency could be determined by comparing the amount of fluorescent protein in a sense codon reassigning system to the amount of wild-type GFP produced in a related, non-reassigning system. Sense codon reassignment as low as 1 part in 1,000 (*in vivo*) and 1 part in 10,000 (purified protein) were measured [100].

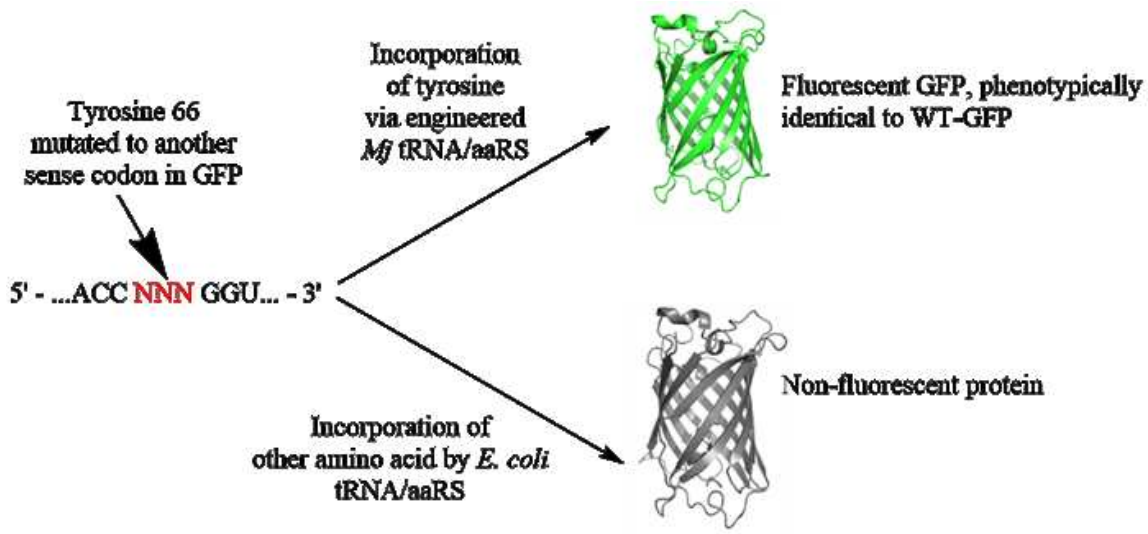

Figure 1.12. System using GFP fluorescence as an indicator of sense codon reassignment efficiency (PDB 2B3P). One mRNA transcript of GFP may be decoded two different ways. In the upper case, the engineered orthogonal *M. jannaschii* tRNA decodes the sense codon and tyrosine is incorporated. In the lower case, the endogenous *E. coli* tRNA decodes the sense codon and incorporates the corresponding amino acid [90, 91, 100].

Using GFP as a reporter protein for codon reassignment allows for measurement of codon reassignment quickly and *in vivo*. The fluorophore of a super folding variant of GFP (sfGFP) is composed of residues threonine 65, tyrosine 66, and glycine 67 (Figure 1.13) [90, 91, 116]. Threonine 65, tyrosine 66, and glycine 67 undergo cyclization followed by dehydration to form a fluorophore essential for fluorescence of the protein. The codon reassignment screen takes advantage of the absolute requirement for a tyrosine residue at position 66 for formation of fluorescent protein. When the tyrosine residue in the fluorophore is changed to one of the other 19 amino acids the resulting protein has greatly reduced and shifted fluorescence [116, 117]. The fluorescence-based screen for measuring sense codon reassignment efficiency to tyrosine is

compatible with 62 of the 64 codons in the genetic code; the two codons specifying tyrosine are the only ones for which reassignment efficiencies cannot be measured.

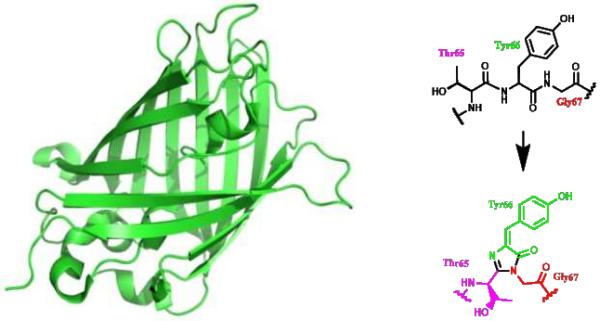

Figure 1.13. Left: Cartoon representation of a crystal structure of super-folding GFP (PDB 2B3P). Right: Formation of the GFP fluorophore [90, 91, 116].

The first four sense codon reassignment efficiencies measured with the engineered *M. jannaschii* tRNA/aaRS pair ranged from 1% (Phe UUU) to 6% (Lys AAG) [100]. Importantly, all codons targeted showed measurable levels of reassignment. Although sense codon reassignment efficiencies were low, any quantifiable level of sense codon reassignment may be used as a starting point for improvement through directed evolution. Sense codon reassignment efficiency of the Lysine AAG codon to tyrosine with the *M. jannaschii* tRNA/aaRS pair was improved by generating a library of tRNA/aaRS variants focused around the tRNA anticodon binding domain in the aaRS and the anticodon stem loop of the tRNA (other than the anticodon) [100]. The guiding idea behind the tRNA/aaRS library was that improvement of the orthogonal tRNA/aaRS interactions would lead to increased aminoacylation efficiency and a higher level of sense codon reassignment. Previous work demonstrated that the anticodon identity was a significant factor in the aminoacylation efficiency of the orthogonal *M. jannaschii* tRNA/aaRS pair [118, 119]. The GFP screen and fluorescence activated cell sorting (FACS) were used to screen and identify library members that showed an increased

level of fluorescence, suggesting an increase sense codon reassignment to tyrosine at the Lys AAG codon. A variant with a 4.9-fold improvement over the starting level of reassignment was identified, supporting the hypothesis that directed evolution of the anticodon binding domain of the orthogonal *M. jannaschii* tRNA/aaRS would lead to improved sense codon reassignment efficiency to tyrosine for the Lys AAG reassignment system.

An additional advantage of the *in vivo* screen for sense codon reassignment was the ability to measure the effects of sense codon reassignment on cell health. During the *in vivo* reassignment of sense codons to tyrosine using the orthogonal *M. jannaschii* tRNA/aaRS pair, every instance of the targeted codon in the proteome of the *E. coli* host may potentially be reassigned to tyrosine. As a result, the structure and function of cellular proteins may be compromised, having a negative effect on cell health. Cell health was measured by dividing the rate of growth (instantaneous doubling time) of a system reassigning a targeted codon to tyrosine to a non-reassigning control [100]. In all reassignment systems the cell health decreased, with the most significant effects observed for the improved Lys AAG and the His CAU reassignments to tyrosine. However, despite significant levels of sense codon reassignment to tyrosine and reduced cell health (slower instantaneous doubling times) the cells still reached a maximum density only slightly lower than a non-reassigning control.

The first study of sense codon reassignment efficiency with the GFP-based screen also identified an unexpected consequence of sense codon reassignment: potential modifications of orthogonal tRNAs by endogenous *E. coli* enzymes. During the evaluation of sense codon reassignment, the engineered *M. jannaschii* $tRNA_{AUG}$ was

able to reassign the His CAU codon with 4% efficiency; however, it was also able to reassign the non-targeted His codon CAC with 3.2% efficiency [100]. None of the other orthogonal engineered *M. jannaschii* tRNAs anticodons measured (Phe UUU, Lys AAG, Asn AAU) showed nearly equal reassignment efficiencies between the targed codon and other closely related non-targeted codons. Fisk and co-workers hypothesized that the first position of the anticodon in the orthogonal *M. jannaschii* $tRNA_{AUG}$ was being modified to inosine, resulting in nearly equal decoding of the CAU and CAC codons [120]. The modification of the first anticodon nucleotide adenine to inosine was proposed because the anticodon stem loop of the orthogonal *M. jannaschii* $tRNA_{AUG}$ was nearly identical to the anticodon stem loop of *E. coli* $tRNA^{Arg2}_{ACG}$, which is known to be modified by the enzyme TadA [121]. Tad A uses the tRNA anticodon stem loop as an important element for substrate recognition, and modifies the first anticodon nucleotide of $tRNA^{Arg2}_{ACG}$ to inosine [121]. Sanger sequencing of cDNA products replicated from the *in vivo* tRNA transcripts of the *M. jannaschii* $tRNA_{AUG}$ was performed and demonstrated that the first anticodon nucleotide was modified to inosine [120].

It was then hypothesized that altering the nucleotides in the anticodon stem loop of the *M. jannaschii* $tRNA_{AUG}$ would prevent it from being recognized as a substrate for TadA, and as a result improve the ability of the tRNA to discriminate between the histidine CAU and CAC codons during sense codon reassignment. A library of *M. jannaschii* $tRNA_{AUG}$ anticodon stem loop variants was generated, and using the GFP screen and FACS sorting, $tRNA_{AUG}$ variants were identified that improved the ability of the *M. jannaschii* $tRNA_{AUG}$ to discriminate between the targeted CAU codon and the non-targeted CAC codon, suggesting that the variants were no longer substrates for

TadA. Furthermore, the identified *M. jannaschii* variants also showed improved reassignment efficiency of the CAU codon [120]. Modification of the orthogonal *M. jannaschii* $tRNA_{AUG}$ by TadA demonstrates that interactions between the host tRNA-modifying enzymes and an orthogonal tRNA may significantly impact sense codon reassignment.

Identifying additional sense codons that have high potential for reassignment is a challenging task because the relative importance of the factors involved in translational fidelity is not well known. Very recently, the sense codon reassignment efficiencies for 20 of the 21 codons read through wobble interactions were measured using the engineered *M. jannaschii* tRNA/aaRS pair and the GFP-based screen [52]. In addition to identifying potential sense codons that could be exploited for genetic code expansion, measurement of the a large subset of sense codons has the potential to shed light on the relative importance of the various factors involved in translational fidelity. Significantly, every codon measured was reassigned to tyrosine with measurable efficiency, ranging from 0.8% (Val GUU) to 41% (Arg CGA) by just providing the orthogonal tRNA with an anticodon that Watson-Crick base paired with the desired codon [52]. Low levels of sense codon reassignment may be improved through directed evolution, and codons that were reassigned at high levels may potentially be directly used for reassignment to a ncAA.

A weak correlation between sense codon reassignment efficiencies and aminoacylation efficiencies of the orthogonal *M. jannaschii* tRNA/aaRS pair was observed, suggesting that while aminoacylation efficiency is likely a contributing factor to sense codon reassignment, it may not be the dominant one [52]. The observed sense

codon reassignment efficiencies are also weakly correlated to the predicted number of competing *E. coli* tRNAs, suggesting that in general orthogonal tRNAs that compete with more abundant *E. coli* tRNAs have lower sense codon reassignment efficiencies.

The observed sense codon reassignment efficiencies corresponded to an estimated 10^5-10^6 substitutions of amino acids in the proteome. Despite a significant number of estimated substitutions in the proteome, the health of the cells during reassignment (calculated as the doubling time of a reassigning system divided by the doubling time of a 100%, non-reassigning system) ranged from 56-97% of the non-reassigning control. The evaluation of cell health during reassignment and the number of estimated substitutions suggests that *E. coli* are broadly tolerable to sense codon reassignment [52].

Measurement of sense codon reassignment efficiencies using the *M. jannaschii* tRNA/aaRS pair identified promising codons for reassignment to ncAAs, and investigated the relative importance of the main factors involved in translational fidelity. For example, the Arg CGA codon was reassigned to tyrosine with 41% efficiency by simply providing an orthogonal tRNA that has an anticodon capable of Watson-Crick base-pairing with the targeted codon. Despite competing against one of the most abundant tRNAs in *E. coli*, the favorable energetics of the orthogonal tRNA anticodon and targeted codon biased incorporation towards tyrosine [52]. The following chapters in this dissertation provide additional sense codon reassignment data to identify additional targets for sense codon reassignment to ncAAs and improve the understanding of the relative importance of various factors involved in translational fidelity.

REFERENCES

1. Berg, J.M., J.L. Tymoczko, and L. Stryer, *Biochemistry*. 5th ed. ed. 2002, New York :: W.H. Freeman.
2. Dobson, C.M., *Protein folding and misfolding.* Nature, 2003. **426**(6968): p. 884.
3. Gupta, B., T.S. Levchenko, and V.P. Torchilin, *Intracellular delivery of large molecules and small particles by cell-penetrating proteins and peptides.* Advanced drug delivery reviews, 2005. **57**(4): p. 637-651.
4. Singer, S.J. and G.L. Nicolson, *The fluid mosaic model of the structure of cell membranes.* Science, 1972. **175**(4023): p. 720-731.
5. Bailey, J.E., *Toward a science of metabolic engineering.* Science, 1991. **252**(5013): p. 1668-1675.
6. Creighton, T.E., *Proteins : structures and molecular properties*. 2013, New York: Freeman.
7. Templin, M.F., et al., *Protein microarrays: promising tools for proteomic research.* Proteomics, 2003. **3**(11): p. 2155-2166.
8. Avouac, J., L. Gossec, and M. Dougados, *Diagnostic and predictive value of anti-cyclic citrullinated protein antibodies in rheumatoid arthritis: a systematic literature review.* Annals of the rheumatic diseases, 2006. **65**(7): p. 845-851.
9. Levitzki, A., *Protein tyrosine kinase inhibitors as novel therapeutic agents.* Pharmacology & therapeutics, 1999. **82**(2-3): p. 231-239.

10. Nangia-Makker, P., et al., *Carbohydrate-binding proteins in cancer, and their ligands as therapeutic agents.* Trends in Molecular Medicine, 2002. **8**(4): p. 187-192.

11. Borrebaeck, C.A., *Precision diagnostics: moving towards protein biomarker signatures of clinical utility in cancer.* Nature Reviews Cancer, 2017. **17**(3): p. 199.

12. Leader, B., Q.J. Baca, and D.E. Golan, *Protein therapeutics: a summary and pharmacological classification.* Nature reviews Drug discovery, 2008. **7**(1): p. 21.

13. Ecker, D.M., S.D. Jones, and H.L. Levine, *The therapeutic monoclonal antibody market.* mAbs, 2015. **7**(1): p. 9-14.

14. Reichert, J.M., *Antibodies to watch in 2017.* mAbs, 2017. **9**(2): p. 167-181.

15. Crick, F., *Central dogma of molecular biology.* Nature, 1970. **227**(5258): p. 561.

16. Watson, J.D. and F.H. Crick, *Molecular structure of nucleic acids.* Nature, 1953. **171**(4356): p. 737-738.

17. Boland, T. and B. Ratner, *Direct measurement of hydrogen bonding in DNA nucleotide bases by atomic force microscopy.* Proceedings of the National Academy of Sciences, 1995. **92**(12): p. 5297-5301.

18. Šponer, J., J. Leszczynski, and P. Hobza, *Electronic properties, hydrogen bonding, stacking, and cation binding of DNA and RNA bases.* Biopolymers, 2001. **61**(1): p. 3-31.

19. Meroueh, M. and C.S. Chow, *Thermodynamics of RNA hairpins containing single internal mismatches.* Nucleic acids research, 1999. **27**(4): p. 1118-1125.

20. Nirenberg, M.W., *The genetic code.* Scientific American, 1963. **208**(3): p. 80-95.

21. Schluenzen, F., et al., *Structure of functionally activated small ribosomal subunit at 3.3 Å resolution.* Cell, 2000. **102**(5): p. 615-623.

22. Gromadski, K.B., T. Daviter, and M.V. Rodnina, *A uniform response to mismatches in codon-anticodon complexes ensures ribosomal fidelity.* Molecular cell, 2006. **21**(3): p. 369-377.

23. Kozak, M., *Initiation of translation in prokaryotes and eukaryotes.* Gene, 1999. **234**(2): p. 187-208.

24. Laursen, B.S., et al., *Initiation of protein synthesis in bacteria.* Microbiology and molecular biology reviews, 2005. **69**(1): p. 101-123.

25. Agirrezabala, X. and J. Frank, *Elongation in translation as a dynamic interaction among the ribosome, tRNA, and elongation factors EF-G and EF-Tu.* Quarterly reviews of biophysics, 2009. **42**(3): p. 159-200.

26. Geggier, P., et al., *Conformational sampling of aminoacyl-tRNA during selection on the bacterial ribosome.* Journal of molecular biology, 2010. **399**(4): p. 576-595.

27. Whitford, P.C., et al., *Accommodation of aminoacyl-tRNA into the ribosome involves reversible excursions along multiple pathways.* Rna, 2010. **16**(6): p. 1196-1204.

28. Chen, J., et al., *Unraveling the dynamics of ribosome translocation.* Current opinion in structural biology, 2012. **22**(6): p. 804-814.

29. Ban, N., et al., *The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution.* Science, 2000. **289**(5481): p. 905-920.

30. LaRiviere, F.J., A.D. Wolfson, and O.C. Uhlenbeck, *Uniform binding of aminoacyl-tRNAs to elongation factor Tu by thermodynamic compensation.* Science, 2001. **294**(5540): p. 165-8.

31. Weijland, A., et al., *Elongation factor Tu: a molecular switch in protein biosynthesis.* Molecular microbiology, 1992. **6**(6): p. 683-688.

32. Schuette, J.C., et al., *GTPase activation of elongation factor EF-Tu by the ribosome during decoding.* The EMBO journal, 2009. **28**(6): p. 755-765.

33. Hopfield, J.J., *KINETIC PROOFREADING - NEW MECHANISM FOR REDUCING ERRORS IN BIOSYNTHETIC PROCESSES REQUIRING HIGH SPECIFICITY.* Proceedings of the National Academy of Sciences of the United States of America, 1974. **71**(10): p. 4135-4139.

34. Ninio, J., *KINETIC AMPLIFICATION OF ENZYME DISCRIMINATION.* Biochimie, 1975. **57**(5): p. 587-595.

35. Chan, P.P. and T.M. Lowe, *GtRNAdb: a database of transfer RNA genes detected in genomic sequence.* Nucleic acids research, 2008. **37**(suppl_1): p. D93-D97.

36. Söll, D., *TRNA : structure, biosynthesis, and function*. 1995, Washington, DC: ASM Press.

37. Kobayashi, T., et al., *Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion.* Nature Structural and Molecular Biology, 2003. **10**(6): p. 425.

38. Dong, H., L. Nilsson, and C.G. Kurland, *Co-variation of trna abundance and codon usage inescherichia coliat different growth rates.* Journal of molecular biology, 1996. **260**(5): p. 649-663.

39. Crick, F.H., *Codon—anticodon pairing: the wobble hypothesis.* Journal of molecular biology, 1966. **19**(2): p. 548-555.

40. Nasvall, S.J., P. Chen, and G.R. Bjork, *The wobble hypothesis revisited: uridine-5-oxyacetic acid is critical for reading of G-ending codons.* RNA, 2007. **13**(12): p. 2151-64.

41. Noguchi, S., et al., *Isolation and characterization of an Escherichia coli mutant lacking tRNA-guanine transglycosylase. Function and biosynthesis of queuosine in tRNA.* Journal of Biological Chemistry, 1982. **257**(11): p. 6544-6550.

42. Kung, F.-L., S. NONEKOWSKI, and G.A. GARCIA, *tRNA–guanine transglycosylase from Escherichia coli: Recognition of noncognate–cognate chimeric tRNA and discovery of a novel recognition site within the TΨC arm of tRNA Phe.* Rna, 2000. **6**(2): p. 233-244.

43. Curran, J.F., *Decoding with the A: I wobble pair is inefficient.* Nucleic acids research, 1995. **23**(4): p. 683-688.

44. Ibba, M., C. Francklyn, and S. Cusack, *The aminoacyl-tRNA synthetases*. 2005, Landes Bioscience :: Georgetown, Tex. :.

45. de Duve, C., *The second genetic code.* Nature, 1988. **333**(6169): p. 117-118.

46. McClain, W.H., *Rules that govern tRNA identity in protein synthesis.* Journal of molecular biology, 1993. **234**(2): p. 257-280.

47. Giegé, R., M. Sissler, and C. Florentz, *Universal rules and idiosyncratic features in tRNA identity.* Nucleic acids research, 1998. **26**(22): p. 5017-5035.

48. Giegé, R., et al., *Crystallogenesis Trends of Free and Liganded Aminoacyl-tRNA Synthetases†.* Crystal Growth & Design, 2008. **8**(12): p. 4297-4306.

49. Wohlgemuth, I., et al., *Evolutionary optimization of speed and accuracy of decoding on the ribosome.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2011. **366**(1580): p. 2979-2986.

50. Gromadski, K.B. and M.V. Rodnina, *Kinetic determinants of high-fidelity tRNA discrimination on the ribosome.* Molecular cell, 2004. **13**(2): p. 191-200.

51. Rodnina, M.V. and W. Wintermeyer, *Fidelity of aminoacyl-tRNA selection on the ribosome: kinetic and structural mechanisms.* Annual review of biochemistry, 2001. **70**(1): p. 415-435.

52. Schmitt, M.A., Biddle, Wil, and John D. Fisk, *Mapping the Plasticity of the E. Coli Genetic Code with Orthgonal Pair Directed Sense Codon Reassignment.* submitted, 2017.

53. Fluitt, A., E. Pienaar, and H. Viljoen, *Ribosome kinetics and aa-tRNA competition determine rate and fidelity of peptide synthesis.* Comput Biol Chem, 2007. **31**(5-6): p. 335-46.

54. Phelps, S.S., O. Jerinic, and S. Joseph, *Universally conserved interactions between the ribosome and the anticodon stem-loop of A site tRNA important for translocation.* Molecular cell, 2002. **10**(4): p. 799-807.

55. Olejniczak, M., et al., *Idiosyncratic tuning of tRNAs to achieve uniform ribosome binding.* Nature Structural and Molecular Biology, 2005. **12**(9): p. 788.

56. Liu, C.-Y., M.T. Qureshi, and T.-H. Lee, *Interaction strengths between the ribosome and tRNA at various steps of translocation.* Biophysical journal, 2011. **100**(9): p. 2201-2208.

57. Thompson, R.C. and P.J. Stone, *Proofreading of the codon-anticodon interaction on ribosomes.* Proceedings of the National Academy of Sciences, 1977. **74**(1): p. 198-202.

58. Liu, C.C. and P.G. Schultz, *Adding new chemistries to the genetic code.* Annual review of biochemistry, 2010. **79**: p. 413-444.

59. Dumas, A., et al., *Designing logical codon reassignment–Expanding the chemistry in biology.* Chemical science, 2015. **6**(1): p. 50-69.

60. Budisa, N., *Engineering the genetic code : expanding the amino acid repertoire for the design of novel proteins*. 2006, Weinheim :: Wiley-VCH.

61. Davis, L. and J.W. Chin, *Designer proteins: applications of genetic code expansion in cell biology.* Nature Reviews Molecular Cell Biology, 2012. **13**(3): p. 168-182.

62. Wang, Q., A.R. Parrish, and L. Wang, *Expanding the genetic code for biological studies.* Chemistry & biology, 2009. **16**(3): p. 323-336.

63. Cowie, D.B. and G.N. Cohen, *BIOSYNTHESIS BY ESCHERICHIA-COLI OF ACTIVE ALTERED PROTEINS CONTAINING SELENIUM INSTEAD OF SULFUR.* Biochimica Et Biophysica Acta, 1957. **26**(2): p. 252-261.

64. Wang, L., et al., *Unnatural amino acid mutagenesis of green fluorescent protein.* The Journal of organic chemistry, 2003. **68**(1): p. 174-176.

65. Xie, J., et al., *The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination.* Nature biotechnology, 2004. **22**(10): p. 1297.

66. Turner, J.M., et al., *Structural plasticity of an aminoacyl-tRNA synthetase active site.* Proceedings of the National Academy of Sciences, 2006. **103**(17): p. 6483-6488.

67. Neumann, H., et al., *Genetically encoding protein oxidative damage.* Journal of the American Chemical Society, 2008. **130**(12): p. 4028-4033.

68. Liu, C.C. and P.G. Schultz, *Recombinant expression of selectively sulfated proteins in Escherichia coli.* Nature biotechnology, 2006. **24**(11): p. 1436.

69. Arbely, E., et al., *Photocontrol of tyrosine phosphorylation in mammalian cells via genetic encoding of photocaged tyrosine.* Journal of the American Chemical Society, 2012. **134**(29): p. 11912-11915.

70. Chin, J.W., et al., *Addition of a photocrosslinking amino acid to the genetic code of Escherichia coli.* Proceedings of the National Academy of Sciences, 2002. **99**(17): p. 11020-11024.

71. Chin, J.W., et al., *Addition of p-Azido-l-phenylalanine to the Genetic Code of Escherichia c oli.* Journal of the American Chemical Society, 2002. **124**(31): p. 9026-9027.

72. Chalker, J.M., G.a.J. Bernardes, and B.G. Davis, *A "tag-and-modify" approach to site-selective protein modification.* Accounts of chemical research, 2011. **44**(9): p. 730-741.

73. van Hest, J. and D.A. Tirrell, *Efficient introduction of alkene functionality into proteins in vivo.* FEBS letters, 1998. **428**(1-2): p. 68-70.

74. Wang, L., et al., *Expanding the genetic code of Escherichia coli.* Science, 2001. **292**(5516): p. 498-500.

75. Furter, R., *Expansion of the genetic code: site-directed p-fluoro-phenylalanine incorporation in Escherichia coli.* Protein Science, 1998. **7**(2): p. 419-426.

76. Tang, Y., et al., *Fluorinated coiled-coil proteins prepared in vivo display enhanced thermal and chemical stability.* Angewandte Chemie, 2001. **113**(8): p. 1542-1544.

77. Yoshikawa, E., et al., *Genetically engineered fluoropolymers. Synthesis of repetitive polypeptides containing p-fluorophenylalanine residues.* Macromolecules, 1994. **27**(19): p. 5471-5475.

78. Link, A.J., M.L. Mock, and D.A. Tirrell, *Non-canonical amino acids in protein engineering.* Current opinion in biotechnology, 2003. **14**(6): p. 603-609.

79. Lin, M.T., et al., *Escherichia coli auxotroph host strains for amino acid-selective isotope labeling of recombinant proteins*, in *Methods in enzymology*. 2015, Elsevier. p. 45-66.

80. Chin, J.W., et al., *An expanded eukaryotic genetic code.* Science, 2003. **301**(5635): p. 964-967.

81. Gorini, L., *Informational suppression.* Annual review of genetics, 1970. **4**(1): p. 107-134.

82. Steege, D.A. and D.G. Söll, *Suppression*, in *Biological regulation and development*. 1979, Springer. p. 433-485.

83. Hayashi, K., et al., *Highly accurate genome sequences of Escherichia coli K-12 strains MG1655 and W3110.* Molecular systems biology, 2006. **2**(1).

84. Lajoie, M.J., et al., *Genomically recoded organisms expand biological functions.* science, 2013. **342**(6156): p. 357-360.

85. Srinivasan, G., C.M. James, and J.A. Krzycki, *Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA.* Science, 2002. **296**(5572): p. 1459-1462.

86. Wang, L., et al., *A New Functional Suppressor tRNA/Aminoacyl− tRNA Synthetase Pair for the in Vivo Incorporation of Unnatural Amino Acids into Proteins.* Journal of the American Chemical Society, 2000. **122**(20): p. 5010-5011.

87. Hoesl, M.G. and N. Budisa, *Recent advances in genetic code engineering in Escherichia coli.* Current opinion in biotechnology, 2012. **23**(5): p. 751-757.

88. Wan, W., J.M. Tharp, and W.R. Liu, *Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool.* Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 2014. **1844**(6): p. 1059-1070.

89. Pastrnak, M., T.J. Magliery, and P.G. Schultz, *A new orthogonal suppressor tRNA/aminoacyl-tRNA synthetase pair for evolving an organism with an expanded genetic code.* Helvetica Chimica Acta, 2000. **83**(9): p. 2277-2286.

90. Pédelacq, J.-D., et al., *Engineering and characterization of a superfolder green fluorescent protein.* Nature biotechnology, 2006. **24**(1): p. 79.

91. Pe, J.-D., et al., *Corrigendum: Engineering and characterization of a superfolder green fluorescent protein.* Nature Biotechnology, 2006. **24**(9): p. 1170.

92. Graille, M., et al., *Molecular basis for bacterial class I release factor methylation by PrmC.* Molecular cell, 2005. **20**(6): p. 917-927.

93. Johnson, D.B., et al., *RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites.* Nature chemical biology, 2011. **7**(11).

94. Mukai, T., et al., *Codon reassignment in the Escherichia coli genetic code.* Nucleic acids research, 2010. **38**(22): p. 8188-8195.

95. Young, T.S., et al., *An Enhanced System for Unnatural Amino Acid Mutagenesis in E. coli.* Journal of Molecular Biology, 2010. **395**(2): p. 361-374.

96. Johnson, D.B., et al., *Release factor one is nonessential in Escherichia coli.* ACS chemical biology, 2012. **7**(8): p. 1337-1344.

97. Mukai, T., et al., *Reassignment of a rare sense codon to a non-canonical amino acid in Escherichia coli.* Nucleic acids research, 2015. **43**(16): p. 8111-8122.

98. Ostrov, N., et al., *Design, synthesis, and testing toward a 57-codon genome.* Science, 2016. **353**(6301): p. 819-822.

99. Kwon, I., K. Kirshenbaum, and D.A. Tirrell, *Breaking the degeneracy of the genetic code.* Journal of the American Chemical Society, 2003. **125**(25): p. 7512-7513.

100. Biddle, W., M.A. Schmitt, and J.D. Fisk, *Evaluating Sense Codon Reassignment with a Simple Fluorescence Screen.* Biochemistry, 2015. **54**(50): p. 7355-64.

101. Lee, B.S., et al., *Incorporation of Unnatural Amino Acids in Response to the AGG Codon.* ACS Chem Biol, 2015. **10**(7): p. 1648-53.

102. Dunin-Horkawicz, S., et al., *MODOMICS: a database of RNA modification pathways.* Nucleic acids research, 2006. **34**(suppl_1): p. D145-D149.

103. OpenWetWare. *Escherichia coli/Codon usage*. 2012 [cited 2018 20 January]; Available from: https://openwetware.org/wiki/Escherichia_coli/Codon_usage.

104. Rozov, A., et al., *Novel base-pairing interactions at the tRNA wobble position crucial for accurate reading of the genetic code.* Nature communications, 2016. **7**: p. 10457.

105. Bohlke, N. and N. Budisa, *Sense codon emancipation for proteome-wide incorporation of noncanonical amino acids: rare isoleucine codon AUA as a target for genetic code expansion.* FEMS microbiology letters, 2014. **351**(2): p. 133-144.

106. Dong, H.J., L. Nilsson, and C.G. Kurland, *Co-variation of tRNA abundance and codon usage in Escherichia coli at different growth rates.* Journal of Molecular Biology, 1996. **260**(5): p. 649-663.

107. Zeng, Y., W. Wang, and W.S.R. Liu, *Towards Reassigning the Rare AGG Codon in Escherichia coli.* Chembiochem, 2014. **15**(12): p. 1750-1754.

108. Biddle, W., Schwark, David G., Schmitt, Margaret A. and John D. Fisk, *Sense Codon Reassignment of the Rare Arginine AGG Codon: Transferability of Improvements Between Orthogonal tRNAs and Synthetases.* submitted, 2018.

109. Shimizu, Y., T. Kanamori, and T. Ueda, *Protein synthesis by pure translation systems.* Methods, 2005. **36**(3): p. 299-304.

110. Carlson, E.D., et al., *Cell-free protein synthesis: applications come of age.* Biotechnology advances, 2012. **30**(5): p. 1185-1194.

111. Russel, M. and P. Model, *Genetic analysis of the filamentous bacteriophage packaging signal and of the proteins that interact with it.* Journal of virology, 1989. **63**(8): p. 3284-3295.

112. Rakonjac, J., J.-n. Feng, and P. Model, *Filamentous phage are released from the bacterial membrane by a two-step mechanism involving a short C-terminal fragment of pIII1.* Journal of molecular biology, 1999. **289**(5): p. 1253-1265.

113. Palomo, J.M., *Solid-phase peptide synthesis: an overview focused on the preparation of biologically relevant peptides.* RSC Advances, 2014. **4**(62): p. 32658-32672.

114. Neckers, D.C., *Solid phase synthesis.* Journal of chemical education, 1975. **52**(11): p. 695.

115. Dawson, P.E. and S.B. Kent, *Synthesis of native proteins by chemical ligation.* Annual review of biochemistry, 2000. **69**(1): p. 923-960.

116. Tsien, R.Y., *The green fluorescent protein*. 1998, Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA.

117. Heim, R. and R.Y. Tsien, *Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer.* Current biology, 1996. **6**(2): p. 178-182.

118. Guo, J., et al., *Evolution of amber suppressor tRNAs for efficient bacterial production of proteins containing nonnatural amino acids.* Angewandte Chemie International Edition, 2009. **48**(48): p. 9148-9151.

119. Wang, N., et al., *Fine-tuning interaction between aminoacyl-tRNA synthetase and tRNA for efficient synthesis of proteins containing unnatural amino acids.* ACS synthetic biology, 2014. **4**(3): p. 207-212.

120. Biddle, W., M.A. Schmitt, and J.D. Fisk, *Modification of orthogonal tRNAs: unexpected consequences for sense codon reassignment.* Nucleic acids research, 2016: p. gkw948.

121. Wolf, J., A.P. Gerber, and W. Keller, *tadA, an essential tRNA-specific adenosine deaminase from Escherichia coli.* The EMBO journal, 2002. **21**(14): p. 3841-3851.

CHAPTER 2

THE EFFECT OF tRNA COMPETITION AND CODON USAGE ON SENSE CODON REASSIGNMENT

2.1 CHAPTER OVERVIEW

The extent to which the degeneracy of the genetic code can be broken is not obvious, and the identification of codons that can be most productively reassigned to an alternate amino acid is challenging. Even after 50 years of intensive study, the relative importance of the various factors that determine the fidelity of translation are largely unknown. This chapter describes the measurement of sense codon reassignment efficiencies to tyrosine for rarely used sense codons in *E. coli* using the orthogonal *Methanocaldococcus* (*M. jannaschii*) tRNA/aaRS pair. Sense codon reassignment was quantified *in vivo* using a previously developed fluorescence-based screen involving gain of GFP flourescence through missense reading. The quantitative measurements of sense codon reassignment efficiencies at 10 of the least frequently used codons augments the previous 20 measurements of sense codon reassignments of *E. coli* wobble codons.

Significantly, every rare codon was partially reassigned to tyrosine with efficiencies ranging from 3.5% to 70%. Furthermore, sense codon reassignment efficiencies were higher for rarely used codons than reassignment efficiencies at wobble codons. In addition, substituting the encoded amino acid for tyrosine at codons throughout the proteome was well-tolerated; only two reassignment systems, Ile AUA

and Leu CUA, caused doubling times to increase more than 2 standard deviations above a non-reassigning, wild-type GFP expression control despite estimates of 100,000s to 1,000,000s of substitutions across the *E. coli* proteome. Analysis of wobble and rare codon data sets together show a weak correlation between sense codon reassignment efficiency to tyrosine and either competing endogenous *E. coli* tRNA abundance or predicted-fold reduction in aminoacylation efficiency. In contrast to the previous data analysis of the wobble codon only data set, analysis of the wobble codon and rare codon data sets together suggests that endogenous *E. coli* tRNA abundance may be more influential than aminoacylation efficiency on sense codon reassignment. Additional sense codon reassignment efficiency data from another orthogonal pair may further assist in determining the relative importance of factors involved in protein translation.

2.2 INTRODUCTION

Protein functions are at the heart of nearly all biological processes, and although sufficient for living systems, the 20 canonical amino acids encode a limited set of chemical functionality. Biological incorporation of non-canonical amino acids (ncAAs) with side chains containing expanded and diverse chemical properties has emerged as a powerful way to modify and extend the properties of proteins. Proteins containing ncAAs have found applications in diverse fields from materials science to biomedicine [1]. To date, much of the focus of genetic code expansion has been on the repurposing of the least commonly used codon in the *E. coli* genome, the amber stop codon UAG, as a sense codon for ncAA incorporation [1, 2]. As 18 of 20 canonical amino acids are encoded by more than one sense codon, breaking the degeneracy of the genetic code

by reassigning the meaning of sense codons has also emerged as a promising strategy for genetic code expansion. Recent reports have targeted both *E. coli* wobble codons and codons that are used infrequently in the *E. coli* genome [3-8].

One of the challenges of genetic code expansion is the fact that protein translation is one of the central processes of life. Nature expends a great deal of energy to maintain the fidelity of translation: 75% of a bacterial cell's energy budget is spent in the production of proteins, and much of that energy is devoted to proofreading at multiple steps in the translation process [9]. The interactions between tRNA and aaRS molecules that drive the fidelity of the genetic code have only partially been mapped, and the space between these interactions, the extent to which additional orthogonal tRNA/aaRS pairs can be added, is largely unknown. The orthogonal translation components are required to interact with dozens of proteins in a balanced way to complete the process of aminoacylating a ncAA onto the tRNA, directing the tRNA to the ribosome for participation in a peptide bond forming reaction, releasing the tRNA, and allowing another aminoacylated tRNA to attach the subsequent amino acid to the growing peptide chain [10]. In the case of amber stop codon suppression, the orthogonal translation machinery competes with the ribosomal release factor (RF1) that recognizes the codon as a stop signal [11]. In the case of sense codon reassignment, the orthogonal translation machinery competes with one (or more) endogenous tRNAs to decode the targeted codon. The points of tension and cases of failure in infiltrating the translation process and reassigning the meaning of sense codons are unclear because the relative importance of the various factors that affect the fidelity of translation is largely unknown [7, 12, 13].

This report experimentally explores the plasticity of the *Escherichia coli* (*E. coli*) genetic code by quantitatively measuring the extent to which the addition of an orthogonal tRNA/aaRS pair can functionally modify protein translation [6]. We have repurposed the *Methanocaldococcus jannaschii* (*M. jannaschii*) tyrosine tRNA/aaRS pair, variants of which are widely employed to insert ncAAs into proteins in response to stop codons, to insert tyrosine in response to several rarely used *E. coli* sense codons. The set of 10 quantitative measurements of sense codon reassignment described here augments the quantitative measurements we previously reported for the reassignment of 20 codons read via wobble interactions in *E. coli* [7]. Together, this suite of data improves the systems level understanding of the space of interactions that determines the fidelity of the code, enables measurements of the fitness penalties imposed by redefining the genetic code in *E. coli*, and provides a map of the promising sense codon targets for genetic code expansion using the *M. jannaschii* orthogonal pair.

The extent to which the *M. jannaschii* orthogonal pair can be used to infiltrate the genetic code of *E. coli* at wobble codons provides quantitative measurements of the tolerance of *E. coli* to genetic code modifications and has practical implications beyond genetic code expansion [14, 15]. One potential advantage of orthogonal pair-directed sense codon reassignment over methods that allow conditional rewriting of the genetic code through either editing-defective aaRSs or residue specific reassignment strategies is that specific *codons* as opposed to *amino acids* may be targeted for reassignment.

Generating a quantitative understanding of the system level effects of altering the reading of individual sense codons, including the extent to which particular sense codons are naturally subject to mistranslation, will enhance the comprehension of the

molecular mechanisms of certain cancers and translation-related diseases [16-18]. For example, orthogonal pair directed sense codon reassignment may be used to further investigate how mis-folding of proteins due to amino acid substitution relate to neurodegenerative disease [16]. In addition, sense codon reassignment could be used to further investigate tRNA-ribosome interactions that have been previously shown to be critical for translational fidelity, or how tRNA abundance and modifications are related to stress signaling and the onset of a number of complex human diseases [17, 18].

2.3 MATERIALS AND METHODS

The *in vivo* fluorescence-based screen used for sense codon reassignment has been previously described [6]. Detailed experimental protocols, including methods for construction of *M. jannaschii* tRNA variants, measurement of system growth and fluorescence, and calculation of codon reassignment efficiencies are included in Appendix 1. The appendix also includes full vector sequences, oligonucleotide primer sequences, detailed cell strain information, and general reagents and materials.

2.4 RESULTS AND DISCUSSION

Reassignment of rarely used sense codons is an attractive strategy for genetic code expansion because it exploits the fact that decoding any mRNA codon requires sampling of tRNAs by the ribosome. Rarely used codons are typically decoded by less abundant tRNAs, which is expected to result in decreased competition between the endogenous tRNA and the introduced orthogonal tRNA for the codon to be reassigned. Additionally, high level substitution at rarely used codons is expected to limit the proteome-wide effects of reassignment.

In order to selectively reassign a codon *in vivo*, an orthogonal tRNA/aminoacyl tRNA synthetase (aaRS) pair is introduced. The orthogonal tRNA is not recognized by the set of *E. coli* aaRSs, and the orthogonal aaRS does not recognize and aminoacylate any of the *E. coli* tRNAs. The vast majority of ncAAs have been introduced into proteins using derivatives of two tRNA/aaRS pairs: the tyrosine tRNA/aaRS pair from *Methanocaldococcus jannaschii* (*M. jannaschii*) and the pyrrolysine tRNA/aaRS pair from the *Methanosarcina* species [19-22]. Variants of these pairs that recognize, aminoacylate, and incorporate over 150 different ncAAs have been developed [1, 11, 23].

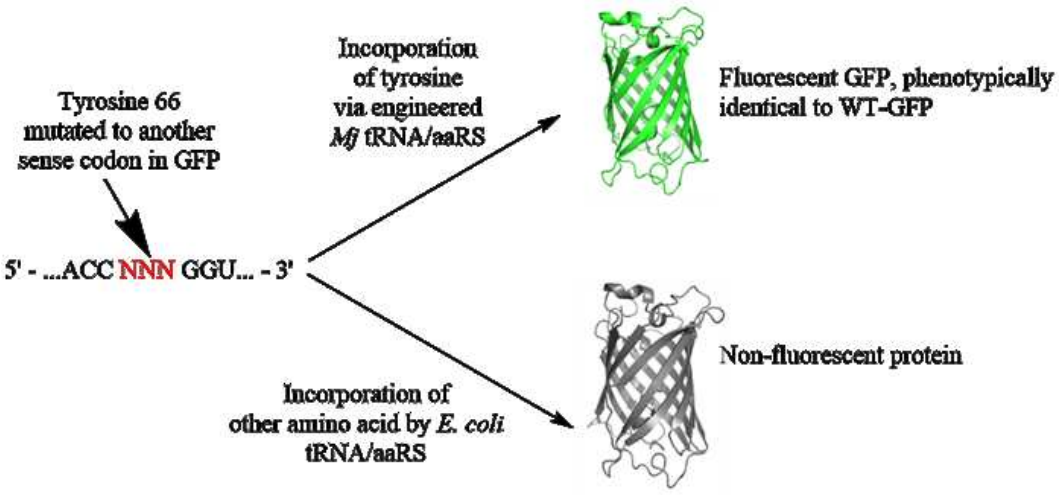

Figure 2.1. System using GFP fluorescence as an indicator of sense codon reassignment efficiency (PDB 2B3P). One mRNA transcript of GFP may be decoded two different ways. In the upper case, the engineered orthogonal *M. jannaschii* tRNA decodes the sense codon and tyrosine is incorporated. In the lower case, the endogenous *E. coli* tRNA decodes the sense codon and incorporates the corresponding amino acid [6, 24].

The efficiency of sense codon reassignment is quantified by measuring the restoration of fluorescence in green fluorescent protein (GFP) reporter variants in which the essential tyrosine residue is encoded by a non-tyrosine codon (Figure 2.1) [6]. Unlike stop codon suppression in which missed incorporations lead to truncated proteins, sense codon reassignment systems produce full-length proteins because incorporations missed by the orthogonal tRNA are read with an endogenous *E. coli* tRNA. The fluorescence-based screen provides a quantitative measure of the sense codon reassignment efficiency at each targeted codon by normalizing the observed GFP fluorescence for a sense codon reassignment measurement to a "100% fluorescence" wild-type GFP with a tyrosine UAC codon at position 66 in the fluorophore. The wild-type GFP control includes a plasmid expressing the *M. jannaschii* aaRS and amber-suppressing tRNA (CUA anticodon) to maintain a metabolic burden on the cells similar to that of the systems that are undergoing sense codon reassignment to tyrosine. The screen integrates the effects of orthogonal *M. jannaschii* tyrosyl aaRS recognition and aminoacylation of the *M. jannaschii* tRNA species with an alternative anticodon and the effects of competition between the altered orthogonal tRNA and *E. coli* tRNA species to decode the codon specifying the essential tyrosine position of GFP.

Rarely used sense codons targeted for reassignment in this evaluation are those codons that are estimated to be translated less frequently than the UAA ochre stop codon over the course of a cell generation. The amber stop codon (UAG) is the least frequently encountered codon in *E. coli* reading frames. An estimated 2.1×10^5 proteins, about 2.1% of the proteins synthesized over a cellular generation, are terminated by

UAG [25, 26]. The UGA (opal) stop codon signals termination for approximately 6.5 times as many proteins over the course of a cell generation. The UAA ochre stop codon signals termination nearly 40 times as frequently as the amber stop codon and terminates approximately 8.4 x 10^6 proteins [25, 26].

Estimates for translation frequency for each codon over the course of a single cell generation were previously calculated and used to estimate the expected numbers of reassignments to tyrosine in the proteome of *E. coli* (Table 2.3) [7, 27-31]. Fisk and co-workers estimated the translation frequency of each codon over the course of a single cell generation using previously published data for the numbers of proteins per cell (5.5 x 10^6), the average size of an *E. coli* protein (300 amino acids), and the codon usage in highly and less expressed genes [7].

Three of the codons that are used less frequently than the ochre stop codon, Cys UGU, Ser AGU, and Arg CGA, are decoded by endogenous tRNAs using non-Watson—Crick base pairing, and their suitability for sense codon reassignment by the *M. jannaschii* tRNA/aaRS pair has been evaluated. Ten additional sense codons belong to the subset of codons translated more frequently than the amber stop codon and less frequently than the ochre stop codon over the course of a cell generation. Each of these codons is decoded by an *E. coli* tRNA with an anticodon capable of Watson—Crick base pairing to the mRNA codon.

tRNA abundance is generally related to codon translation frequency, and most of the rarely used codons are decoded by less abundant *E. coli* tRNAs. However, there are several exceptions to general rule; the *E. coli* tRNA which allows incorporation of histidine in response to CAU and CAC codons is one of the least abundant *E. coli*

tRNAs, but histidine is approximately 200-fold more prevalent in proteins than the ochre stop codon [25, 26]. In contrast, the rarely used arginine CGA codon (1.75 times UAA) is decoded by one of the most abundant *E. coli* tRNAs, $tRNA_{ICG}$ [25, 26]. In this case, the I34/A3 pairing required for the *E. coli* tRNA to translate the CGA codon is energetically-weak, even compared to other wobble interactions [32]. The following sections discuss the full suite of sense codons evaluated for reassignment by anticodon variants of the *M. jannaschii* tyrosine tRNA/aaRS pair with regard to (1) the significance of sense codon reassignment efficiencies, (2) discrimination between targeted and non-targeted codons (3) cell health effects, (4) the predicted loss in aminoacylation efficiency due to changing the identity of the *M. jannaschii* tRNA anticodon, and (5) the number of competing endogenous *E. coli* tRNAs.

Significance of Sense Codon Reassignment Efficiencies

Significantly, every *E. coli* sense codon analyzed may be partially reassigned by supplying an orthogonal tRNA with an anticodon that interacts with the targeted codon through Watson—Crick base pairing (Table 2.1). Reassignment efficiencies of rarely used codons range from 3.5% (8 out of every 1000 incorporations at the targeted codon) to 70% (700 out of every 1000 incorporation events at the targeted codon). Presumably, sense codons that are reassigned to tyrosine with higher efficiency will be more productive for reassignment to ncAAs; however, any sense codon for which tyrosine incorporation is detected is a starting point for improvement by directed evolution. Directed evolution of the *M. jannaschii* tRNA anticodon loop/aaRS anticodon binding domain allowed nearly quantitative reassignment of the AGG codon [8].

In general, the reassignment efficiencies of rarely used, non-wobble codons are higher than the reassignment efficiencies of *E. coli* wobble codons. Ten of the 30 sense codons evaluated are reassigned with efficiencies of nearly 20% or more; eight of the codons in this class are rare codons (Arg CGA is both a wobble codon and a rarely used codon). Only 2 rarely used codons, Gly GGA and Cys UGU (also a wobble codon), are reassigned with efficiencies in the bottom half (Table 2.1). The observation that the orthogonal tRNAs supplied to reassign rarely used codons which compete against an *E. coli* tRNA with an identical anticodon tend to be more readily reassignable than *E. coli* wobble codons suggests that factors such as effective concentration of the orthogonal tRNA relative to the amount of endogenous tRNA competing to decode the targeted codon may be more influential upon sense codon reassignment than the energetics of codon-anticodon interaction. The guiding idea behind targeting wobble codons is that differences in codon-anticodon binding energies between introduced orthogonal tRNAs capable of Watson—Crick base pairing and endogenous tRNAs that utilize wobble base pairing can be harnessed to bias the incorporation of amino acids [3, 7].

Rare Proline CCC and Rare Glycine GGG

Despite multiple attempts, and in contrast to the vast majority of over 30 tRNA anticodon sequences created, the *M. jannaschii* $tRNA^{Opt}_{GGG}$ to reassign Pro CCC could not be synthesized without point mutations. Mutagenesis using an XhoI-containing starting vector with two different primers (the method used to generate every other tRNA utilized, including *M. jannaschii* $tRNA^{Opt}_{AGG}$ to reassign Pro CCU) resulted in vectors with single nucleotide deletions at a variety of positions within the tRNA. Many

Table 2.1. Sense codon reassignment efficiencies to tyrosine in *E. coli* using the *M. jannaschii* orthogonal pair [7].

tRNA Anticodon	Amino acid targeted	Sense codon targeted	Reassignment of targeted codon	Competing *E. coli* tRNAs[a]	*E. coli* Codon Usage[b]	Predicted fold reduction in aminoacylation efficiency[c]
$tRNA^{Opt}_{AAA}$	Phe	Phe TTT	**3.2%**	4440	0.57	864
$tRNA^{Opt}_{AAG}$	Leu	Leu CTT	**4.2%**	7400[a]	0.10	3456
$tRNA^{Opt}_{AAU}$	Ile	Ile ATT	**4.3%**	17390	0.51	1037
$tRNA^{Opt}_{AAC}$	Val	Val GTT	**0.8%**	23680	0.26	3024
$tRNA^{Opt}_{AGA}$	Ser	Ser TCT	**19.6%**	9620	0.15	639
$tRNA^{Opt}_{AGG}$	Pro	Pro CCT	**12.2%**	5920	0.16	2556
$tRNA^{Opt}_{AGU}$	Thr	Thr ACT	**10.5%**	9250	0.17	767
$tRNA^{Opt}_{AGC}$	Ala	Ala GCT	**1.3%**	19240	0.16	2237
$tRNA^{Opt}_{AUG}$	His	His CAT	**6.1%**	2960	0.57	360
$tRNA^{Opt}_{AUU}$	Asn	Asn AAT	**7.5%**	5550	0.45	108
$tRNA^{Opt}_{AUC}$	Asp	Asp GAT	**3.7%**	11100	0.63	315
$tRNA^{Opt}_{ACA}$	Cys	Cys TGT	**3.5%**	6660	0.45	747
$tRNA^{Opt}_{ACU}$	Ser	Ser AGT	**9.1%**	5180	0.15	896
$tRNA^{Opt}_{ACC}$	Gly	Gly GGT	**1.1%**	18500	0.34	2615
$tRNA^{Opt}_{CAC}$	Val	Val GTG	**2.1%**	17760	0.37	3260
$tRNA^{Opt}_{CGC}$	Ala	Ala GCG	**9.2%**	16280	0.36	2410
$tRNA^{Opt}_{CUU}$	Lys	Lys AAG	**7.7%**	8140	0.23	116
$tRNA^{Opt}_{CUC}$	Glu	Glu GAG	**19.8%**	22570	0.31	340
$tRNA^{Opt}_{GCG}$	Arg	Arg CGC	**7.6%**	22200	0.40	33
$tRNA^{Opt}_{UCG}$	Arg	Arg CGA	**40.8%**	22200	0.06	365
$tRNA^{Opt}_{UAG}$	Leu	Leu CUA	**24.3%**	2960	0.04	422
$tRNA^{Opt}_{UAU}$	Ile	Ile AUA	**70.0%**	17390[b]	0.07	127
$tRNA^{Opt}_{UGA}$	Ser	Ser UCA	**24.7%**	6290	0.12	78
$tRNA^{Opt}_{UGU}$	Thr	Thr ACA	**10.9%**	4440	0.13	94
$tRNA^{Opt}_{GCA}$	Cys	Cys UGC	**20.2%**	6660	0.55	8.3
$tRNA^{Opt}_{CCG}$	Arg	Arg CGG	**53.0%**	2220	0.10	3220
$tRNA^{Opt}_{UCU}$	Arg	Arg AGA	**40.7%**	2960	0.04	110
$tRNA^{Opt}_{CCU}$	Arg	Arg AGG	**50.6%**	5190	0.02	966
$tRNA^{Opt}_{UCC}$	Gly	Gly GGA	**4.8%**	10360[a]	0.11	320
$tRNA^{Opt}_{GGG}$	Pro	Pro CCC	**N/A**[d]	3700	0.12	28
$tRNA^{Opt}_{CCC}$	Gly	Gly GGG	**N/A**[d]	10360[a]	0.15	

a The number of competing *E. coli* tRNAs was determined from [33]. For reassignment of Ile AUU and AUA, competing *E. coli* tRNAs are treated collectively because they could not be separated. For reassignment of Gly GGA, competing *E. coli* tRNAs are also treated collectively [33].
b As a fraction of the total number of codons encoding a given canonical amino acid. Codon usage is based on http://openwetware.org/wiki/Escherichia_coli/Codon_usage.
c Relative to the change from GUA (tyrosine) to CUA for amber stop codon suppression. The predicted reduction in aminoacylation efficiency of the M. jannaschii tyrosyl tRNA with a given anticodon is calculated based on an assumption of additivity of the effect of measured, single nucleotide mutations [34].
d No sense codon reassignment efficiency was able to be measured because the orthgonal *M. jannaschii* tRNA could not be constructed.

variants included the appropriate anticodon and anticodon loop, but had single nucleotide mutations/deletions at other positions in the tRNA. tRNA gene assembly with TBIO yielded similar results, as did all DNA isolated from viable clones after inserting a *de novo* synthesized tRNA gene into the vector. Similar difficulty was noted during attempts to synthesize the *M. jannaschii* $tRNA^{Opt}_{CCC}$ to reassign the rarely used Gly GGG codon.

Interestingly, several other *M. jannaschii* $tRNA^{Opt}$ variants that differed at only a single nucleotide relative to either the GGG or CCC anticodon could be generated quickly and easily without point mutations (e.g. $tRNA^{Opt}_{CGC}$, $tRNA^{Opt}_{CCG}$, $tRNA^{Opt}_{GCG}$). High levels of frameshifting leading to fatal translation errors is one possible explanation for the extreme difficulty in generating these two tRNAs and could offer an evolutionary reason why the CCC and GGG codons are rarely used in *E. coli* [35].

Two clones with single nucleotide mutations within the $tRNA^{Opt}_{GGG}$ expression cassette but outside the tRNA itself were isolated from viable colonies. A variant with a point mutation/deletion in the terminator failed to reassign the Pro CCC codon. The limit of detection for the in cell assay is 0.15%. A variant with a single nucleotide deletion in the promoter reassigned Pro CCC at 16.6%. For this reason, the *M. jannaschii* $tRNA^{Opt}_{GGG}$ with a single nucleotide mutation in the *proK* promoter is not included in subsequent analysis of the suite of sense codon reassignment efficiencies by *M. jannaschii* tRNA/aaRS variants.

Discrimination

Ten of 13 rarely used *E. coli* codons are decoded by an endogenous tRNA that Watson—Crick base pairs to the codon. 8 of these 10 tRNAs are responsible for

decoding multiple *E. coli* codons. 12 of 16 codons boxes include an endogenous U34 tRNA. U34 is rarely unmodified, as unmodified uridine is expected to pair with all four nucleotides in the third position of the codon. U34 modifications modulate the base pairing ability of the tRNA. The modification of U34 correlates with whether the tRNA is part of a codon 2-box or a codon 4-box. The 5-methylaminomethyl-2-thiouridine modification at position 34 restricts decoding to the A- and G-ending codons and is found in U34 tRNAs for mostly 2 codon boxes, including lysine and glutamic acid [36, 37]. tRNAs with uridine 5-oxyacetic acid at position 34 expand their decoding to include the U-ending codon [36, 37]. Several U34 tRNAs in 4-codon boxes include this modification and read the U, A and G-ending codons. UAA and UGA are stop codons; no tRNAs with a U34 anticodon exist in either of those codon boxes. The arginine 4-box utilizes an A34 tRNA modified to inosine to decode the U-, C-, and A-ending codons. No U34 tRNA is present in this box in *E. coli*.

As predicted by the wobble rules, most of the introduced U34 tRNAs reassign the G-ending codon in addition to the A-ending codon which they target [38]. Discrimination ratios vary from nearly 50:50 in the case of *M. jannaschii* $tRNA^{Opt}_{UCG}$ for the Arg CGA codon to about 80:20 for *M. jannaschii* $tRNA^{Opt}_{UGA}$, *M. jannaschii* $tRNA^{Opt}_{UAG}$, and *M. jannaschii* $tRNA^{Opt}_{UCU}$ which target the Ser UCA, Leu CTA, and Arg AGA codons respectively (Table 2.2). The better-than-expected discrimination between Ile AUA and Met AUG codons by *M. jannaschii* $tRNA^{Opt}_{UAU}$ will be discussed subsequently in a section on sense codon reassignment of the Ile AUA codon. Given the absence of A34 tRNAs to decode U-ending codons, *E. coli* encodes 15 G34 tRNAs to decode U- and C-ending codons. These codons are typically very commonly used; only 3 C-ending

codons are infrequently used. Not surprisingly, the introduced G34 tRNAs are able to wobble to the U-ending codon in those boxes. *M. jannaschii* $tRNA^{Opt}_{GCA}$ reassigns the Cys UGC codon at 20.2% and the other Cys codon, UGU at 17.3%. *M. jannaschii* $tRNA^{Opt}_{GCG}$ reassigns Arg CGU and Arg CGC nearly equivalently at about 7.5%.

C34 tRNAs are not expected to decode codons other than the G-ending codon for which they are specific. In the analysis of reassignment of all *E. coli* wobble codons, poor discrimination was observed between targeted G- and non-targeted A-ending codons by introduced C34 tRNAs for both alanine and valine [7]. Discrimination between the targeted G- and non-targeted A-ending codons by introduced C34 tRNAs for both lysine and glutamic acid was much better, at least 96:4. Two of the rare arginine codons, AGG and CGG are targeted by introducing a C34 tRNA. In both cases, discrimination between the targeted G- and non-targeted A-ending codons by introduced C34 tRNAs is greater than 95:5. The majority of introduced C34 tRNAs appear to reassign only the targeted G-ending sense codon; the valine and alanine cases remain anomalous. One potential hypothesis for the poorly observed discrimination by the orthogonal C34 Ala and Val *M. jannaschii* tRNAs is the “two out of three hypothesis, which suggests that codons containing strong G/C base pairing interactions between the tRNA anticodon and the codon for the first two positions of the codon may be decoded by base-pairing mismatches at the third position of the codon [39, 40]. However, the good discrimination by the orthogonal *M. jannaschii* $tRNA^{Opt}_{CCG}$ thwarts, to some extent, the 2 out of 3 hypothesis (Table 2.2).

Table 2.2. Discrimination of closely-related codons by introduced orthogonal *M. jannaschii* tRNAs (wobble codons measured previously [7]).

tRNA Anticodon	Amino acid targeted	Sense codon targeted	Nucleotide at Position 34	Reassignment observed at YZU[a]	Reassignment observed at YZC	Reassignment observed at YZA	Reassignment observed at YZG
$tRNA^{Opt}_{AAA}$	Phe	Phe TTT	**A**	***3.2%***	**B.D.**[b]	-----[c]	-----
$tRNA^{Opt}_{AAG}$	Leu	Leu CTT	**A**	***4.2%***	1.3%	B.D.	-----
$tRNA^{Opt}_{AAU}$	Ile	Ile ATT	**A**	***4.3%***	0.4%	B.D.	0.3%
$tRNA^{Opt}_{AAC}$	Val	Val GTT	**A**	***0.8%***	0.6%	B.D.	0.3%
$tRNA^{Opt}_{AGA}$	Ser	Ser TCT	**A**	***19.6%***	1.3%	-----	-----
$tRNA^{Opt}_{AGG}$	Pro	Pro CCT	**A**	***12.2%***	10.2%	0.9%	-----
$tRNA^{Opt}_{AGU}$	Thr	Thr ACT	**A**	***10.5%***	1.3%	-----	-----
$tRNA^{Opt}_{AGC}$	Ala	Ala GCT	**A**	***1.3%***	1.1%	B.D.	B.D.
$tRNA^{Opt}_{AUG}$	His	His CAT	**A**	***6.1%***	2.9%	B.D.	-----
$tRNA^{Opt}_{AUU}$	Asn	Asn AAT	**A**	***7.5%***	B.D.	-----	-----
$tRNA^{Opt}_{AUC}$	Asp	Asp GAT	**A**	***3.7%***	B.D.	B.D.	B.D.
$tRNA^{Opt}_{ACA}$	Cys	Cys TGT	**A**	***3.5%***	B.D.	-----	-----
$tRNA^{Opt}_{ACU}$	Ser	Ser AGT	**A**	***9.1%***	B.D.	-----	-----
$tRNA^{Opt}_{ACC}$	Gly	Gly GGT	**A**	***1.1%***	0.5%	-----	-----
$tRNA^{Opt}_{CAC}$	Val	Val GTG	**C**	0.3%	B.D.	0.4%	***2.1%***
$tRNA^{Opt}_{CGC}$	Ala	Ala GCG	**C**	0.3%	0.8%	1.1%	***9.2%***
$tRNA^{Opt}_{CUU}$	Lys	Lys AAG	**C**	-----	-----	B.D.	***7.7%***
$tRNA^{Opt}_{CUC}$	Glu	Glu GAG	**C**	B.D.	B.D.	0.8%	***19.8%***
$tRNA^{Opt}_{GCG}$	Arg	Arg CGC	**G**	7.5%	***7.6%***	B.D.	B.D.
$tRNA^{Opt}_{UCG}$	Arg	Arg CGA	**U**	0.5%	0.4%	***40.8%***	33.0%
$tRNA^{Opt}_{UAG}$	Leu	Leu CUA	**U**	15.0%	1.1%	***24.3%***	6.5%
$tRNA^{Opt}_{UAU}$	Ile	Ile AUA	**U**	4.4%	0.4%	***70.0%***	3.5%
$tRNA^{Opt}_{UGA}$	Ser	Ser UCA	**U**	-----	-----	***24.7%***	7.1%
$tRNA^{Opt}_{UGU}$	Thr	Thr ACA	**U**	-----	-----	***10.9%***	1.6%
$tRNA^{Opt}_{GCA}$	Cys	Cys UGC	**G**	17.3%	***20.2%***	-----	-----
$tRNA^{Opt}_{CCG}$	Arg	Arg CGG	**C**	BD	BD	2.8%	***53.0%***
$tRNA^{Opt}_{UCU}$	Arg	Arg AGA	**U**	-----	-----	***40.7%***	10.4%
$tRNA^{Opt}_{CCU}$	Arg	Arg AGG	**C**	-----	-----	BD	***50.6%***
$tRNA^{Opt}_{UCC}$	Gly	Gly GGA	**U**	-----	-----	***4.8%***	2.3%
$tRNA^{Opt}_{GGG}$	Pro	Pro CCC	**C**	-----	***-----***	-----	-----
$tRNA^{Opt}_{CCC}$	Gly	Gly GGG	**G**	-----	-----	-----	***-----***

Anticodons with A or G at position 34 were typically evaluated against codons ending in U and C; anticodons with U or C at position 34 were typically evaluated against codons ending in A and G. Reassignment efficiencies emboldened and italicized are those for the codon targeted for reassignment, which fully base pairs to the orthogonal tRNA via Watson-Crick interactions

a For continuity, "Y" and "Z" are used to represent the first two positions of the codon in the column headers. The sense codons evaluated with a single tRNA in each row are determined by substituting the complement of the nucleobases in the second and third anticodon positions for Y and Z. The third codon position is specified in the column headers. For example, row 1 shows sense codon reassignment by the M. barkeri tRNAPylAAA. Data in the YZU column are for reassignment at the UUU codon. Data in the YZC column are for reassignment at the UUC codon.

b B.D. indicates that the codon was evaluated with the specified tRNA, and the measurement was below the detection limit of the in cell assay (0.15%).

c "----" indicates that the codon was not evaluated for reassignment by the specified tRNA.

Case of rare isoleucine

The *M. jannaschii* $tRNA^{Opt}_{UAU}$ reassigns the AUA codon with the highest efficiency of all sense codons evaluated; 70 of 100 incorporations in response to the AUA codon in the GFP reporter are made by the orthogonal tRNA. Significantly, this high level reassignment is facilitated simply by altering the anticodon of the orthogonal tRNA to Watson—Crick base pair with the AUA codon and is achieved in rich media without genomic modifications to the *E. coli* strain. Previous work has shown that directed evolution of the aaRS anticodon binding domain and tRNA anticodon loop is able to improve reassignment of a given codon by the *M. jannaschii* tRNA/aaRS pair [6, 8]. Presumably, a similar approach could be used to increase reassignment of the Ile AUA codon.

The isoleucine/methionine box is the only sense codon box in which the A- and G-ending codons encode different amino acids. Nature has evolved several tactics for minimizing mistranslation of the AUA/AUG codons. In *E. coli*, the Ile AUA codon is read by a tRNA with a modified C34 anticodon. The modification of C34 to lysidine allows the tRNA to read the Ile AUA codon and prevents decoding of the Met AUG codon [41]. Eukaryotes typically utilize an inosine-modified A34 tRNA to read the AUU, AUC, and AUA codons, similar to the way *E. coli* uses inosine to decode the CGU, CGC, and CGA arginine codons [41]. The orthogonal tRNA introduced to target the rare isoleucine AUA codon includes a UAU anticodon. The wobble rules suggest that regardless of modification, a U34 tRNA should read both the A- and G-ending codons. 70 of 100 incorporations in response to the AUA codon in the fluorophore position of GFP are tyrosine as a result of the *M. jannaschii* $tRNA^{Opt}_{UAU}$. In contrast, fewer than 4 of 100

incorporations in response to the AUG codon in the fluorophore position of GFP are tyrosine as a result of the *M. jannaschii* $tRNA^{Opt}_{UAU}$ (Table 2.2). Given the strength of the G/U wobble interaction, one would expect the reassignment efficiencies at the AUA and AUG codons to be more similar. The G/U wobble is borne out in the reassignment of other G-ending codons by U34 tRNAs. One potential hypothesis that would explain the better than expected discrimination by the orthogonal *M. jannaschii* $tRNA^{Opt}_{UAU}$ is the modification of anticodon nucleotide U34 to lysidine. Although the endogenous *E. coli* enzyme normally modifies C to L, a modification of U34 to a lysidine-like nucleotide modification might be possible.

Furthermore, competing tRNA abundance does not seem to be sufficient to explain the high discrimination between the AUA and AUG codons by *M. jannaschii* $tRNA^{Opt}_{UAU}$. Only regular Met tRNAs are to be counted as competing for that AUG codon at the critical fluorophore position in GFP. In that event, we estimate 3,700 *E. coli* tRNA Met competing for the AUG codon [33]. Based on estimates of the number of times a codon is translated over the course of a cellular generation, the AUA codon accounts for approximately 3.1% of the Ile residues incorporated into proteins.

Case of cysteine

Cysteine represents an interesting case study for the factors that affect protein translation. As cysteine is a rarely used amino acid in *E. coli* proteins, both cysteine codons, UCU and UCG are used infrequently. A single tRNA with a GCA anticodon is responsible for decoding both codons, through either a G34/C3 canonical Watson—Crick interaction or a G34/U3 wobble interaction. The cysteine codons are used with approximate frequency throughout the genome: UGU is used 45% of the time; UGC is

used 55% of the time. Propagating codon distribution across highly expressed versus less expressed proteins to estimate the number of times each codon is translated during a single cell generation skews the ratios slightly. UGU is responsible for about 36% of Cys residues, and UGC encodes about 64% of Cys residues incorporated during a single cell generation.

The hypothesis behind targeting *E. coli* wobble codons for breaking the degeneracy of the genetic code is that the introduction of an orthogonal tRNA with an anticodon that Watson—Crick base pairs to the targeted codon would have improved energetics of codon-anticodon interactions, resulting in an improved rate and probability of acceptance on the ribosome. The orthogonal tRNA would then be able to compete favorably against the endogenous tRNA [3, 7]. Introduction of *M. jannaschii* $tRNA^{Opt}_{ACA}$ to target the Cys wobble codon UGU resulted in 3.5% reassignment to tyrosine, one of the lowest observed reassignment efficiencies. Both the estimated number of competing tRNAs and reduction in aminoacylation efficiency as a result of changing the anticodon of the orthogonal tRNA fall in the middle of the range of values predicted. One hypothesis for the low incorporation efficiency is related to the functional significance of cysteine (His also falls into this category). Cysteine is important for protein folding and stability because of the ability form disulfide bonds. Replacement of Cys residues by tyrosine may result in an outsized effect on the health of the host cells and place a ceiling on the ability to reassign Cys codons.

Introduction of an orthogonal $tRNA^{Opt}_{GCA}$ reassigned the UGC codon to tyrosine at 20.2% and the UGU codon to tyrosine at 17.3% as a G34/U3 wobble, and despite replacing about 10% of cysteine residues with tyrosine, these *E. coli* remain about 90%

as fit as non-reassigning, wild-type GFP expressing control cells (Tables 2.1-2.3). The initial hypothesis of a ceiling on Cys reassignment due to outsized growth defects therefore does not hold.

A second hypothesis for the lower than expected reassignment of UGU by $tRNA^{Opt}_{ACA}$ is the benefit of improving anticodon-codon interactions by supplying a Watson—Crick base pairing tRNA may be outweighed by the benefit of presumed increased effective concentration of a wobble base pairing tRNA as a result of predicted efficiency of aminoacylation. The *M. jannaschii* $tRNA^{Opt}_{GCA}$ is expected to be aminoacylated very well, approximately 77-fold better than *M. jannaschii* $tRNA^{Opt}_{ACA}$ for Cys UGU reassignment and 10-fold better than the *M. jannaschii* $tRNA^{Opt}_{GUA}$ for amber stop codon suppression [7, 34]. The *M. jannaschii* $tRNA^{Opt}_{GCA}$ decodes the UGU codon via a G34/U3 wobble approximately 5-fold better than the *M. jannaschii* $tRNA^{Opt}_{ACA}$ decodes the UGU codon via a A34/U3 Watson—Crick interaction. Aside from predicted aminoacylation efficiency of the two orthogonal tRNAs, the other factors expected to play a role in reassignment efficiency (e.g. number of competing tRNAs, expression level of orthogonal tRNA, amount of GFP reporter mRNA) are expected to be approximately the same. The potentially higher concentration of aminoacylated $tRNA^{Opt}_{GCA}$ relative to $tRNA^{Opt}_{ACA}$ might explain the unexpected differences in sense codon reassignment efficiencies by the two orthogonal *M. jannaschii* tRNAs.

Effect of Substitution of Tyrosine for Encoded Amino Acid on Cell Health

The orthogonal pair-directed sense codon reassignments to tyrosine described here enable codon-specific amino acid substitutions across the proteome of *E. coli* and may be used as a tool to investigate the ability of cellular proteins and more broadly

cellular systems, to tolerate amino acid substitutions to tyrosine. The majority of systems reassigning wobble codons to tyrosine had instantaneous doubling times that were within 2 standard deviations of a non-reassigning, wild-type GFP expressing control, despite estimated 100,000s to 1,000,000s of tyrosine substitutions across the proteome, suggesting that *E.coli* are very tolerant to amino acid substitutions to tyrosine across the proteome (Figure 2.2, Table 2.3) [7]. The additional 100,000 to 1,000,000 directed missense substitution per cell generation represent between 2x and 10x the expected background level of missense incorporation.

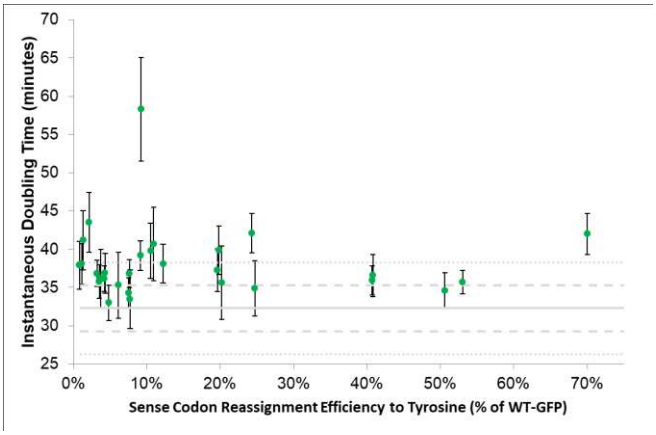

Figure 2.2. Average instantaneous doubling time of each sense codon reassigning system plotted against the sense codon reassignment efficiency for that system. The straight grey line at 32.3 minutes represents the instantaneous doubling time for the non-reassigning, wild type GFP culture (GFP with a Tyr codon at position 66, amber-reassigning tRNA with a CUA anticodon). The grey dashed lines represent the standard deviation (± 3.0 minutes) of the instantaneous doubling time for the non-reassigning, wild type GFP control. The grey dotted lines represent the second standard deviation for the non-reassigning, wild type GFP control.

Table 2.3. Cell health and growth profiles for sense codon reassigning systems (wobble codon data previously determined [7]).

tRNA Anticodon	Sense Codon Targeted	Instantaneous Doubling Time (min)	Relative System Fitness	Estimated Number of Proteome-Wide Substitutions per Cell Generation
$tRNA^{Opt}_{AAA}$	Phe UUU	36.85 ± 1.73	0.88	6.51E+05
$tRNA^{Opt}_{AAG}$	Leu CUU	36.14 ± 1.68	0.90	5.50E+05
$tRNA^{Opt}_{AAU}$	Ile AUU	36.87 ± 2.59	0.88	1.78E+06
$tRNA^{Opt}_{AAC}$	Val GUU	37.90 ± 3.12	0.85	6.04E+05
$tRNA^{Opt}_{AGA}$	Ser UCU	37.27 ± 2.80	0.87	4.82E+06
$tRNA^{Opt}_{AGG}$	Pro CCU	38.12 ± 2.56	0.85	1.52E+06
$tRNA^{Opt}_{AGU}$	Thr ACU	39.82 ± 3.61	0.81	3.70E+06
$tRNA^{Opt}_{AGC}$	Ala GCU	41.19 ± 3.89	0.79	1.06E+06
$tRNA^{Opt}_{AUG}$	His CAU	35.30 ± 4.34	0.92	1.46E+06
$tRNA^{Opt}_{AUU}$	Asn AAU	34.28 ± 2.01	0.94	1.28E+06
$tRNA^{Opt}_{AUC}$	Asp GAU	36.18 ± 3.80	0.89	1.41E+06
$tRNA^{Opt}_{ACA}$	Cys UGU	35.79 ± 2.19	0.90	1.57E+05
$tRNA^{Opt}_{ACU}$	Ser AGU	39.17 ± 1.95	0.83	6.04E+05
$tRNA^{Opt}_{ACC}$	Gly GGU	38.07 ± 2.64	0.85	1.05E+06
$tRNA^{Opt}_{CAC}$	Val GUG	43.52 ± 3.93	0.74	9.13E+05
$tRNA^{Opt}_{CGC}$	Ala GCG	58.30 ± 6.80	0.56	4.50E+06
$tRNA^{Opt}_{CUU}$	Lys AAG	33.47 ± 3.86	0.97	2.75E+06
$tRNA^{Opt}_{CUC}$	Glu GAG	39.87 ± 3.17	0.81	6.18E+06
$tRNA^{Opt}_{GCG}$	Arg CGC	36.82 ± 1.83	0.88	8.80E+06
$tRNA^{Opt}_{UCG}$	Arg CGA	36.57 ± 2.77	0.88	2.44E+06
$tRNA^{Opt}_{UAG}$	Leu CUA	42.1 ± 2.6	0.77	2.17E+06
$tRNA^{Opt}_{UAU}$	Ile AUA	42.0 ± 2.7	0.77	5.16E+06
$tRNA^{Opt}_{UGA}$	Ser UCA	34.9 ± 3.6	0.93	1.61E+06
$tRNA^{Opt}_{UGU}$	Thr ACA	40.7 ± 4.8	0.80	6.97E+05
$tRNA^{Opt}_{GCA}$	Cys UGC	35.6 ± 4.8	0.91	2.38E+06
$tRNA^{Opt}_{CCG}$	Arg CGG	35.7 ± 1.5	0.91	2.09E+06
$tRNA^{Opt}_{UCU}$	Arg AGA	35.9 ± 1.9	0.90	7.80E+05
$tRNA^{Opt}_{CCU}$	Arg AGG	34.6 ± 2.3	0.94	3.17E+05
$tRNA^{Opt}_{UCC}$	Gly GGA	33.0 ± 2.3	0.98	4.29E+05
$tRNA^{Opt}_{GGG}$	Pro CCC	----[a]	----	----
$tRNAOpt_{CCC}$	Gly GGG	----[a]	----	----

a The orthogonal tRNA with the indicated anticodon was unable to be successfully constructed.

The majority of systems reassigning rare codons to tyrosine also had instantaneous doubling times that were within 2 standard deviations of the non-reassigning, wild-type GFP expressing control (Figure 2.2, Table 2.3). Reassignment of the rare Ile AUA codon and reassignment of the rare Leu CUA codon both had instantaneous doubling times that were more than 2 standard deviations away from the non-reassigning wild-type GFP expressing control (42.0 and 42.1 minutes, vs. 32.3 minutes for control, respectively).

Ile AUA is the most efficiently reassigned codon at 70.0%, and Leu CUA is reassigned with the seventh highest efficiency at 24.3%. The growth rates for both systems are about 25% longer than wild type (relative fitness of 0.770) during the exponential phase, with instantaneous doubling times of about 42 minutes. The carrying capacity of the cells that are reassigning Ile AUA to tyrosine is reduced to about half of that of the non-reassigning control system, and one fifth of that of the non-reassigning control system for Leu CUA. Leu and Ile are expected to be buried within proteins most of the time, and are not typically surface-exposed. Consequently, replacement of leucine or isoleucine with tyrosine may significantly affect the hydrophobic cores of proteins, negatively affecting structure and function.

The second to fifth most efficiently reassigned sense codons are the four rare arginine codons CGA, CGG, AGA and AGG. Systems reassigning rare Arg codons to tyrosine have instantaneous doubling times during the exponential phase that are within 2 standard deviations of the non-reassigning control, despite having reassignment efficiencies ranging from 40%-53%. Many Arg residues are surface-exposed or surface-adjacent to avoid burying the positively-charged side chain in the hydrophobic cores of

proteins. Substitution of tyrosine at Arg codons may be more tolerated because of the relative tolerance to mutation of surface exposed amino acid residues.

Differences between "wobble only" and "all" sense codon evaluations

The suite of sense codon reassignment efficiency measurements gathered using closely related systems essentially controls for many of the expected contributing factors that are identical between systems with different anticodons. The comparison between the natural system and an essentially constant ruler in the form of the orthogonal pair enables further investigation of the *in vivo* importance of tRNA abundance and aminoacylation level in determining translational fidelity. Analysis of the sense codon reassignment efficiencies of 20 *E. coli* wobble codons showed that reductions in aminoacylation efficiency of the orthogonal tRNA as a result of changing the anticodon to Watson—Crick base pair with different sense codon targets are weakly correlated (R^2=0.15) to sense codon reassignment efficiencies (Figure 2.3 a). Furthermore, almost no apparent correlation (R^2=0.01) between sense codon reassignment efficiency and number of competing endogenous *E. coli* tRNAs was observed (Figure 2.3 c). The R^2 terms represent the relative importance of each factor in orthogonal pair directed sense codon reassignment. Therefore, analysis of sense codon reassignment efficiencies at the 20 *E. coli* wobble codons suggested that reductions in aminoacylation efficiency of the orthogonal tRNA as a result of changing the anticodon to Watson—Crick base pair with different sense codon targets was more significant in determining sense codon reassignment efficiency than the number of endogenous tRNAs competing against the orthogonal tRNA for a given sense codon [7].

The inclusion of additional sense codon reassignment efficiencies for rarely used *E. coli* codons resulted in the opposite order of importance for aminoacylation efficiency and competing endogenous tRNA codoncentration for orthogonal pair directed sense codon reassignment. Analysis of sense codon reassignment efficiencies for wobble and rare sense codons together showed almost no correlation (R^2=0.04) between sense codon reassignment efficiency and reductions in aminoacylation efficiency of the orthogonal tRNA as a result of changing the sequence of the tRNA anticodon (Figure 2.3 b). On the other hand, a stronger correlation (R^2=0.20) is observed between sense codon reassignment efficiency and number of competing endogenous tRNAs when wobble and rarely used codons are considered together.

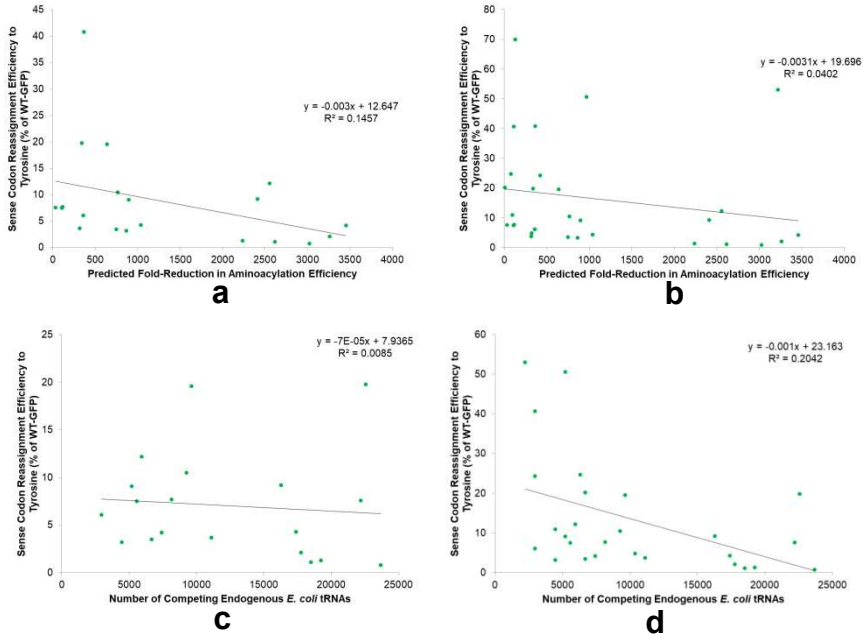

Figure 2.3. Top: sense codon reassignment efficiencies vs. predicted-fold reduction in aminoacylation efficiency as a result of changing the tRNA anticodon for (a) wobble codons only and (b) wobble and rarely used codons. Bottom: sense codon reassignment efficiencies vs. number of competing endogenous *E. coli* tRNAs for (c) wobble codons only and (d) wobble and rarely used codons.

2.5 CONCLUSIONS

The extent to which the degeneracy of the genetic code can be broken has not been systematically investigated. The 10 quantitative measurements at rare codons described here augment the previous 20 measurements made at wobble codons. Each *E. coli* rare codon is capable of reassignment to tyrosine at a measurable efficiency, ranging from 3.5% to 70% by simply providing an orthogonal tRNA with an anticodon capable of Watson-Crick base pairing with the codon targeted for reassignment.

Substituting tyrosine for the encoded amino acid in sense codon reassignment systems is well-tolerated, despite 100,000s to 1,000,000s of predicted substitutions across the *E. coli* proteome. Substitution of amino acids typically buried in the hydrophobic cores of proteins (such as isoleucine and leucine) with tyrosine may be less tolerated than substitution of surface-exposed amino acids such as arginine.

In addition to revealing additional promising sense codon targets not previously considered for future non-canonical amino acid incorporation using the orthgonal *M. jannaschii* tRNA/aaRS, the evaluation of sense codon reassignment at rare codons further contributes to understanding the *in vivo* importance of the factors involved in determining translational fidelity. Analysis of sense codon reassignment to tyrosine at wobble and rare codons as one data set reveals that both aminoacylation efficiency and competing tRNA abundance are not strongly correlated to sense codon reassignment efficiency, suggesting that they are not dominant factors in orthogonal pair directed sense codon reassignment.

The strategy of employing an orthogonal tRNA/aaRS to reassign sense codons coupled with a screen that allows quantitative evaluation of sense codon reassignment

should be generalizable to other orthogonal tRNA/aaRS pairs, allowing for the generation of large data sets that can be used to further understand the relative importance of the biological interactions that are involved in protein translation. Chapters 3-4 of this thesis describe the engineering of an orthogonal tyrosine-incorporating pyrrolysine *Methanosarcina barkeri* (*M. barkeri*) tRNA/aaRS pair compatible with the fluorescence-based screen for sense codon reassignment. Chapter 5 compares sense codon reassignment efficiencies to tyrosine using the engineered orthogonal *M. barkeri* pair with the sense codon reassignment efficiencies to tyrosine with the orthogonal *M. jannaschii* pair.

REFERENCES

1. Dumas, A., et al., *Designing logical codon reassignment–Expanding the chemistry in biology.* Chemical science, 2015. **6**(1): p. 50-69.
2. Lajoie, M.J., et al., *Genomically recoded organisms expand biological functions.* science, 2013. **342**(6156): p. 357-360.
3. Kwon, I., K. Kirshenbaum, and D.A. Tirrell, *Breaking the degeneracy of the genetic code.* Journal of the American Chemical Society, 2003. **125**(25): p. 7512-7513.
4. Mukai, T., et al., *Reassignment of a rare sense codon to a non-canonical amino acid in Escherichia coli.* Nucleic acids research, 2015. **43**(16): p. 8111-8122.
5. Lee, B.S., et al., *Incorporation of Unnatural Amino Acids in Response to the AGG Codon.* ACS Chem Biol, 2015. **10**(7): p. 1648-53.
6. Biddle, W., M.A. Schmitt, and J.D. Fisk, *Evaluating Sense Codon Reassignment with a Simple Fluorescence Screen.* Biochemistry, 2015. **54**(50): p. 7355-64.
7. Schmitt, M.A., Biddle, Wil, and John D. Fisk, *Mapping the Plasticity of the E. Coli Genetic Code with Orthgonal Pair Directed Sense Codon Reassignment.* submitted, 2017.
8. Biddle, W., Schwark, David G., Schmitt, Margaret A. and John D. Fisk, *Sense Codon Reassignment of the Rare Arginine AGG Codon: Transferability of Improvements Between Orthogonal tRNAs and Synthetases.* submitted, 2018.
9. Gingold, H. and Y. Pilpel, *Determinants of translation efficiency and accuracy.* Molecular systems biology, 2011. **7**(1): p. 481.

10. Fluitt, A., E. Pienaar, and H. Viljoen, *Ribosome kinetics and aa-tRNA competition determine rate and fidelity of peptide synthesis.* Comput Biol Chem, 2007. **31**(5-6): p. 335-46.

11. Liu, C.C. and P.G. Schultz, *Adding new chemistries to the genetic code.* Annual review of biochemistry, 2010. **79**: p. 413-444.

12. Drummond, D.A. and C.O. Wilke, *The evolutionary consequences of erroneous protein synthesis.* Nature Reviews Genetics, 2009. **10**(10): p. 715.

13. Biddle, W., M.A. Schmitt, and J.D. Fisk, *Modification of orthogonal tRNAs: unexpected consequences for sense codon reassignment.* Nucleic acids research, 2016: p. gkw948.

14. Döring, V., et al., *Enlarging the amino acid set of Escherichia coli by infiltration of the valine coding pathway.* Science, 2001. **292**(5516): p. 501-504.

15. Ruan, B., et al., *Quality control despite mistranslation caused by an ambiguous genetic code.* Proceedings of the National Academy of Sciences, 2008. **105**(43): p. 16502-16507.

16. Drummond, D.A. and C.O. Wilke, *Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution.* Cell, 2008. **134**(2): p. 341-352.

17. Schimmel, P. and M. Guo, *A tipping point for mistranslation and disease.* Nature Structural and Molecular Biology, 2009. **16**(4): p. 348.

18. Kirchner, S. and Z. Ignatova, *Emerging roles of tRNA in adaptive translation, signalling dynamics and disease.* Nature Reviews Genetics, 2015. **16**(2): p. 98.

19. Polycarpo, C.R., et al., *Pyrrolysine analogues as substrates for pyrrolysyl-tRNA synthetase.* Febs Letters, 2006. **580**(28-29): p. 6695-6700.

20. Srinivasan, G., C.M. James, and J.A. Krzycki, *Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA.* Science, 2002. **296**(5572): p. 1459-1462.

21. Wang, L., et al., *A New Functional Suppressor tRNA/Aminoacyl− tRNA Synthetase Pair for the in Vivo Incorporation of Unnatural Amino Acids into Proteins.* Journal of the American Chemical Society, 2000. **122**(20): p. 5010-5011.

22. Wang, L., et al., *Expanding the genetic code of Escherichia coli.* Science, 2001. **292**(5516): p. 498-500.

23. Wan, W., J.M. Tharp, and W.R. Liu, *Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool.* Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 2014. **1844**(6): p. 1059-1070.

24. Pédelacq, J.-D., et al., *Engineering and characterization of a superfolder green fluorescent protein.* Nature biotechnology, 2006. **24**(1): p. 79.

25. OpenWetWare. *Escherichia coli/Codon usage*. 2012 [cited 2018 20 January]; Available from: https://openwetware.org/wiki/Escherichia_coli/Codon_usage.

26. Hayashi, K., et al., *Highly accurate genome sequences of Escherichia coli K-12 strains MG1655 and W3110.* Molecular systems biology, 2006. **2**(1).

27. Neidhardt, F.C., *Escherichia coli and Salmonella typhimurium : cellular and molecular biology*. 1997, Washington, D.C.: American Society for Microbiology.

28. Milo, R., *What is the total number of protein molecules per cell volume? A call to rethink some published values.* Bioessays, 2013. **35**(12): p. 1050-1055.

29. Brocchieri, L. and S. Karlin, *Protein length in eukaryotic and prokaryotic proteomes.* Nucleic acids research, 2005. **33**(10): p. 3390-3400.

30. Nakamura, Y., T. Gojobori, and T. Ikemura, *Codon usage tabulated from international DNA sequence databases: status for the year 2000.* Nucleic acids research, 2000. **28**(1): p. 292-292.

31. Li, G.-W., et al., *Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources.* Cell, 2014. **157**(3): p. 624-635.

32. Curran, J.F., *Decoding with the A: I wobble pair is inefficient.* Nucleic acids research, 1995. **23**(4): p. 683-688.

33. Dong, H., L. Nilsson, and C.G. Kurland, *Co-variation of trna abundance and codon usage inescherichia coliat different growth rates.* Journal of molecular biology, 1996. **260**(5): p. 649-663.

34. Fechter, P., et al., *Major tyrosine identity determinants in Methanococcus jannaschii and Saccharomyces cerevisiae tRNATyr are conserved but expressed differently.* The FEBS Journal, 2001. **268**(3): p. 761-767.

35. Gamper, H.B., et al., *Maintenance of protein synthesis reading frame by EF-P and m(1)G37-tRNA.* Nature Communications, 2015. **6**.

36. Agris, P.F., *Decoding the genome: a modified view.* Nucleic acids research, 2004. **32**(1): p. 223-238.

37. Agris, P.F., F.A. Vendeix, and W.D. Graham, *tRNA's wobble decoding of the genome: 40 years of modification.* J Mol Biol, 2007. **366**(1): p. 1-13.

38. Crick, F.H., *Codon—anticodon pairing: the wobble hypothesis.* Journal of molecular biology, 1966. **19**(2): p. 548-555.

39. Lagerkvist, U., *" Two out of three": An alternative method for codon reading.* Proceedings of the National Academy of Sciences, 1978. **75**(4): p. 1759-1762.

40. Lehmann, J. and A. Libchaber, *Degeneracy of the genetic code and stability of the base pair at the second position of the anticodon.* RNA, 2008. **14**(7): p. 1264-1269.

41. Soma, A., et al., *An RNA-modifying enzyme that governs both the codon and amino acid specificities of isoleucine tRNA.* Molecular cell, 2003. **12**(3): p. 689-698.

CHAPTER 3

A SIMPLE FLUORESCENCE-BASED SCREEN FOR SENSE CODON REASSIGNMENT UTILIZING THE *M. Barkeri* tRNA/aaRS PAIR

3.1 CHAPTER OVERVIEW

Expanding the genetic code through orthogonal pair directed sense codon reassignment is an emerging technique for multi-site non-canonical amino acid incorporation into proteins. However, the relative importance of *in vivo* factors affecting sense codon reassignment are not known, making orthogonal pair and targeted codon selection challenging. Recent efforts have investigated the ability to break the degeneracy of the genetic code with the orthogonal *Methanocaldococcus jannaschii* tRNA/aminoacyl-tRNA synthetase (aaRS) pair and further understand the relative importance of the factors involved in translational fidelity using an *in vivo* fluorescence-based screen.

In order to use the fluorescence-based screen for sense codon reassignment with the other commonly used pyrrolysyl tRNA/aaRS from the *Methanosarcina* species (*M. barkeri* or *M. mazei*), a *M. barkeri* aaRS variant capable of aminoacylating tyrosine was engineered. The tyrosine-aminoacylating *M. barkeri* aaRS variant was identified from a library of $4.5 \cdot 10^7$ aaRS variants through fluorescence activated cell sorting, and is capable of suppressing an amber stop codon at the fluorophore tyrosine position of GFP at 18.4% of wild-type GFP in *E. coli* DH10B, of comparable efficiency to other engineered *M. barkeri* pyrrolysyl aaRS variants. Using the orthogonal *M. barkeri* pair the

sense codon reassignment efficiencies to tyrosine at a selection of eight sense codons representing the range of reassignment efficiencies observed with the orthogonal *M jannaschii* pair were measured. For each sense codon measurable reassignment efficiency above the limit of detection was observed. However, the dynamic range of the fluorescence-based screen was significantly reduced when used with the engineered tyrosine-incorporating *M. barkeri* pair.

3.2 INTRODUCTION

The 20 natural amino acids encode a limited set of chemical functionalities. Encoding non-canonical amino acids (ncAAs) with chemical functionalities not contained in the natural set is a powerful strategy for modifying and extending the properties of proteins [1, 2]. The most widely employed technology for genetic code expansion involves the suppression of amber stop codons.[1, 2] Variants of just two orthogonal tRNA/aaRS pairs, the *Methanocaldococcus jannaschii* tyrosine (*M. jannaschii* Tyr) and the *Methanosarcina* species pyrrolysine pair, have been used to incorporate the vast majority of non-canonical amino acids into proteins in response to the amber stop codon [3, 4]. Both orthogonal pairs have been subjected to repeated directed evolution experiments to select variants that are capable of incorporating over 150 different ncAAs into proteins [1]. Although great strides have been made allowing incorporation of non-canonical amino acids in response to the amber stop codon, the broader use of orthogonal pair systems in the reassignment of sense codons is much less well studied [5-11].

The reassignment of the meaning of sense codons is constrained by the fact that all 64 codon triplets have an assigned function. The constraint imposed by the lack of

free codons is mitigated by the degeneracy of the code; 61 sense codons are used to specify the 20 canonical amino acids. While the genetic code is nearly universal, the translational apparatuses of different organisms have diverged considerably [12, 13]. Variations in genome compositions, codon usage frequencies, and the complements of adapter tRNA molecules used to translate the code suggest a high degree of plasticity is possible in translation systems. Breaking the degeneracy of the genetic code by reassigning the meaning of sense codons has the potential to expand the genetic code to 22 (or more) amino acids, greatly increasing the encodable properties of proteins. The interactions between tRNA and aaRS molecules that drive the fidelity of the genetic code have only partially been mapped, and the space between these interactions, the extent to which orthogonal tRNA/aaRS pairs can be added, is largely unknown. As a result, predicting which sense codons are amenable to reassignment and which orthogonal translation machinery is best suited for the task is challenging.

Recent experiments have described the potential of sense codon reassignment in *E. coli* using the *M. jannaschii* tyrosine orthogonal pair evaluated with a green fluorescent protein reporter [11, 14-16]. The green fluorescent protein (GFP) based screen exploits the absolute requirement of an active site tyrosine in the maturation of the GFP fluorophore. Residues 65-67 of super folder GFP specify the Thr-Tyr-Gly sequence that auto-catalytically reacts to produce the extended conjugated fluorophore. Replacement of tyrosine at position 66 with any other natural amino acid effectively abolishes the fluorescence of the GFP reporter protein [17, 18]. The screen for sense codon reassignment quantifies the ability of the tyrosine-charging *M. jannaschii* aaRS to recognize *M. jannaschii* tRNAs with altered anticodons and the extent to which tyrosine-

aminoacylated *M. jannaschii* tRNAs can compete with endogenous *E. coli* tRNAs to decode specific sense codons.

The evaluation of sense codon reassignment using the *M. jannaschii* orthogonal pair system has been investigated at more than 30 codons in *E. coli* utilizing the GFP screen [[14], Chapter 2]. Importantly, every sense codon evaluated could be reassigned by simply providing an orthogonal tRNA with an anticodon capable of Watson—Crick base pairing to the targeted codon. Directed evolution to improve the interaction between the aaRS and its cognate tRNA was shown to be an effective strategy for improving the efficiency of sense codon reassignment by orthogonal pairs with altered anticodon sequences [11]. In addition to facilitating ncAA incorporation for genetic code expansion orthogonal pair directed sense codon reassignment is broadly applicable as a tool for evaluating the factors affecting translational efficiency in the context of wholly functional translation systems in living cells.

Motivated by sense codon reassignment to ncAAs, but directed at a deeper understanding of the factors influencing the fidelity of translation, this chapter describes the expansion of the GFP based screen for quantification of sense codon reassignment to the other most commonly used orthogonal pair system for genetic code expansion of amber stop codons, the pyrrolysine orthogonal pair from *M. barkeri*. This chapter describes the selection of a tyrosine incorporating variant of the *M. barkeri* pyrrolysine orthogonal pair system through directed evolution and high-throughput library screening. The tyrosine incorporating variant was applied to evaluate sense codon reassignment using the GFP based fluorescence screen. The *Methanosarcina barkeri* pyrrolysine orthogonal translation machinery differs from the *M. jannaschii* system in

that the anticodon sequence is not important for aminoacylation. Comparison between the two orthogonal pair systems will enable the isolation of the relative importance of aminoacylation efficiency to fidelity of the genetic code and to genetic code reassignment.

In the *M. jannaschii* system, the anticodon of the tRNA is a critical identity element for effective recognition and aminoacylation by the aaRS [19]. The tRNA anticodon changes required to target specific sense codons affect aminoacylation and may also change the anticodon stem loop geometry, impacting decoding on the ribosome. Lower effective concentrations of aminoacylated tRNAs compete less efficiently against endogenous *E. coli* tRNAs to decode the targeted codon. In contrast, the *M. barkeri* Pyl aaRS does not utilize the anticodon of the tRNA as one of its identity elements [20-22]. As a result, the effective concentrations of aminoacylated tRNAs is expected to be similar across systems in which the tRNAs are engineered with different anticodons to target different sense codons.

The absence of tRNA anticodon sequence from the identity elements that specify aaRS recognition for the *M. barkeri* Pyl orthogonal pair not only removes one of the expected barriers to efficient sense codon reassignment, but also provides a second unique data set to assist in unraveling the *in vivo* importance of contributing factors in determining translation efficiency. Understanding the interactions that drive the fidelity of the genetic code and the limits to which modifications can be made without breaking the translational system has practical implications for understanding the molecular mechanisms of evolution as well as expanding the set of encodable amino acids. A comparison of the relative reassignment efficiencies of a given sense codon to tyrosine

by both the *M. jannaschii* tRNA/aaRS pair and the *M. barkeri* tRNA/aaRS pair should shed light on the contribution of aminoacylation to the efficiency to codon translation.

3.3 MATERIALS AND METHODS

The gain of function GFP fluorescence-based screen for sense codon reassignment has been described [11]. Detailed experimental protocols, including methods for library construction and high throughput screening, measurement of system growth and fluorescence, and calculation of codon reassignment efficiencies are included in Appendix 1. The appendix also includes full vector sequences, oligonucleotide primer sequences, detailed cell strain information, and general reagents and materials.

3.4 RESULTS AND DISCUSSION

Directed evolution of a tyrosine-incorporating variant of the M. barkeri Pyl tRNA/aaRS pair

Variants of the *M. barkeri* Pyl aaRS which aminoacylate phenylalanine (Phe), o-methyltyrosine (OmeY) and 4-bromo/iodo-phenylalanine (pBr/pIO) have been reported [23-26]. A comparison of the positions within the amino acid binding domain of the *M. barkeri* Pyl aaRS which were mutated in aaRS variants that enabled the aminoacylation of amino acids with side chains similar to tyrosine were used to design a large library of aaRS variants. The comparison suggested that positions Ala267, Tyr271, Leu274, Asn311, Cys313 and Val367 should be included in the library to select variants capable of activating tyrosine (Figure 3.1). Each selected position varied from the wild-type Pyl aaRS sequence in either the Phe, OmeY or pBr/pIO activating aaRSs. The degenerate codon NNK (encodes all 20 amino acids and the amber stop codon) was used at each

of the 6 library positions. The library had a theoretical diversity of $1.07 \cdot 10^9$ at the DNA level, $6.4 \cdot 10^7$ at the protein level including variants that contain an amber stop codon in the protein sequence.

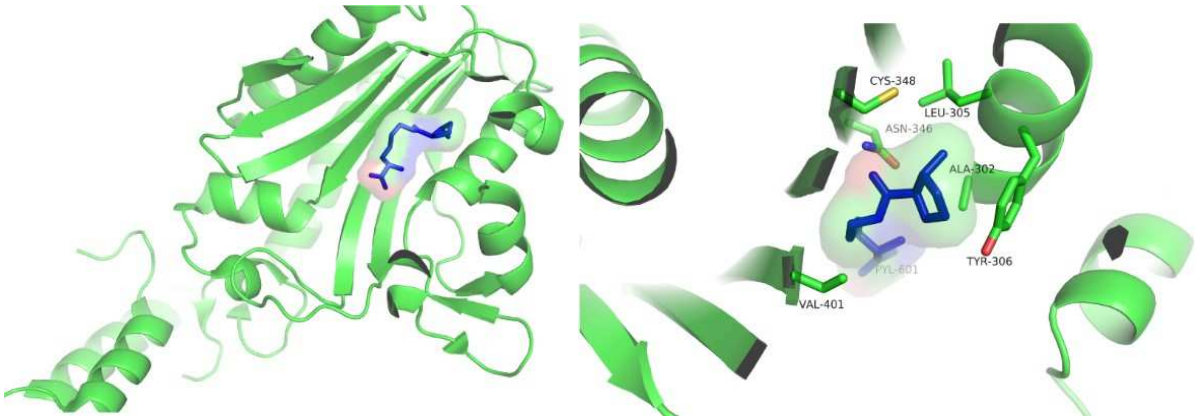

Figure 3.1 Left: crystal structure of the *M. mazei* Pyl aaRS complexed with pyrrolysine highlighted in blue (PDB 2ZCE) [27]. Bottom: active site of the *M. mazei* Pyl aaRS with library positions highlighted.

The starting material for library generation included three KpnI restriction sites within the amino acid binding domain of the aaRS, one of which introduced an amber stop codon into the aaRS gene. Transformation of the mutagenized DNA into cells yielded $1.01 \cdot 10^9$ transformants; a small portion of the transformation was plated to determine the transformation efficiency and mutagenesis rate. Twenty-four colonies from the aliquot of library cells plated were analyzed by colony PCR, and 1 out of 24 had all KpnI sites removed, suggesting that three mutagenic library primers had modified the template DNA. One out of 24 colonies suggests that approximately $4.2 \cdot 10^7$ of the transformants were true library members in which all 3 mutagenic primers hit. The library transformation was amplified for 2 hours (~3 doubling times), and the library DNA

was then isolated. The isolated DNA was subjected to digestion with KpnI to remove non-mutated starting material. Importantly, none of the target DNA sequences in the library included a KpnI site; so no library members were removed during digestion. Following digestion, the DNA was transformed into cells harboring a UAG stop codon at the fluorophore position of GFP. A small portion of the second transformation was plated to determine the transformation efficiency and frequency of full library members. The transformation for the second library contained $2.7 \cdot 10^8$ independent transformants. Twenty four colonies from a small aliquot of the second library were analyzed by colony PCR and sequencing. 19 of the 24 colonies tested were full library members that had all 3 mutagenic primers hit the template. The library cells were amplified overnight in the presence of 1 mM IPTG to induce the expression of the orthogonal translational components and GFP. These cells were evaluated for incorporation of tyrosine in response to UAG by the *M. barkeri* Pyl aaRS variants by fluorescence-activated cell sorting (FACS).

In the first round of FACS, approximately $1.2 \cdot 10^8$ cells were screened. Most of the library cells did not have detectable fluorescence above that of a non-fluorescent control, suggesting that the *M. barkeri* Pyl aaRS variants in those cells were not able to aminoacylate tyrosine onto the corresponding $tRNA^{Pyl}_{CUA}$ (Figure 3.2). Clones representing the top 1% of fluorescence (0.03% of the total population) were collected and amplified overnight. Approximately $1.8 \cdot 10^7$ of these cells were reexamined with FACS. A small population of cells with fluorescence greater than that of the non-fluorescent control population was during the screening of the amplified clones from the first round of FACS (Figure 3.2 inset). $1.3 \cdot 10^4$ of these cells (approximately 0.02% of the

population) were collected and amplified. A selection of these cells was plated directly for analysis of individual clones.

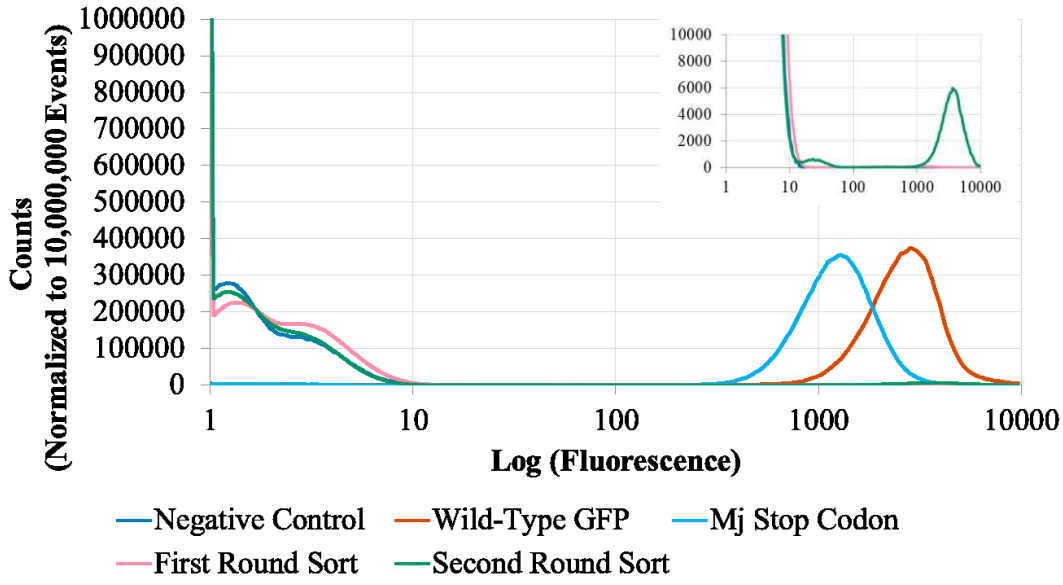

Figure 3.2. Normalized counts vs. Log (Fluorescence) for FACS sorting of the *M. barkeri* aaRS library. The counts for each sample were normalized to 10,000,000 events to compare control and library traces on the same plot. The inset expands the region between 1 and 10,000 counts to show the small population of cells with fluorescence greater than that of the non-fluorescent control population.

Individual clones selected from the portion of the library plated directly after the second round of FACS sorting were evaluated for ability of the orthogonal Pyl *M. barkeri* $tRNA^{Pyl}_{CUA}$ and aaRS variant to incorporate tyrosine in response to an amber stop codon UAG at position 66. Multiple clones were identified that had GFP fluorescence ranging from 0.6% to 71.1% of a wild-type GFP expressing control. For ten of the fluorescent variants, the plasmid harboring the orthogonal translational components was isolated and co-transformed back into *E. coli* DH10B with an unselected GFP reporter.

Isolating the plasmid containing the orthogonal translational components enabled sequencing of the Pyl *M. barkeri* aaRS to determine the identity of the amino acids in the binding pocket, and also allowed for the variant to be re-tested to ensure that the variants were not false-positives. False positive variants may be a result of carry-through of controls expressing wild-type GFP during the sorting process, a library cell containing a GFP gene that had reverted back to wild-type (only 1 nucleotide needs to be mutated at position 66 to revert back to wild-type: UAG to UAC or UAU) or a mutation that results in over expression of GFP.

The majority of the variants analyzed, 9 of the 10, were false positives (were not fluorescent during the second test for amber stop codon suppression) and were not carried forward. One variant remained fluorescent – indicating that the aaRS was able to aminoacylate tyrosine onto the corresponding $tRNA^{Pyl}_{CUA}$. The *M. barkeri* tRNA/aaRS variant identified was able to suppress an amber stop codon at position 66 of GFP to tyrosine with 8.2% efficiency of a wild-type GFP expressing, non-suppressing control. The identified tyrosine-aminoacylating Pyl *M. barkeri* aaRS variant contains 4 mutations relative to the wild-type pyrrolysine-charging aaRS: Ala 267 Lys, Leu 274 Met, Asn 311 Ala, and Cys 313 Glu (Table 3.1). The Asn 311 Ala mutation is not surprising because the same mutation is also observed in the Phe and OmeY aaRS variants, and mutating the side chain to a smaller amino acid likely widens the binding pocket to accommodate the phenol side chain of tyrosine (Figure 3.1) [23]. The Cys 313 Glu mutation may potentially shorten the depth of the long binding pocket to better fit the smaller tyrosine, and mutation of Cys 313 to a larger amino acid is also observed in the Phe, OmeY, and pBr/pIO variants (Figure 3.1) [23-26].

Table 3.1 Active site sequences of Pyl aaRS Variants

	Amino Acid Position					
Identity of Pyl-aaRS	**267**	**271**	**274**	**311**	**313**	**367**
Wild-Type [28]	Ala	Tyr	Leu	Asn	Cys	Val
Phenylalanine [25, 26]	Ala	Tyr	Leu	Ala	Leu	Val
	Ala	Tyr	Leu	Ala	Lys	Val
O-methyl-L-tyrosine [24]	Thr	Tyr	Leu	Val	Trp	Leu
4-Bromo-L-phenylalane [25, 26]	Ala	Leu	Ser	Ser	Met	Leu
4-Iodo-L-phenylalanine [25, 26]						
Tyrosine	Lys	Tyr	Met	Ala	Glu	Val

Improving the Incorporation Efficiency of Tyr by the M. barkeri aaRS Variant

The apparent 8.2% efficiency of incorporation of Tyr of the first identified *M. barkeri* aaRS variant represents a starting point for further improvement. Unlike the *M. janaschii* tRNA/aaRS pair that naturally aminoacylates tyrosine, the evolved tyrosine-incorporating *M. barkeri* tRNA/aaRS pair is not expected to be a highly efficient version of the enzyme. Engineered aaRSs are typically not as efficient as natural evolved aaRSs; efficiencies of amino acid incorporation in response to stop codons by engineered aaRSs average about 25% per position [3, 29]. In the vector used to generate and sort the aaRS library, the transcription of the mRNA for the *M. barkeri* aaRS variant was driven by an *lpp* promoter. The vector construct used for evaluation of sense codon reassignment using the *M. janaschii* tRNA/aaRS pair relies upon the stronger *tac* promoter to drive transcription of the aaRS. Swapping the *lpp* promoter for the *tac* promoter more than doubled the incorporation efficiency of tyrosine in response to amber stop codons by the *M. barkeri* tRNA/aaRS pair to 18.4% (Figure 3.3).

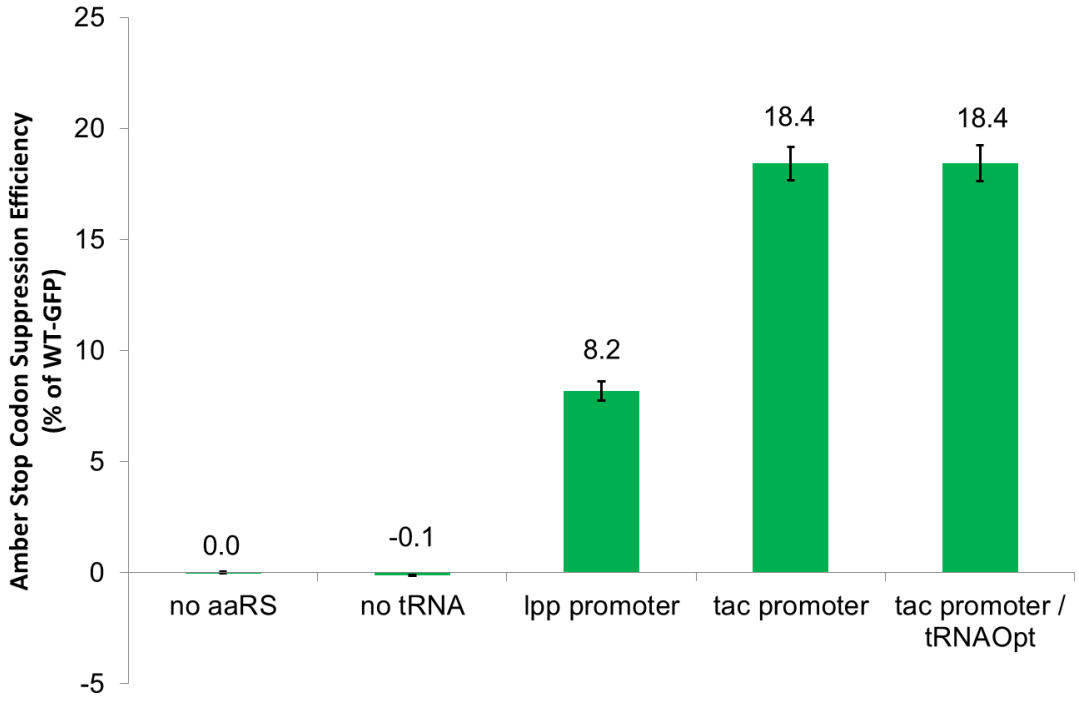

Figure 3.3 Amber stop codon reassignment efficiency to tyrosine at position 66 relative to wild-type GFP using the engineered *M. barkeri* tyrosyl aaRS pair. Replacing the *lpp* promoter that drives the aaRS with a stronger *tac* promoter more than doubled the incorporation efficiency. Replacing the $tRNA^{Pyl}$ with $tRNA^{Pyl\text{-}Opt}$ did not significantly improve incorporation efficiency.

Evaluation of Pyl tRNA Variants to Improve Incorporation Efficiency

Soll and co-workers reported a variant of the Pyl *M. barkeri* tRNA which improved incorporation efficiencies of ncAAs in response to amber stop codons when used with different *M. barkeri* aaRS variants [30]. Soll and co-workers hypothesized that changing the sequence of the acceptor and T stems of $tRNA^{Pyl}$ would improve the aminoacylation efficiency of the orthogonal pair and improve interaction between the orthogonal tRNA and elongation-factor Tu. Soll and coworkers generated $tRNA^{Pyl}$ variants that had the targeted mutations in the acceptor and T-stems were constructed and evaluated

individually. The top mutations in the acceptor and T-stems were combined, generating $tRNA^{Pyl\text{-}Opt}_{CUA}$. The amber stop codon suppression efficiencies improved 1.3-1.7 fold (suppression efficiencies ranged from ~3% to ~30% of the wild-type protein, depending on the location of the amber stop codon and the *M. barkeri* aaRS variant used) when $tRNA^{Pyl\text{-}Opt}_{CUA}$ was used.

Thermally balanced inside out (TBIO) PCR was used to synthesize $tRNA^{Pyl\text{-}Opt}_{CUA}$. The optimized tRNA was then used with the identified tyrosine-incorporating Pyl *M. barkeri* aaRS variant to suppress an amber stop codon at position 66 in GFP. In contrast to the 1.3-1.7 fold improvements in amber stop codon suppression efficiencies observed by Soll and co-workers with $tRNA^{Pyl\text{-}Opt}_{CUA}$, no improvement in amber stop codon suppression efficiency was observed when used with the identified tyrosine-incorporating Pyl *M. barkeri* aaRS variant (Figure 3.3).

Evaluation of Amino Acid Specificity of M. barkeri aaRS variant

The orthogonality of the tyrosine incorporating aaRS variant was evaluated. Changing the amino acid specificity of the aaRS did not break orthogonality. No incorporation of tyrosine in response to an amber stop codon was observed when a vector containing the engineered aaRS but no tRNA was used, meaning that the aaRS did not charge endogenous *E. coli* tRNAs with any observable amount of tyrosine (Figure 3.3).

Given that the *M. barkeri* aaRS variant that aminoacylates phenylalanine onto its corresponding tRNA was used to guide the design of the library for the selection of a tyrosine incorporating variant, examination of the possibility that the evolved tyrosine-incorporating variant also incorporated phenylalanine was investigated. The

fluorescence-based screen does not, however, report on aminoacylation of other canonical amino acids onto the tRNA. Incorporation of either phenylalanine or histidine (the amino acids most likely to be accepted by an enzyme binding pocket that also accepts Tyr) by the *M. barkeri* tRNA would result in proteins with greatly decreased and shifted fluorescence.

Electrospray ionization mass spectrometry (ESI-MS) was used to evaluate the identity of the amino acids incorporated in response to a single UAG codon in the Z domain of protein A [31]. The Z domain is a small, soluble, 8.3 kDa, three-helix protein that has been previously used to evaluate codon reassignment. The phenylalanine codon at position 5 in the native Z domain was mutated to UAC to provide a protein sample with tyrosine at that position or UAG to evaluate the *M. barkeri* aaRS variant identified in the library (Appendix 1). The UAC codon-containing variant or UAG codon containing Z domain variant was expressed in DH10B cells also expressing the *M. barkeri* machinery to reassign the UAG codons to tyrosine. Z domain proteins were purified and analyzed using ESI-MS of the intact protein followed by deconvolution of the mass spectra using the Maximum Entropy algorithm (MassHunter Software, Agilent Technologies).

The calculated mass of the Z domain peptide with tyrosine at position 5 is 8315 Da, and the +14 Da second peak observed for all parent peaks is likely due to a methylation of the Z-domain. These masses are observed for the Z domain protein expressed from a gene including a tyrosine codon at position 5 (Figure 3.4).

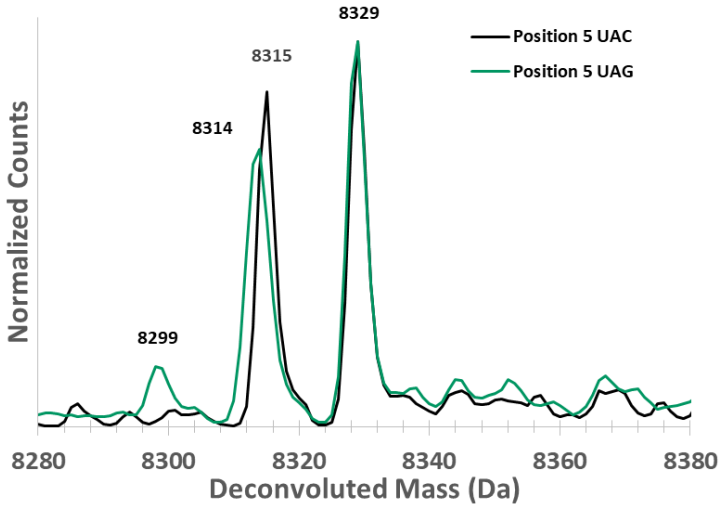

Figure 3.4 ESI-MS of purified Z-domain expressed from a gene containing a UAC codon at position 5 (black) or a UAG codon at position 5 (green).

These two peaks are also readily apparent in the ESI-MS analysis of Z domain protein containing a UAG amber stop codon at position 5, offering a second indication that the identified *M. barkeri* aaRS variant is capable of aminoacylating tyrosine onto its cognate tRNA. A second parent peak at 8299 Da was also observed in this sample. The -16 Da peak represents incorporation of phenylalanine in response to the UAG codon. Methylation of the Z domain with Phe at position 5 would be observed at 8313 Da. The broader peak at approximately 8314 Da represents methylated Z domain with Phe at position 5 and unmethylated Z domain with Tyr at position 5, as evidenced by the other peaks at 8299 Da and 8329 Da in this sample. Although not quantitative, the size of the peaks that indicate incorporation of Phe relative to those peaks that indicate incorporation of Tyr suggest that Phe is not the predominant amino acid recognized by the *M. barkeri* aaRS variant. A small amount of Phe incorporation by the *M. barkeri* tRNA/aaRS pair will result in some production of non-fluorescent GFP in response to the targeted codon, leading to a slightly lower calculation of reassignment efficiency for

that codon. However, the comparative map of the most fruitful sense codons for reassignment by the *M. barkeri* tRNA/aaRS pair will not be affected by a small amount of Phe incorporation during each evaluation.

Sense Codon Reassignment Utilizing the Tyrosine-Incorporating M. barkeri tRNA/aaRS Pair

The *M. barkeri* aaRS amino acid binding domain library was generated in DH10B *E. coli*, but further analysis of the codon reassignment efficiency of tyrosine-incorporating *M. barkeri* tRNA/aaRS variants was carried out in *E. coli* SB3930 to allow direct comparison to ~~our~~ the suite of sense codon reassignment data by variants of the *M. jannaschii* tyrosyl tRNA/aaRS pair. The efficiency of reassignment decreased slightly when comparable systems are evaluated in SB3930 relative to DH10B *E. coli*. The *M. barkeri* tRNA/aaRS variant incorporated tyrosine in response to an amber stop codon with 12.6% efficiency in SB3930 compared to 17.8% in DH10B (Figure 3.5).

A selection of sense codons were evaluated for reassignment using the fluorescence-based screen and the tyrosine-charging *M. barkeri* aaRS with its cognate tRNA. The eight sense codons selected for initial evaluation span the range of reassignment efficiencies observed using the *M. jannaschii* tRNA/aaRS pair and include both *E. coli* wobble codons and rarely used codons. The guiding idea behind targeting wobble codons for genetic code expansion is that differences in codon-anticodon binding energies between introduced orthogonal tRNAs capable of Watson—Crick base pairing and endogenous tRNAs that utilize wobble base pairing can be harnessed to bias the incorporation of amino acids. Reassignment of rarely used sense codons exploits the fact that decoding any mRNA codon requires sampling of tRNAs by the

ribosome. Rarely used codons are typically decoded by less abundant tRNAs, which is expected to result in lower competition between the endogenous tRNA and the introduced orthogonal tRNA for the codon to be reassigned.

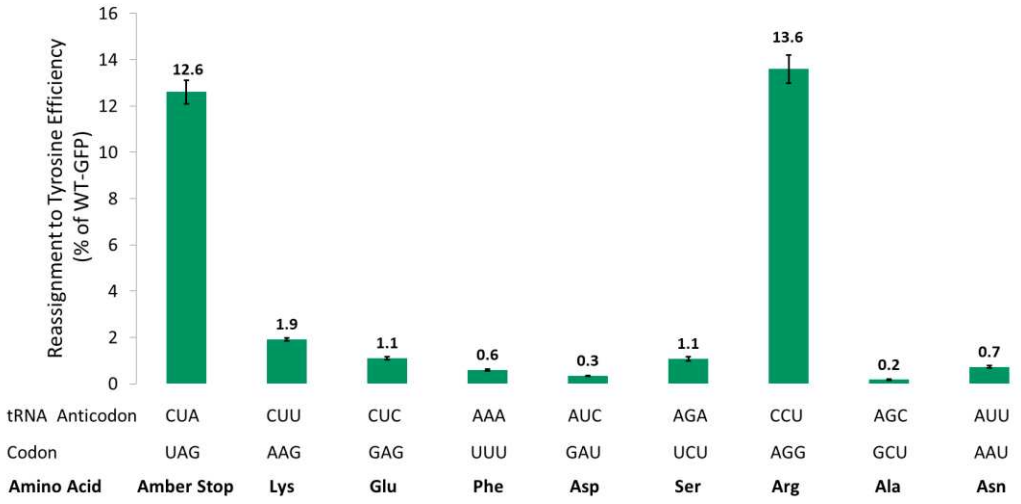

Figure 3.5 Reassignment of sense codons at position 66 in GFP to tyrosine as a percentage of wild-type GFP using the identified tyrosine-incorporating *M. barkeri* aaRS variant and its corresponding $tRNA^{Pyl}$ in *E. coli* SB3930. All reassigning systems had detectable fluorescence above background.

The evolved *M. barkeri* tRNA/aaRS pair reassigned the selected codons to tyrosine with efficiencies between 0.2% and 13.6% in *E. coli* SB3930 (Figure 3.5). Based on the variation across 6 clones of the non-reassigning negative control, the limit of detection for the fluorescence-based screen using the *M. barkeri* tRNA/aaRS pair is 0.14%. Reassignment of the amber stop codon with 12.6% efficiency in SB3930 cells reduces the dynamic range of the fluorescence-based screen. Apparent poorly reassigned codons were included to evaluate the lower detection limit of reassignment

using the evolved aaRS. Fluorescence as a result of incorporation of Tyr in response to the Ala GCU codon, one of the least efficiently reassigned sense codons with the *M. jannaschii* system at 1.3%., was above the limit of detection with the engineered tyrosine-incorporating *M. barkeri* tRNA/aaRS based on the non-reassigning control.

Although the engineered tyrosine-incorporating orthogonal *M. barkeri* pair reassigned the selected codons with efficiencies above the limit of detection, four of the 8 codons had reassignment efficiencies that were below the limit of quantification: Ala GCU, Asp GAU, Phe UUU, and Asn AUU. In contrast, none of 8 sense codon reassignment efficiencies to tyrosine measured with the orthogonal *M. jannaschii* tRNA/aaRS pair were below the limit of quantification of the fluorescence-based screen (Table 3.4) [Chapter 2, [14]]. Improvement of the orthogonal tyrosine-incorporating *M. barkeri* pair through directed evolution may increase the dynamic range of the fluorescence-based screen to allow for reassignment measurements of the full suite of sense codon to be made, with all of them potentially above the limit of quantification.

Table 3.4 Codon Reassignment Efficiencies for Selected Codons with the *M. barkeri* and *M. jannaschii* Orthogonal Pairs

Orthogonal tRNA	Sense Codon Targeted	Reassignment Efficiency with *M. barkeri* (% WT-GFP)	Reassignment Efficiency with *M. jannaschii* (% WT-GFP)
$tRNA_{CUA}$	Amber UAG	12.6	86.0
$tRNA_{CUU}$	Lys AAG	1.9	7.7
$tRNA_{CUC}$	Glu GAG	1.1	19.8
$tRNA_{AAA}$	Phe UUU	0.6	3.2
$tRNA_{AUC}$	Asp GAU	0.3	3.7
$tRNA_{AGA}$	Ser UCU	1.1	19.6
$tRNA_{CCU}$	Arg AGG	13.6	40.7
$tRNA_{AGC}$	Ala GCU	0.2	1.3
$tRNA_{AUU}$	Asn AAU	0.7	7.5

3.5 CONCLUSIONS

An emerging technique for genetic code expansion is orthogonal pair directed sense codon reassignment. However, identifying which sense codons are well suited for reassignment with a particular orthogonal pair remains challenging, because the relative importance of the factors involved in translational fidelity remain largely unknown. This chapter describes the construction of an orthogonal tyrosine-incorporating Pyl *M. barkeri* pair compatible with a previously developed fluorescence-based screen for sense codon reassignment. The engineered orthogonal *M. barkeri* pair is capable of incorporating tyrosine in response an amber stop codon at the critical fluorophore tyrosine residue with 18.4% efficiency of a wild-type GFP in *E. coli* DH10B, and 12.6% of wild-type GFP in the strain of *E. coli* used for sense codon reassignment measurements to tyrosine with the orthogonal *M. jannaschii* pair.

The engineered tyrosine-incorporating *M. barkeri* pair was shown to be orthogonal; the *M. barkeri* tRNA was not aminoacylated by any *E. coli* tRNAs, and the *M. barkeri* aaRS variant did not aminoacylate any *E. coli* tRNAs. Furthermore, the substrate specificity of the engineered pair was characterized by electrospray ionization (ESI) mass spectrometry of an intact Z-domain of protein A produced by suppressing an amber stop codon with the engineered tyrosine-incorporating *M. barkeri* pair, and suggested that a small amount of phenylalanine was also incorporated in response to the amber stop codon by the orthogonal pair.

The *M. barkeri* tRNA/aaRS pair evolved for tyrosine incorporation was then applied to the fluorescence-based screen, enabling evaluation of the reassignment of eight sense codons as carried out using the orthogonal *M. jannaschii* tRNA/aaRS pair.

Reassignment efficiencies to tyrosine with the *M. barkeri* pair was above the limit of detection for each of the eight sense codons. However, the dynamic range of the fluorescence-based screen was significantly reduced when used with the *M. barkeri* pair. Improvement of the orthogonal tyrosine-incorporating *M. barkeri* pair by directed evolution is necessary to increase the dynamic range of the fluorescence-based screen.

As the biotechnological applications of systems with expanded genetic codes continue to expand in scope and complexity (e.g. 22 amino acid genetic codes and beyond), a quantitative, holistic understanding of the biological interactions that govern protein translation will foster the rational design of orthogonal translation components that behave in predictable ways.

REFERENCES

1. Dumas, A., et al., *Designing logical codon reassignment–Expanding the chemistry in biology.* Chemical science, 2015. **6**(1): p. 50-69.
2. Liu, C.C. and P.G. Schultz, *Adding new chemistries to the genetic code.* Annual review of biochemistry, 2010. **79**: p. 413-444.
3. Wang, L., et al., *Expanding the genetic code of Escherichia coli.* Science, 2001. **292**(5516): p. 498-500.
4. Srinivasan, G., C.M. James, and J.A. Krzycki, *Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA.* Science, 2002. **296**(5572): p. 1459-1462.
5. Lajoie, M.J., et al., *Genomically recoded organisms expand biological functions.* science, 2013. **342**(6156): p. 357-360.
6. Ostrov, N., et al., *Design, synthesis, and testing toward a 57-codon genome.* Science, 2016. **353**(6301): p. 819-822.
7. Kwon, I., K. Kirshenbaum, and D.A. Tirrell, *Breaking the degeneracy of the genetic code.* Journal of the American Chemical Society, 2003. **125**(25): p. 7512-7513.
8. Zeng, Y., W. Wang, and W.R. Liu, *Towards reassigning the rare AGG codon in Escherichia coli.* ChemBioChem, 2014. **15**(12): p. 1750-1754.
9. Mukai, T., et al., *Reassignment of a rare sense codon to a non-canonical amino acid in Escherichia coli.* Nucleic acids research, 2015. **43**(16): p. 8111-8122.

10. Lee, B.S., et al., *Incorporation of Unnatural Amino Acids in Response to the AGG Codon.* ACS Chem Biol, 2015. **10**(7): p. 1648-53.

11. Biddle, W., M.A. Schmitt, and J.D. Fisk, *Evaluating Sense Codon Reassignment with a Simple Fluorescence Screen.* Biochemistry, 2015. **54**(50): p. 7355-64.

12. Agris, P.F., *Decoding the genome: a modified view.* Nucleic acids research, 2004. **32**(1): p. 223-238.

13. Koonin, E.V. and A.S. Novozhilov, *Origin and evolution of the genetic code: the universal enigma.* IUBMB life, 2009. **61**(2): p. 99-111.

14. Schmitt, M.A., Biddle, Wil, and John D. Fisk, *Mapping the Plasticity of the E. Coli Genetic Code with Orthgonal Pair Directed Sense Codon Reassignment.* submitted, 2017.

15. Biddle, W., Schwark, David G., Schmitt, Margaret A. and John D. Fisk, *Sense Codon Reassignment of the Rare Arginine AGG Codon: Transferability of Improvements Between Orthogonal tRNAs and Synthetases.* submitted, 2018.

16. Biddle, W., M.A. Schmitt, and J.D. Fisk, *Modification of orthogonal tRNAs: unexpected consequences for sense codon reassignment.* Nucleic acids research, 2016: p. gkw948.

17. Heim, R. and R.Y. Tsien, *Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer.* Current biology, 1996. **6**(2): p. 178-182.

18. Tsien, R.Y., *The green fluorescent protein*. 1998, Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA.

19. Fechter, P., et al., *Major tyrosine identity determinants in Methanococcus jannaschii and Saccharomyces cerevisiae tRNATyr are conserved but expressed differently.* The FEBS Journal, 2001. **268**(3): p. 761-767.

20. Ambrogelly, A., et al., *Pyrrolysine is not hardwired for cotranslational insertion at UAG codons.* Proceedings of the National Academy of Sciences, 2007. **104**(9): p. 3141-3146.

21. Herring, S., et al., *The amino-terminal domain of pyrrolysyl-tRNA synthetase is dispensable in vitro but required for in vivo activity.* FEBS Lett, 2007. **581**(17): p. 3197-203.

22. Nozawa, K., et al., *Pyrrolysyl-tRNA synthetase-tRNA(Pyl) structure reveals the molecular basis of orthogonality.* Nature, 2009. **457**(7233): p. 1163-7.

23. Ko, J.H., et al., *Pyrrolysyl-tRNA synthetase variants reveal ancestral aminoacylation function.* FEBS Lett, 2013. **587**(19): p. 3243-8.

24. Takimoto, J.K., et al., *Stereochemical basis for engineered pyrrolysyl-tRNA synthetase and the efficient in vivo incorporation of structurally divergent non-native amino acids.* ACS Chem Biol, 2011. **6**(7): p. 733-43.

25. Wang, Y.S., et al., *The de novo engineering of pyrrolysyl-tRNA synthetase for genetic incorporation of L-phenylalanine and its derivatives.* Mol Biosyst, 2011. **7**(3): p. 714-7.

26. Wang, Y.S., et al., *A rationally designed pyrrolysyl-tRNA synthetase mutant with a broad substrate spectrum.* J Am Chem Soc, 2012. **134**(6): p. 2950-3.

27. Yanagisawa, T., et al., *Crystallographic studies on multiple conformational states of active-site loops in pyrrolysyl-tRNA synthetase.* J Mol Biol, 2008. **378**(3): p. 634-52.

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

29. Young, T.S., et al., *An enhanced system for unnatural amino acid mutagenesis in E. coli.* Journal of molecular biology, 2010. **395**(2): p. 361-374.

30. Fan, C., et al., *Rationally evolving tRNAPyl for efficient incorporation of noncanonical amino acids.* Nucleic Acids Res, 2015. **43**(22): p. e156.

31. Braisted, A.C. and J.A. Wells, *Minimizing a binding domain from protein A.* Proceedings of the National Academy of Sciences, 1996. **93**(12): p. 5688-5692.

CHAPTER 4

DIRECTED EVOLUTION OF A TYROSINE-INCORPORATING PYRROLYSINE *M. BARKERI* tRNA/aaRS PAIR FOR QUANTITATIVE MEASUREMENT OF SENSE CODON REASSIGNMENT IN *E. COLI* USING A SIMPLE FLUORESCENCE-BASED SCREEN

4.1 CHAPTER OVERVIEW

Genetic code expansion for incorporation of non-canonical amino acids (ncAAs) into proteins is a powerful strategy for increasing the chemical diversity in proteins. Recently the most widely employed method of genetic code expansion utilizes an engineered orthogonal tRNA/aminoacyl-tRNA synthetase (aaRS) pair to direct incorporation of ncAAs at an amber stop codon. Further expansion of the genetic code to 22 or more amino acids requires that additional codons be reassigned. Breaking the degeneracy of the genetic code at sense codons with an orthogonal tRNA/aaRS is an emerging technique for genetic code expansion.

The previous chapter described the engineering of an orthogonal tyrosine-incorporating pyrrolysyl *Methanosarcina barkeri* (*M. barkeri*) tRNA/aaRS pair compatible with a fluorescence-based screen for sense codon reassignment efficiency. The engineered *M. barkeri* pair was capable of incorporating tyrosine in response to an amber stop codon at position 66 in GFP with 12.6% efficiency of wild-type GFP in *E. coli* SB3930. Measurable sense codon reassignment to tyrosine above the limit of detection was also observed at 8 sense codons with the engineered *M. barkeri* pair; however, the

dynamic range of the fluorescence-based screen was lower than when used with another orthogonal tRNA/aaRS pair from *Methanocaldococcus jannaschii* (*M. jannaschii*).

This chapter describes the improvement of amber stop codon suppression efficiency with the tyrosine-incorporating orthogonal *M. barkeri* pair through directed evolution. After 2 rounds of error-prone PCR of the tyrosine-incorporating *M. barkeri* aaRS and high-throughput screening for improved amber stop codon suppression by fluorescence-activated cell sorting, a variant that was capable of suppressing an amber stop codon at position 66 in GFP with 92.5% efficiency in *E. coli* DH10B was identified. In addition to the 5.2-fold improvement in amber stop codon suppression efficiency, the engineered *M. barkeri* pair showed improved substrate specificity compared to the starting tyrosine-incorporating *M. barkeri* pair.

The improved orthogonal tyrosine-incorporating *M. barkeri* pair was then used to measure codon reassignment efficiencies to tyrosine at the amber stop codon and eight sense codons in *E. coli* SB3930. Although the amber stop codon suppression efficiency decreased from 92.5% in *E. coli* DH10B to 65.1% in *E. coli* SB3930, the dynamic range of the fluorescence-based screen was improved. Reassignment efficiencies at the eight sense codons with the improved orthogonal *M. barkeri* pair were all above the level of quantification, and were improved from 3.1-10.6 fold. The improved orthogonal *M. barkeri* pair should be capable of making the same suite of sense codon reassignment measurements that were made with the orthogonal *M. jannaschii* pair. Generation of a second sense codon reassignment data set with the improved orthogonal *M. barkeri*

pair will identify promising sense codons for reassignment to ncAAs, and may also assist in further understanding of the factors involved in the process of translation.

4.2 INTRODUCTION

The 20 canonical amino acids are sufficient for living systems, but encode a limited set of chemical functionalities. Incorporation of non-canonical amino acids (ncAAs) into proteins that contain chemistries not found in nature significantly expands the ability to understand and modify proteins. Multiple methods for the incorporation of ncAAs have been developed, the most widely used of which is the suppression of amber stop codons by an engineered orthogonal tRNA and aminoacyl tRNA synthetase (aaRS) pair [1-3]. The most commonly used orthogonal pairs for amber stop codon suppression are the *Methanocaldococcus jannaschii* (*M. janaschii*) tyroysyl tRNA/aaRS and the *Methanosarcina barkeri* (*M. barkeri*) pyrrolysyl tRNA/aaRS. These pairs have been repeatedly engineered, and have been used to incorporate over 150 different ncAAs in response to the amber stop codon [4].

In vivo incorporation of multiple ncAAs at multiple sites remains a significant challenge. Although significant improvements have been made for *in vivo* systems that incorporate ncAAs in response to the amber stop codon by genomic recoding and removal of ribosomal release factor 1, these approaches are not easily extendable to other codons [5, 6]. Further expansion of the genetic code to 22 or more amino acids requires the reassignment of additional codons. The genetic code is constrained by the fact that all 64 codons in the genetic code have assigned functions. Although each codon has an assigned function, the genetic code is degenerate. The 20 amino acids

are encoded by 61 of the 64 codons, and reassigning one or more of the 61 sense codons expand the genetic code to 22 or more amino acids [7-10].

The Fisk lab recently developed a method to quickly measure sense codon reassignment efficiency *in vivo*. The method is based on the absolute requirement of tyrosine at position 66 for fluorophore maturation in green fluorescent protein (GFP) [7]. The fluorescence-based screen was used to measure sense codon reassignment efficiencies at 30 sense codons in *E. coli* using the orthogonal *M. janaschii* tyrosyl tRNA/aaRS pair [[11],Chapter 2]. Importantly, all codons measured were reassigned with efficiencies ranging from 0.8% to 70% of a wild-type GFP expressing, non-reassigning control. Systems with high levels of reassignment efficiency to tyrosine may be used directly for ncAA incorporation and lower levels of sense codon reassignment may be improved through directed evolution.

In addition to mapping the landscape of sense codon reassignment efficiencies for the *M. janaschii* orthogonal pair, the fluorescence-based screen has been applied to further understand the relative importance of the factors affecting the process of translation in *E. coli.* Measurements of sense codon reassignment efficiencies with the *M. janaschii* orthogonal pair suggested that orthogonal pair aminoacylation efficiency and competing tRNA concentrations both weakly correlate with sense codon reassignment efficiencies [Chapter 2].

Increased understanding of the relative importance of aminoacylation efficiency may be achieved by collecting additional data from another orthogonal pair system. The other commonly used system for ncAA introduction at amber codons, the pyrrolysine tRNA/aaRS from the *Methanosarcina* species, is an attractive option for multiple

reasons. Unlike the *M. janaschii* pair, the anticodon of $tRNA^{Pyl}$ is not a critical identity element for recognition by the corresponding aaRS (changing the tRNA anticodon sequence of the *M. janaschii* tRNA is expected to decrease aminoacylation efficiency in most cases), and comparison of differences in sense codon reassignment efficiencies between the two systems may shed light on the role of aminoacylation efficiency in sense codon reassignment by an orthogonal pair [12, 13].

The orthogonal *M. barkeri* pair was not initially compatible with the fluorescence-based screen for sense codon reassignment. Chapter 3 described the engineering of a tyrosine-incorporating *M. barkeri* pyrrolysyl tRNA/aaRS pair compatible with the fluorescence-based screen. The engineered *M. barkeri* aaRS aminoacylated tyrosine onto the corresponding tRNA, and was capable of suppressing an amber stop codon at the fluorophore position 66 in GFP at 12.6% of a wild-type non-suppressing control in *E. coli* SB3930 [Chapter 3]. The orthogonal pair was then used to measure sense codon reassignment at 8 different sense codons that represented the range of efficiencies observed with the orthogonal *M. janaschii* pair [11]. All 8 sense codon reassignment measurements were above the limit of detection at the Watson-Crick base-pairing codon. However, the dynamic range of the fluorescence-based screen with the engineered *M. barkeri* tRNA/aaRS pair is smaller than when used with the *M. janaschii* tRNA/aaRS pair, and 4 of the 8 measurements were below the limit of quantification.

The smaller dynamic range of the GFP screen is potentially related to reduced stability, low aminoacylation efficiency or polysubstrate specificity of the engineered tyrosine-incorporating *M. barkeri* aaRS. The engineered tyrosine-incorporating *M. barkeri* tRNA/aaRS pair, while of similar efficiency to other engineered *M. janaschii* and

M. barkeri pairs, is approximately 4-fold less efficient than the *M. janaschii* tyrosyl tRNA/aaRS pair that is a naturally evolved tyrosl aaRS [11]. Improving the stability or the aminoacylation efficiency of the engineered aaRS may improve the dynamic range of the fluorescence-based screen by increasing the cellular concentration of the charged $tRNA^{Pyl}$. Additionally, a small amount of phenylanaine was also charged onto $tRNA^{Pyl}$. The sensitivity of the fluorescence-based sense codon reassignment screen with the engineered *M. barkeri* orthogonal pair may be improved by increasing the specificity of the aaRS.

Directed evolution strategies to improve enzyme function involve rounds of diversification and selection based on function [14]. The originally selected tyrosine activating pyrrolysine aaRS was selected from a focused library with diversity at just six sites within the amino acid binding pocket of the enzyme. To improve the function of the initially selected tyrosine activating variant additional random mutations were introduced and the resulting libraries were selected using the fluorescence activated cell sorting (FACS) based GFP reporter screen [7]. A library of tyrosine-incorporating Pyl *M. barkeri* aaRS variants generated by error-prone PCR (EP-PCR) may potentially contain variants with improved amber suppression abilities. Random mutagenesis by EP-PCR is an attractive method because it makes mutations throughout the sequence of the gene coding for the aaRS [15]. Mutations outside of the amino acid binding pocket that improve the stability, function, or specificity of the aaRS and would otherwise be difficult to identify may be made. An EP-PCR library of aaRSs may also be constructed quickly using well-documented EP-PCR and cloning methods, and the fluorescence-based

screen developed previously may be used in a high-throughput format to identify clones with improved amber stop codon suppression abilities [7, 16-19].

Two rounds of random mutagenesis by error-prone PCR and high-throughput screening by fluorescence activated cell sorting (FACS) yielded a variant for which amber stop codon suppression efficiency was improved 5.2-fold over the starting tRNA/aaRS pair in *E. coli* SB3930. The improved orthogonal pair is capable of suppressing an amber stop codon at the fluorophore tyrosine 66 in GFP at 65% efficiency of a wild-type GFP expressing, non-suppressing control, a level that is near the suppression capabilities of the orthogonal *M. janaschii* pair. In addition, the specificity of the orthogonal was improved. The orthogonal *M. barkeri* pair was then used to measure sense codon reassignment at the same set of sense codons previously measured with the original engineered tyrosine-incorporating *M. barkeri* tRNA/aaRS pair. Sense codon reassignment efficiencies were improved 3.1-10.6 fold with the improved orthogonal pair, and should be capable of making the same sense codon reassignment measurements that were made with the orthogonal *M. janaschii* pair.

4.3 MATERIALS AND METHODS

The gain of function through codon reassignment fluorescence-based screen has been described [7]. Detailed protocols, including methods for error-prone PCR and library construction, high-throughput screening by fluorescence activated cell sorting (FACS), measurement of system growth and fluorescence, calculation of sense codon reassignment efficiencies and mass spectrometry are included in the supporting information. The supporting information also includes full vector sequences,

oligonucleotide primer sequences, detailed cell strain information, and general reagents and materials.

4.4 RESULTS AND DISCUSSION

Directed evolution of a tyrosine-incorporating variant of the M. barkeri Pyl tRNA/aaRS pair for improved amber stop codon suppression efficiency

The template used for library generation was the tyrosine-incorporating *M. barkeri* Pyl aaRS previously described in Chapter 3. The aaRS sequence was amplified from the plasmid using two different EP-PCR conditions that were predicted to produce an average of 1-4 nucleotide mutations per 1000 [16-19]. Each mutagenic PCR product was then restricted and ligated into a plasmid backbone that contained the gene for the corresponding *M. barkeri* $tRNA^{Pyl}_{CUA}$ to generate two libraries. Library 1 contained *M. barkeri* aaRS variants generated by an EP-PCR method using $MnCl_2$ and unbalanced concentrations of each of the four nucleotides [16]. Library 2 contained *M. barkeri* aaRS variants generated by an EP-PCR method with unbalanced concentrations of each of the four nucleotides [17-19].

The mutagenic PCR products replaced an inactive version of the *M. barkeri* aaRS that contained 3 KpnI restriction sites. To reduce the ability of the inactive aaRS to ligate back into the plasmid, the backbone was digested with KpnI (in addition to the enzymes used to insert the library of aaRSs). The plasmids containing the aaRS libraries were transformed into cells harboring a UAG stop codon at the fluorophore position of GFP. A small portion of each library transformation was plated onto LB agar plates to determine the number of transformants and mutation frequency of each EP-PCR reaction. Each library was then split in half; one half of each library was expressed

overnight at 37^0C (libraries 1a and 2a), and the other half of each library was expressed at 42^0C (libraries 1b and 2b). The idea behind expression at a higher temperature was to potentially identify *M. barkeri* aaRS clones containing stabilizing mutations. The following morning, each library aliquot was frozen at -80^0C for selection by fluorescence assisted cell sorting (FACS).

The first round of EP-PCR yielded $1.2 \cdot 10^7$ unique transformants for library 1, and $1.7 \cdot 10^7$ unique transformants for library 2. Twelve colonies were randomly selected from the transformation efficiency plates of each library for analysis by colony PCR. All 24 colonies analyzed by colony PCR were library members (not cut by KpnI). The PCR products from 10 of the colonies from each library were sent for DNA sequencing. An average mutation rate of 1.3 mutations per 1,000 nucleotides was observed for library 1 (EP-PCR using $MnCl_2$), and an average mutation rate of 2.2 mutations per 1,000 nucleotides was observed for library 2 (EP-PCR using unbalanced nucleotides only). The cells containing the libraries of mutagenized aaRSs were then evaluated for incorporation of tyrosine in response to UAG codon at position 66 in a GFP reporter by high-throughput FACS sorting.

In the first round of fluorescence-activated cell sorting (FACS), approximately $1.0 \cdot 10^7$ cells were screened for libraries 1a, 1b, 2a, and 2b. Fluorescent clones representing the top 1% (approximately 0.4% of each population) were collected. A portion of the collected cells for each library was plated directly onto plates for analysis of individual clones. The remaining amount was amplified overnight, and the plasmid containing the orthogonal *M. barkeri* translational machinery components was extracted to perform a second round of EP-PCR on the pool of clones from the first round.

The following day, twelve fluorescent colonies from each of the four plates (48 colonies total) were selected and the *in vivo* fluorescence-based screen was used to analyze the amber stop codon suppression efficiency. The range of amber stop codon suppression efficiencies for the 48 individual clones was from below the detection limit to 42% of wild-type GFP (Figure 4.1, Table 4.1). The plasmids containing the orthogonal translational components for the 12 individual clones with the highest amber stop codon suppression efficiencies were isolated and sequenced (Table 4.1).

Table 4.1. Amber stop codon suppression efficiencies and mutations relative to the starting sequence of the top 12 clones identified from the first round of EP-PCR and selection (a=expression at 37^0C, b=expression at 42^0C)

Clone Name / Library	Amino Acid Mutations Relative to Starting *M. barkeri* Tyr Variant	Amber Stop Codon Suppression Efficiency of Single Clone (% of WT-GFP)
Clone G5 / Library 1a	Ile 26 Val, Cys 346 Ser	28.9
Clone C6 / Library 1a	Met 17 Leu	36.3
Clone G6 / Library 1a	Thr 269 Ser, Glu 302 Ala	39.5
Clone B7 / Library 1b	Pro 186 Leu	25.1
Clone B8 / Library 1b	None	19.9
Clone E8 / Library 1b	Cys 319 Ser	22.9
Clone G9 / Library 2a	Thr 269 Ser, Ser 416 Cys	33.4
Clone E10 / Library 2a	Val 370 Ile	30.1
Clone G10 / Library 2a	Asn 202 Thr, Ile 385 Phe	26.4
Clone C11 / Library 2b	Asp 351 Glu	23.5
Clone G11 / Library 2b	Thr 269 Ala, Phe 399 Tyr	42.3
Clone D12 / Library 2b	Glu 170 Gly, Thr 190 Ala	22.0

In the 12 most efficient clones identified after one round of EP-PCR and selection, the 3 of the 4 most efficient clones (G6, G9, and G11) had a mutation at position 269, suggesting that it is an important residue for either structure or function of the tyrosine-incorporating *M. barkeri* aaRS variant (Table 4.1).

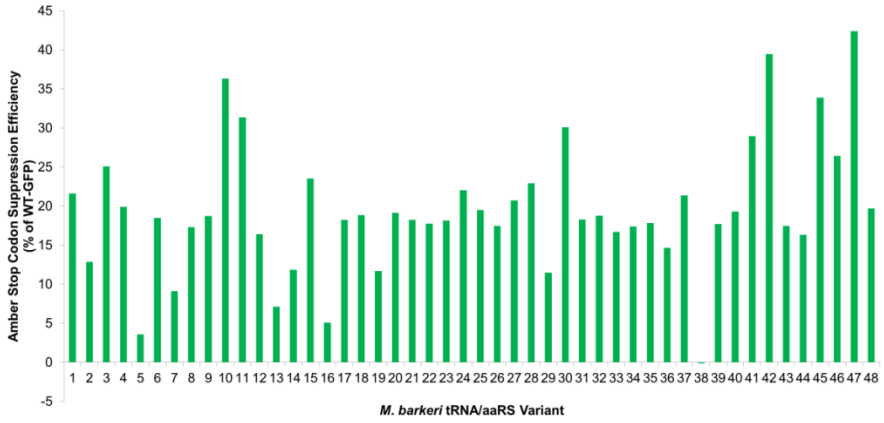

Figure 4.1. Amber stop codon suppression efficiency (% of WT-GFP) for 48 individual clones identified after 1 round of EP-PCR and selection. No error bars are present because this analysis is of 48 potentially different clones.

Position 269 is located in a helix that lines the amino acid binding pocket, and may re-position the helix to better accommodate tyrosine (Figure 4.2) [12]. Other mutations that appear to be either the structural or functional improvements of the tyrosine-incorporating *M. barkeri* aaRS variant include position Met 17 Leu, and Val 370 Ile.

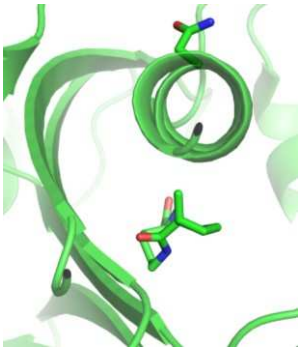

Figure 4.2. PyMOL cartoon of the active site of the *M. mazei* pyrrolysyl-aaRS showing pyrrolysine in the active site, and position 269 (PDB 2ZCE) [12]. The *Methanosarcina mazei* aaRS is used here because no crystal structure of the *M. barkeri* pyrrolysyl aaRS currently exists; position 269 is Thr in the *M. barkeri* aaRS, and Asn in the *M. mazei* aaRS.

For the second round of EP-PCR, the pool of clones isolated from the first round of FACS from libraries 2a and 2b, and the 2 clones showing the highest amber stop codon suppression efficiency (Clones G6 and G11) from the first round of EP-PCR and selection were used as the starting templates, and the EP-PCR protocol that generated on average 2.2 mutations per 1,000 nucleotides was used to generate four new libraries of *M. barkeri* aaRS variants (Library2a.2, Library 2b.2, Library G6.2, and Library G11.2) [17-19]. Each mutagenic PCR product was restricted and ligated into the same plasmid backbone used to construct the first set of libraries. The plasmids containing the aaRS libraries were transformed into cells harboring a UAG stop codon at the fluorophore position of GFP. A small portion of each library transformation was plated onto LB agar plates to determine the number of unique transformants for each library, and the remaining amount of each library was expressed overnight at 37^0C. The following morning, aliquots of each library were saved at -80^0C for future use. The library transformations yielded $4.0 \cdot 10^6$ unique transformants for library 2a.2, $9.6 \cdot 10^6$ unique transformants for library 2b.2, $4.8 \cdot 10^6$ unique transformants for library G6.2, and $5.2 \cdot 10^6$ unique transformants for library G11.2. In the second round of FACS, approximately $1.0 \cdot 10^7$ cells were screened for each library, and fluorescent clones representing the top 1% (approximately 0.4% of each population) of each library were collected. A portion of the collected cells was plated directly for analysis of individual clones, and the remaining amount of collected cells were amplified overnight and frozen at -80^0C.

The following day, fifteen fluorescent colonies from each of the four plates (60 colonies total) were selected and the *in vivo* fluorescence-based screen was used to analyze the amber stop codon suppression efficiency to tyrosine. The range of amber

stop codon suppression efficiencies for the 60 individual clones was from below the detection limit to 92.5% of wild-type GFP (Figure 4.3). The plasmids containing the orthogonal translational components for the 10 individual clones with the highest amber stop codon suppression efficiencies were isolated and sequenced.

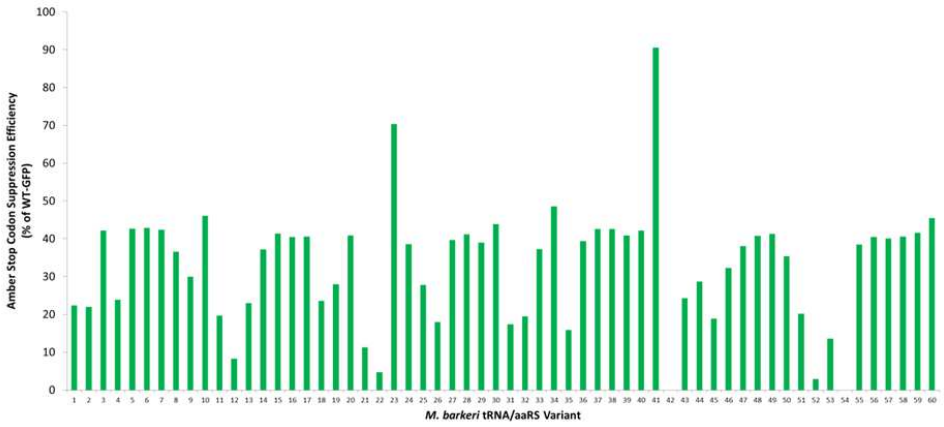

Figure 4.3. Amber stop codon suppression efficiency (% of WT-GFP) for 60 individual clones identified after 2 rounds of EP-PCR and selection. No error bars are present because this analysis is of 60 potentially different clones.

In the 10 most efficient clones identified after two rounds of EP-PCR and selection, the two most efficient clones (F3 and D5) both had a Met 265 Thr mutation, suggesting that it is an important residue for either structure or function of the tyrosine-incorporating *M. barkeri* aaRS variant (Table 4.2). The Met 265 residue that is mutated to Thr in clones F3 and D5 is positioned in the amino acid binding pocket near the amino and carboxy termini of the amino acid, and may aid in binding of the amino acid (Figure 4.4) [12]. The role of the other amino acid mutations in the top 2 identified clones after 2 rounds of EP-PCR and selection are not clear have not been further investigated yet.

Table 4.2. Amber stop codon suppression efficiencies and mutations relative to the starting sequence of the top 10 clones identified from the second round of EP-PCR and selection.

Clone Name / Library	Amino Acid Mutations Relative to Starting *M. barkeri* Tyr Variant	Amber Stop Codon Suppression Efficiency of Single Clone (% of WT-GFP)
Clone F3 / Library 2a.2	Ser 121 Cys, Lys 125 Arg, Met 265 Thr, Glu 409 Val	92.5
Clone D5 / Library 2b.2	Met 265 Thr, Asn 323 Thr	64.9
Clone B5 / Library 2b.2	Thr 130 Pro, Ser 168 Thr, Thr 269 Ala	37.6
Clone E6 / Library 2b.2	Thr 269 Ser, Asp 335 Glu	47.1
Clone B7 / Library G6.2	Thr 269 Ser, Glu 302 Ala	37.7
Clone B8 / Library G6.2	Thr 269 Ser, Glu 302 Ala,	38.3
Clone E10 / Library G11.2	Ser 158 Gly, Thr 269 Ala, Phe 399 Tyr	41.7
Clone G11 / Library G11.2	Thr 269 Ala, Phe 399 Tyr	40.2
Clone B12 / Library G11.2	Thr 269 Ala, Phe 399 Tyr	39.8
Clone G12 / Library G11.2	Thr 269 Ala, Phe 399 Tyr	42.0

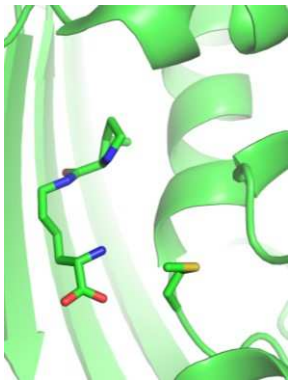

Figure 4.4. PyMOL cartoon of the active site of the *M. mazei* pyrrolysyl-aaRS showing pyrrolysine in the active site, and Met 265 (PDB 2ZCE) [12]. The *Methanosarcina mazei* aaRS is used here because no crystal structure of the *M. barkeri* pyrrolysyl aaRS currently exists.

After only 2 rounds of EP-PCR and high-throughput selection by FACS, the ability of the improved tyrosine-incorporating Pyl *M. barkeri* aaRS variant and corresponding tRNA to suppress an amber stop codon was improved 5.2-fold compared to the starting tyrosine-incorporating Pyl *M. barkeri* pair (Figure 4.5). The fact that multiple variants were identified after 2 rounds of EP-PCR and high-throughput selection suggests that additional rounds of EP-PCR and screening may yield variants with further improvements. Future work may focus on further investigating the relative importance of the mutational positions identified.

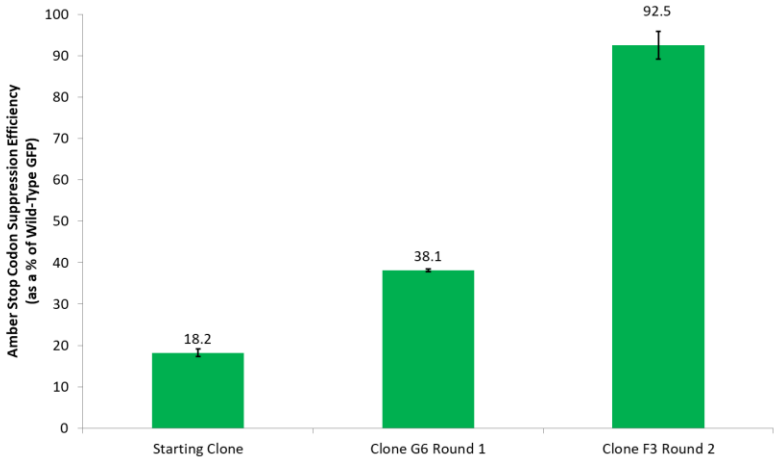

Figure 4.5. Amber stop codon suppression efficiencies of the most efficient *M. barkeri* aaRS clones identified as a percentage of wild-type GFP vs. round of mutagenesis and screening.

Evaluation of amino acid specificity of the improved M. barkeri aaRS variant

The improved *M. barkeri* Pyl aaRS variant incorporated tyrosine in response to an amber stop codon with 92.5% efficiency in *E. coli* DH10B. However, because the *in vivo* screen is based upon the restoration of GFP fluorescence and is specific for tyrosine, it does not provide a measure of the specificity of the improved tyrosine-

incorporating *M. barkeri* aaRS. It is possible for the improved variant to also aminoacylate other canonical amino acids onto the $tRNA^{Pyl}_{CUA}$, as was observed with the starting *M. barkeri* variant. The most likely canonical amino acids that could also serve as substrates for the improved tyrosine-incorporating *M. barkeri* aaRS variant include phenylalanine, histidine, and tryptophan [20].

The improved tyrosine-incorporating *M. barkeri* variant's substrate specificity was further investigated by using electrospray ionization mass spectrometry (ESI-MS) to identify the amino acids incorporated in response to a single UAG codon at position 5 in the Z-domain of protein A [21]. The Z-domain of protein A is a small, soluble 8.3 kDa protein that was used in previous experiments to identify the amino acids incorporated as a result of UAG suppression with the original tyrosine-incorporating *M. barkeri* aaRS variant [8, 21, 22]. The improved *M. barkeri* aaRS variant was used to suppress a UAG codon at position 5 in the Z-domain of protein A. A positive control Z-domain protein (with a tyrosine codon UAC at position 5) was also produced in the presence of the improved *M. barkeri* aaRS variant. Purified, intact Z-domain proteins produced from the positive control and amber suppression systems were analyzed by ESI-MS. The spectra were deconvoluted using the maximum entropy method (MassHunter Software, Agilent Technologies). The deconvoluted Z-domain spectra from the positive control and amber suppression systems were compared to each other and to Z-domain of protein A spectra generated from systems that used the original tyrosine-incorporating *M. barkeri* aaRS variant.

The predicted mass of the Z-domain of protein A with tyrosine at position 5 is 8315 Da. The +14 peak observed at 8329 Da in the positive-control sample is likely a

result of a methylation, potentially at a lysine residue [23]. The 8315 and 8329 peaks are also observed in the UAG suppression system, strongly suggesting that the engineered *M. barkeri* aaRS variants direct incorporation of tyrosine in response to a UAG codon (Figure 4.6, left). Significantly, the -16 peak at 8299 Da representing incorporation of phenylalanine in response to a UAG codon that is observed in the spectra of the Z-domain of protein A produced via UAG suppression with the original tyrosine-incorporating *M. barkeri* aaRS variant is absent when the improved tyrosine-incorporating *M. barkeri* aaRS variant is used (Figure 4.6, right). The absence of the 8299 Da peak shows that in addition to improving the overall amber stop codon suppression efficiency of the tyrosine-incorporating *M. barkeri* variant, the substrate specificity of the aaRS has also been improved.

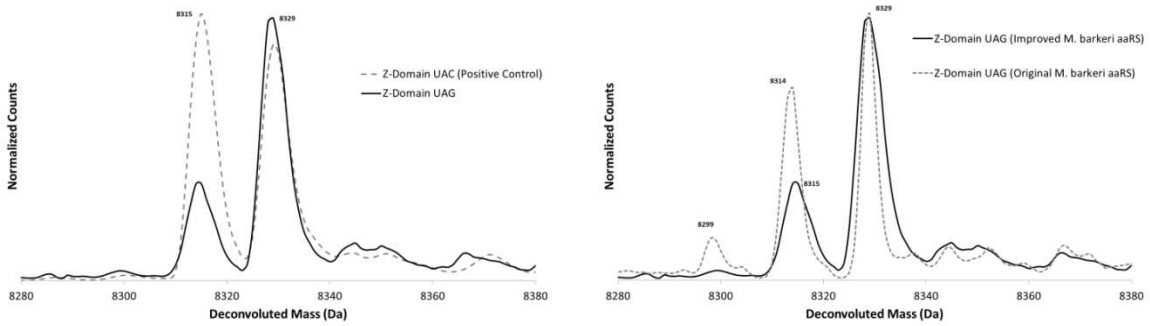

Figure 4.6. ESI-MS of intact Z-domain proteins. Left: Overlay of Z-domain Tyr (dashed) and Z-Domain UAG (solid). Right: Overlay of Z-domain UAG produced with the original M. *barkeri* aaRS (dashed) and Z-domain UAG produced with the improved *M. barkeri* aaRS (solid).

Sense codon reassignment utilizing the improved tyrosine-incorporating M. barkeri aaRS variant

The original tyrosine-incorporating *M. barkeri* Pyl aaRS was capable of suppressing an amber stop codon at the tyrosine 66 in GFP with 92.5% efficiency in *E. coli* DH10B. The EP-PCR aaRS libraries were generated and screened in *E coli* DH10B, but further sense codon reassignment efficiencies of the improved tyrosine-incorporating Pyl *M. barkeri* tRNA/aaRS variant was carried out in *E. coli* SB3930 to allow direct comparison to the suite of sense codon reassignment data by variants of the *M. jannaschii* tyrosyl tRNA/aaRS pair. The amber stop codon suppression efficiency of the improved tyrosine-incorporating Pyl *M. barkeri* tRNA/aaRS decreased when identical systems were used from 92.5% in DH10B to 65.2% in SB3930 (Figure 4.7). The observed decrease in amber stop codon suppression efficiency may result from differences in expression of orthogonal translational components, reporter proteins, or endogenous translational machinery between the two strains. The choice of *E. coli* host strain may significantly affect codon reassignment efficiencies, and future experiments may address the variability in codon reassignment efficiencies for identical systems in multiple strains of *E. coli*.

The same 8 sense codons used to evaluate the original tyrosine-incorporating Pyl *M. barkeri* tRNA/aaRS were selected for evaluating the improved tyrosine-incorporating Pyl *M. barkeri* tRNA/aaRS using the fluorescence-based screen The 8 codons span the range of reassignment efficiencies observed using the *M. jannaschii* tRNA/aaRS pair and include both *E. coli* wobble codons and rarely used codons.

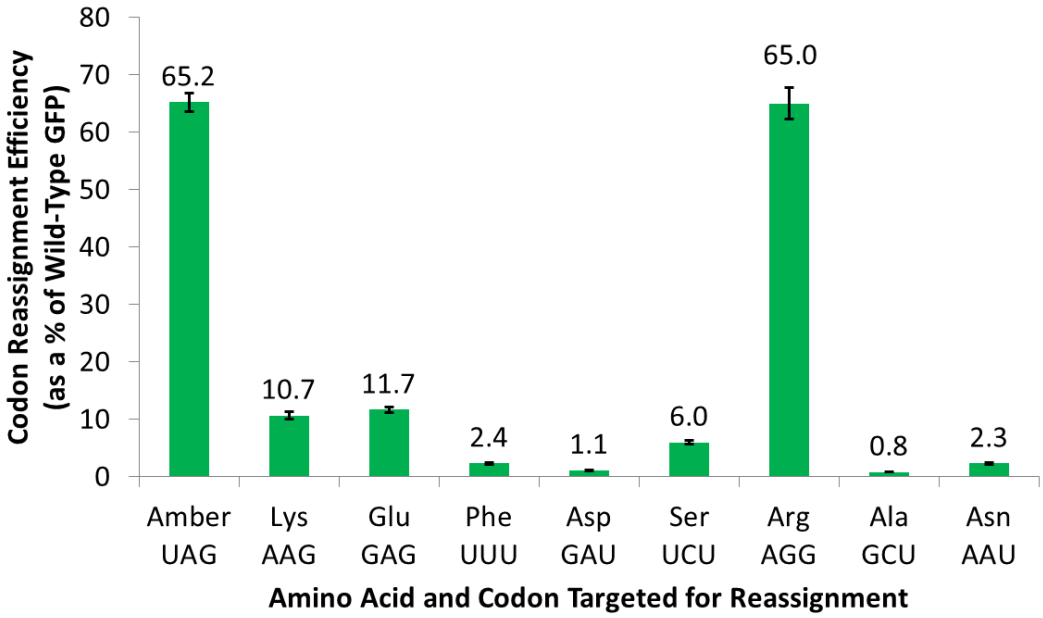

Figure 4.7. Sense codon reassignment efficiency with the improved tyrosine-incorporating Pyl *M. barkeri* tRNA/aaRS as a percentage of wild-type GFP vs. the amino acid and specified codon targeted for reassignment.

The improved *M. barkeri* tRNA/aaRS pair reassigned the selected codons to tyrosine with efficiencies between 0.8% and 65.0% in *E. coli* SB3930 (Figure 4.7, Table 4.3). Based on the variation across 6 clones of the non-reassigning negative control, the limit of detection for the fluorescence-based screen using the improved *M. barkeri* tRNA/aaRS pair is 0.14%. The dynamic range of the fluorescence-based screen with improved tyrosine-incorporating Pyl *M. barkeri* tRNA/aaRS pair is sufficient to make the same set of sense codon reassignment measurements that were made with the *M. janaschii* Tyr tRNA/aaRS pair.

Table 4.3. Codon reassignment efficiencies for the starting tyrosine-incorporating Pyl *M. barkeri* aaRS, the improved *M. barkeri* aaRS variant, and the fold-increase in reassignment efficiencies between the two aaRSs. Measurements were made in *E. coli* SB3930. Error bars are the standard deviation of 6 colonies.

tRNA Anticodon	Codon Targeted	Reassignment of Codon with the Starting aaRS (% of WT-GFP)	Reassignment of Codon with the Improved aaRS (% of WT-GFP)	Fold Increase
$tRNA^{Pyl}_{CUA}$	Amber UAG	12.6 ± 0.5	65.2 ± 1.6	5.2 ± 0.2
$tRNA^{Pyl}_{CCU}$	Lys AAG	1.9 ± 0.1	10.7 ± 0.6	5.6 ± 0.4
$tRNA^{Pyl}_{CUC}$	Glu GAG	1.1 ± 0.1	11.7 ± 0.5	10.6 ± 0.8
$tRNA^{Pyl}_{AAA}$	*Phe UUU	0.6 ± 0.0	2.6 ± 0.1	3.9 ± 0.3
$tRNA^{Pyl}_{AUC}$	*Asp GAU	0.3 ± 0.0	1.1 ± 0.1	3.3 ± 0.2
$tRNA^{Pyl}_{AGA}$	Ser UCU	1.1 ± 0.1	6.0 ± 0.3	5.6 ± 0.5
$tRNA^{Pyl}_{CCU}$	Arg AGG	13.6 ± 0.6	65.0 ± 2.8	4.8 ± 0.3
$tRNA^{Pyl}_{AGC}$	*Ala GCU	0.2 ± 0.0	0.9 ± 0.1	4.8 ± 0.9
$tRNA^{Pyl}_{AUU}$	*Asn AAU	0.7 ± 0.1	2.3 ± 0.1	3.1 ± 0.3

* Indicates the measurement of sense codon reassignment efficiency with the starting aaRS was above the limit of detection, but below the limit of quantification.

The fold-increase in codon reassignment efficiencies between the starting tyrosine-incorporating Pyl *M. barkeri* aaRS and the improved tyrosine incorporating Pyl *M. barkeri* aaRS ranged from 3.1 to 10.6 fold. Sense codon reassignment efficiencies measured with the original aaRS for the Phe UUU, Asp GAU, Ala GCU, and Asn AAU codons were below the limit of quantification [24]. After improvement, sense codon reassignment efficiencies measured at all 8 codons were above the limit of quantification. The improved orthogonal tyrosine-incorporating *M. barkeri* pair should be able to make the same set of measurements that were previously made with the orthogonal *M. jannaschii* pair.

4.5 CONCLUSIONS

A tyrosine-incorporating Pyl *M. barkeri* aaRS variant that was capable of suppressing an amber stop codon at position 66 in GFP with 65.2% efficiency in *E. coli* SB3930 was identified after two rounds of error-prone PCR and high-throughput screening. The 65.2% efficiency was a 5.2-fold improvement over the starting aaRS, and is near the efficiency of the other commonly used orthogonal pair from *M. janaschii*. Furthermore, the specificity of the tyrosine-incorporating *M. barkeri* aaRS was improved compared to the starting aaRS; no dectable amount of phenylalanine was present in the ESI-MS spectrum of an intact Z-domain protein with an amber stop codon suppressed at position 5 with the improved tyrosine incorporating Pyl *M. barkeri* variant.

The improved tyrosine-incorporating Pyl *M. barkeri* pair was then used along with the fluorescence-based screen to measure reassignment efficiencies at 8 sense codons. The sense codon reassignment measurements ranged from 0.8% to 65.0% of a wild-type GFP expressing, non-reassigning control. The limit of detection with the improved orthogonal pair is 0.14%, and the dynamic range of the fluorescence-based screen with the improved pair is sufficient to make the same sense codon measurements that were made with the *M. janaschii* tRNA/aaRS pair.

Systems with expanded genetic codes that contain 22 amino acids or beyond open the door for increasingly useful and complex biotechnological applications. In order to build efficient systems with 22 or more amino acid genetic codes, an improved understanding of the factors that affect protein translation and how they relate to orthogonal pair directed codon reassignment is needed. Measurement of sense codon reassignment efficiencies with the orthogonal *M. barkeri* and *M. janaschii* pairs using the

fluorescence-based screen has quickly identified multiple promising codons for reassignment to ncAAs.

REFERENCES

1. Furter, R., *Expansion of the genetic code: site-directed p-fluoro-phenylalanine incorporation in Escherichia coli.* Protein Science, 1998. **7**(2): p. 419-426.
2. Wang, L., et al., *Expanding the genetic code of Escherichia coli.* Science, 2001. **292**(5516): p. 498-500.
3. Srinivasan, G., C.M. James, and J.A. Krzycki, *Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA.* Science, 2002. **296**(5572): p. 1459-1462.
4. Dumas, A., et al., *Designing logical codon reassignment–Expanding the chemistry in biology.* Chemical science, 2015. **6**(1): p. 50-69.
5. Lajoie, M.J., et al., *Genomically recoded organisms expand biological functions.* science, 2013. **342**(6156): p. 357-360.
6. Ostrov, N., et al., *Design, synthesis, and testing toward a 57-codon genome.* Science, 2016. **353**(6301): p. 819-822.
7. Biddle, W., M.A. Schmitt, and J.D. Fisk, *Evaluating Sense Codon Reassignment with a Simple Fluorescence Screen.* Biochemistry, 2015. **54**(50): p. 7355-64.
8. Lee, B.S., et al., *Incorporation of Unnatural Amino Acids in Response to the AGG Codon.* ACS Chem Biol, 2015. **10**(7): p. 1648-53.
9. Kwon, I., K. Kirshenbaum, and D.A. Tirrell, *Breaking the degeneracy of the genetic code.* Journal of the American Chemical Society, 2003. **125**(25): p. 7512-7513.

10. Mukai, T., et al., *Reassignment of a rare sense codon to a non-canonical amino acid in Escherichia coli.* Nucleic acids research, 2015. **43**(16): p. 8111-8122.

11. Schmitt, M.A., Biddle, Wil, and John D. Fisk, *Mapping the Plasticity of the E. Coli Genetic Code with Orthgonal Pair Directed Sense Codon Reassignment.* submitted, 2017.

12. Yanagisawa, T., et al., *Crystallographic studies on multiple conformational states of active-site loops in pyrrolysyl-tRNA synthetase.* J Mol Biol, 2008. **378**(3): p. 634-52.

13. Ambrogelly, A., et al., *Pyrrolysine is not hardwired for cotranslational insertion at UAG codons.* Proceedings of the National Academy of Sciences, 2007. **104**(9): p. 3141-3146.

14. Hibbert, E.G. and P.A. Dalby, *Directed evolution strategies for improved enzymatic performance.* Microbial Cell Factories, 2005. **4**(1): p. 29.

15. Cirino, P.C., K.M. Mayer, and D. Umeno, *Generating mutant libraries using error-prone PCR*, in *Directed evolution library creation*. 2003, Springer. p. 3-9.

16. Miyazaki, K., et al., *Directed evolution study of temperature adaptation in a psychrophilic enzyme.* J Mol Biol, 2000. **297**(4): p. 1015-26.

17. Arnold, F.H. and J.C. Moore, *Optimizing industrial enzymes by directed evolution*, in *New Enzymes for Organic Synthesis*. 1997, Springer. p. 1-14.

18. Moore, J.C. and F.H. Arnold, *Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents.* Nature biotechnology, 1996. **14**(4): p. 458.

19. Hanson-Manful, P. and W.M. Patrick, *Construction and analysis of randomized protein-encoding libraries using error-prone PCR.* Methods Mol Biol, 2013. **996**: p. 251-67.

20. Takimoto, J.K., et al., *Stereochemical basis for engineered pyrrolysyl-tRNA synthetase and the efficient in vivo incorporation of structurally divergent non-native amino acids.* ACS Chem Biol, 2011. **6**(7): p. 733-43.

21. Braisted, A.C. and J.A. Wells, *Minimizing a binding domain from protein A.* Proceedings of the National Academy of Sciences, 1996. **93**(12): p. 5688-5692.

22. Biddle, W., Schwark, David G., Schmitt, Margaret A. and John D. Fisk, *Sense Codon Reassignment of the Rare Arginine AGG Codon: Transferability of Improvements Between Orthogonal tRNAs and Synthetases.* submitted, 2018.

23. Apostol, I., et al., *Recombinant protein sequences can trigger methylation of N-terminal amino acids in Escherichia coli.* Protein Science, 1995. **4**(12): p. 2616-2618.

24. Zorn, M.E., R.D. Gibbons, and W.C. Sonzogni, *Evaluation of approximate methods for calculating the limit of detection and limit of quantification.* Environmental science & technology, 1999. **33**(13): p. 2291-2295.

CHAPTER 5

ANALYSIS OF THE RELATIVE IMPORTANCE OF THE FACTORS INVOLVED IN TRANSLATION BY PARTIAL SENSE CODON REASSIGNMENT WITH AN ENGINEERED ORTHOGONAL *M. BARKERI* PAIR

5.1 CHAPTER OVERVIEW

The extent to which the degeneracy of the genetic code can be broken is not obvious, and identifying codons well-suited for reassignment to non-canonical amino acids (ncAAs) by an orthogonal tRNA and aminoacyl tRNA-synthetase (aaRS) pair remains challenging. Translation is a complex process governed by the multiple balanced interactions of the many tRNAs, ribosomes, elonogation factors, aaRSs, and tRNA modifying enzymes involved. The relative quantitative importance of the various molecular interactions involved in protein translation is not well known. Here we describe a new suite of sense codon reassignment measurements using an engineered tyrosine incorporating variant of the orthogonal *M. barkeri* pyrrolysine pair. The sense codon reassignment measurements are directed at evaluating the most productive places to infiltrate the *E. coli* genetic code using the *M. barkeri* orthogonal pair and to improve the quantitative understanding of the molecular interactions that determine the fidelity of translation.

Importantly, every sense codon measured had quantifiable reassignment efficiency to tyrosine with the orthogonal *M. barkeri* pair, ranging from 0.9% to 70.3%. The *M. barkeri* sense codon reassignment measurements strongly correlated with the

sense codon reassignment measurements made using the orthogonal *M. jannaschii* pair. The sense codon reassignment efficiencies were similar between the two systems and suggested combinations of codon reassignments that would maximize the efficiencies of both pairs in 23 amino acid genetic codes.

Accumulation of a second large sense codon reassignment data set and comparison to reassignment efficiencies with the orthogonal *M. jannaschii* pair enabled further analysis of the relative importance of the factors involved in protein translation. As with the orthogonal *M. jannaschii* pair, sense codon reassignment efficiencies did not strongly correlate with the number of competing endogenous tRNAs unless the subset of rarely used codons was considered, suggesting that competing endogenous tRNA concentration is not a dominant factor in protein translation. In addition, the difference between sense codon reassignment efficiencies of the two orthogonal pairs was not strongly correlated to the predicted fold-reduction in aminoacylation efficiency of the *M. jannaschii* pair as a result of changing the anticodon of the *M. jannaschii* tRNA, suggesting that aminoacylation efficiency is also not a dominant factor in protein translation.

5.2 INTRODUCTION

Proteins are vital for just about every biological process, and have applications in diverse fields including biomedicine and materials science. Nature uses 20 canonical amino acids as building blocks for proteins (selenocysteine and pyrrolysine are also ribosomally encoded amino acids, but are rarely used). Biologically incorporating non-canonical amino acids (ncAAs) containing diverse chemistries not found in nature is a powerful strategy for further understanding, modifying, and expanding the properties of

proteins. The most widely developed and used method for expanding the genetic code to include a 21st amino acid has been directing incorporation of the ncAA in response to the amber stop (UAG) codon using an engineered orthogonal tRNA and aminoacyl tRNA synthetase (aaRS) pair [1-3]. The two most commonly employed orthogonal pairs are the tyrosyl tRNA/aaRS from *Methanocaldococcus jannaschii* (*M. jannaschii*) and the pyrrolysyl (Pyl) tRNA/aaRS from the *Methanosarcina* species (*M. barkeri* or *M. mazei*), and the aaRSs from each pair have been repeatedly engineered to generate variants capable of incorporating over 150 different ncAAs [1]. To further expand the genetic code to contain 22 or more amino acids, incorporating ncAAs in response to other codons in the genetic code is required. There are 61 sense codons in the genetic code that signal the incorporation of the 20 canonical amino acids; 18 of the 20 amino acids are encoded by more than one sense codon (methionine and tryptophan are each encoded by a single codon). Breaking the degeneracy of the genetic code by reassigning sense codons to ncAAs with orthogonal tRNA/aaRS pairs is an emerging technique for expanding the genetic code [4-9].

One of the challenges of expanding the genetic code is the precise understanding of the interactions between the orthogonal tRNA/aaRS pair and the cell's endogenous translational machinery. Cells devote a large amount of their energy to producing proteins and maintaining the fidelity of the translation process [10]. The process of translation involves dozens of proteins interacting in a balanced manner. The factors that may affect orthogonal pair directed sense codon reassignment include, but are not limited to, tRNA modifications, competing endogenous tRNA concentration, efficiency of tRNA aminoacylation, elongation factor binding efficiency, and tRNA

anticodon-codon binding energies. Because the relative importance of these and other factors involved in translational fidelity is largely unknown, the selection of a sense codon for reassignment and the orthogonal tRNA/aaRS pair best suited for the targeted codon is challenging.

To quickly probe the possibilities of sense codon reassignment, Fisk and co-workers repurposed the orthogonal *M. jannaschii* pair and developed a simple fluorescence-based screen using GFP as the reporter protein [5]. Similar gain of function through missense incorporation screens have been employed to estimate the background level of missense errors in cells [11-13]. The fluorescence-based screen was used to evaluate the reassignment of 30 *E. coli* codons to tyrosine with the orthogonal *M. jannaschii* pair [[6], Chapter 2]. Importantly, every sense codon evaluated was partially reassigned to tyrosine by simply providing an orthogonal tRNA with an anticodon capable of Watson—Crick base pairing to the targeted codon, and promising codons not previously considered for reassignment to ncAAs were identified. In addition, the relative *in vivo* importance of the factors involved in the process of translation with regards to sense codon reassignment including competing endogenous tRNA concentrations, aminoacylation efficiency, tRNA anticodon-codon interactions, and tRNA modifications were evaluated [6]. No dominant factor was revealed through the sense codon reassignment measurements. Additional reassignment measurements with a second orthogonal pair may help to identify new codons for reassignment to ncAAs, and further aid in dissecting the relative importance of the *in vivo* factors involved in translation.

In this chapter we describe a second unique set of sense codon reassignment efficiency measurements for the 30 sense codons with an engineered orthogonal tyrosine-incorporating Pyl *M. barkeri* pair and the fluorescence-based screen. The second set of data identifies promising codons for reassignment to ncAAs with the orthogonal *M. barkeri* pair. In addition, combining the sense codon reassignment efficiency data from the orthogonal *M. jannaschii* and *M. barkeri* pairs allows for further dissection of the factors involved in translational fidelity and paves a path towards 22 amino acid genetic codes where more than one sense codon is reassigned to a ncAA. The *M. jannaschii* and *M. barkeri* orthogonal pairs were evaluated using the same vector and reporter constructs in the same cells strains, controlling for a majority of the factors that affect competition in these systems. A significant difference between the two orthogonal pair systems is that the anticodon sequence is a recognition element for the *M. jannaschii* but not the *M. barkeri* system [14-17]. Comparisons between reassignment across the system should serve to isolate contribution of the aminoacylation efficiency to translation fidelity.

5.3 MATERIALS AND METHODS

The fluorescence-based screen for sense codon reassignment has been described [5]. Detailed experimental protocols, including methods for mutagenesis of orthogonal *M. janaschii* tRNA variants, measurement of system growth and fluorescence, and calculation of sense codon reassignment efficiencies are included in Appendix 1. The appendix also includes full vector sequences, oligonucleotide primer sequences, detailed cell strain information, and general reagents and materials.

5.4 RESULTS AND DISCUSSION

Measurement of sense codon reassignment efficiencies with the orthogonal tyrosine-incorporating pyrrolysine *M. barkeri* pair were made at the same 30 targeted sense codons analyzed with the orthogonal *M. jannaschii* pair, including wobble, rare and discrimination codons [[6], Chapter 2]. The idea behind targeting wobble codons for reassignment is that differences in codon-anticodon binding energies between introduced orthogonal tRNAs designed to Watson—Crick base pair at all three codon bases and endogenous tRNAs that utilize wobble base pairing at the third codon position can be used to bias incorporation of the amino acid towards the introduced orthogonal pair [4]. The idea behind targeting of rarely used codons in the *E. coli* genome for reassignment is that endogenous tRNAs that read rarely used codons are typically present in lower concentrations in the cell, and because the ribosome requires sampling of the tRNA pool before decoding any mRNA codon, a highly expressed orthogonal tRNA may be able to out-compete the less abundant endogenous tRNAs [7, 9, 18, 19].

Sense codon reassignment efficiencies to tyrosine by the engineered orthogonal *M. barkeri* pair were measured using a previously developed *in vivo* GFP fluorescence-based screen capable of measuring sense codon reassignment as low as 1 part in 1000 of natural GFP protein [5]. The screen uses the fact that tyrosine is essential for the fluorescence of GFP [20, 21]. The codon specifying the critical tyrosine residue in the reporter GFP gene is replaced with the sense codon targeted for reassignment. When GFP is expressed, the mRNA transcript may be read in one of two ways [5]. The orthogonal *M. barkeri* tRNA may decode the sense codon at the critical tyrosine position

in GFP and incorporate tyrosine, resulting in a phenotypically wild-type GFP or the endogenous *E. coli* tRNA may decode the sense codon at the critical tyrosine position and incorporate an amino acid other than tyrosine, resulting in a non-fluorescent protein (Figure 5.1) [5, 20, 21]. Sense codon reassignment efficiency may then be quantified as the amount of GFP fluorescence from a sense codon reassigning system divided by the amount of GFP fluorescence in a wild-type GFP expressing control where the key tyrosine position is encoded by a tyrosine codon.

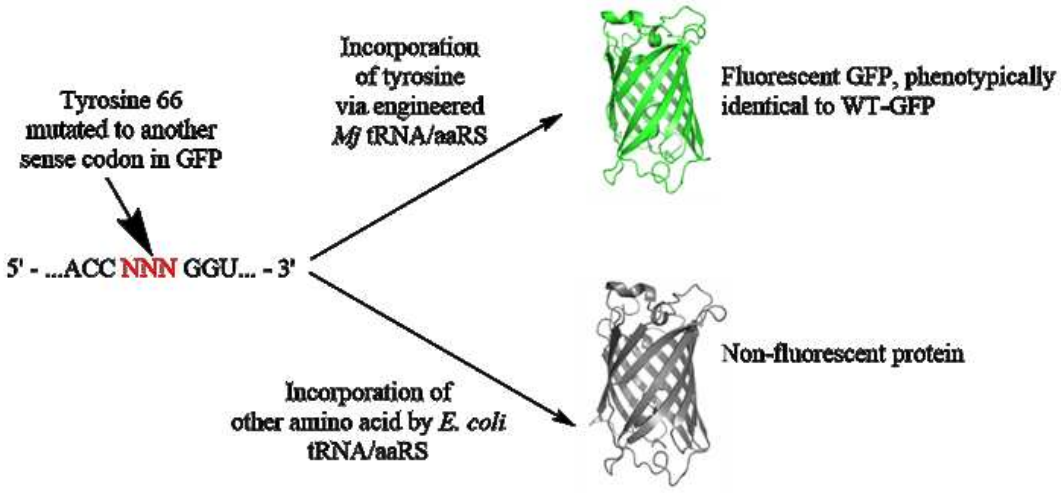

Figure 5.1. System using GFP fluorescence as an indicator of sense codon reassignment efficiency (PDB 2B3P). One mRNA transcript of GFP may be decoded two different ways. In the upper case, the engineered orthogonal *M. janaschii* tRNA decodes the sense codon and tyrosine is incorporated. In the lower case, the endogenous *E. coli* tRNA decodes the sense codon and incorporates the corresponding amino acid [5, 22].

The orthogonal pair induced sense codon reassignments to tyrosine occur proteome-wide in the host *E. coli*, allowing for an investigation of the impact of sense codon reassignment to tyrosine on cell health. Cell health was quantified as relative cell

fitness by dividing the instantaneous doubling time during the exponential phase of growth in a sense codon reassigning system compared by the instantaneous doubling time of a non-reassigning, wild-type GFP expressing control. Fitness effects were compared to the estimated number of substitutions to tyrosine occurring in the course of a single cell generation ~~was~~ calculated for each sense codon reassigning system, using estimates for translation frequency of each codon [6, 23-27].

We have generated two large, closely related data sets of sense codon reassignment efficiencies with the acquisition of the suite of sense codon reassignment efficiencies to tyrosine using the orthogonal tyrosine-incorporating pyl *M. barkeri* pair and the previously acquired measurements with the orthogonal *M. jannaschii* pair. The large data sets may be used to dissect how the interactions between the dozens of proteins involved in translation and the introduced orthogonal pair affects sense codon reassignment. The only differences between the *M. barkeri* and *M. jannaschii* reassignment systems were the identities of the orthogonal tRNAs and aaRSs; all other variables that may potentially affect sense codon reassignment efficiency (not related to the process of translation, such as plasmid copy number, promoter identities or cell line) were the same between the two systems.

The following sub-sections compare the suite of sense codon reassignments with the engineered tyrosine-incorporating pyl *M. barkeri* pair to the *M. jannaschii* pair with regard to (1) the significance of reassignment efficiencies, (2) discrimination of closely related codons by the orthogonal tRNA, (3) cell health effects resulting from reassignment, (4) the effect of competing endogenous *E. coli* tRNAs on reassignment efficiency and (5) the impact of aminoacylation on sense codon reassignment.

Significance of sense codon reassignment efficiencies with the engineered tyrosine-incorporating Pyl M. barkeri tRNA/aaRS

Every *E. coli* sense codon analyzed was capable of partial reassignment to tyrosine using the engineered orthogonal tyrosine-incorporating Pyl *M. barkeri* tRNA/aaRS pair (Table 5.1). Reassignment efficiencies to tyrosine using the *M. barkeri* tRNA/aaRS pair range from 0.9% (9 out of 1000 incorporation events at the targeted codon) to 70.3% (703 out of 1000 incorporation events at the targeted codon). Just as with sense codon reassignment to tyrosine with the orthogonal *M. jannaschii* tRNA/aaRS, sense codons that are reassigned to tyrosine by the engineered orthogonal *M. barkeri* tRNA/aaRS with higher efficiencies may be more easily reassigned to a ncAA; however, any quantifiable amount of sense codon reassignment to tyrosine may be used as a starting point for improvement through directed evolution. Recently, directed evolution of the *M. jannaschii* tRNA anticodon loop/aaRS anticodon binding domain allowed nearly quantitative reassignment of the AGG codon [7].

Each value in Table 5.1 represents the average reassignment efficiency to tyrosine of at least 6 individual clones. Reassignment measurements were consistent across experiments on multiple days. The overall standard deviation in reassignment percentage for more than 80 individual non-reassigning clones (*M. barkeri* system) across 13 days is 0.023%, making the detection limit of the in cell fluorescence assay for sense codon reassignment efficiency 0.14%. The two codons least efficiently reassigned by the orthogonal engineered tyrosine-incorporating *M. barkeri* tRNA/aaRS, Val GUU and Ala GCU, are measured at 0.9% efficiency, above the limit of quantification of the screen.

Table 5.1. Sense codon reassignment efficiencies to tyrosine with the *M. barkeri* and *M. jannaschii* orthgonal tRNA/aaRS pairs [Chapter 2, [6]].

tRNA Anticodon	Sense Codon Targeted	Reassignment of Targeted Codon with *M. barkeri*	Reassignment of Targeted Codon with *M. jannaschii*	*E. coli* Codon Usage[a]	Competing *E. coli* tRNAs[b,c]
$tRNA_{AAA}$	Phe UUU	2.3%	3.2%	0.57	4440
$tRNA_{AAG}$	Leu CUU	2.9%	4.2%	0.10	7400
$tRNA_{AAU}$	Ile AUU	1.5%	4.3%	0.51	17390
$tRNA_{AAC}$	Val GUU	0.9%	0.8%	0.26	23680
$tRNA_{AGA}$	Ser UCU	6.0%	19.6%	0.15	9620
$tRNA_{AGG}$	Pro CCU	5.0%	12.2%	0.16	5920
$tRNA_{AGU}$	Thr ACU	3.6%	10.5%	0.17	9250
$tRNA_{AGC}$	Ala GCU	0.9%	1.3%	0.16	19240
$tRNA_{AUG}$	His CAU	3.7%	6.1%	0.57	2960
$tRNA_{AUU}$	Asn AAU	2.6%	7.5%	0.45	5550
$tRNA_{AUC}$	Asp GAU	1.1%	3.7%	0.63	11100
$tRNA_{ACA}$	Cys UGU	5.1%	3.5%	0.45	6660
$tRNA_{ACU}$	Ser AGU	15.6%	9.1%	0.15	5180
$tRNA_{ACC}$	Gly GGU	4.0%	1.1%	0.34	18500
$tRNA_{CAC}$	Val GUG	2.1%	2.1%	0.37	17760
$tRNA_{CGC}$	Ala GCG	1.7%	9.2%	0.36	16280
$tRNA_{CUU}$	Lys AAG	10.9%	7.7%	0.23	8140
$tRNA_{CUC}$	Glu GAG	11.7%	19.8%	0.31	22570
$tRNA_{GCG}$	Arg CGC	1.3%	7.6%	0.40	22200
$tRNA_{UCG}$	Arg CGA	44.8%	40.8%	0.06	22200
$tRNA_{UAG}$	Leu CUA	7.0%	24.3%	3.85	2960
$tRNA_{UAU}$	Ile AUA	41.4%	70.0%	0.07	17390
$tRNA_{UGA}$	Ser UCA	9.5%	24.7%	0.12	6290
$tRNA_{GGG}$	Pro CCC	N/A[d]	N/A[d]	0.12	3700
$tRNA_{UGU}$	Thr ACA	7.5%	10.9%	0.13	4440
$tRNA_{GCA}$	Cys UGC	6.4%	20.2%	0.55	6660
$tRNA_{CCG}$	Arg CGG	70.3%	53.0%	0.1	2220
$tRNA_{UCU}$	Arg AGA	16.8%	40.7%	0.04	2960
$tRNA_{CCU}$	Arg AGG	65.1%	50.6%	0.02	5190
$tRNA_{UCC}$	Gly GGA	2.8%	4.8%	0.11	10360
$tRNA_{CCC}$	Gly GGG	N/A[d]	N/A[d]	0.15	10360

a As a fraction of the total number of codons encoding a given canonical amino acid. Codon usage is based on http://openwetware.org/wiki/Escherichia_coli/Codon_usage.

b Number of competing tRNAs calculated utilizing reported tRNA-to-ribosome ratios for cells with a 37 minute doubling time and an estimated 37,000 ribosomes [23].

c Number of competing tRNAs for Ile AUU and AUA, and for Gly GGA and GGG were treated collectively because they could not be separated [28].

d The orthogonal tRNA with the indicated anticodon was unable to be successfully constructed for either orthogonal pair.

Reassignment of the wobble codons to tyrosine with the engineered orthogonal Pyl *M. barkeri* pair ranged from 0.9% (Val GUU and Ala GCU) to 44.8% (Arg CGA). The Arg CGA codon is the most efficiently reassigned wobble codon for both the orthogonal *M. barkeri* pair (44.8%) and the orthogonal *M. jannaschii* pair (40.8%) despite competition against one of the most abundant tRNAs in *E. coli*, $tRNA^{Arg2}_{ICG}$ [28]. The Arg CGA codon is read in *E. coli* through an I34/A3 base pairing interaction, was previously hypothesized to be much less energetically favorable than a Watson-Crick U34/A3 interaction by an orthogonal tRNA [29]. The high levels of reassignment to tyrosine at the Arg CGA codon by the *M. barkeri* pair supports the hypothesis that codon-anticodon energetics play a role in tRNA selection [4, 6]. High levels of reassignment for sense codons that may have significantly improved anticodon-codon interactions over the endogenous tRNA might be transferrable across multiple orthogonal pairs.

The rank ordering of wobble codon reassignment efficiencies are relatively consistent between the orthogonal *M. barkeri* and *M. jannaschii* pairs (Figure 5.2). Close rank ordering of reassignment efficiencies between the orthogonal *M. barkeri* and *M. jannaschii* pairs suggests that the interactions between the orthogonal translational machinery and the endogenous translational machinery are similar across systems; consequently, reassignment of these codons to ncAAs may be achieved with either orthogonal pair.

The reassignment to tyrosine at the Ala GCG (5 for *M. barkeri*, 15 for *M. jannaschii*) and Arg CGC (3 for *M. barkeri*, 12 for *M. jannaschii*) codons are significantly better with the orthogonal *M. jannaschii* pair. The reassignment to tyrosine at the Gly

GGU (12 for *M. barkeri*, 2 for *M. jannaschii*) and the Cys UGU (14 for *M. barkeri*, 6 for *M. jannaschii*) codons are significantly better with the orthogonal *M. barkeri* pair. The differences in rank ordering between the orthogonal pairs for the Ala GCG, Arg CGC, Gly GGU, and Cys UGU codons may be due to multiple factors, including amount of orthogonal tRNA aminoacylated with tyrosine and energies of tRNA anticodon-codon interactions at the A site in the ribosome.

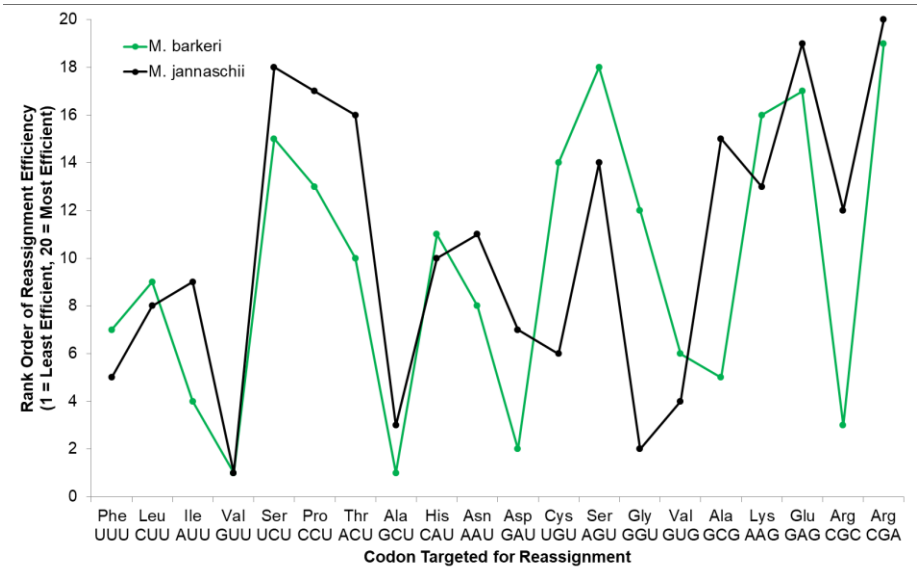

Figure 5.2. Rank ordering of sense codon reassignment efficiencies at the 20 wobble codons vs. the codon targeted for reassignment for the orthogonal *M. barkeri* and *M. jannaschii* pairs. 1 is the least efficient and 20 is the most efficient.

Reassignment of rarely used codons with the engineered orthogonal *M. barkeri* pair ranged from 2.8% (Gly GGU) to 70.3% (Arg CGG). In general, rarely used codons are more efficiently reassigned than wobble codons with the *M. barkeri* pair, a trend also observed with the orthogonal *M. jannaschii* pair [Chapter 2]. The rank ordering of reassignment efficiencies at rarely used codons are relatively consistent between the two orthogonal pairs (Figure 5.3). The four most efficiently reassigned rare codons for

both orthogonal pairs, although not ranked in the same order, are the Ile AUA, Arg CGA (also a wobble codon), Arg CGG, and Arg AGG codons. Furthermore, the least efficiently reassigned rarely used codons (Cys UGU and Gly GGA) are the same for the orthogonal *M. barkeri* and *M. jannaschii* pairs.

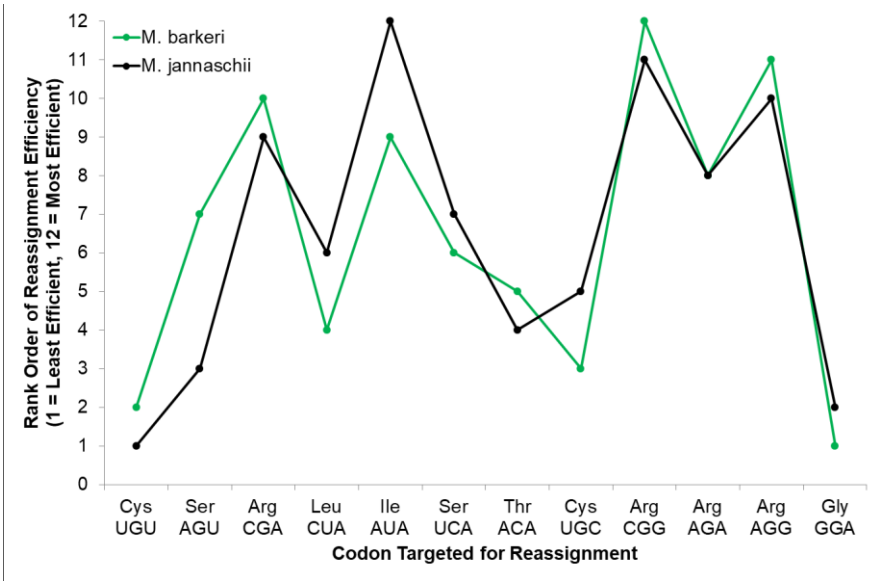

Figure 5.3. Rank ordering of sense codon reassignment efficiencies at the 13 rarely used codons vs. the codon targeted for reassignment for the orthogonal *M. barkeri* and *M. jannaschii* pairs. 1 is the least efficient, and 12 is the most efficient.

Despite multiple attempts and in contrast to the other 30 tRNA anticodon sequences generated for the *M. barkeri* system, the $tRNA^{Pyl}_{GGG}$ to reassign the rarely used Pro CCC codon and the $tRNA^{Pyl}_{CCC}$ to reassign the Gly GGG codon could not be constructed. The same difficulties were previously observed when attempting to construct the analgous *M. jannaschii* $tRNA^{Opt}$ variants [Chapter 2]. The hypothesis that high levels of frameshifting caused by the *M. jannaschii* $tRNA^{Opt}$ variants with CCC and GGG anticodons leading to fatal translation errors also applies to the orthogonal *M.*

barkeri tRNA variants [Chapter 2]. A tRNA modification of the base after the codon, m1-G37, is essential for reading the proline CCC codon [30].

When all measured sense codons are considered, reassignment efficiencies with the orthogonal *M. barkeri* pair are strongly correlated (R^2=0.74) to the reassignment efficiencies with the *M. jannaschii* pair (Figure 5.4). The strong correlation for all sense codons supports the conjecture that the interactions between the dozens of proteins involved in translation and the orthogonal *M. barkeri* translational components are behaving in the same manner as in reassignment systems using the *M. jannaschii* orthogonal pair.

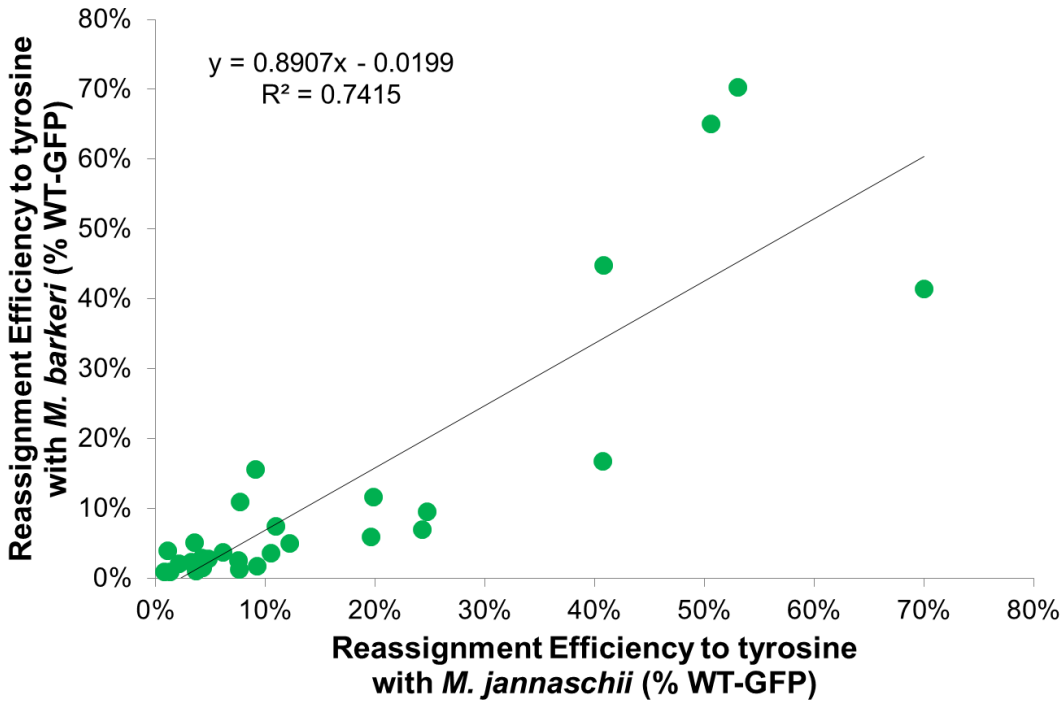

Figure 5.4. Sense codon reassignment efficiency to tyrosine with the engineered orthogonal *M. barkeri* tRNA/aaRS vs. sense codon reassignment efficiency to tyrosine with the *M. jannaschii* tRNA/aaRS.

Discrimination of closely related codons by the orthogonal tRNA

The extent to which a particular codon targeted for reassignment as opposed to closely related codons is decoded by an orthogonal tRNA is also important for reassignment related to genetic code expansion. One motivation behind using orthogonal pair directed sense codon reassignment is that the incorporation of a desired ncAA is codon specific, leaving the other closely related codons to specify incorporation of a natural amino acid. Discrimination between the 30 targeted and closely related codons by orthogonal *M. barkeri* tRNAs was evaluated (Table 5.2).

The wobble rules suggest that reassignment introduced by an A34 tRNA should be most efficient at the targeted U ending codon, and the order of efficiencies should be U>C>>A>G [31]. Discrimination between the targeted U-ending codons and the closely related C-ending codons for the orthogonal A34 *M. barkeri* tRNAs follows the same trends as with the *M. jannaschii* orthogonal A34 tRNAs (Tables 5.2, 5.3). Orthogonal *M. barkeri* A34 tRNAs for the Phe, Ile, Ser, Thr, Asn, Cys, Ser, and Gly wobble codons strongly discriminate between the targeted codon and the closely-related C-ending codons; reassignment values at the non-targeted C-ending codons are near or below the *in vivo* detection limit of 0.14% (Tables 5.2, 5.3). Reassignment of the Ser AGU codon with the orthogonal *M. barkeri* pair is very promising, just as with the *M. jannaschii* orthogonal pair, because reassignment of the targeted codon is 15.6% and no measurable level of reassignment at the non-targeted AGC codon was observed [6].

Poor discrimination between the orthogonal *M. barkeri* A34 tRNAs and non-targeted codons occurs for the Leu, Val, Pro, Asp, Ala, and His wobble codons. Discrimination ratios vary from nearly 50:50 for the targeted versus non-targeted codons

(tRNA$^{Pyl}_{AAC}$ targeting the Val GUU codon) to 85:15 for the targeted versus non-targeted codon (tRNA$^{Pyl}_{AUC}$ targeting the Asp GAU codon). The poor discrimination trend for these codons was observed for the comparable orthogonal *M. jannaschii* tRNAs.

One potential hypothesis for poor observed discrimination for some of the A34 orthogonal *M. barkeri* tRNAs is modification of nucleotides in the anticodon by endogenous *E. coli* enzymes. Fisk and co-workers previously demonstrated that poor discrimination between the targeted CAU and non-targeted CAC codons by the orthogonal *M. jannaschii* tRNA$^{Opt}_{AUG}$ was a result of nucleotide A34 in the anticodon of the orthogonal tRNA being modified to inosine by the endogenous *E. coli* enzyme TadA [32]. The enzyme TadA modifies position A34 of the *E. coli* tRNA$^{Arg2}_{ACG}$ to inosine, and uses the anticodon stem loop of the tRNA as an important substrate recognition element [33]. The nearly identical anticodon stem loop sequences of the orthogonal *M. jannaschii* tRNA$^{Opt}_{AUG}$ and the *E. coli* tRNA$^{Arg2}_{ICG}$ likely allowed the *M. jannaschii* tRNA to be recognized as a substrate for *E. coli* TadA. As a result, A34 of the *M. jannaschii* tRNA$^{Opt}_{AUG}$ was modified to inosine. Fisk and co-workers also demonstrated that altering the nucleotides in the anticodon stem loop of the orthogonal *M. jannaschii* tRNA$^{Opt}_{AUG}$ prevented it from being recognized as a substrate for TadA, and also improved sense codon reassignment efficiency at the targeted CAU codon [34].

Partial inosine modification at position 34 is one potential reason for the orthogonal *M. barkeri* A34 tRNAs showing poor discrimination; I34/U3 and I34/C3 pairings are energetically favorable. Furthermore, the anticodon stem loop of the orthogonal *M. barkeri* tRNA is very similar to the anticodon stem loops of the orthogonal

Table 5.2. Discrimination of closely-related codons by introduced orthogonal *M. barkeri* tRNAs.

tRNA Anticodon	Amino acid targeted	Sense codon targeted	Reassignment observed at YZU[a]	Reassignment observed at YZC	Reassignment observed at YZA	Reassignment observed at YZG
$tRNA^{Pyl}_{AAA}$	Phe	PheUUU	***2.3%***	B.D.[b]	----	----
$tRNA^{Pyl}_{AAG}$	Leu	Leu CUU	***2.9%***	0.7%	B.D.	B.D.
$tRNA^{Pyl}_{AAU}$	Ile	Ile AUU	***1.5%***	B.D.	B.D.	0.1%
$tRNA^{Pyl}_{AAC}$	Val	Val GUU	***0.9%***	0.5%	0.2%	0.2%
$tRNA^{Pyl}_{AGA}$	Ser	Ser UCU	***6.0%***	0.3%	----	----
$tRNA^{Pyl}_{AGG}$	Pro	Pro CCU	***5.0%***	3.5%	0.4%	---
$tRNA^{Pyl}_{AGU}$	Thr	Thr ACU	***3.6%***	B.D.	B.D.	B.D.
$tRNA^{Pyl}_{AGC}$	Ala	Ala GCU	***0.9%***	0.7%	B.D.	B.D.
$tRNA^{Pyl}_{AUG}$	His	His CAU	***3.7%***	1.1%	----	----
$tRNA^{Pyl}_{AUU}$	Asn	Asn AAU	***2.6%***	B.D.	----	----
$tRNA^{Pyl}_{AUC}$	Asp	Asp GAU	***1.1%***	0.2%	----	----
$tRNA^{Pyl}_{ACA}$	Cys	Cys UGU	***5.1%***	B.D.	----	----
$tRNA^{Pyl}_{ACU}$	Ser	Ser AGU	***15.6%***	B.D.	----	----
$tRNA^{Pyl}_{ACC}$	Gly	Gly GGU	***4.0%***	0.2%	0.1%	1.0%
$tRNA^{Pyl}_{CAC}$	Val	Val GUG	0.2%	0.2%	0.2%	***2.1%***
$tRNA^{Pyl}_{CGC}$	Ala	Ala GCG	0.2%	0.3%	0.3%	***1.7%***
$tRNA^{Pyl}_{CUU}$	Lys	Lys AAG	----	----	B.D.	***10.9%***
$tRNA^{Pyl}_{CUC}$	Glu	Glu GAG	----	----	0.2%	***11.7%***
$tRNA^{Pyl}_{GCG}$	Arg	Arg CGC	2.1%	***1.3%***	B.D.	B.D.
$tRNA^{Pyl}_{UCG}$	Arg	Arg CGA	0.1%	0.1%	***44.8%***	29.4%
$tRNA^{Pyl}_{UAG}$	Leu	Leu CUA	----	----	***7.0%***	0.4%
$tRNA^{Pyl}_{UAU}$	Ile	Ile AUA	----	----	***41.4%***	0.3%
$tRNA^{Pyl}_{UGA}$	Ser	Ser UCA	----	----	***9.5%***	1.4%
$tRNA^{Pyl}_{GGG}$	Pro	Pro CCC	----	***----***	----	----
$tRNA^{Pyl}_{UGU}$	Thr	Thr ACA	----	----	***7.5%***	0.5%
$tRNA^{Pyl}_{GCA}$	Cys	Cys UGC	6.3%	***6.4%***	----	----
$tRNA^{Pyl}_{CCG}$	Arg	Arg CGG	B.D.[b]	B.D.	0.1%	***70.3%***
$tRNA^{Pyl}_{UCU}$	Arg	Arg AGA	----	----	***16.8%***	1.3%
$tRNA^{Pyl}_{CCU}$	Arg	Arg AGG	----	----	B.D.	***65.1%***
$tRNA^{Pyl}_{UCC}$	Gly	Gly GGA	----	----	***2.8%***	0.8%
$tRNA^{Pyl}_{CCC}$	Gly	Gly GGG	----	----	----	***----***

Anticodons with A or G at position 34 were typically evaluated against codons ending in U and C; anticodons with U or C at position 34 were typically evaluated against codons ending in A and G. Reassignment efficiencies emboldened and italicized are those for the codon targeted for reassignment, which fully base pairs to the orthogonal tRNA via Watson-Crick interactions

a For continuity, "Y" and "Z" are used to represent the first two positions of the codon in the column headers. The sense codons evaluated with a single tRNA in each row are determined by substituting the complement of the nucleobases in the second and third anticodon positions for Y and Z. The third codon position is specified in the column headers. For example, row 1 shows sense codon reassignment by the M. barkeri tRNAPylAAA. Data in the YZU column are for reassignment at the UUU codon. Data in the YZC column are for reassignment at the UUC codon.

b B.D. indicates that the codon was evaluated with the specified tRNA, and the measurement was below the detection limit of the in cell assay (0.14%).

c "----" indicates that the codon was not evaluated for reassignment by the specified tRNA.

Table 5.3. Discrimination of closely-related codons by introduced orthogonal *M. jannaschii* tRNAs [Chapter 2, [6]].

tRNA Anticodon	Amino acid targeted	Sense codon targeted	Reassignment observed at YZU[a]	Reassignment observed at YZC	Reassignment observed at YZA	Reassignment observed at YZG
$tRNA^{Opt}_{AAA}$	Phe	Phe UUU	***3.20%***	**B.D.**[b]	**-----**[c]	**-----**
$tRNA^{Opt}_{AAG}$	Leu	Leu CUU	***4.20%***	1.30%	B.D.	-----
$tRNA^{Opt}_{AAU}$	Ile	Ile AUU	***4.30%***	0.40%	B.D.	0.30%
$tRNA^{Opt}_{AAC}$	Val	Val GUU	***0.80%***	0.60%	B.D.	0.30%
$tRNA^{Opt}_{AGA}$	Ser	Ser UCU	***19.60%***	1.30%	-----	-----
$tRNA^{Opt}_{AGG}$	Pro	Pro CCU	***12.20%***	10.20%	0.90%	-----
$tRNA^{Opt}_{AGU}$	Thr	Thr ACU	***10.50%***	1.30%	-----	-----
$tRNA^{Opt}_{AGC}$	Ala	Ala GCU	***1.30%***	1.10%	B.D.	B.D.
$tRNA^{Opt}_{AUG}$	His	His CAU	***6.10%***	2.90%	B.D.	-----
$tRNA^{Opt}_{AUU}$	Asn	Asn AAU	***7.50%***	B.D.	-----	-----
$tRNA^{Opt}_{AUC}$	Asp	Asp GAU	***3.70%***	B.D.	B.D.	B.D.
$tRNA^{Opt}_{ACA}$	Cys	Cys UGU	***3.50%***	B.D.	-----	-----
$tRNA^{Opt}_{ACU}$	Ser	Ser AGU	***9.10%***	B.D.	-----	-----
$tRNA^{Opt}_{ACC}$	Gly	Gly GGU	***1.10%***	0.50%	-----	-----
$tRNA^{Opt}_{CAC}$	Val	Val GUG	0.30%	B.D.	0.40%	***2.10%***
$tRNA^{Opt}_{CGC}$	Ala	Ala GCG	0.30%	0.80%	1.10%	***9.20%***
$tRNA^{Opt}_{CUU}$	Lys	Lys AAG	-----	-----	B.D.	***7.70%***
$tRNA^{Opt}_{CUC}$	Glu	Glu GAG	B.D.	B.D.	0.80%	***19.80%***
$tRNA^{Opt}_{GCG}$	Arg	Arg CGC	7.50%	***7.60%***	B.D.	B.D.
$tRNA^{Opt}_{UCG}$	Arg	Arg CGA	0.50%	0.40%	***40.80%***	33.00%
$tRNA^{Opt}_{UAG}$	Leu	Leu CUA	15.0%	1.1%	***24.3%***	6.5%
$tRNA^{Opt}_{UAU}$	Ile	Ile AUA	4.4%	0.4%	***70.0%***	3.5%
$tRNA^{Opt}_{UGA}$	Ser	Ser UCA	-----	-----	***24.7%***	7.1%
$tRNA^{Opt}_{GGG}$	Pro	Pro CCC	----	**----**	----	----
$tRNA^{Opt}_{UGU}$	Thr	Thr ACA	-----	-----	***10.9%***	-----
$tRNA^{Opt}_{GCA}$	Cys	Cys UGC	17.3%	***20.2%***	-----	-----
$tRNA^{Opt}_{CCG}$	Arg	Arg CGG	BD	BD	2.8%	***53.0%***
$tRNA^{Opt}_{UCU}$	Arg	Arg AGA	-----	-----	***40.7%***	10.4%
$tRNA^{Opt}_{CCU}$	Arg	Arg AGG	-----	-----	BD	50.6%
$tRNA^{Opt}_{UCC}$	Gly	Gly GGA	-----	-----	***4.8%***	2.3%
$tRNA^{Opt}_{CCC}$	Gly	Gly GGG	----	----	----	**----**

Anticodons with A or G at position 34 were typically evaluated against codons ending in U and C; anticodons with U or C at position 34 were typically evaluated against codons ending in A and G. Reassignment efficiencies emboldened and italicized are those for the codon targeted for reassignment, which fully base pairs to the orthogonal tRNA via Watson-Crick interactions

a For continuity, "Y" and "Z" are used to represent the first two positions of the codon in the column headers. The sense codons evaluated with a single tRNA in each row are determined by substituting the complement of the nucleobases in the second and third anticodon positions for Y and Z. The third codon position is specified in the column headers. For example, row 1 shows sense codon reassignment by the M. barkeri tRNAPylAAA. Data in the YZU column are for reassignment at the UUU codon. Data in the YZC column are for reassignment at the UUC codon.

b B.D. indicates that the codon was evaluated with the specified tRNA, and the measurement was below the detection limit of the in cell assay (0.15%).

c "----" indicates that the codon was not evaluated for reassignment by the specified tRNA.

M. jannaschii tRNA^{Opt} and *E. coli* tRNA^{Arg2} (Figure 5.5) [33-35]. Future work may address the identities of the nucleotide at position 34 of poorly discriminating A34 orthogonal *M. barkeri* tRNAs.

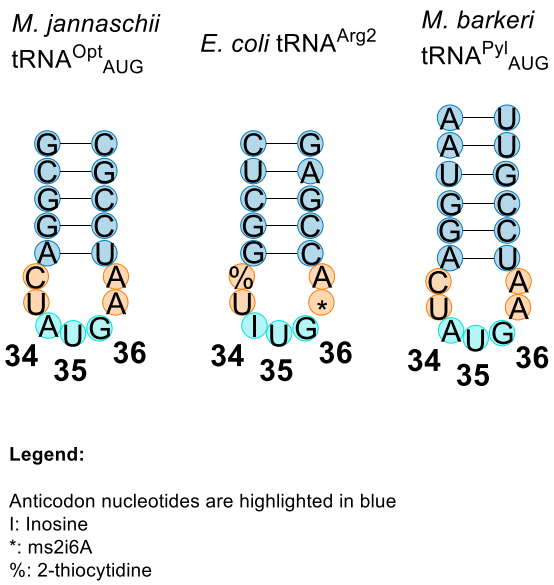

Figure 5.5. Anticodon stem loops of the *M. jannaschii* tRNA$^{Opt}_{AUG}$, *E. coli* tRNA^{Arg2}, and *M. barkeri* tRNA$^{Pyl}_{AUG}$ [33-35].

The wobble rules suggest that reassignment introduced by a C34 orthogonal tRNA should only read the targeted G-ending codon [31]. Discrimination by orthogonal C34 tRNAs for Lys, Glu, and Arg all showed discrimination ratios between targeted and non-targeted codons of 98:2 or greater (Table 5.2). Discrimination ratios between targeted and non-targeted codons was 96:4 or greater for the same set of orthogonal *M. jannaschii* tRNAs (Table 5.3) [Chapter 2,[6]]. Discrimination between G and A-ending Ala and Val codons by the orthogonal *M. barkeri* tRNAs was unexpectedly less efficient, and also observed with the same set of orthogonal *M. jannaschii* tRNAs (Tables 5.2, 5.3). Previously, Fisk and co-workers hypothesized that poor discrimination by the orthogonal *M. jannaschii* tRNAs that target the Ala GCG and Val GUG codons may be a

result of an unknown tRNA modification or the “2 out of 3” decoding hypothesis [36, 37]. The “2 out of 3” decoding hypothesis suggests that codons that have G/C pairs at the first 2 positions of the codon may be more easily decoded by base-pairing mismatches at the third position of the codon. Future work may address why poor discrimination by the orthogonal *M. barkeri* and *M. jannaschii* tRNAs targeting the Ala GCG and Val GUG codons is observed.

Two orthogonal *M. barkeri* tRNAs contained G at position 34: $tRNA^{Pyl}_{GCA}$ that decodes the Cys UGC codon, and $tRNA^{Pyl}_{GCG}$ that decodes the Arg CGC codon. Both orthogonal *M. barkeri* tRNAs that contain G at position 34 decode non-targeted codons through a G-U wobble interaction between G34 of the tRNA and U3 of the codon, and is not surprising given the strength of the G-U wobble interaction (Table 5.2). The orthogonal *M. barkeri* $tRNA^{Pyl}_{GCA}$ decodes the targeted UGC (6.4%) and non-targeted UGU (6.3%) codons with nearly equal efficiency. The orthogonal *M. barkeri* $tRNA^{Pyl}_{GCG}$ decodes the targeted CGC (1.3%) and non-targeted CGU (2.1%) as well. Similar discrimination by the analogous orthogonal *M. jannaschii* tRNAs was observed (Table 5.3) [Chapter 2, [6]].

According to the wobble rules, tRNAs that contain U at position 34 would be expected to pair with all four nucleotides at the third position of the anticodon. As expected, most of the orthogonal U34 *M. barkeri* tRNAs reassign the G-ending codon in addition to the A-ending codon which they target [31]. Discrimination of the orthogonal U34 *M. barkeri* tRNAs between the targeted A-ending codon and the non-targeted G-ending codon varies from nearly 60:40 in the case of *M. barkeri* $tRNA^{Pyl}_{UCG}$, which targets the Arg CGA codon to about 95:5 for the *M. barkeri* $tRNA^{Pyl}_{UAU}$ which decodes

the Ile AUA codon (Table 5.2). Very similar discrimination trends are observed for orthogonal U34 *M. jannaschii* tRNAs (Table 5.3) [Chapter 2, [6]]. The high level of discrimination between the targeted Ile AUA and non-targeted Met AUG codon by the orthogonal *M. barkeri* $tRNA^{Pyl}_{UAU}$ is also observed with the orthogonal *M. jannaschii* $tRNA^{Opt}_{UAU}$ [Chapter 2]. The high level of discrimination by orthogonal tRNAs that target the Ile AUA codon is surprising because the U at position 34 would be expected to wobble base-pair with G at position 3. One possible hypothesis for the high level of discrimination observed is modification of U34 in the orthogonal tRNA to lysidine (L). In *E. coli*, the Ile AUA codon is read by an endogenous tRNA with C34 modified to lysidine [38]. The modification of C34 to L allows the tRNA to decode the Ile AUA codon and prevents it from reading the Met AUG codon. The lysidine modification involves the formation of a Schiff base between the epsilon amino group of lysine and the C2 carbonyl of cytosine catalyzed by the enzyme TilS. Although modification to lysidine normally occurs at C, a similar modification of U may also be possible.

Cell health effects as a result of orthogonal pair directed sense codon reassignment to tyrosine

The orthogonal pair directed sense codon reassignment efficiencies with the engineered orthogonal tyrosine-incorporating Pyl *M. barkeri* pair presented here resulted in codon-specific amino acid substitutions from the canonically-encoded amino acid to tyrosine across the proteome of *E. coli*. The *in vivo* reassignment systems were also used as a tool to investigate the ability of *E. coli* to tolerate large numbers of amino acid substitutions. Quantification of the effect of sense codon reassignment on cell health was achieved by determining the instantaneous doubling times of systems

undergoing sense codon reassignment to tyrosine with the orthogonal *M. barkeri* pair and comparing them to the doubling time a non-reassigning, wild-type GFP expressing control. The majority of systems undergoing sense codon reassignment to tyrosine had doubling times that were just outside of two standard deviations from the non-reassigning, wild-type control, even though estimated 100,000s to 1,000,000s of amino acid substitutions to tyrosine were occurring (Figure 5.6, Table 5.4). The orthogonal pair codon-specific amino acid substitutions represent between 2X and 10X the expected background level of natural amino acid substitutions due to errors in translation [39, 40].

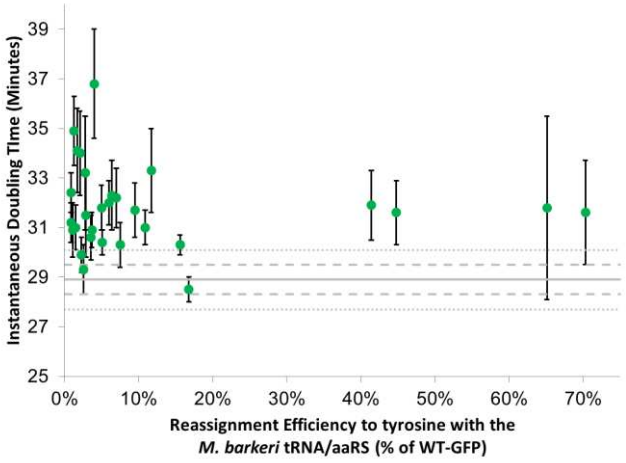

Figure 5.6. Average instantaneous doubling time of each sense codon reassigning system plotted against the sense codon reassignment efficiency for that system (with the *M. barkeri* tRNA/aaRS). The straight grey line at 28.9 minutes represents the instantaneous doubling time for a non-reassigning, wild type GFP culture (GFP with a tyrosine codon at position 66, amber-reassigning tRNA with a CUA anticodon). The larger grey dotted line represents the standard deviation (± 0.6 minutes) of the instantaneous doubling time for the control. The smaller grey dotted line represents the second standard deviation for the control.

Table 5.4. Cell health and growth profiles for sense codon reassigning systems with the *M. barkeri* tRNA/aaRS.

tRNA Anticodon	Sense Codon Targeted	Instantaneous Doubling Time (min)	Relative System Fitness	Estimated Number of Proteome-Wide Substitutions per Cell Generation
$tRNA^{Pyl}_{AAA}$	Phe UUU	29.9 ± 0.7	0.97	4.68E+05
$tRNA^{Pyl}_{AAG}$	Leu CUU	31.5 ± 1.7	0.92	3.61E+05
$tRNA^{Pyl}_{AAU}$	Ile AUU	31.0 ± 0.9	0.93	5.37E+05
$tRNA^{Pyl}_{AAC}$	Val GUU	31.2 ± 0.8	0.93	7.68E+05
$tRNA^{Pyl}_{AGA}$	Ser UCU	32.0 ± 0.9	0.90	1.46E+06
$tRNA^{Pyl}_{AGG}$	Pro CCU	31.8 ± 0.9	0.91	6.03E+05
$tRNA^{Pyl}_{AGU}$	Thr ACU	30.6 ± 0.9	0.94	1.07E+06
$tRNA^{Pyl}_{AGC}$	Ala GCU	32.4 ± 0.8	0.89	7.20E+05
$tRNA^{Pyl}_{AUG}$	His CAU	30.9 ± 0.7	0.94	7.38E+05
$tRNA^{Pyl}_{AUU}$	Asn AAU	29.3 ± 1.0	0.99	4.43E+05
$tRNA^{Pyl}_{AUC}$	Asp GAU	30.9 ± 1.1	0.93	5.04E+05
$tRNA^{Pyl}_{ACA}$	Cys UGU	30.4 ± 0.5	0.95	2.29E+05
$tRNA^{Pyl}_{ACU}$	Ser AGU	30.3 ± 0.4	0.95	1.03E+06
$tRNA^{Pyl}_{ACC}$	Gly GGU	36.8 ± 2.2	0.79	2.84E+06
$tRNA^{Pyl}_{CAC}$	Val GUG	34.0 ± 1.7	0.85	8.25E+05
$tRNA^{Pyl}_{CGC}$	Ala GCG	34.1 ± 1.7	0.85	9.98E+05
$tRNA^{Pyl}_{CUU}$	Lys AAG	31.0 ± 0.7	0.93	3.90E+06
$tRNA^{Pyl}_{CUC}$	Glu GAG	33.3 ± 1.7	0.87	3.44E+06
$tRNA^{Pyl}_{GCG}$	Arg CGC	34.9 ± 1.4	0.83	2.12E+06
$tRNA^{Pyl}_{UCG}$	Arg CGA	31.6 ± 1.3	0.91	2.02E+06
$tRNA^{Pyl}_{UAG}$	Leu CUA	32.2 ± 1.2	0.90	5.23E+05
$tRNA^{Pyl}_{UAU}$	Ile AUA	31.9 ± 1.4	0.91	1.32E+06
$tRNA^{Pyl}_{UGA}$	Ser UCA	31.7 ± 1.1	0.91	5.38E+05
$tRNA^{Pyl}_{GGG}$	Pro CCC	----[a]	----	----
$tRNA^{Pyl}_{UGU}$	Thr ACA	30.3 ± 0.9	0.95	5.38E+05
$tRNA^{Pyl}_{GCA}$	Cys UGC	32.3 ± 1.4	0.89	7.96E+05
$tRNA^{Pyl}_{CCG}$	Arg CGG	31.6 ± 2.1	0.91	2.68E+06
$tRNA^{Pyl}_{UCU}$	Arg AGA	28.5 ± 0.5	1.01	2.94E+05
$tRNA^{Pyl}_{CCU}$	Arg AGG	31.8 ± 3.7	0.91	4.08E+05
$tRNA^{Pyl}_{UCC}$	Gly GGA	33.2 ± 2.3	0.87	2.07E+05
$tRNA^{Pyl}_{CCC}$	Gly GGG	----[a]	----	----

a The orthogonal tRNA with the indicated anticodon was unable to be successfully constructed.

The rank ordering of the relative cell fitness measurements for the two orthogonal pairs was relatively consistent (Figure 5.7). Cell fitness was calculated by dividing the instantaneous doubling time of a reassigning system by the instantaneous doubling time of a non-reassigning, wild-type GFP expressing control [6]. The two codons targeted for reassignment that had the most significant differences between relative fitness measurements are Gly GGA and Arg AGG. However, the reassignment efficiencies of the Gly GGA and Arg AGG codons are very similar for the two systems (Table 5.1). Codon context effects (discussed further in Chapter 6) may be one reason for the observed similarity in sense codon reassignment efficiency but difference in cell fitness.

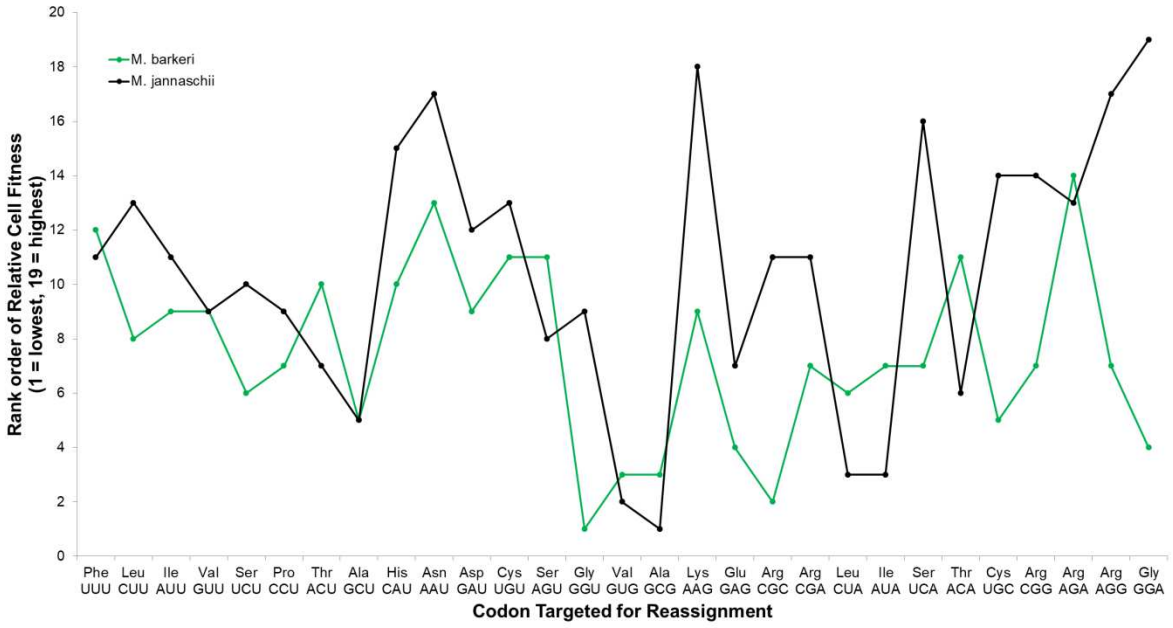

Figure 5.7. Rank order of relative cell fitness vs. codon targeted for reassignment for each orthogonal pair. Cell fitness was calculated by dividing the inverse of the instantaneous doubling time for a reassigning system by the inverse of the instantaneous doubling time for a non-reassigning, wild-type GFP expressing control [6]. 1 = lowest relative fitness, 19 = highest relative fitness.

Effect of competing endogenous E. coli tRNAs on sense codon reassignment

The ribosome and related translational components have evolved to interact with multiple diverse tRNA species in a balanced manner. Elongation factor Tu binds all of the tRNA species nearly equally, and the ribosome interacts with the tRNA at positions outside of the anticodon to normalize the difference between A/U and G/C base pairing interactions [41-44]. If all other factors than the concentration of the tRNA species were equal, it would be expected that the frequencies of amino acid incorporation would be directly related to the concentrations of their aminoacylated tRNAs. Sense codon reassignment efficiencies with an orthogonal pair would then be expected to decrease with increasing numbers of competing orthogonal tRNAs.

The effect of competing endogenous tRNAs on sense codon reassignment efficiency with the orthogonal *M. barkeri* pair is similar to the effect observed with the orthogonal *M. jannaschii* pair [Chapter 2]. A very weak correlation between sense codon reassignment efficiency and endogenous tRNA concentration (R^2=0.17) is observed when all sense codons are considered, and almost no correlation is observed between sense codon reassignment efficiency and endogenous tRNA concentration when sense codons read through a wobble interaction (R^2 = 0.06) are considered (Figure 5.8 a, b). A correlation between sense codon reassignment and endogenous tRNA concentration (R^2=0.42) is observed when the subset of rarely used codons is considered (Figure 5.8 c).

The strong correlation observed when rarely used codons are considered may have may be dependent on the reassignment efficiencies observed at the rarely used Arg CGG and Arg AGG codons (70.3% and 65.1%, respectively). When those data

points are removed from the set of rarely used codons, the correlation (R^2=0.39) between sense codon reassignment efficiency and endogenous tRNA concentration remains (Figure 5.8 d).

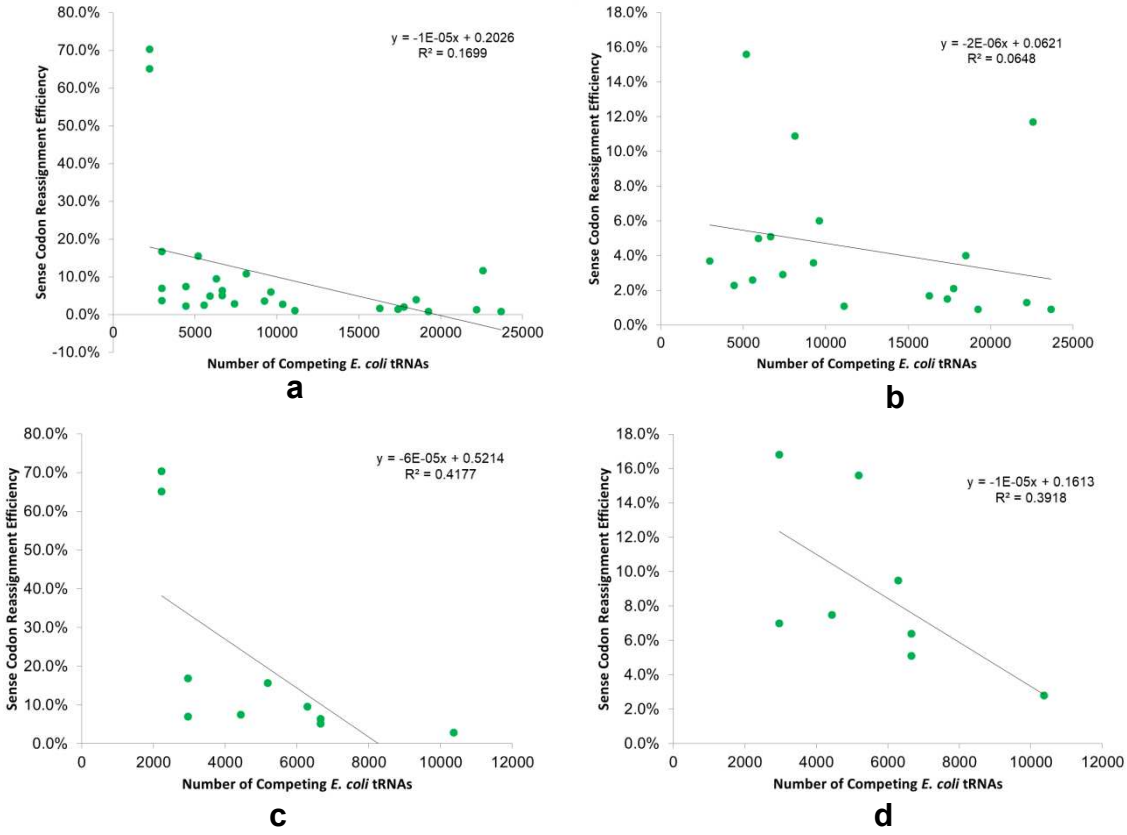

Figure 5.8. Sense codon reassignment efficiency to tyrosine with the engineered orthogonal *M. barkeri* tRNA/aaRS vs. number of competing *E. coli* tRNAs for (a) all sense codons (b) wobble codons (c) rare codons and (d) rare codons with Arg CGG and AGG removed. Note: reassignment of the Arg CGA and Ile AUA codons was excluded from the regressions. Reassignment of the Arg CGA codon was removed because it is a special case where extremely favorable tRNA anticodon-codon energetics is expected. Reassignment of the Ile AUA was removed because the exact number of competing endogenous *E. coli* tRNAs is not known [28].

The correlation between sense codon reassignment efficiency with the orthogonal *M. jannaschii* pair and number of competing *E. coli* tRNAs is also only observed for the rarely used codon subset [Chapter 2]. The fact that for both orthogonal pairs correlation between competing endogenous tRNA concentration and sense codon reassignment efficiency is only apparent when rarely used codons is considered suggests that competing endogenous tRNA concentration is not likely a dominant factor in sense codon reassignment.

The effect of aminoacylation efficiency on orthogonal pair directed sense codon reassignment

The efficiency of the aminoacylation reactions between aaRSs and their cognate tRNAs indirectly affect the process of translation, as they determine the concentrations of aminoacylated tRNAs available in the cell. The critical sequence and structure elements of a tRNA that enable recognition by its cognate aaRS are termed identity elements [45, 46]. Identity elements for most tRNA/aaRS pairs are located in the tRNA acceptor stem or anticodon stem loop [45, 46]. Changing the nucleotide sequence in these regions of a tRNA may affect aminoacylation efficiency by its cognate aaRS, and therefore the effective concentration of aminoacylated tRNA available to the cell.

The anticodon sequence of the orthogonal *M. jannaschii* tRNA is important for recognition by its corresponding *M. jannaschii* aaRS [14]. Using previous partial measurements of aminoacylation efficiencies of *M. jannaschii* tRNA variants with altered anticodon sequences, the decrease in recognition of the orthogonal *M. jannaschii* tRNA by its corresponding aaRS as a result of changing the tRNA anticodon sequence for each of the 30 sense codons targeted for reassignment to tyrosine was estimated

[Chapter 2,[6, 14]]. The sense codon reassignment efficiencies with the orthogonal *M. jannaschii* pair did not strongly correlate with the predicted loss in aminoacylation efficiency due to changing the sequence of the tRNA anticodon, suggesting that aminoacylation efficiency is a contributing factor in sense codon reassignment, but not a dominant factor.

Unlike the orthogonal *M. jannaschii* aaRS, the orthogonal *M. barkeri* aaRS does not use the anticodon of its corresponding tRNA as an important identity element [15, 47, 48]. As a result, the concentrations of aminoacylated orthogonal *M. barkeri* tRNAs is expected to be similar across systems that have different anticodon sequences for targeting different sense codons for reassignment. Comparison of the sense codon reassignment data for the *M. jannaschii* and *M. barkeri* orthogonal pairs may allow for further understanding of the importance of aminoacylation efficiency in sense codon reassignment.

An initial analysis of the two data sets together shows that the difference in sense codon reassignment efficiency between the two orthogonal pairs (defined as *M. barkeri* – *M. jannaschii*) does not correlate strongly to the predicted loss in *M. jannaschii* aminoacylation efficiency as a result of changing the anticodon sequence when all codons (R^2=0.16) or the subset of wobble codons (R^2=0.01) are considered (Figure 5.9 a,b). When the subset of rarely used codons is considered, a correlation between the difference in sense codon reassignment of orthogonal systems and predicted loss in *M. jannaschii* aminoacylation efficiency as a result of changing the tRNA anticodon sequence (R^2=0.533) is observed, suggesting that sense codon reassignment efficiency decreases as aminoacylation efficiency decreases (Figure 5.9 c). As a whole,

aminoacylation efficiency may affect sense codon reassignment efficiency, but is not likely a dominant factor. Currently ongoing and future work will focus on a deeper data analysis of the two data sets, including the consideration of additional subsets of data and methods to normalize the data.

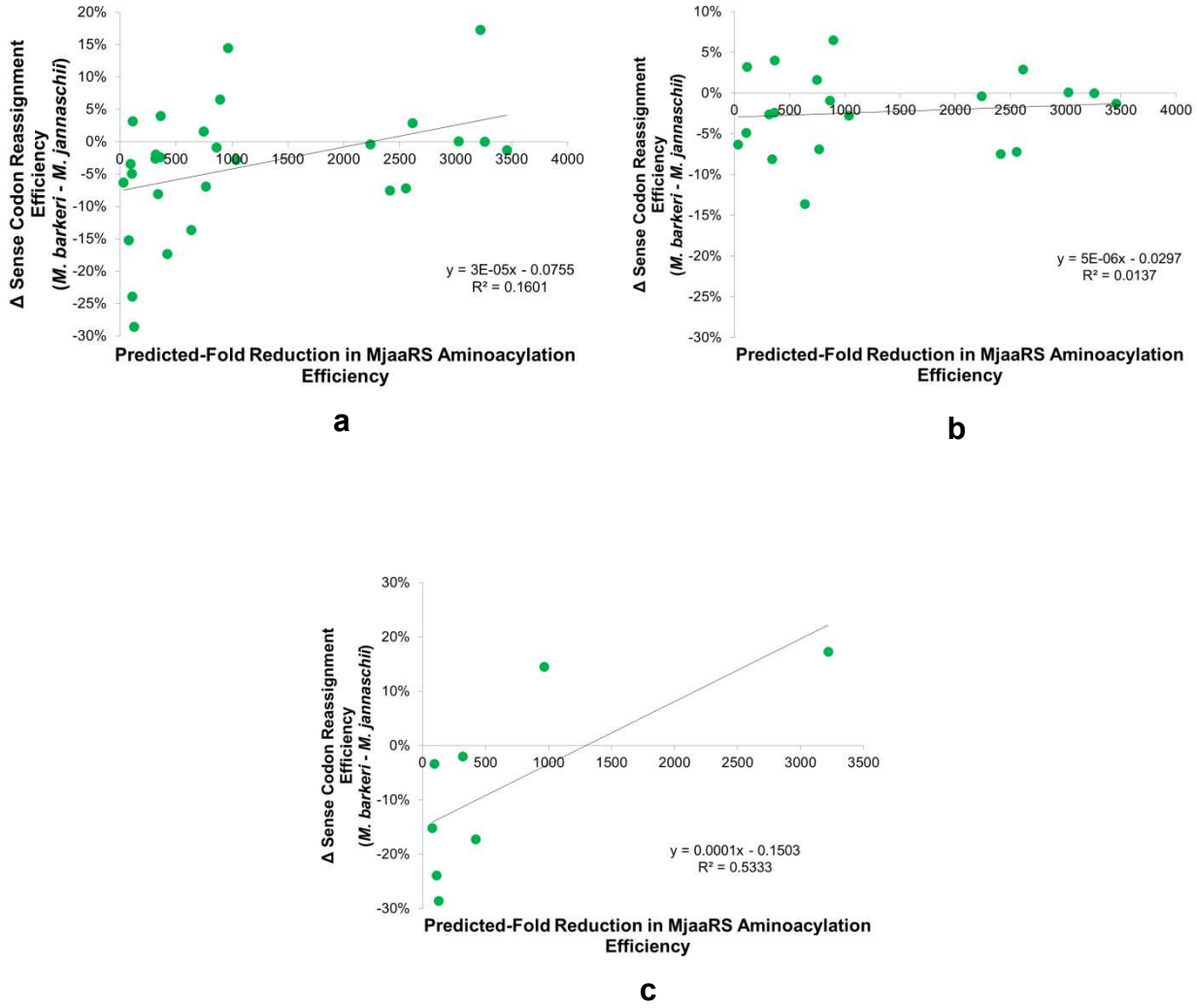

Figure 5.9. Change in sense codon reassignment efficiency (*M. barkeri – M. jannaschii*) vs. predicted-fold reduction in *M. jannaschii* aminoacylation efficiency as a result of changing the anticodon for (a) all sense codons, (b) codons read in *E. coli* through a wobble interaction, and (c) codons used rarely in *E. coli*.

5.5 CONCLUSIONS

The extent to which the degeneracy of the genetic code can be broken with an engineered orthogonal tyrosine-incorporating *M. barkeri* pair at 30 different sense codons was investigated. Significantly, each of the 30 codons is capable of reassignment to tyrosine by the *M. barkeri* pair at a measurable efficiency, ranging from 0.9% to 70.3% by simply providing an orthogonal tRNA with an anticodon capable of Watson-Crick base pairing with the codon targeted for reassignment. The sense codon reassignment efficiencies to tyrosine with the orthogonal *M. barkeri* pair correlate strongly to the sense codon reassignment efficiencies to tyrosine with the orthogonal *M. jannaschii* pair. In addition, the abilities of the orthogonal *M. barkeri* tRNAs to discriminate between the codon targeted for reassignment and closely related codons were very similar to those observed for the orthogonal *M. jannaschii* tRNAs.

The sense codon reassignment efficiencies to tyrosine with the orthogonal *M. barkeri* pair were also used to further investigate the *in vivo* importance of the factors involved in sense codon reassignment. Correlation between sense codon reassignment efficiency with the orthogonal *M. barkeri* pair and the number of competing endogenous tRNAs was only evident when the rarely used subset of sense codons was considered, suggesting that while competing endogenous tRNA competitions may affect sense codon reassignment efficiency, it is not a dominant factor. An initial analysis of the differences between sense codon reassignment efficiencies with the *M. barkeri* pair and sense codon reassignment efficiencies with the *M. jannaschii* pair were not strongly correlated to the predicted reduction in *M. jannaschii* aminoacylation efficiency due to

changing the anticodon of the *M. jannaschii* tRNA, suggesting that aminoacylation efficiency is also not a dominant factor in sense codon reassignment efficiency.

REFERENCES

1. Dumas, A., et al., *Designing logical codon reassignment–Expanding the chemistry in biology.* Chemical science, 2015. **6**(1): p. 50-69.
2. Wan, W., J.M. Tharp, and W.R. Liu, *Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool.* Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 2014. **1844**(6): p. 1059-1070.
3. Liu, C.C. and P.G. Schultz, *Adding new chemistries to the genetic code.* Annual review of biochemistry, 2010. **79**: p. 413-444.
4. Kwon, I., K. Kirshenbaum, and D.A. Tirrell, *Breaking the degeneracy of the genetic code.* Journal of the American Chemical Society, 2003. **125**(25): p. 7512-7513.
5. Biddle, W., M.A. Schmitt, and J.D. Fisk, *Evaluating Sense Codon Reassignment with a Simple Fluorescence Screen.* Biochemistry, 2015. **54**(50): p. 7355-64.
6. Schmitt, M.A., Biddle, Wil, and John D. Fisk, *Mapping the Plasticity of the E. Coli Genetic Code with Orthgonal Pair Directed Sense Codon Reassignment.* submitted, 2017.
7. Biddle, W., Schwark, David G., Schmitt, Margaret A. and John D. Fisk, *Sense Codon Reassignment of the Rare Arginine AGG Codon: Transferability of Improvements Between Orthogonal tRNAs and Synthetases.* submitted, 2018.
8. Lee, B.S., et al., *Incorporation of Unnatural Amino Acids in Response to the AGG Codon.* ACS Chem Biol, 2015. **10**(7): p. 1648-53.

9. Mukai, T., et al., *Reassignment of a rare sense codon to a non-canonical amino acid in Escherichia coli.* Nucleic acids research, 2015. **43**(16): p. 8111-8122.

10. Gingold, H. and Y. Pilpel, *Determinants of translation efficiency and accuracy.* Molecular systems biology, 2011. **7**(1): p. 481.

11. Ruan, B.F., et al., *Quality control despite mistranslation caused by an ambiguous genetic code.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(43): p. 16502-16507.

12. Kramer, E.B. and P.J. Farabaugh, *The frequency of translational misreading errors in E. coli is largely determined by tRNA competition.* Rna-a Publication of the Rna Society, 2007. **13**(1): p. 87-96.

13. Parker, J., *ERRORS AND ALTERNATIVES IN READING THE UNIVERSAL GENETIC-CODE.* Microbiological Reviews, 1989. **53**(3): p. 273-298.

14. Fechter, P., et al., *Major tyrosine identity determinants in Methanococcus jannaschii and Saccharomyces cerevisiae tRNATyr are conserved but expressed differently.* The FEBS Journal, 2001. **268**(3): p. 761-767.

15. Ambrogelly, A., et al., *Pyrrolysine is not hardwired for cotranslational insertion at UAG codons.* Proceedings of the National Academy of Sciences, 2007. **104**(9): p. 3141-3146.

16. Nozawa, K., et al., *Pyrrolysyl-tRNA synthetase–tRNA Pyl structure reveals the molecular basis of orthogonality.* Nature, 2009. **457**(7233): p. 1163.

17. Yanagisawa, T., et al., *Crystallographic studies on multiple conformational states of active-site loops in pyrrolysyl-tRNA synthetase.* Journal of molecular biology, 2008. **378**(3): p. 634-652.

18. Zeng, Y., W. Wang, and W.R. Liu, *Towards reassigning the rare AGG codon in Escherichia coli.* ChemBioChem, 2014. **15**(12): p. 1750-1754.

19. Lee, B.S., et al., *Incorporation of unnatural amino acids in response to the AGG codon.* ACS chemical biology, 2015. **10**(7): p. 1648-1653.

20. Heim, R. and R.Y. Tsien, *Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer.* Current biology, 1996. **6**(2): p. 178-182.

21. Tsien, R.Y., *The green fluorescent protein*. 1998, Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA.

22. Pédelacq, J.-D., et al., *Engineering and characterization of a superfolder green fluorescent protein.* Nature biotechnology, 2006. **24**(1): p. 79.

23. Neidhardt, F.C., *Escherichia coli and Salmonella typhimurium : cellular and molecular biology*. 1997, Washington, D.C.: American Society for Microbiology.

24. Milo, R., *What is the total number of protein molecules per cell volume? A call to rethink some published values.* Bioessays, 2013. **35**(12): p. 1050-1055.

25. Brocchieri, L. and S. Karlin, *Protein length in eukaryotic and prokaryotic proteomes.* Nucleic acids research, 2005. **33**(10): p. 3390-3400.

26. Nakamura, Y., T. Gojobori, and T. Ikemura, *Codon usage tabulated from international DNA sequence databases: status for the year 2000.* Nucleic acids research, 2000. **28**(1): p. 292-292.

27. Li, G.-W., et al., *Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources.* Cell, 2014. **157**(3): p. 624-635.

28. Dong, H., L. Nilsson, and C.G. Kurland, *Co-variation of trna abundance and codon usage inescherichia coliat different growth rates.* Journal of molecular biology, 1996. **260**(5): p. 649-663.

29. Curran, J.F., *Decoding with the A: I wobble pair is inefficient.* Nucleic acids research, 1995. **23**(4): p. 683-688.

30. Gamper, H.B., et al., *Maintenance of protein synthesis reading frame by EF-P and m(1)G37-tRNA.* Nature Communications, 2015. **6**.

31. Crick, F.H., *Codon—anticodon pairing: the wobble hypothesis.* Journal of molecular biology, 1966. **19**(2): p. 548-555.

32. Biddle, W., M.A. Schmitt, and J.D. Fisk, *Modification of orthogonal tRNAs: unexpected consequences for sense codon reassignment.* Nucleic Acids Res, 2016. **44**(21): p. 10042-10050.

33. Wolf, J., A.P. Gerber, and W. Keller, *tadA, an essential tRNA-specific adenosine deaminase from Escherichia coli.* The EMBO journal, 2002. **21**(14): p. 3841-3851.

34. Biddle, W., M.A. Schmitt, and J.D. Fisk, *Modification of orthogonal tRNAs: unexpected consequences for sense codon reassignment.* Nucleic acids research, 2016: p. gkw948.

35. Théobald-Dietrich, A., et al., *Atypical archaeal tRNA pyrrolysine transcript behaves towards EF-Tu as a typical elongator tRNA.* Nucleic acids research, 2004. **32**(3): p. 1091-1096.

36. Lagerkvist, U., *" Two out of three": An alternative method for codon reading.* Proceedings of the National Academy of Sciences, 1978. **75**(4): p. 1759-1762.

37. Lehmann, J. and A. Libchaber, *Degeneracy of the genetic code and stability of the base pair at the second position of the anticodon.* RNA, 2008. **14**(7): p. 1264-1269.

38. Soma, A., et al., *An RNA-modifying enzyme that governs both the codon and amino acid specificities of isoleucine tRNA.* Molecular cell, 2003. **12**(3): p. 689-698.

39. Kramer, E.B. and P.J. Farabaugh, *The frequency of translational misreading errors in E. coli is largely determined by tRNA competition.* RNA, 2007. **13**(1): p. 87-96.

40. Parker, J., *Errors and alternatives in reading the universal genetic code.* Microbiological reviews, 1989. **53**(3): p. 273.

41. Gromadski, K.B., T. Daviter, and M.V. Rodnina, *A uniform response to mismatches in codon-anticodon complexes ensures ribosomal fidelity.* Molecular cell, 2006. **21**(3): p. 369-377.

42. LaRiviere, F.J., A.D. Wolfson, and O.C. Uhlenbeck, *Uniform binding of aminoacyl-tRNAs to elongation factor Tu by thermodynamic compensation.* Science, 2001. **294**(5540): p. 165-8.

43. Phelps, S.S., O. Jerinic, and S. Joseph, *Universally conserved interactions between the ribosome and the anticodon stem-loop of A site tRNA important for translocation.* Molecular cell, 2002. **10**(4): p. 799-807.

44. Olejniczak, M., et al., *Idiosyncratic tuning of tRNAs to achieve uniform ribosome binding.* Nature Structural and Molecular Biology, 2005. **12**(9): p. 788.

45. McClain, W.H., *Rules that govern tRNA identity in protein synthesis.* Journal of molecular biology, 1993. **234**(2): p. 257-280.

46. Giegé, R., M. Sissler, and C. Florentz, *Universal rules and idiosyncratic features in tRNA identity.* Nucleic acids research, 1998. **26**(22): p. 5017-5035.

47. Herring, S., et al., *The amino-terminal domain of pyrrolysyl-tRNA synthetase is dispensable in vitro but required for in vivo activity.* FEBS Lett, 2007. **581**(17): p. 3197-203.

48. Nozawa, K., et al., *Pyrrolysyl-tRNA synthetase-tRNA(Pyl) structure reveals the molecular basis of orthogonality.* Nature, 2009. **457**(7233): p. 1163-7.

CHAPTER 6

PRECISE MEASURMENT OF AMBER STOP CODON REASSIGNMENT EFFICIENCIES IN SUPER-FOLDER GFP WITH THE *M. jannaschii* tRNA/aaRS PAIR

6.1 CHAPTER OVERVIEW

The most widely used method for site-specific incorporation of ncAAs into proteins is the reassignment of an amber stop codon (UAG) using engineered orthogonal tRNA/ aminoacyl tRNA synthetase (aaRS) pairs. The most commonly employed orthogonal pairs are derived from the *Methanocaldococcus jannaschii* (*M. jannaschii*) tyrosyl and the *Methanosarcinae* pyrrolysyl tRNAs and aaRSs [1-2]. Multiple factors affect the efficiency of ncAA incorporation at UAG codons. Orthogonal tRNA/aaRS pairs evolved to recognize different ncAAs exhibit broadly different efficiencies; individual orthogonal pairs also exhibit varying efficiencies depending on site of the UAG codon in the target protein. Variability across pairs is related to the differing enzymatic efficiencies of evolved variants and variability in the cellular availability of different ncAAs. The variability in the efficiency of incorporation at different sites within a target protein has long been recognized as a component of natural nonsense suppression systems but the causes of so called sequence context effects are not clear. The relative importance of sequence context effects for ncAA incorporation have not been systematically studied. More generally, the relative quantitative contributions of the multiple factors affecting the fidelity of translation have

not been clearly established and nonsense suppression systems provide a powerful tool for investigating factors affecting translational fidelity including context effects.

This chapter describes the use of GFP reporters to 1) quantify the sequence context effects of amber codon reassignment using the *M. jannaschii* orthogonal pair system, and 2) to evaluate the contribution of release factor 1 to context effects by examining the relative efficiency of nonsense suppression in the genomically recoded *E. coli* strain where all the instances of the amber stop codon have been converted to other stop signals and peptide release factor 1, which recognizes amber codons, has been deleted.

6.2 INTRODUCTION

To a first approximation the process of translation involves the direct reading of 3 nucleotide codons in mRNA by 3 nucleotide anticodons of tRNAs employing standard Watson-Crick base-pairing rules. The simple picture of translation is incomplete as system composition and multiple specific molecule modifications contribute to alter the reading of individual codons. Certain tRNA modifications are known to alter codon-anticodon interactions and more generally tRNA concentrations, the energy of codon-anticodon interactions, and other effects including tRNA-ribosome interactions, extended tRNA interactions with mRNA, and tRNA-tRNA interaction on the ribosome. The effects of modified bases on tRNA structure also contribute towards determining codon specific translation rates. The relative quantitative contributions of the multiple factors affecting the fidelity of translation have not been clearly established.

The experimental consequence of the many contributing factors is that the efficiency of codon reading is dependent on the mRNA sequence contexts in which a codon appears. Context effects are of particular importance for genetic code expansion

studies as the choice of substitution position within a target protein can have a large effect on the efficiency of non-canonical amino acid (ncAA) introduction. To better understand the factors influencing context effects on the efficiency of reassigning of amber stop (UAG) codons in *E. coli* using the *Methanocaldococcus jannaschii* (*M. jannaschii*) orthogonal tRNA/aaRS system the efficiency of suppression of individual and combinations of amber stop codons in GFP and two related *E. coli* strains was quantified.

Context effects on the efficiency of nonsense suppression were recognized early in the biochemical analysis of translation [3-4]. The majority of experimental work on codon context effects has involved looking at context effects of natural nonsense suppressor tRNAs, but context effects are also evident in sense codon reading [5]. The most well studied systems for context specific reading of sense codons involve programed reading frame shifts and gain of function missense mutations [5-7]. Codon context preferences have additionally been identified through sequence analysis [8-11].

The degree and direction of context effects on amber stop codon suppression are variable and depends on the specific suppressor tRNA employed. Data on multiple natural and engineered amber stop codon suppressors inserting different amino acids into the same seven test positions in a Lac repressor protein showed efficiencies at ranged over approximately 10 fold [12-15]. Each position in the Lac repressor protein was suppressed with a different average efficiency. Each amber suppressor tRNA studied showed a different range and average amber stop codon suppression efficiency across the various positions in the Lac repressor protein; furthermore, the different

suppressor tRNAs showed different patterns of relative increases and decreases of efficiency at the different positions (Table 6.1 and Figure 6.1) [12,13].

Table 6.1. Ranges of amber stop codon suppressor tRNA efficiencies at different test positions in the LacZ reporter protein. Data shown is from [12].

Suppressor tRNA Identity	Lowest UAG Suppression Efficiency	Highest UAG Suppression Efficiency	Average UAG Suppression Efficiency
Ile1	1	10	6
Su2(Gln)	1	26	13
Lys	9	29	17
His	1	35	18
Glu	1	44	23
Su1(ser)	6	54	27
mMet	5	45	31
Cys	17	54	39
Ala2	8	83	44
Arg	11	55	45
Ile2	6	67	48
Su3(Tyr)	11	100	52
Gly2	39	67	54
Val	18	83	58
Su6(Leu)	30	100	64
Gly1	24	100	65
Phe	48	100	77

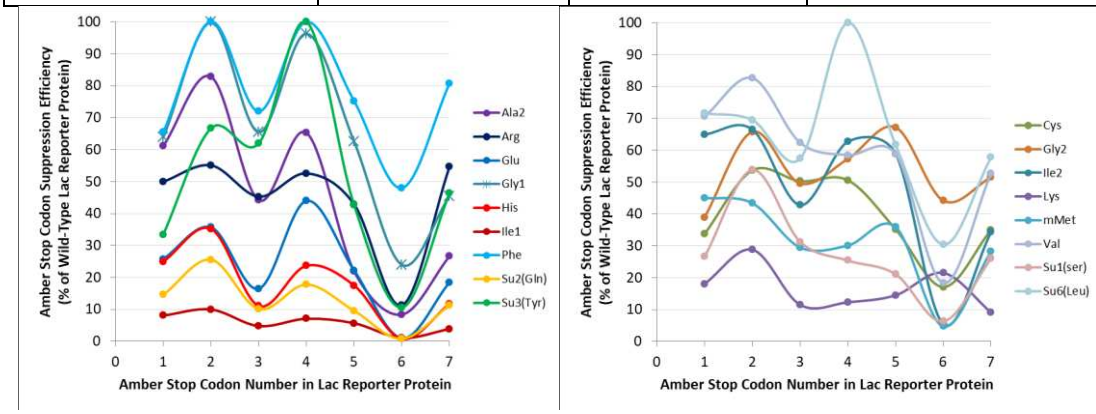

Figure 6.1. Patterns of measured amber stop codon suppression efficiencies of groups of amber stop codon tRNA suppressors at 7 test positions in the LacZ reporter protein. Data is from [12,13]. Left: Amber stop codon tRNA supressors with a similar pattern of response. Right: Amber stop codon tRNA supressors with variable responses.

Factors including tRNA-mRNA interactions (outside of the codon), release factor-mRNA interactions, tRNA-tRNA interactions on the ribosome, and modified nucleotides in the tRNA have been shown to contribute to context effects. However, a clear picture of the relative importance of these and other contributing factors has not emerged (Figure 6.2) reviewed in [16-18]).

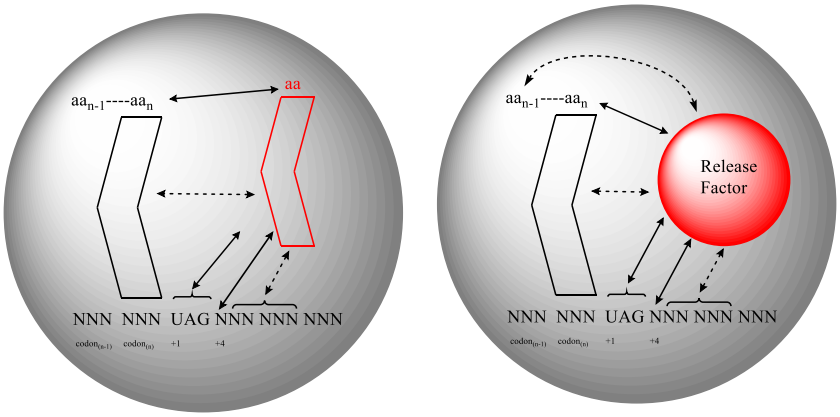

Figure 6.2. Interactions between a suppressor tRNA, realease factor and mRNA at the ribosomal A site that have been shown (dark lines) or may be (dashed lines) important in determining sequence context effects. The extent to which individual elements contribute and the importance of differential effects in specific contexts are not known.

Preferred codon contexts are evident through sequence analysis; however, the function of preferred contexts is not clear [8-9,19-20]. Although some codon pairs are not found in protein coding genes, and different patterns of codon usage appear in highly and weakly expressed genes, simple rules regarding sequence preferences around individual codons are not evident [19,21]. Identifying signatures of codon context bias through bioinformatics analysis is confounded by the necessity of correcting for codon

usage, G-C content and amino acid pair frequencies in proteins and that contexts that increase and decrease reading efficiency could be functionally selected [19,20,22-23].

In the case of amber suppressor tRNAs, the identity of the nucleotides immediately prior to and after the targeted codon have been shown to affect reading efficiency [24-25]. The bases in the vicinity of the amber stop codon are numbered such that the amber stop codon positions are +1,+2, +3, the 5' nucleotides are assigned negative numbers and the 3' nucleotides are assigned positive numbers. Evidence exists both for and against stop codon context effects being related to an extended interaction of a tRNA with mRNA sequence [24-26], through indirect effects of tRNA-ribosome interactions [27-28], and the context dependent efficiency of termination factors [18,29].

Three recent works have tangentially addressed the extent of context effects for the both *M. jannaschii* and *Methanosarcina mazei* (*M. mazei*) orthogonal tRNA/aaRS systems used in genetic code expansion [30-32]. In a study describing an improved vector system for the expression of orthogonal tRNA and aaRSs, Young and co-workers measured the incorporation of fifteen non-canonical amino acids at three amber stop codon positions in GFP. The position specific reassignment efficiencies measured in the study are complicated by the fact that each evaluated system employed a different engineered variant of the *M. jannaschii* aaRS (each has a different ncAA as a substrate). The engineered *M. jannaschii* aaRSs have different solubilities and cellular availabilities, and the produced GFP reporter proteins contained different ncAAs incorporated in response to the amber stop codons. Different ncAAs in the reporter GFP protein may affect folding; all systems, however, were normalized to a single 100%

reference protein. The ncAA incorporation efficiencies varied with the particular evolved *M. jannaschii* aaRS used and with the position of the amber stop codon in the GFP reporter protein. The three positions chosen for amber stop codon suppression showed an approximate 10 fold range of efficiencies, similar to those observed by Kelina and co-workers for natural suppressors tRNAs. The average incorporation efficiencies at position 3, 151 and 133 of GFP across the set of 15 different ncAAs and associated evolved aaRS examined were 16%, 39% and 28% respectively. Examples of ranges of incorporation at position 3, 151 and 133 in GFP for several representative ncAAs are 25%, 85%, and 56% for p-Azidophenylalanine, 8%, 12% and 8% for p-propargyloxyphenylalanine, 23%, 54% and 27% for 2-naphthylalaline and 3%, 39% and 13% for 2,2'-bipyridinylalanine.

Summerer and co-workers selected optimized sequence contexts for amber stop codon suppression using the both the *M. janasschii* and *M. mazei* orthogonal tRNA/aaRS systems as a means to efficiently introduce ncAAs into proteins in the form of an appended tag [33]. Libraries of N-terminal tags containing an amber stop codon between two randomized codons (NNN-NNN-UAG-NNN-NNN) upstream of a chloramphenicol acetyltransferase resistance gene were evaluated for amber stop codon suppression efficiency. The libraries were expressed in cells also expressing ncAA incorporating aaRS variants and iteratively plated onto and scraped from chloramphenicol plates. After 5-11 rounds of selection, evaluating multiple clones through sequencing after each round, the nucleotide sequences upstream and downstream of the amber stop codon in the N-terminal tag converged.

The nucleotide sequence contexts selected for ncAA incorporation using the *M. jannaschii* and *M. mazei* systems were similar, but not identical. Both selections found a strong preference for adenine (A) in the position immediately following the UAG stop codon, as had been previously observed for natural amber stop codon suppression systems [24-25]. In addition strong preference for A at -2, -5, +6 and +7 positions relative to the amber stop codon was also identified. The screen with the *M. jannaschii* tRNA/aaRS pair also identified thymine (T) at the -1 position. Oddly, the authors discuss the selected sequences in terms of the amino acid context of the appended peptide as opposed to the mRNA sequence context of the N-terminal tag. The majority of the proposed mechanisms of codons context effects are related to mRNA sequence effects rather than the amino acid sequence [16,17]. The sequence contexts selected for tyrosine and lysine sense codons were decidedly different than from the contexts selected for amber stop codons, potentially suggesting selection against release factor mRNA interactions in the amber stop codon cases. The selected sequences increased the apparent efficiency of amber stop codon suppression 3 to 5 fold relative to randomly chosen unselected sequences, measured by protein expression yield. The utility of the data presented for evaluating context effects on the orthogonal tRNA/aaRS systems is limited by the small number of selected sequence (less than 10 per selection) and an incomplete analysis of the efficiency of amber stop codon suppression for the selected sequences [33].

Codon context effects were also examined for the orthogonal pyrrolysyl *M. mazei* tRNA/aaRS pair by Xia, Zhou and co-workers using a β-galactosidase assay [34]. In a similar fashion, a library of sequences was generated containing randomized codons

upstream and downstream of the UAG codon (NNN TAG NNN). The library of sequences was inserted into the lacZα gene at a known permissive site. The library was expressed in the presence of plasmids encoding *M. mazei* pyrrolysine aaRSs evolved to incorporate bioorthogonally reactive ncAAs and the ncAAs. The efficiency of amber stop codon suppression was evaluated qualitatively by the intensity of the library colonies' blue color on X-gal plates. Approximately 100 colonies each from the qualitative strongly, weakly and non-suppressing groups were sequenced. For the strongly and weakly suppressing groups a preference for A at +4 and C at +5 positions were observed. Weaker consensus sequences were evident at additional positions in the strongly suppressing clones but not in the weakly suppressing clones. An inverse bias against A at +4 and C at +5 was observed in the non-suppressing colonies. The overall consensus sequence for strong amber stop codon suppression was AAT TAG ACT. The consensus sequence selected by the Summerer and co-workers analysis, (A/C)A(G/T) TAG AA(G/T) contained the same preference, with the caveat that additional stronger consensuses were found at positions -5, -6, +7,+8 and +9 than at the sites analyzed by Xia, Zhou and co-workers. The context effects of sequence variation around the amber codon were quantified using the β-galactosidase assay for a set of 90 sequences. The data are difficult to interpret as many single mutants are not present and the amber stop codon incorporation efficiencies are not reported as absolute efficiencies (relative to 0% and 100% references) but rather normalized to the selected consensus sequence. The data does show that that the range of observed context effects is from 0.5 - 100% of the most efficient sequence selected. In cases where comparable single and double mutants were evaluated, the effects of the mutations on

amber stop codon suppression efficiency appear to be largely non-additive. A separate quantitative assay was employed to look at the set of 18 single site mutants away from the consensus sequence; this data indicates that the context was exquisitely sensitive to the exact sequence as any change at the +1 or -1 positions resulted in an average 33% reduction in amber stop codon reassignment efficiency, and any change in sequence at +2 or -2 individually caused a 25% reduction in efficiency.

6.3 MATERIALS AND METHODS

The fluorescence-based screen for amber stop codon suppression efficiency has been described [36]. Detailed experimental protocols, including methods for construction of amber stop codon-containing GFP variants, measurement of system growth and fluorescence, and calculation of codon reassignment efficiencies are included in Appendix 1. The appendix also includes full vector sequences, oligonucleotide primer sequences, detailed cell strain information, and general reagents and materials.

6.4 RESULTS AND DISCUSSION

Quantification of Amber codon Context Effects in GFP

In order to better understand the extent of context effects in amber stop codon suppression using the orthogonal *M. jannaschii* tyrosyl tRNA/aaRS pair, we employ a fluorescence-based screen to quantify amber stop codon reassignment efficiency at 8 different tyrosine codons in superfolder green fluorescent protein (GFP). In order remove effects related to protein folding and performance differences, the suppression of amber stop codons was done with a tyrosine incorporating orthogonal *M. jannaschii* tRNA/aaRS pair at tyrosine positions in GFP. Tyrosine codons in the GFP gene were

individually substituted by amber codons such that the protein produced by amber stop codon suppression would be phenotypically identical to wild-type GFP, allowing fluorescence to be a direct measurement of amber stop codon suppression efficiency. The only difference between each GFP reporter was the location of the amber stop codon, and as a result all variables that affect amber stop codon suppression efficiency other than mRNA structure and codon context were eliminated. The identity of and the expression levels of the orthogonal *M. jannaschii* tyrosyl tRNA/aaRS, the promoter driving the GFP reporter gene, and the cell lines used (cellular environment) were all consistent between systems.

The measured amber stop codon suppression efficiencies for the single amber stop codon variants were then used to predict the suppression efficiencies of GFP variants containing multiple amber stop codons; the hypothesis that the suppression of multiple amber stop codons is independent and additive (based on single UAG variants) was then evaluated by comparing the predicted and measured amber stop codon suppression efficiencies for 14 multiple amber stop codon containing GFP variants. Finally, the context effects were evaluated in an *E. coli* strain where the competing release factor 1 has been removed in order to assess the contribution of release factor mRNA interactions in sequence context effects.

Evaluation of tyrosine positions in GFP as reporters of amber stop codon suppression efficiency

In order to use GFP fluorescence as a quantitative measure of amber stop codon suppression efficiency, two requirements must be met. First, GFP produced from a gene containing an amber stop codon must not be altered in structure, folding or

fluorescence compared to the wild type GFP acting as the 100% fluorescence reference. To meet this requirement, the 9 tyrosine codons throughout the GFP gene were chosen as the positions for mutation to the amber stop codon such that the orthogonal *M. jannaschii* tRNA/aaRS pair would restore wild type protein sequence through the incorporation of tyrosine at the UAG codons.

The second requirement for the choice of test codon positions is that failed instances of amber reassignment must lead to truncated, non-fluorescent products. A variant containing an amber stop codon that results in a truncated product that has measurable levels of GFP fluorescence (without the orthogonal *M. jannaschii* tRNA/aaRS present) cannot be used as a reporter for amber stop codon suppression efficiency, as fluorescence would not be indicative of amber stop codon suppression by the orthogonal *M. jannaschii* tRNA/aaRS pair. The GFP protein sequence contains 9 tyrosine residues. Each of the single tyrosine codon to UAG codon GFP variants was constructed and transformed into *E. coli* DH10B cells without the orthogonal *M. jannaschii* tRNA/aaRS pair. In each system, GFP production was induced and GFP fluorescence was monitored (Figure 6.3).

No significant fluorescence from the GFP products for 8 of the 9 single amber stop codon variants was observed. One variant, Tyr 237 UAG, showed GFP fluorescence above wild-type GFP. Position 237 is the second to last residue in the GFP sequence, and in the construct used is followed only by a lysine residue and a hexa-histidine tag. It is not surprising that truncating GFP at position 237 results in a system where wild-type levels of GFP fluorescence are produced. Position 237 was

therefore eliminated from the group of single stop codon GFP variants used to measure the range of amber stop codon suppression efficiencies.

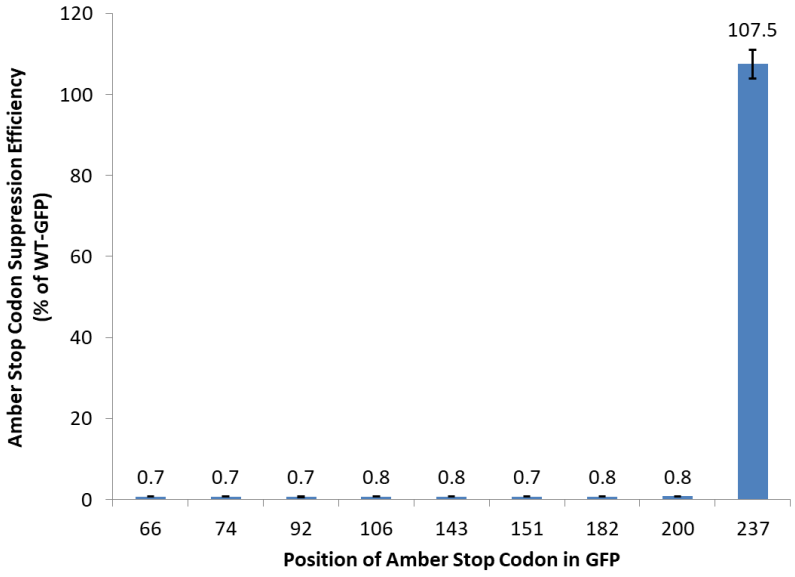

Figure 6.3. Incorporation efficiencies of tyrosine at UAG codons in GFP when no orthogonal tRNA/aaRS pair is present vs. position of UAG codon in GFP.

The background levels of amber stop codon suppression observed are higher than background levels of sense codon reassignment observed in related GFP reporters (Evaluating reading NNN codons as Tyr at position 66) [35, 36]. The higher background levels suggest a fair degree of background missense reassignment (~0.5%) in the cells absent a dedicated amber stop codon suppressor tRNA. The fluorescence of the GFP gene with an amber stop codon at position 66 is particularly high as only suppression of this amber stop codon as tyrosine should lead to any measurable fluorescence. The high relative fluorescence of the Tyr 66 UAG mutant relative to Tyr 66 sense codon mutants (~0.3% across 20 wobble codons [35]) suggests that the primary missense amber stop codon suppressor in *E. coli* DH10Bs is the tyrosine tRNA.

Measurement of reassignment efficiencies for single UAG GFP variants using the M. jannaschii tyrosyl tRNA/aaRS pair

Amber stop codon reassignment efficiency was measured with the *M. jannaschii* tyrosyl tRNA/aaRS pair for each of the 8 GFP variants in *E. coli* DH10B cells. The measured incorporation efficiencies, reported as the cell density corrected fluorescence of the stop codon suppressing system divided by the cell density corrected fluorescence of a system in which wild-type GFP was expressed, ranged from 51%-116% (Figure 6.4). The sequence contexts of the various positions are indicated in Table 6.2. The sequence contexts include cases were the -1 bases are U, C, and A and the +1 bases are C, A, and G.

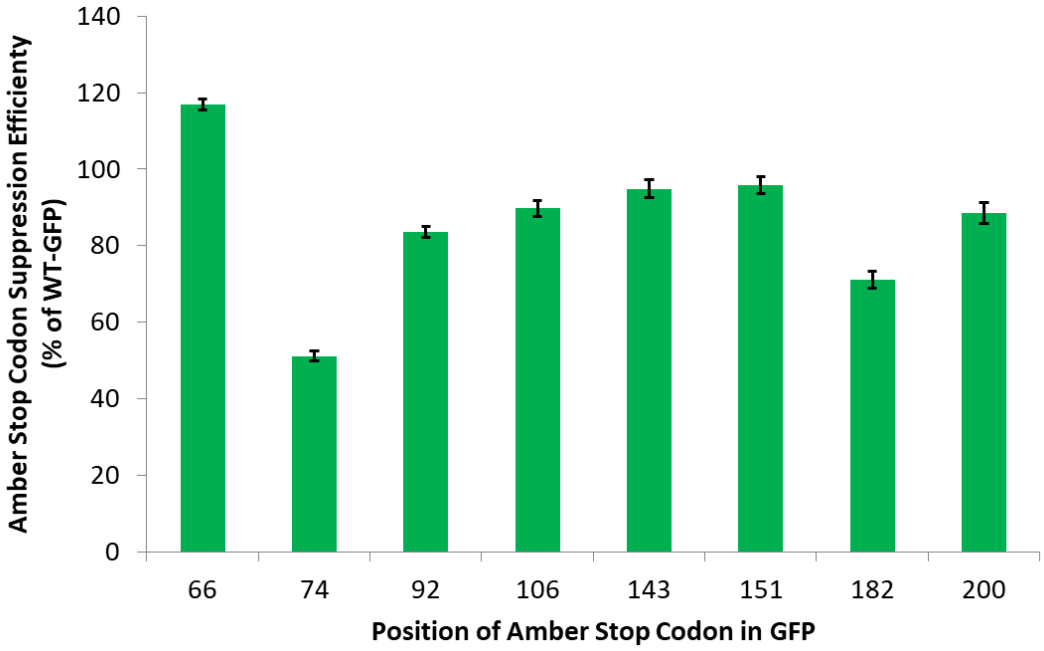

Figure 6.4. Amber stop codon reassignment efficiencies to tyrosine using the *M. jannaschii* tyrosyl tRNA/aaRS vs. position of UAG codon in GFP.

Table 6.2. Sequences directly upstream and downstream of UAG codons in 8 single stop codon GFP variants, and the suppression efficiencies of the amber stop codons with the *M. jannaschii* tyrosyl tRNA/aaRS in *E. coli* DH10B.

Position of Amber Stop Codon in GFP	Sequence including 2 codons 5' and 3' of UAG codon	Amber Stop Codon Suppression Efficiency in *E. coli* DH10B
66	CUG ACC UAG GGC GUC	116.9 ± 1.5
74	UCC CGU UAG CCG GAC	51.2 ± 1.4
92	GAA GGC UAG GUA CAG	83.6 ± 1.3
106	GGG ACC UAG AAA ACC	89.7 ± 2.1
143	CUC GAA UAG AAC UUC	94.9 ± 2.3
151	AAC GUA UAG AUC ACG	95.9 ± 2.2
182	GAC CAC UAG CAG CAG	71.1 ±2.3
200	AAC CAC UAG CUG UCC	88.5 ±2.8

The nucleotide after the UAG codon was the first sequence context effect to be recognized [15]. If the nucleotide following the amber codon is an A or G it can potentially base pair with the universally conserved U33 of tRNAs. For the SupE amber suppressor tRNA, changing the nucleotide following the amber stop codon from C to A caused a 10 fold increase in the efficiency of suppression [14,15]. This trend was supported in a large study employing 42 different amber stop codon contexts in a LacI-LacZ fusion protein where amber stop codon suppression leads to functional β-galatosidase activity [14]. The amber stop codon suppression efficiencies were generally the greatest in contexts with a +1 A.

Three of the four least efficient amber stop codon suppressions with the orthogonal *M. jannaschii* tRNA/aaRS were in the context of a +1 C and three of the four most efficient amber stop codon suppressions occurred in a context with a +1 A, consistent with previous findings for SupE and other amber suppressor tRNAs. In the 8 positions studied there is not a clear trend of effect related to the nucletodie at the -1

position. Further, the present study employs a highly efficient orthogonal tRNA/aaRS pair for which the effect of aa-tRNA on amber stop codon suppression efficiency has not been directly studied.

Measurements of amber stop codon suppression efficiencies for multiple UAG GFP variants in E. coli DH10B

The predominant proposed mechanisms for context effects all involve local interactions and suggest that the effects of context on multiple incorporations should be additive and independent. The synergy / independence of multiple amber stop codon mutations has not been subjected to wide spread systematic analysis and is potentially useful in evaluating the contribution of local and global effects on codon context variation. To be clear, additive and independent interactions imply that if two different amber stop codons are individually suppressed with 50% efficiency then those two amber stop codons as a double mutant would be expected to be suppressed at 25% efficiency. The amber stop codon reassignment efficiency for 14 different double, triple, quadruple and quintuple tyrosine to amber stop codon GFP mutants were measured (Table 6.2).

The measured reassignment efficiencies for the multisite amber mutants were compared to predicted reassignments based on the efficiencies of single site mutants. Both the measured and predicted reassignment efficiencies show that with increasing numbers of UAG codons in GFP, the overall UAG reassignment efficiencies decrease (Figure 6.5). For all but one of the variants the measured amber stop codon reassignment efficiencies were lower than the predicted efficiencies, suggesting that a least one contributing component is exerting a synergistic as opposed to independent

effect. Possible explanations (unrelated to local context) that could result in synergic interactions involve differential ribosome drop off frequencies or mRNA degradation times related to overall translation efficiency [37, 38] (Figure 6.5).

Table 6.2. Measured and predicted amber stop codon suppression efficiencies for GFP variants with multiple UAG codons with the *M. jannaschii* tyrosyl tRNA/aaRS in *E. coli* DH10B.

Number of Amber Stop Codons	Position of Amber Stop Codons	Predicted Amber Stop Codon Suppression Efficiency (% of WT-GFP) in *E. coli* DH10B	Measured Amber Stop Codon Suppression Effiency (% of WT-GFP)
2	Y74, Y182	36.4 ± 1.5	28.7 ± 1.1
2	Y66, Y143	110.9 ± 3.1	70.6 ± 3.1
2	Y66, Y74	59.8 ± 1.8	33.6 ± 0.7
2	Y74, Y143	48.5 ± 1.8	58.8 ± 1.3
2	Y74, Y200	45.3 ± 1.9	37.9 ± 0.9
2	Y74, Y92	42.8 ± 1.3	33.9 ± 1.4
2	Y143,Y200	84.0 ± 3.3	68.9 ± 3.5
2	Y143,Y182	67.4 ± 2.7	53.9 ± 2.2
2	Y106,Y143	85.2 ± 2.9	66.6 ± 2.4
2	Y92,Y143	79.3 ± 2.3	57.9 ± 2.2
3	Y74, Y92, Y200	37.8 ± 1.7	26.2 ± 0.9
3	Y74, Y182, Y200	32.2 ± 1.7	24.7 ± 0.8
4	Y74, Y92, Y106, Y200	34.0 ± 1.7	19.8 ± 0.3
5	Y74, Y92, Y106, Y182, Y200	24.1 ± 1.4	13.2 ± 0.3

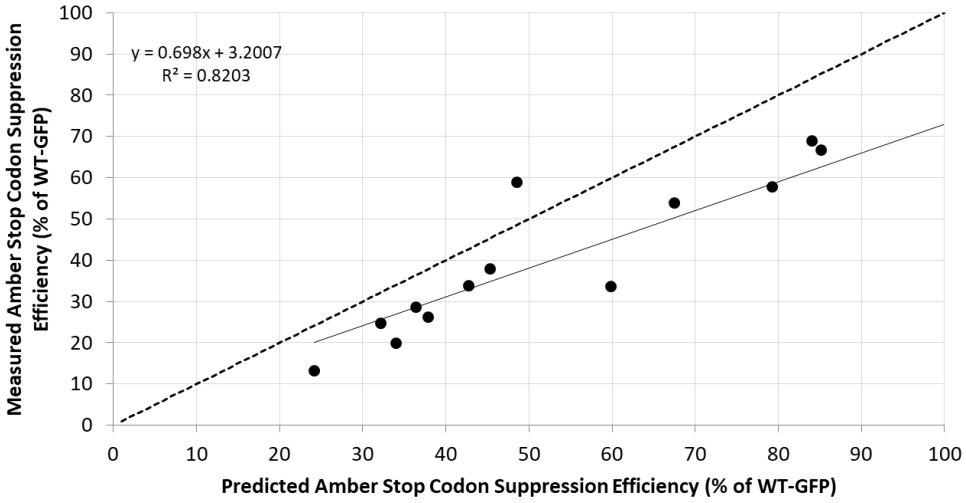

Figure 6.5. Measured vs. predicted UAG reassignment efficiencies for GFP variants containing multiple UAG codons in *E. coli* DH10B.

Role of Ribosomal Release Factor 1 in Determining Amber Stop Codon Context

One of the predominant hypothesis about the source of context effects posits that differential interactions between the protein release factor and the mRNA contexts of stop codons is responsible for observed context effects. Separate studies both support and refute the role of release factors in determining context effects [15,16]. In order to evaluate the contribution of release factor-mRNA interactions in determining context effects for amber stop codon suppression with the orthogonal *M. jannaschii* tRNA/aaRS pair, the same set of single and multiple amber stop codon GFP mutants were evaluated in a genomically recoded *E. coli* strain, C321.ΔRF1, in which all instances of the UAG codon in the genome were replaced with alternate termination signal (UAA) and the gene coding for ribosomal release factor 1 (RF1) was removed [39]. The measured amber stop codon suppression efficiencies for the single mutants with the *M. jannaschii* tRNA/aaRS in *E. coli* C321.ΔRF1 are listed in table 6.3 and shown graphically in Figure 6.6.

Table 6.3. Amber Stop Codon Suppression efficiencies with the *M. jannaschii* tyrosyl tRNA/aaRS and mRNA sequences surrounding the UAG codons in the 8 singe amber stop codon GFP variants in *E. coli* C321.ΔRF1.

Position of UAG Codon	Sequence including 2 codons 5' and 3' of UAG codon	Reassignment Efficiency in *E. coli* DH10B	Reassignment Efficiency in C321.ΔRF1
66	CUG ACC UAG GGC GUC	116.9 ± 1.5	96.8 ± 2.3
74	UCC CGU UAG CCG GAC	51.2 ± 1.4	75.5 ± 2.8
92	GAA GGC UAG GUA CAG	83.6 ± 1.3	76.4 ± 1.9
106	GGG ACC UAG AAA ACC	89.7 ± 2.1	81.5 ± 2.0
143	CUC GAA UAG AAC UUC	94.9 ± 2.3	101.9 ± 3.5
151	AAC GUA UAG AUC ACG	95.9 ± 2.2	97.8 ± 2.2
182	GAC CAC UAG CAG CAG	71.1 ±2.3	89.9 ± 2.0
200	AAC CAC UAG CUG UCC	88.5 ±2.8	103.7 ± 2.4

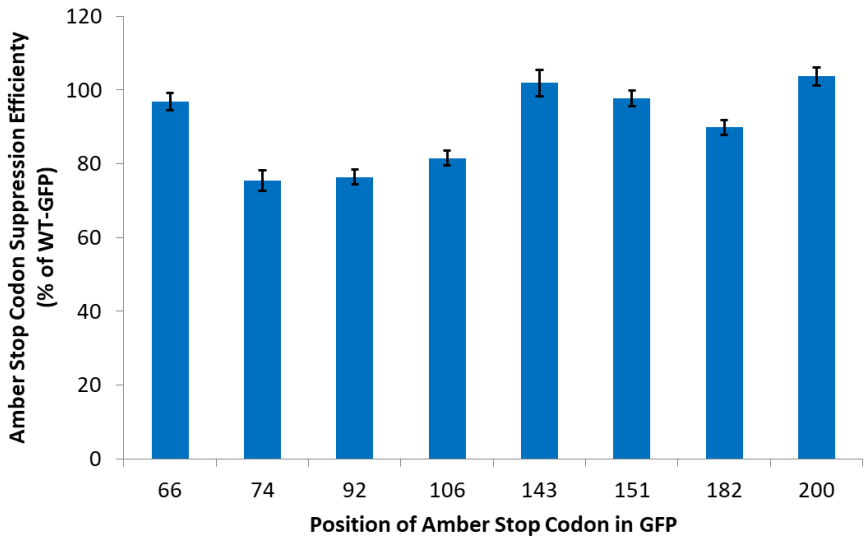

Figure 6.6. Amber stop codon suppression efficiencies to tyrosine using the *M. jannaschii* tyrosyl tRNA/aaRS vs. position of UAG codon in GFP in *E. coli* C321.ΔRF1.

The single amber stop codon GFP variants were capable of being suppressed with the orthogonal *M. jannaschii* tyrosyl tRNA/aaRS pair with 75%-103% efficiencies in the genomically recoded *E. coli strain,* compared to 51%-116% in *E. coli* DH10Bs.

The general trend of the relative efficiency at the various positions was similar in the two systems (Figure 6.7). Figure 6.7 shows the rank ordering of the position of the most efficiently suppressed amber stop codon (1) to the position of the least efficiently suppressed amber stop codon (6). There were significant differences in observed amber stop codon suppression efficiencies between the two cell lines. The least efficiently suppressed amber stop codon positions in *E coli* DH10B cells (74, 182, and 200) were all significantly improved in the RF1 deleted cell line (+24%, +18%, and +15%, respectively). However, four of the positions tested (66, 92, 106, 143) showed

decreases in absolute amber stop codon suppression efficiency upon removal of RF1 (-20, -7%,-8%,-7%, respectively).

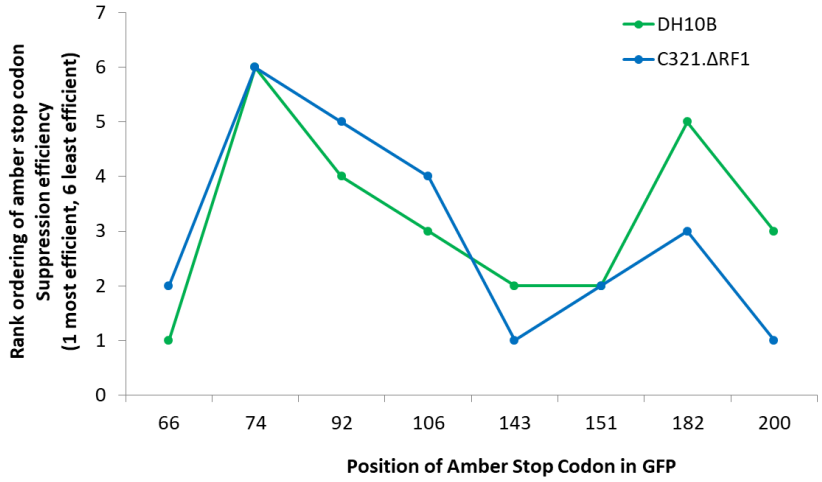

Figure 6.7. Rank ordering of amber stop codon reassignment efficiencies with the orthogonal *M. jannaschii* tyrosyl tRNA/aaRS in *E. coli* DH10B and C321.ΔRF1.

The +4 sequence contexts for the sets of codons that showed both increased and decreased amber stop codon suppression efficiencies upon removal of RF1 were consistent. All of the sequence contexts that showed increased amber stop codon suppression efficiency upon RF1 deletion had a +4 C and either U or C at the -1 position relative to the amber stop codon. The sequence contexts that showed reduced amber stop codon suppression efficiency upon RF1 removal contained a -1 C and +4 G or A. The two positions that were least impacted by RF1 removal contained -1 and +4 A. Although the data set is small, the consistency of the sequence contexts and direction of changes in suppression efficiency suggest that RF1 does play a role in determining context effects.

Measurements of amber stop codon suppression efficiencies for multiple UAG containing GFP variants

The measured vs predicted amber stop codon suppression efficiencies for the multisite amber stop codon mutants were also generated for *E. coli* C321.ΔRF1 (Table 6.4). A similar trend to *E. coli* DH10B is observed in *E. coli* C321.ΔRF1. The measured and predicted amber stop codon suppression efficiencies indicate that increasing numbers of UAG codons in GFP decreases the overall amber stop codon suppression efficiency (Figure 6.8). In the RF1 deleted cell line, the measured amber stop codon suppression efficiencies were consistently lower than the efficiencies predicted for non-interacting independent sites, suggesting that a least one component contributing to context effects is exerting a synergistic as opposed to independent effect.

Table 6.4 Measured and predicted reassignment efficiencies for GFP variants with multiple UAG codon in *E. coli* DH10B and C321.ΔRF1.

Number of Amber Stop Codons	Position of Amber Stop Codons	Predicted Amber Stop Codon Suppression Efficiency (% of WT-GFP) in *E. coli* DH10B	Measured Amber Stop Codon Suppression Effiency (% of WT-GFP) in *E. coli* DH10B	Predicted Amber Stop Codon Suppression (% of WT-GFP) in *E. coli* C321.ΔRF1	Measured Amber Stop Codon Suppression (% of WT-GFP) in *E. coli* C321.ΔRF1
2	Y74, Y182	36.4 ± 1.5	28.7 ± 1.1	67.8 ± 3.0	50.9 ± 1.9
2	Y66, Y143	110.9 ± 3.1	70.6 ± 3.1	98.6 ± 4.1	60.7 ± 1.7
2	Y66, Y74	59.8 ± 1.8	33.6 ± 0.7	73.0 ± 3.3	51.2 ± 1.6
2	Y74, Y143	48.5 ± 1.8	58.8 ± 1.3	76.9 ± 3.9	56.7 ± 5.8
2	Y74, Y200	45.3 ± 1.9	37.9 ± 0.9	78.3 ± 3.5	70.9 ± 1.8
2	Y74, Y92	42.8 ± 1.3	33.9 ± 1.4	57.7 ± 2.6	50.9 ± 2.9
2	Y143,Y200	84.0 ± 3.3	68.9 ± 3.5	105.7 ± 4.4	76.8 ± 2.7
2	Y143,Y182	67.4 ± 2.7	53.9 ± 2.2	91.5 ± 3.7	68.8 ± 1.8
2	Y106,Y143	85.2 ± 2.9	66.6 ± 2.4	83.0 ± 3.5	61.0 ± 2.2
2	Y92,Y143	79.3 ± 2.3	57.9 ± 2.2	77.8 ± 3.3	51.6 ± 1.7
3	Y74, Y92, Y200	37.8 ± 1.7	26.2 ± 0.9	59.8 ± 3.0	56.0 ± 2.8
3	Y74, Y182, Y200	32.2 ± 1.7	24.7 ± 0.8	70.3 ± 3.5	66.0 ± 1.5
4	Y74, Y92, Y106, Y200	34.0 ± 1.7	19.8 ± 0.3	48.7 ± 2.8	51.3 ± 1.5
5	Y74, Y92, Y106, Y182, Y200	24.1 ± 1.4	13.2 ± 0.3	43.8 ± 2.7	40.0 ± 1.5

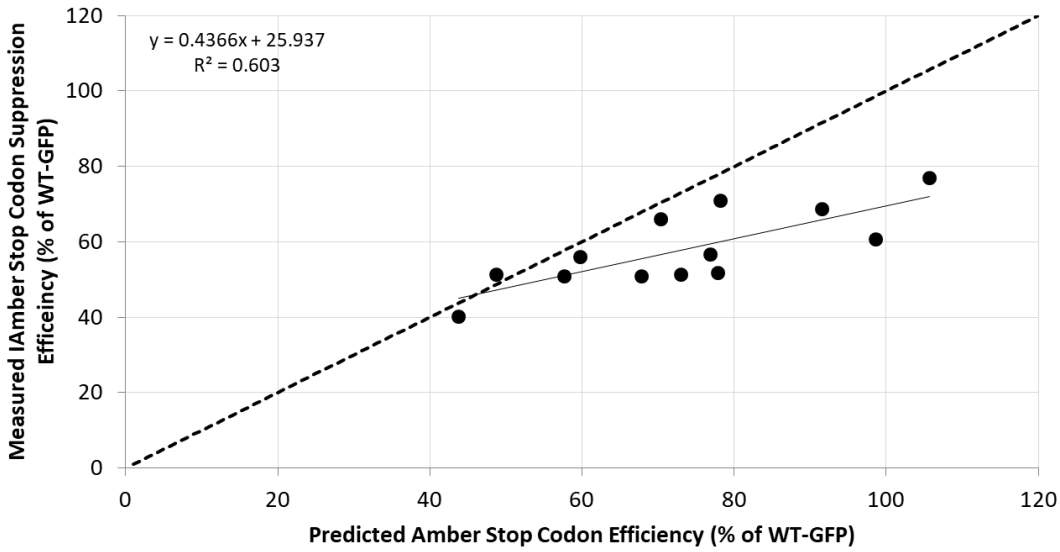

Figure 6.8 Measured UAG suppression efficiencies vs. predicted UAG suppression efficiencies to tyrosine for multiple UAG containing GFP variants with the *M. jannaschii* tyrosyl tRNA/aaRS pair in *E. coli* C321.ΔRF1.

6.5 CONCLUSIONS

The application of a GFP fluorescence-based screen to quantitatively measure amber stop codon suppression efficiency to tyrosine with the orthogonal *M. jannaschii* tyrosyl tRNA/aaRS was described. Amber stop codon reassignment efficiencies ranged from 51%-116% in *E. coli* DH10B and 75%-103% in the genomically recoded *E. coli* that does not contain ribosomal release factor 1 for eight single stop codon GFP variants. The only difference between each of the GFP variants was the location of the UAG codon, demonstrating that local codon context is an important factor for UAG suppression efficiency with an orthogonal tRNA/aaRS pair. Measurement of suppression efficiencies for 14 multiple UAG-containing GFP variants showed that the effects of individual UAG codons are not strictly independent and additive, and that some factors affecting context are non-local and negatively synergistic. The comparison

of context effects between a normal cell line and a strain where the competing release factor RF1 for amber stop codon reading has been deleted supports the conclusion that context effects are in part related to the context preferences of the release factors, but clearly involve other contributions.

The present study is a first small step towards the application of orthogonal tRNA/aaRS pairs to understanding context effects on both sense and nonsense codon reassignment. Because of the complexity involved, a thorough understanding of context effects will require the generation of a large data set of well controlled experiments. The orthogonal tRNA/aaRS systems will be particularly useful in generating datasets because they can be manipulated and evaluated in the context of constant natural translation systems. The next step in evaluating the *M. jannaschii* context effects is a much broader study involving libraries of sequences around the amber stop codon, with precise quantification of efficiencies and related measurements to control for protein sequence changes on quantification, expanding on studies involving natural suppressor tRNAs and orthogonal tRNA/aaRS pairs. We have generated several orthogonal tRNA/aaRS systems that incorporate tyrosine at varying efficiencies. One question that has not been previously quantitatively evaluated is how important the concentration of competing nonsense suppressor is to observed context effects. The efficiency of orthogonal tRNA/aaRS pairs evolved for non-canonical amino acid incorporation varies considerably and the extent that codon context effects vary with competing component concentrations may be the deciding factor determining whether ncAAs may be introduced into diverse proteins.

REFERENCES

1. Liu, Chang C., and Peter G. Schultz. "Adding new chemistries to the genetic code." *Annual review of biochemistry* 79 (2010): 413-444.
2. Wan, Wei, Jeffery M. Tharp, and Wenshe R. Liu. "Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool." *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* 1844.6 (2014): 1059-1070.
3. Wan, Wei, Jeffery M. Tharp, and Wenshe R. Liu. "Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool." *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* 1844.6 (2014): 1059-1070.
4. Yahata, Hiroko, Yoshiko Ocada, and Akira Tsugita. "Adjacent effect on suppression efficiency." *Molecular and General Genetics MGG* 106.3 (1970): 208-212.
5. Murgola, Emanuel J., Frances T. Pagel, and Kathryn A. Hijazi. "Codon context effects in missense suppression." *Journal of molecular biology* 175.1 (1984): 19-27.
6. Farabaugh, Philip J., and Glenn R. Björk. "How translational accuracy influences reading frame maintenance." *The EMBO journal* 18.6 (1999): 1427-1434.
7. Kramer, Emily B., and Philip J. Farabaugh. "The frequency of translational misreading errors in *E. coli* is largely determined by tRNA competition." *Rna* 13.1 (2007): 87-96.
8. Yarus, M., and L. S. Folley. "Sense codons are found in specific contexts." *Journal of molecular biology* 182.4 (1985): 529-540.

9. Behura, Susanta K., and David W. Severson. "Codon usage bias: causative factors, quantification methods and genome-wide patterns: with emphasis on insect genomes." *Biological Reviews* 88.1 (2013): 49-61.

10. Behura, Susanta K., and David W. Severson. "Bicluster pattern of codon context usages between flavivirus and vector mosquito Aedes aegypti: relevance to infection and transcriptional response of mosquito genes." *Molecular genetics and genomics* 289.5 (2014): 885-894.

11. Chevance, Fabienne FV, Soazig Le Guyon, and Kelly T. Hughes. "The effects of codon context on *in vivo* translation speed." *PLoS genetics* 10.6 (2014): e1004392.

12. Kleina, Lynn G., et al. "Construction of *Escherichia coli* amber suppressor tRNA genes: II. Synthesis of additional tRNA genes and improvement of suppressor efficiency." *Journal of molecular biology* 213.4 (1990): 705-717.

13. Normanly, Jennifer, et al. "Construction of *Escherichia coli* amber suppressor tRNA genes: III. Determination of tRNA specificity." *Journal of molecular biology* 213.4 (1990): 719-726.

14. Bossi, Lionello. "Context effects: translation of UAG codon by suppressor tRNA is affected by the sequence following UAG in the message." *Journal of molecular biology* 164.1 (1983): 73-87.

15. Bossi, Lionello, and John R. Roth. "The influence of codon context on genetic code translation." *Nature* 286.5769 (1980): 123.

16. Buckingham, R. H. "Codon context." *Experientia* 46.11-12 (1990): 1126-1133.

17. Brar, Gloria A. "Beyond the triplet code: context cues transform translation." *Cell* 167.7 (2016): 1681-1692.

18. Tate, Warren P., and Sally A. Mannering. "Three, four or more: the translational stop signal at length." *Molecular microbiology*21.2 (1996): 213-219.

19. Tats, Age, Tanel Tenson, and Maido Remm. "Preferred and avoided codon pairs in three domains of life." *BMC genomics*9.1 (2008): 463.

20. Moura, Gabriela R., et al. "Species-specific codon context rules unveil non-neutrality effects of synonymous mutations." *PloS one* 6.10 (2011): e26817.

21. Das, Debojyoti, et al. "Occurrence of all nucleotide combinations at the third and the first positions of two adjacent codons in open reading frames of bacteria." *Current Science*90.1 (2006): 22-24.

22. Czech, Andreas, et al. "Silent mutations in sight: co-variations in tRNA abundance as a key to unravel consequences of silent mutations." *Molecular Biosystems* 6.10 (2010): 1767-1772.

23. Zhang, Gong, Magdalena Hubalewska, and Zoya Ignatova. "Transient ribosomal attenuation coordinates protein synthesis and co-translational folding." *Nature Structural and Molecular Biology* 16.3 (2009): 274.

24. Yarus, Michael. "Translational efficiency of transfer RNA's: uses of an extended anticodon." *Science* 218.4573 (1982): 646-652.

25. Ayer, Don, and Michael Yarus. "The context effect does not require a fourth base pair." *Science* 231.4736 (1986): 393-395.

26. Curran, James F., et al. "Selection of aminoacyl-tRNAs at sense codons: the size of the tRNA variable loop determines whether the immediate 3′ nucleotide to the codon has a context effect." *Nucleic acids research* 23.20 (1995): 4104-4108.

27. Murgola, Emanuel J., et al. "Mutant 16S ribosomal RNA: a codon-specific translational suppressor." *Proceedings of the National Academy of Sciences* 85.12 (1988): 4162-4165.

28. Olejniczak, Mikołaj, et al. "Idiosyncratic tuning of tRNAs to achieve uniform ribosome binding." *Nature Structural and Molecular Biology* 12.9 (2005): 788.

29. Pedersen, W. Todd, and James F. Curran. "Effects of the nucleotide 3′ to an amber codon on ribosomal selection rates of suppressor tRNA and release factor-1." *Journal of molecular biology* 219.2 (1991): 231-241.

30. Xu, Huan, et al. "Re-exploration of the Codon Context Effect on Amber Codon-Guided Incorporation of Noncanonical Amino Acids in *Escherichia coli* by the Blue–White Screening Assay." *ChemBioChem* 17.13 (2016): 1250-1256.

31. Young, Travis S., et al. "An enhanced system for unnatural amino acid mutagenesis in *E. coli*." *Journal of molecular biology* 395.2 (2010): 361-374.

32. Zheng, Yunan, et al. "Performance of optimized noncanonical amino acid mutagenesis systems in the absence of release factor 1." *Molecular BioSystems* 12.6 (2016): 1746-1749.

33. Pott, Moritz, Moritz Johannes Schmidt, and Daniel Summerer. "Evolved sequence contexts for highly efficient amber suppression with noncanonical amino acids." *ACS chemical biology* 9.12 (2014): 2815-2822.

34. Xu, Huan, et al. "Re-exploration of the Codon Context Effect on Amber Codon-Guided Incorporation of Noncanonical Amino Acids in *Escherichia coli* by the Blue–White Screening Assay." *ChemBioChem* 17.13 (2016): 1250-1256.

35. Schmitt, Margaret A., Biddle, Wil, and John D. Fisk, “Mapping the Plasticity of the *E. coli* Genetic Code with Orthogonal Pair Directed Sense Codon Reassignment.” *submitted*, 2017.

36. Biddle, Wil, Margaret A. Schmitt, and John D. Fisk. "Evaluating sense codon reassignment with a simple fluorescence screen." *Biochemistry* 54.50 (2015): 7355-7364.

37. Deutscher, Murray P. "Degradation of RNA in bacteria: comparison of mRNA and stable RNA." *Nucleic acids research* 34.2 (2006): 659-666.

38. Sin, Celine, Davide Chiarugi, and Angelo Valleriani. "Quantitative assessment of ribosome drop-off in *E. coli*." *Nucleic acids research* 44.6 (2016): 2528-2537.

39. Lajoie, Marc J., et al. "Genomically recoded organisms expand biological functions." *science* 342.6156 (2013): 357-360.

CHAPTER 7

FUTURE DIRECTIONS

7.1 CHAPTER OVERVIEW

This dissertation contains 5 chapters that describe the use of orthogonal translation systems, directed evolution, and quantitative *in vivo* fluorescence measurements as tools for quantitatively evaluating the relative quantitative importance of the factors involved in translation in *E. coli* and developing systems for the biosynthetic incorporation of multiple non-canonical amino acids (ncAAs) at multiple sites into proteins. Multiple promising codons for reassignment to ncAAs using either the orthogonal *Methanocaldococcus jannaschii* (*M. jannaschii*) pair or the orthogonal *Methanosarcina barkeri* (*M. barkeri*) pair were identified using a previously developed fluorescence-based screen. The codon reassignment measurements also allowed for an investigation of the relative quantitative importance of the factors involved in protein translation. Additional future directions have also been identified and are described briefly below. Future directions include 1) multi-site ncAA incorporation at multiple codons, 2) further investigation of the factors involved in translation, 3) evaluation of the importance of *E. coli* cell line for codon reassignment, and 4) improvement of engineered orthogonal synthetases by directed evolution.

7.2 MULTI-SITE ncAA INCORPORATION AT MULTIPLE CODONS

One potential future direction is to incorporate multiple ncAAs at multiple sense codons using the orthogonal *M. jannaschii* and *M. barkeri* pairs. The map of sense

codon reassignment efficiencies to tyrosine with the orthogonal *M. jannaschii* pair identified multiple promising targets for reassignment to ncAAs. Presumably, codons that were reassigned to tyrosine with high efficiencies will also be efficiently reassigned to ncAAs with the orthogonal *M. jannaschii* and *M. barkeri* pairs. Combining the orthogonal *M. barkeri* and *M. jannaschii* systems may allow for efficient reassignment of two codons to ncAAs, resulting in 22 amino acid genetic codes.

In order to expand the genetic code to 22 amino acids, multiple requirements must be met. First, each orthogonal pair must be engineered to direct incorporation of the desired ncAA at the targeted codon. The *M. jannaschii* tRNA anticodon must be mutated to Watson-Crick base pair with the desired codon, and the active site of the aaRS must aminoacylate the desired ncAA onto the corresponding tRNA. The same modifications must be made to the orthogonal *M. barkeri* tRNA/aaRS. We have already constructed over 30 orthogonal *M. jannaschii* tRNAs and over 30 orthogonal *M. barkeri* tRNAs that may be used for reassignment of sense codons. In addition, the *M. jannaschii* and *M. barkeri* aaRSs have been repeatedly engineered for incorporation of over 150 ncAAs, and the desired aaRS sequences may be constructed by site-directed mutagenesis of aaRS sequences that we already have [1-3]. Second, a vector system that efficiently expresses both orthogonal tRNA/aaRS pairs must be constructed. The ideal system set-up for co-expression of the orthogonal pairs is unknown, and will require optimization. Third, a method for evaluating the reassignment efficiency of the targeted codons by the orthogonal pairs must be constructed. Evaluation of 22-amino acid genetic code systems could be achieved by (but are not limited to) mass spectrometry or GFP fluorescence.

7.3 FURTHER INVESTIGATION OF THE FACTORS INVOLVED IN TRANSLATION

Another future direction is a deeper investigation of the relative quantitative importance of the factors involved in translation using measurements of orthogonal pair directed sense codon reassignment efficiencies. The analysis presented in chapter 5 is not exhaustive, and further dissection of the large data sets collected is needed. One analysis that is currently underway and may provide further information on the relative quantitative importance of the factors involved in translation is a principle component analysis of the two large data sets of sense codon reassignment efficiencies.

In addition to a deeper analysis of the sense codon reassignment efficiency data sets, additional data may be collected. Less commonly used orthogonal pairs exist, and measuring sense codon reassignment efficiencies with them may augment the previously collected data sets from the orthogonal *M. jannaschii* and *M. barkeri* pairs [4-7]. Using additional orthogonal tRNA/aaRS pairs to make additional sense codon reassignment efficiency measurements requires the replacement of the orthogonal tRNA/aaRS pairs in our current system, and may potentially require the modification of the fluorescence-based screen to be compatible.

The analysis of context effects presented in chapter 6 represent the first step towards understanding how context affects orthogonal pair directed codon reassignment. Collection of amber stop codon suppression efficiencies using other orthogonal tRNA/aaRS pairs and a fluorescence-based screen will generate a large data set that may be used to quantitatively evaluate context effects. Using other orthogonal tRNA/aaRS pairs to make additional amber stop codon suppression efficiency measurements requires the replacement of the orthogonal *M. jannaschii* pair

in our current system, and may potentially require the modification of the fluorescence-based screen to be compatible. In addition, the measurement of context effects may be extended to sense codons.

7.4 EVALUATION OF THE IMPORTANCE OF *E. COLI* CELL LINE FOR CODON REASSIGNMENT

A fourth future direction is to investigate orthogonal pair directed codon reassignment efficiencies in many different strains of *E. coli*. Chapters 3-6 of this dissertation described orthogonal pair directed codon reassignment efficiencies in multiple host *E. coli* strains, including DH10B, SB3930, and the genomically recoded strain containing no UAG codons or peptide release factor 1. Differences in codon reassignment efficiencies were observed at the same codons in different strains of *E. coli*, despite the fact that identical orthogonal translation systems were used. The observed differences in codon reassignment efficiencies may be due to differences in expression of the orthogonal translation components or reporter protein, or differences in the endogenous translational components of the host strains.

Measurement of *in vivo* codon reassignment efficiencies with orthogonal pairs in many different strains of *E. coli* may be significant for a couple of reasons. First, it could provide a map of codon reassignment efficiencies using orthogonal pairs in different commonly used strains of *E. coli*. Researchers may be able to use the map to select the strains of *E. coli* best suited for use with their orthogonal pairs. Second, by making identical sets of codon reassignment measurements in many closely related strains of *E. coli*, additional factors important for translation may be identified. Furthermore, the relative quantitative importance of those factors may also be investigated using the

fluorescence-based screen. The Keio collection contains a large number of closely related strains of *E. coli* (each strain differs from a parent strain by one gene deletion) and may be an ideal set to make codon reassignment efficiency measurements with [8].

7.5 IMPROVEMENT OF ENGINEERED ORTHOGONAL SYNTHETASES BY DIRECTED EVOLUTION

A fourth potential direction is to use error-prone PCR and high-throughput fluorescence activated cell sorting to improve the efficiency of incorporation for previously evolved orthogonal tRNA/aaRS pairs. Efficiencies of orthogonal aaRSs evolved for ncAA incorporation vary from about 10% to 50%, and are on average 25% [9]. In general, aaRSs evolved for ncAAs are not improved further after initial identification.

The method used in Chapter 4 to improve the efficiency of the engineered orthogonal tyrosine-incorporating *M. barkeri* tRNA/aaRS pair is similar to an approach used to improve the amber stop codon suppression efficiency of an engineered orthogonal *O*-methyl-L-tyrosine (OmeY) *M. jannaschii* pair [10]. The approach by Kuhn and co-workers involved error-prone PCR of the aaRS and rounds of selection for improved variants by fluorescence activated cell sorting, and could be quickly applied to improve other previously evolved orthogonal tRNA/aaRS pairs.

REFERENCES

1. Liu, C.C. and P.G. Schultz, *Adding new chemistries to the genetic code.* Annual review of biochemistry, 2010. **79**: p. 413-444.
2. Wan, W., J.M. Tharp, and W.R. Liu, *Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool.* Biochim Biophys Acta, 2014. **1844**(6): p. 1059-70.
3. Dumas, A., et al., *Designing logical codon reassignment–Expanding the chemistry in biology.* Chemical science, 2015. **6**(1): p. 50-69.
4. Liu, D.R. and P.G. Schultz, *Progress toward the evolution of an organism with an expanded genetic code.* Proceedings of the National Academy of Sciences, 1999. **96**(9): p. 4780-4785.
5. Chatterjee, A., H. Xiao, and P.G. Schultz, *Evolution of multiple, mutually orthogonal prolyl-tRNA synthetase/tRNA pairs for unnatural amino acid mutagenesis in Escherichia coli.* Proceedings of the National Academy of Sciences, 2012. **109**(37): p. 14841-14846.
6. Hughes, R.A. and A.D. Ellington, *Rational design of an orthogonal tryptophanyl nonsense suppressor tRNA.* Nucleic Acids Res, 2010. **38**(19): p. 6813-30.
7. Furter, R., *Expansion of the genetic code: site-directed p-fluoro-phenylalanine incorporation in Escherichia coli.* Protein Science, 1998. **7**(2): p. 419-426.
8. Baba, T., et al., *Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection.* Molecular systems biology, 2006. **2**(1).

9. Young, T.S., et al., *An enhanced system for unnatural amino acid mutagenesis in E. coli.* Journal of molecular biology, 2010. **395**(2): p. 361-374.

10. Kuhn, S.M., et al., *Engineering of an orthogonal aminoacyl-tRNA synthetase for efficient incorporation of the non-natural amino acid O-methyl-L-tyrosine using fluorescence-based bacterial cell sorting.* Journal of molecular biology, 2010. **404**(1): p. 70-87.

APPENDIX 1

MATERIALS AND METHODS

A.1 DNA MANIPULATION MATERIALS

All enzymes (restriction enzymes, polymerases, kinase, and ligase) were purchased from New England Biolabs (Ipswitch, MA). ATP was purchased from Fisher Scientific (BP413-25) and dNTP's (N0447S) were purchased from New England Biolabs. Oligonucleotides for mutagenesis and primers for amplification of DNA were ordered from Integrated DNA Technologies (Coralville, IA). Plasmid extraction from cells was achieved using a GeneJET plasmid miniprep kit (Thermo Scientific K0503), and other reactions (kunkel mutagenesis, restriction digests) were cleaned up using a GeneJET PCR purification kit (Thermo Scientific K0702). A M13 spin kit was used for isolation of single-stranded deoxy-uridine containing DNA (ssdU-DNA) was purchased from Qiagen (27704).

A.2 ANTIBIOTICS AND PROTEIN INDUCTION AGENTS

Spectinomycin (Enzo Life Science, BML-A281) was used to maintain the vectors harboring the genes for the orthogonal translational components. Carbenicillin (PlantMedia, 40310000-2) was used to maintain the vectors harboring the genes for the GFP and Z-domain reporter proteins. Kanamycin (Amresco 0408-10G), chloramphenicol (Fisher BP0904-100), and uridine (Acros 14077-0250) were used during production of ssdU-DNA. Isopropyl-beta-D-thiogalactoside (IPTG, Gold Bio, I2481C5) was used to induce the GFP reporter gene, Z-domain reporter gene, and the

orthogonal aaRS gene. Each antibiotic was prepared separately by dissolving the solid in ultra-pure water to a concentration of 50 mg/mL, filter-sterilized with a 0.2 μm filter (VWR 28145-501), and used at a final concentration of 50 μg/mL, with the exception of chloramphenicol, which was dissolved in ethanol and used at a concentration of 5 μg/mL. IPTG was prepared by dissolving solid in ultra-pure water to a concentration of 0.5 M, filter-sterilized with a 0.2 μm filter, and used at a final concentration of 1 mM. Uridine was dissolved in ultra-pure water to a concentration of 5 mg/mL, filter-sterlized with a 0.2 μm filter, and used at a final concentration of 0.25 μg/mL.

A.3 CELL CULTURE MEDIUM

Liquid lysogeny broth (LB) (10 g/L tryptone, 5 g/L yeast extract, and 5 g/L NaCl, pH adjusted to ~7.0 using 1 M NaOH) and solid LB agar plates (15 g/L agar) were used for growing *E. coli*, with the exception of cultures for production of ssdU-DNA. These cultures used 2X YT (16 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl, pH adjusted to ~7.0 using 1 M NaOH). All media contained antibiotics and IPTG (where indicated) at the concentrations described above. Super optimal broth with catabolite repression (SOC) was prepared by adding filter-sterilized 20% w/v glucose to super optimal broth (SOB) (20 g/L tryptone, 5 g/L yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, and 10 mM $MgSO_4$).

A.4 *E. COLI* STRAINS

CJ236: FΔ*(HindIII)::cat* (Tra^+ Pil^+ Cam^R)*/ ung*-1 *rel*A1 *dut*-1 *thi*-1 *spo*T1 *mcr*A

Source: New England Biolabs

DH10B: F^- mcrA Δ(*mrr-hsd*RMS-*mcr*BC) Φ80d*lac*ZΔM15 Δ*lac*X74 *end*A1 *rec*A1 *deo*R Δ(*ara,leu*)7697 *ara*D139 *gal*U *gal*K *nup*G *rps*L λ^-

Source: Invitrogen

SB3930: λ^- Δ*his*B463

Source: Yale CGSC

A.5 PREPARATION OF COMPETENT *E. COLI*

Each cell strain used was made electrocompetent in-house according to the protocol described in Sambrook and Russell [1]. Transformation efficiencies for each electrocompetent strain were approximately 10^9 transformants per µg of super-coiled plasmid DNA.

A.6 ISOLATION OF SINGLE-STRANDED DEOXY-URIDINE CONTAINING DNA (SSDU-DNA) FOR KUNKEL MUTAGENESIS

Preparation of single-stranded deoxy-uridine containing DNA (ssdU-DNA) was prepared using a method adapted from Sidhu and Weiss [2]. Cultures of *E. coli* CJ236 cells containing the phagemid to be mutated were grown to an OD_{600} of 0.5 and then infected with M13K07 helper phage (New England Biolabs) at a multiplicity of infection of 10:1. After 30 minutes of infection at 37 °C, 225 rpm, the culture was transferred into 50-125 mL of pre-warmed 2X YT medium containing spectinomycin, kanamycin, chloramphenicol, and uridine. The cultures were then grown overnight at 30 °C, 225 rpm. The following day cultures were pelleted at 17,000 x *g* at 4 °C for 20 minutes in a Sorvall RC 6+ centrifuge with a Thermo FIBERLite F14-6X250y rotor. After centrifugation, the supernatant was decanted into another centrifuge bottle containing $1/5^{th}$ volume of autoclaved, pre-chilled 20% 8,000 molecular weight polyethylene glycol and 2.5 M NaCl in deionized water. The mixture of PEG/NaCl and supernatant was then inverted ~50X and incubated on ice for at least 2 hours to precipitate phage

particles. The phage were then isolated by centrifuging the solution at 17,000 x *g*, 4 °C for 20 minutes in a Sorvall RC6+ centrifuge. The supernatant was decanted, and the centrifuge bottle was spun at the conditions described above for an additional minute to collect any remaining supernatant, which was then removed using a pipette. The phage pellet was re-suspended in 1.5 mL of Tris buffered saline (pH = 7.4). The solution was then centrifuged at 17,000 x *g*, room temperature for 5 minutes to remove any insoluble debris. ssdU-DNA was then isolated from the phage particles using a M13 spin kit according to manufacturer's instructions and eluted in 100 µL of the kit's elution buffer.

A.7 KUNKEL STYLE SITE-DIRECTED MUTAGENESIS (NON-LIBRARY MUTAGENESIS)

Mutagenesis for routine cloning, including mutation of tRNA anticodons was achieved through kunkel style site-directed mutagenesis [3]. Mutagenic primers were phosphorylated for 1.5 hours at 37 °C using T4 polynucleotide kinase (1 unit). Following phosphorylation, primers were annealed to either 0.5 or 1.0 µg of ssdU template DNA at a 3:1 or molar ratio by incubating at 90 °C for 2 minutes, 53 °C for 3 minutes, and 25 °C for 5 minutes. After annealing, ATP (1 mM), dNTP's (1 mM each), T7 DNA polymerase (3.5 units), and T4 DNA ligase (10 units) were added to the tube, gently mixed, and extended overnight at room temperature. The following day reactions were put directly onto ice and transformed into electrocompetent *E. coli* DH10B without clean-up.

A.8 QUIKCHANGE STYLE SITE-DIRECTED MUTAGENESIS (NON-LIBRARY MUTAGENESIS)

Mutagenesis for routine cloning, including the mutation of tRNA anticodons, GFP fluorophore codons, and the codon for position 5 of the Z-domain of protein A was achieved by using a protocol modified from the Stratagene QuikChange Multi-Site Directed Mutagenesis kit. Briefly, 125 ng of mutagenic primer (one or two primers used per site mutated) , 100 ng of template plasmid DNA, Q5 High Fidelity Polymerase Buffer (1X final concentration), dNTP's (200 µM final concentration), Q5 DNA polymerase (0.5 units), were combined into a 200 µL PCR tube. Ultrapure water was added to adjust the final reaction volume to 25 µL. Reactions were then subjected to thermocycling in either a PeqStar (peqLab) or MjMini (Bio-Rad) thermocycler. After initial denaturing at 98 °C for 1 minute, 25 cycles of the following were performed: 98 °C for 20 seconds, 72 °C for 20 seconds, and 72 °C for 15 second/kb of template plasmid. After 25 cycles were finished, a final 5 minute extension at 72 °C was performed, and the reactions were then cooled down to 4 °C in the thermocycler. Once thermocycling was finished, 1 µL (20 units) of DpnI restriction enzyme was added to the reaction tube, and incubated for 2 hours at 37 °C. Reactions were then transformed directly into electrocompetent *E. coli* DH10B.

A.9 CONSTRUCTION OF THE PYL *M. BARKERI* AARS LIBRARY BY KUNKEL STYLE SITE-DIRECTED MUTAGENESIS FOR IDENTIFICATION OF A TYROSINE-INCORPORATING VARIANT (CHAPTER 3)

Mutagenic library primers in Table A.12 were phosphorylated for 1.5 hours at 37 °C using T4 polynucleotide kinase (20 units) in a 200 µL PCR tube. Following

phosphorylation, primers were annealed to 5.0 µg of ssdU pDGS_Ultra-Inact-CUA template at a 3:1 molar ratio by incubating at 90 °C for 2 minutes, and then 70 °C for 20 seconds, and decreasing temperature by 1 °C every 20 seconds until the temperature reached 20 °C. The tube was incubated at 20 °C for 2 minutes. After annealing of the mutagenic library primers to the ssdU template DNA, ATP (2.0 mM), dNTP's (0.8 mM each), T7 DNA polymerase (15 units), and T4 DNA Ligase (600 units) were added to the tube, gently mixed, and extended overnight at room temperature. The following morning, the reaction was cleaned up using a PCR spin kit, and eluted in 55 µL of ultrapure water. The reaction was then transformed into electrocompetent *E. coli* DH10B to remove the ssdU-DNA template. The transformation was recovered in 6 mL of SOC media without antibiotics at 37 °C with shaking at 225 rpm for 1 hour. At the end of 1 hour, a 50 µL aliquot was plated onto LB agar plates with appropriate antibiotics to determine the number of transformants and mutation efficiency (1.01×10^8 transformants, with 1/24 a full library member by colony PCR). The remainder of the recovery was diluted into LB with antibiotic (160 mL total volume) and grown at 37 °C with shaking at 225 rpm for 2 hours, 20 minutes (equivalent to the starting OD_{600} doubling 2 times. Plasmid DNA was then extracted from the culture, yielding 15.3 µg. The plasmid DNA was then digested with KpnI-HF (72 units) for 2 hours at 37 °C to remove un-mutated or partially mutated template. After digestion, the reaction was cleaned up with 2 PCR spin kit columns and eluted in 55 µL of ultrapure water per column. The DNA was then transformed into electrocompetent *E. coli* DH10B containing the reporter plasmid pGFP66uag and recovered in 50 mL of SOC medium at 37 °C with shaking at 225 rpm for 1 hour. At the end of 1 hour, a 50 µL aliquot was

removed and dilutions of the aliquot were plated to determine transformation efficiency and library efficiency (2.7 x 10^8 transformants, 19/24 were full library members). The remaining amount of recovery was diluted into LB media with antibiotics and IPTG (600 mL total volume) to express the aaRS and GFP reporter protein. The expression was allowed to go overnight at 37 °C with shaking at 225 rpm. The following morning, aliquots containing 0.3 OD's of the library were frozen at -80 °C in 35% glycerol for future use.

A.10 FLOW CYTOMETRY AND FLUORESCENCE ACTIVATED CELL SORTING (FACS) FOR IDENTIFICATION OF A TYROSINE-INCORPORATING PYL *M. BARKERI* AARS VARIANT (CHAPTER 3)

Tubes containing 0.3 OD's of *E. coli* in 35% glycerol were removed from the -80 °C freezer and thawed on ice. Tubes were then centrifuged at 8,000 x *g*, room-temperature for 5 minutes. Supernatant was pipetted off, and 1 mL of sterile 0.9% NaCl in deionized water was added to each tube. Cells were then re-suspended by gentle pipetting, and put back on ice. Flow cytometry and FACS were performed on a Dako-Cytomation MoFlo Legacy instrument using the 488 nm laser line with a 530/40 nm band-pass filter and the 70 µm flow cell tip. For each sample, including controls, cells were initially gated on forward and side scatter to remove clumps and non-cell particles. Cells were then gated on GFP fluorescence, and sorting was performed on the purify 1,2 mode at approximately 21,000 events/second. Controls for each experiment included a 100% positive control (*E. coli* DH10B containing pDGS_Ultra-lpp-Inact-CUA and pGFP66tat) and a 0% negative control (*E. coli* DH10B containing pDGS_Ultra-lpp-Inact-CUA and pGFP66uag). For each control, approximately 1.0 x 10^6 cells were

analyzed. For the first library sort, 1.203 x 10^8 library cells were analyzed and 3.1 x 10^4 library cells with greater than 12.0 arbitrary fluorescence units (representing the top 1% fluorescent clones, approximately 0.02% of the total number of cells analyzed) were collected into LB. The collected library cells were then diluted directly into LB with appropriate antibiotics and IPTG to a final volume of 6 mL, and grown overnight at 37 °C with shaking at 225 rpm. The following day, aliquots of the saturated culture of FACS-enriched library cells (0.3 OD's) were frozen at -80 °C for performing a second round of enrichment. During the second round of enrichment, 1.8 x 10^7 cells from the first enrichment were analyzed, and 1.3 x 10^4 cells that showed arbitrary fluorescence greater than 15.6 (greater than the non-fluorescent control population) and less than 270.0 (less than a population that was suspected to be wild-type contamination) were collected into LB. Cells from the second round of enrichment were diluted into LB containing antibiotics and IPTG, and grown overnight at 37 °C with shaking at 225 rpm. The following morning, aliquots containing 0.3 OD's of cells were frozen at -80 °C for additional rounds of FACS enrichment, and a portion of the culture was also plated onto LB agar plates containing antibiotics and IPTG for individual clone analysis. Individual clones were initially screened using the *in vivo* fluorescence assay described below (no replicates were possible). After initial screening, clones were further analyzed by isolating the plasmid containing the engineered *M. barkeri* translational components, re-transforming them with the pGFP66uag reporter plasmid, and screening again with *in vivo* fluorescence assay. One clone, G4, showed incorporation through all tests and was subjected to further analysis.

A.11 CONSTRUCTION OF A *M. BARKERI* AARS LIBRARY BY ERROR-PRONE PCR FOR IMPROVEMENT OF THE TYROSINE-INCORPORATING PYL *M. BARKERI* AARS VARIANT AND FLOW CYTOMETRY AND FLUORESCENCE ACTIVATED CELL SORTING (FACS) FOR IDENTIFICATION OF AN IMPROVED TYROSINE-INCORPORATING PYL *M. BARKERI* AARS VARIANT (CHAPTER 4)

The template used for library generation was the tyrosine-incorporating *M. barkeri* Pyl aaRS described in Chapter 3. The aaRS sequence was amplified from the plasmid using two different EP-PCR conditions that were predicted to produce an average of 1-4 nucleotide mutations per 1000 [4-7]. Each mutagenic PCR product was then restricted and ligated into a plasmid backbone that contained the gene for the corresponding *M. barkeri* $tRNA^{Pyl}_{CUA}$ to generate two libraries. Library 1 contained *M. barkeri* aaRS variants generated by an EP-PCR method using $MnCl_2$ and unbalanced concentrations of each of the four nucleotides (Tables A.1, A.2) [4]. Library 2 contained *M. barkeri* aaRS variants generated by an EP-PCR method with only unbalanced concentrations of each of the four nucleotides (Tables A.3, A.4) [5-7].

Table A.1. EP-PCR Conditions for Library 1.

Component	Final Concentration
UP H20	-
10X Taq Polymerase Buffer	1X Concentrated
10 mM dNTP's	0.2 mM
100 mM dCTP	1.0 mM
100 mM dTTP	1.0 mM
50 uM Fwd Primer	1 uM
50 uM Rev Primer	1 uM
Template	105.0 ng
50 mM $MgCl_2$	Brings total to 7 mM
10 mM $MnCl_2$	Brings total to 0.15 mM
Taq DNA Polymerase	0.05 U/uL
Total Volume	-

Table A.2. EP-PCR Thermocycling for Library 1.

Cycle	Temperature (^{0}C)	Time
Initial Denature	95	1 min
Start Loop, 6X	N/A	N/A
Denature	95	30 sec
Anneal	58	30 sec
Extend	72	1 min, 45 sec
End Loop, 6X	N/A	N/A
Final Extension	72	5 min
Storage	4	Hold

Table A.3. EP-PCR Conditions for Library 1.

Component	Final Concentration
UP H20	-
10X Taq Polymerase Buffer	Normal Taq Conditions
10 mM dNTP's	0.2 mM
100 mM dCTP	1.0 mM
100 mM dTTP	1.0 mM
10 uM MbaaRS_EcoRI_Fwd	0.4 uM
10 uM WB Pmr IH	0.4 uM
Template (mp0408-4) (1 ng/uL dilution)	10.0 ng
50 mM $MgCl_2$	Brings total to 7 mM
10 mM $MnCl_2$	-
Taq DNA Polymerase	0.05 U/uL
Total Volume	-

Table A.4. EP-PCR Thermocycling for Library 2.

Cycle	Temperature (^{0}C)	Time
Initial Denature	95	1 min
Start Loop, 25X	N/A	N/A
Denature	95	30 sec
Anneal	58	30 sec
Extend	72	1 min, 30 sec
End Loop, 25X	N/A	N/A
Final Extension	72	5 min
Storage	4	Hold

The mutagenic PCR products replaced an inactive version of the *M. barkeri* aaRS that contained 3 KpnI restriction sites. To reduce the ability of the inactive aaRS to ligate back into the plasmid, the backbone was digested with KpnI (in addition to the enzymes used to insert the library of aaRSs). The plasmids containing the aaRS libraries were transformed into cells harboring a UAG stop codon at the fluorophore position of GFP. A small portion of each library transformation was plated onto LB agar plates to determine the number of transformants and mutation frequency of each EP-PCR reaction. Each library was then split in half; one half of each library was expressed overnight at 37^0C (libraries 1a and 2a), and the other half of each library was expressed at 42^0C (libraries 1b and 2b). The idea behind expression at a higher temperature was to potentially identify *M. barkeri* aaRS clones containing stabilizing mutations. The following morning, each library aliquot was frozen at -80^0C for selection by fluorescence assisted cell sorting (FACS).

The first round of EP-PCR yielded $1.2 \cdot 10^7$ unique transformants for library 1, and $1.7 \cdot 10^7$ unique transformants for library 2. Twelve colonies were randomly selected from each of the plates that were used to determine the transformation efficiency of each library were selected for further analysis by colony PCR. All 24 colonies analyzed by colony PCR were confirmed as library members and not the inactive starting material. 10 of the colony PCR products from each library were sent for DNA sequencing, and an estimated average mutation rate of 1.3 mutations per 1,000 nucleotides was observed for library 1, and an estimated average mutation rate of 2.2 mutations per 1,000 nucleotides was observed for library 2. The cells containing the libraries of mutagenized

aaRSs were then evaluated for incorporation of tyrosine in response to UAG codon at position 66 in a GFP reporter by high throughput screening.

In the first round of fluorescence-activated cell sorting (FACS), approximately $1.0 \cdot 10^7$ cells were screened for libraries 1a, 1b, 2a, and 2b. Fluorescent clones representing the top 1% (approximately 0.4% of each population) were collected. A portion of the collected cells for each library was plated directly onto separate plates for analysis of individual clones. The remaining amount was amplified overnight, and the plasmid containing the orthogonal *M. barkeri* translational machinery components was extracted to perform a second round of EP-PCR on the pool of clones from the first round.

For the second round of EP-PCR, the pool of clones isolated from the first round of FACS from libraries 2a and 2b, and the 2 showing the highest amber stop codon suppression efficiency from the first round of EP-PCR and selection were used as the starting templates, and the EP-PCR protocol that generated on average 2.2 mutations per 1,000 nucleotides was used to generate four new libraries of *M. barkeri* aaRS variants (Library2a.2, Library 2b.2, Library G6.2, and Library G11.2) [5-7]. Each mutagenic PCR product was restricted and ligated into the same plasmid backbone used to construct the first set of libraries. The plasmids containing the aaRS libraries were transformed into cells harboring a UAG stop codon at the fluorophore position of GFP. A small portion of each library transformation was plated onto LB agar plates to determine the number of unique transformants for each library, and the remaining amount of each library was expressed overnight at 37^0C. The following morning, aliquots of each library were saved at -80^0C for future use. The library transformations

yielded $4.0 \cdot 10^6$ unique transformants for library 2a.2, $9.6 \cdot 10^6$ unique transformants for Library 2b.2, $4.8 \cdot 10^6$ unique transformants for library G6.2, and $5.2 \cdot 10^6$ unique transformants for library G11.2. In the second round of FACS, approximately $1.0 \cdot 10^7$ cells were screened, and fluorescent clones representing the top 1% (approximately 0.4% of each population) of each library were collected. A portion of the collected cells was plated directly for analysis of individual clones, and the remaining amount of collected cells were amplified overnight and frozen at -80^0C.

A.12 *IN VIVO* MEASUREMENT OF TYROSINE INCORPORATION USING SUPER-FOLDER GFP AS A REPORTER PROTEIN (ALL CHAPTERS)

Superfolder green fluorescent protein (GFP) reporter plasmids (pGFP66xxx where xxx represents the codon at position 66 of GFP) were co-transformed with pDGS_Ultra-Tyr-yyy (where yyy indicates the anticodon of the orthogonal *M. barkeri* or *M. jannaschii* tRNA) vectors containing the engineered orthogonal translational components. The co-transformations were plated onto LB agar plates containing the appropriate antibiotics and grown overnight at 37 °C. The following morning, six colonies from each co-transformation were picked into individual wells of a 96-well plate containing 200 µL of LB with appropriate antibiotics. Cultures were monitored until they were grown to saturation (usually 6-8 hours) at 37 °C with shaking in a Bio-Tek Synergy H1 plate reader. The cultures were then diluted 10-fold into LB media containing appropriate antibiotics and 1 mM IPTG in a 96-well Fluorotrac 200 clear-bottom 96-well plate (Greiner 655096). IPTG was added to induce expression of the aaRS and GFP reporter. Cultures were monitored in a BioTek Synergy H1 plate reader with continuous shaking at 37 °C. Every 20 minutes the OD_{600} and fluorescence were measured.

Fluorescence was measured with an excitation wavelength of 485 nm, and detection at 515 nm with an 8-nm band pass. The OD_{600} was blanked to wells containing only LB media with appropriate antibiotics and IPTG. The fluorescence of each well was subtracted from a negative control containing pDGS-Ultra-Tyr-CUA and pGFP66aag or pDGS-Ultra-Tyr-XhoI and pGFP66uag. The negative subtracted fluorescence of each well was then divided by the OD_{600} of each well to obtain a corrected GFP RFU value (fluorescence units per optical density at 600 nm). The GFP RFU values for each well over the time span of 8-12 hours were averaged, and divided by the 8-12 hour GFP RFU values for a 100% GFP control (pDGS_Ultra-Tyr-CUA and pGFP66tat). Resulting values for incorporation are expressed as a percentage of the 100% GFP control. Stop codon and sense codon reassignment values for each of the 6 colonies from each system were calculated individually using the average 100% GFP RFU value and then averaged. Instantaneous doubling times for each colony were calculated using the following equation, and then averaged [8, 9]:

$$OD_{600,b} = OD_{600,a} \times e^{\frac{\ln 2}{T_d} t}$$

$$T_d = \frac{\ln 2 \times t}{\ln OD_{600,b} - \ln OD_{600,a}}$$

T_d is the doubling time in minutes

t is the amout of time elapsed between measuring $OD_{600,b}$ and $OD_{600,a}$

A.13 Z-DOMAIN OF PROTEIN A EXPRESSION AND PURIFICATION (CHAPTERS 3 AND 4)

Z-Domain of Protein A reporter plasmids pSPEL253-F5Y (100% positive control) and pSPEL253-F5Amber (test case and negative control) were co-transformed with vectors containing the orthogonal engineered *M. barkeri* translational components,

pDGS_Ultra_MbTyr_CUA (100% positive control and test case) and pDGS_Ultra_MbTyr_notRNA (negative control) into *E. coli* DH10B [10, 11]. After overnight growth, 3 colonies from each transformation plate were inoculated into 8-mL liquid cultures containing appropriate antibiotics and grown overnight at 37 °C, 225 rpm to saturation. The following morning, the OD_{600} of each culture was measured. A portion of each culture was taken and added to 40 mL of pre-warmed LB with antibiotics and IPTG (an approximately 20-fold dilution, with a starting OD_{600} of approximately 0.1) to express the engineered aaRS and the reporter protein. Proteins were expressed for 12 hours at 37 °C and 225 rpm. At the end of protein expression, the OD_{600} of each culture was measured, and 30 OD's of cells were pelleted at 17,000 x *g*, 4 °C for 30 minutes in a Sorvall RC+ centrifuge. The supernatant was discarded, and the cell pellets were frozen at -20 °C overnight. The following morning the cell pellets were thawed at room temperature and 1.8 mL of B-Per lysis reagent (Thermo Scientific 78248, prepared according to manufacturer's instructions) was added to each cell pellet. After 30 minutes of lysis at room temperature, the lysis was clarified by centrifugation at 17,000 x *g*, 4 °C for 1 hour in a Sorvall RC6+ centrifuge. Immediately after centrifugation the clarified lysate for each sample was transferred to a clean tube. 20 µL of each sample was saved for running on an SDS-PAGE gel. The Z-Domain of Protein A was then purified from the remaining amount of clarified lysate for each sample using Qiagen Ni-NTA spin columns (Qiagen 31014) according to the manufacturer's instructions. 20 µL from each step in the purification process was saved for running on an SDS-PAGE gel. Each column was eluted twice with 250 µL of NPI-500 buffer and combined into one tube (500 µL total). After elution, the protein samples

were buffer-exchanged into HPLC-grade water using a 3,000 molecular weight cutoff column according to manufacturer's instructions (Millipore UF500396). The final volume of each protein sample after buffer exchange was 300 µL.

A.14 ELECTROSPRAY IONIZATION MASS SPECTROMETRY (CHAPTERS 3 AND 4)

Electrospray ionization mass spectrometry (ESI-MS) was used to evaluate the identity of the amino acids incorporated in response to a single UAG codon at position 5 in the purified, intact Z-domain of protein A. All samples were run by the staff in the central instrument facility (CIF) at Colorado State University. The mass spectra was deconvoluted using a maximum entropy algorithm (MassHunter Software, Agilent Technologies).

A.15 CONSTRUCTION OF TRNA/AARS TRANSLATIONAL COMPONENT VECTORS

The wild-type *M. barkeri* aaRS DNA sequence was constructed using thermally balanced inside-out PCR (TBIO PCR) [11]. Briefly, each quarter was synthesized by TBIO using the primers in Table A.10. Quarters 1 and 2 and quarters 3 and 4 were then overlap-extended to generate halves. Due to difficulties in performing overlap extension of the two halves, each half was ligated into a pEVOL backbone independently. An SpeI restriction site that was introduced when the halves were ligated was then removed by kunkel-style site directed mutagenesis. The *M. barkeri* tRNA and $tRNA^{Opt}$ (under control of the proK promoter and terminator) were also synthesized as cassettes by TBIO PCR using the primers in Table A.11 [11]. The *M. barkeri* tRNA was cloned into a pEVOL backbone using restriction sites ApaL1 and XhoI. The *M. barkeri* translational components were verified as functional before additional cloning. The

tRNA was then cloned into a pWB_Ultra backbone using two PstI sites at the ends of the cassette [8]. An inactive version of the *M. barkeri* aaRS was generated by introducing KpnI restriction sites by kunkel mutagenesis. The inactive *M. barkeri* aaRS had 2 EcoRI restriction sites in the DNA sequence removed by replacing the relevant codons with synonymous codons by kunkel mutagenesis. The tyrosine 349 codon was also mutated to a phenylalanine codon using Quikchange style site directed mutagenesis (the tyr349phe mutation was previously demonstrated to increase aminoacylation efficiency) [12]. The aaRS was then cloned into the pWB-Ultra backbone using restriction sites EcoRI and NotI. The sequences for *M. barkeri* aaRS variants *M. barkeri* tRNA genes are given below. The vector backbone sequence has been described previously, and one full sequence is also provided below for convenience [8, 13]. All *M. barkeri* tRNA variants described in chapters 3-4 were constructed by non-library site-directed kunkel mutagenesis of pDGS_Ultra-lpp-Tyr-XhoI. After mutagenesis colonies were screened by colony PCR amplification of the tRNA and aaRS genes from 12 individual colonies, digested with XhoI to test for removal of the XhoI site (indicating successful mutagenesis at the anticodon of the tRNA). 1-2 clones for each variant that yielded a correct colony PCR product after digestion with XhoI and agarose gel electrophoresis were grown up overnight at 37 °C, 225 rpm. The following day the pULTRA plasmids from overnight cultures were extracted from *E. coli* DH10B using the miniprep kit according to manufacturer's instructions. The plasmid DNA concentration was then quantified using 2.5 µL spots of plasmid DNA on the Take3 plate and the DNA quantification protocol on the BioTek

Synergy H1 plate reader. Plasmid samples were then sequenced by the Sanger sequencing method by Genewiz (South Plainfield, NJ) before further use.

Construction of the pWB-Ultra backbone and associated cloning of the orthogonal *M. jannaschii* translational components has been described previously [8]. All *M. jannaschii* tRNA variants described in Chapter 2 were constructed by non-library site-directed kunkel mutagenesis of pWB_Ultra-Tyr-XhoI. After mutagenesis colonies were screened by colony PCR amplification of the tRNA and aaRS genes from 12 individual colonies, digested with XhoI to test for removal of the XhoI site (indicating successful mutagenesis at the anticodon of the tRNA). 1-2 clones for each variant that yielded a correct colony PCR product after digestion with XhoI and agarose gel electrophoresis were grown up overnight at 37 °C, 225 rpm. The following day the pULTRA plasmids from overnight cultures were extracted from *E. coli* DH10B using the miniprep kit according to manufacturer's instructions. The plasmid DNA concentration was then quantified using 2.5 µL spots of plasmid DNA on the Take3 plate and the DNA quantification protocol on the BioTek Synergy H1 plate reader. Plasmid samples were then sequenced by the Sanger sequencing method by Genewiz (South Plainfield, NJ) before further use.

A.16 CONSTRUCTION OF VECTORS CONTAINING GFP VARIANTS OR Z-DOMAIN OF PROTEIN A VARIANTS

All GFP variants were constructed from pGFP-66tat using either non-library kunkel-style site-directed mutagenesis or Quikchange-style site-directed mutagenesis. The backbone construction of the starting material pGFP66tat has been described previously, and the full sequence of which is provided below for convenience [8]. After

transformation of the mutagenesis reaction, the recovery was plated onto LB agar plates with carbenecillin and IPTG. The following day clones that were non-fluorescent (indicating that either the fluorophore tyrosine codon had been mutated to the desired codon or the appropriate tyrosine codon had been mutated to a UAG codon, except for the tyrosine at position 237 – these variants remained fluorescent even after mutagenesis of the tyrosine codon to a UAG codon) were grown up overnight at 37 °C, 225 rpm in LB medium with carbenecillin. The following day, plasmid DNA was extracted, quantified, and sequenced as described with the pULTRA vectors.

All Z-domain of protein A plasmid variants were constructed from a starting pSPEL253 (Z-domain of Protein A with an AGG codon at the phe position 5) vector using Quikchange-style non-library site directed mutagenesis [10]. The sequence of the entire starting vector is provided for convenience below. After transformation of the mutagenesis reaction and growth overnight on LB agar plates with carbenecillin, four clones of each mutagenesis experiment were selected at random and grown up overnight at 37 °C, 225 rpm. The following day the Z-domain of protein A plasmids were extracted from the overnight cultures, the DNA was quantified and then sequenced as described above with the pULTRA vectors. Colonies were selected at random for these plasmids because no fluorescence or restriction digestion screens to test for successful mutagenesis were available.

A.17 DNA SEQUENCES AND PLASMID MAPS (5’ -> 3’)

DNA sequences containing the M. barkeri orthogonal translational components

pDGS-ULTRA-lpp-Inact-CUA (5536 nt) [8, 14].

```
GCGCTGCGGACACATACAAAGTTACCCACAGATTCCGTGGATAAGCAGGGGACTAACATGTGAG
GCAAAACAGCAGGGCCGCGCCGGTGGCGTTTTTCCATAGGCTCCGCCCTCCTGCCAGAGTTCAC
ATAAACAGACGCTTTTCCGGTGCATCTGTGGGAGCCGTGAGGCTCAACCATGAATCTGACAGTA
```

CGGGCGAAACCCGACAGGACTTAAAGATCCCCACCGTTTCCGGCGGGTCGCTCCCTCTTGCGCT
CTCCTGTTCCGACCCTGCCGTTTACCGGATACCTGTTCCGCCTTTCTCCCTTACGGGAAGTGTG
GCGCTTTCTCATAGCTCACACACTGGTATCTCGGCTCGGTGTAGGTCGTTCGCTCCAAGCTGGG
CTGTAAGCAAGAACTCCCCGTTCAGCCCGACTGCTGCGCCTTATCCGGTAACTGTTCACTTGAG
TCCAACCCGGAAAAGCACGGTAAAACGCCACTGGCAGCAGCCATTGGTAACTGGGAGTTCGCAG
AGGATTTGTTTAGCTAAACACGCGGTTGCTCTTGAAGTGTGCGCCAAAGTCCGGCTACACTGGA
AGGACAGATTTGGTTGCTGTGCTCTGCGAAAGCCAGTTACCACGGTTAAGCAGTTCCCCAACTG
ACTTAACCTTCGATCAAACCACCTCCCCAGGTGGTTTTTTCGTTTACAGGGCAAAAGATTACGC
GCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACTGAACCGCTCTAGATTTCAGT
GCAATTTATCTCTTCAAATGTAGCACCTGAAGTCAGCCCCATACGATATAAGTTGTAATTCTCA
TGTTAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGG
TCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCT
TTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCG
GTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGC
CCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCG
AAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTAT
CCCACTACCGAGATGTCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCA
GCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCAT
GGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGA
TTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGC
CCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACC
GTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCC
GGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGA
TCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCT
TCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCG
ACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTT
TGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCAC
TTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAA
GAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATT
GACTCTCTTCCGGGCGCTATCATGCCAAGCTTTGAGCGGAAGAGCGCCTGATGCGGTATTTTCT
CCTTACGCATCTGTGCGGTATTTCACACCGCATACAGATCCTGACGCGCCCTGTAGCGGCGCAT
TAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCC
CGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTA
AATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTG
ATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTT
GGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCG
GTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGA
TTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAGGATCTGTATGGTGCAC
TCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTG
ACTGCCTCGACCTGCAGCTAGCAAAAAAGCCTGCTCGTTGAGCAGGCTTTTCGAATTTGGCGGA
AACCCCGGGAATCTAACCCGGCTGAACGGATTTAGAGTCCATTCGATCTACATGATCAGGTTCC
CAATGCGGGGCGCATCTTACTGCGCAGATACGCCCTCGTCAATCCCTTAATAGCAAAATGCCTT
TTGATCGGCGAGAAAGTCAGCGGATCCATCTGCAGATCTCCCGGGTAGGCTACTAGTGTGGCGC
CCCATCAAAAAAATATTCTCAACATAAAAAACTTTGTGTAATACTTGTAACGCTGAGTTTACGC
TTTGAGGAAGAATTCATGGATAAAAAACCGCTGGATGTGCTGATTAGCGCGACCGGCCTGTGGA
TGAGCCGTACCGGCACCCTGCATAAAATCAAACATCATGAAGTGAGCCGCAGCAAAATCTATAT
TGAAATGGCGTGCGGCGATCATCTGGTGGTGAACAACAGCCGTAGCTGCCGTACCGCGCGTGCG

TTTCGTCATCATAAATACCGCAAAACCTGCAAACGTTGCCGTGTGAGCGATGAAGATATCAACA
ACTTTCTGACCCGTAGCACCGAAAGCAAAAACAGCGTGAAAGTGCGTGTGGTGAGCGCGCCGAA
AGTGAAAAAAGCGATGCCGAAAAGCGTGAGCCGTGCGCCGAAACCGCTGGAAAATAGCGTGAGC
GCGAAAGCGAGCACCAACACCAGCCGTAGCGTTCCGAGCCCGGCGAAAAGCACCCCGAACAGCA
GCGTTCCGGCGTCTGCGCCGGCACCGAGCCTGACCCGCAGCCAGCTGGATCGTGTGGAAGCGCT
GCTGTCTCCGGAAGATAAAATTAGCCTGAACATGGCGAAACCGTTTCGTGAACTGGAACCGGAA
CTGGTGACCCGTCGTAAAAACGATTTTCAGCGCCTGTATACCAACGATCGTGAAGATTATCTGG
GCAAACTGGAACGTGATATCACCAAATTTTTTGTGGATCGCGGCTTTCTGGAAATTAAAAGCCC
GATTCTGATTCCGGCGGAATATGTGGAACGTATGGGCATTAACAACGACACCGAACTGAGCAAA
CAAATTTTCCGCGTGGATAAAAACCTGTGCCTGCGTCCGATGCTGGCCCCGACCCTGTAGGTAC
CTCTGCGTAAACTGGATCGTATTCTGCCGGGTCCGATCAAAATTTTTGAAGTGGGCCCGTGCTA
TCGCAAAGAAAGCGATGGCAAAGAACACCTGGAAGAATTTACCATGGTTAACGGTACCCAAATG
GGCAGCGGCTGCACCCGTGAAAACCTGGAAGCGCTGATCAAAGAATTTCTGGATTATCTGGAAA
TCGACTTCGAAATTGTGGGCGATAGCTGCATGGTGTTTGGCGATACCCTGGATATTATGCATGG
CGATCTGGAACTGAGCAGCGCGGGTACCGGTCCGGTTAGCCTGGATCGTGAATGGGGCATTGAT
AAACCGTGGATTGGCGCGGGTTTTGGCCTGGAACGTCTGCTGAAAGTGATGCATGGCTTCAAAA
ACATTAAACGTGCGAGCCGTAGCGAAAGCTACTATAACGGCATTAGCACGAACCTGTAAGCGGC
CGCGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGTT
AATTAACCTAGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACG
GGTCTTGAGGGGTTTTTTGCTGAAACCTCAGGCATTTGAGAAGCACACGGTCACACTGCTTCCG
GTAGTCAATAAACCGGTAAACCAGCAATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAAC
CGACGACCGGGTCATCGTGGCCGGATCTTGCGGCCCCTCGGCTTGAACGAATTGTTAGACATTA
TTTGCCGACTACCTTGGTGATCTCGCCTTTCACGTAGTGGACAAATTCTTCCAACTGATCTGCG
CGCGAGGCCAAGCGATCTTCTTCTTGTCCAAGATAAGCCTGTCTAGCTTCAAGTATGACGGGCT
GATACTGGGCCGGCAGGCGCTCCATTGCCCAGTCGGCAGCGACATCCTTCGGCGCGATTTTGCC
GGTTACTGCGCTGTACCAAATGCGGGACAACGTAAGCACTACATTTCGCTCATCGCCAGCCCAG
TCGGGCGGCGAGTTCCATAGCGTTAAGGTTTCATTTAGCGCCTCAAATAGATCCTGTTCAGGAA
CCGGATCAAAGAGTTCCTCCGCCGCTGGACCTACCAAGGCAACGCTATGTTCTCTTGCTTTTGT
CAGCAAGATAGCCAGATCAATGTCGATCGTGGCTGGCTCGAAGATACCTGCAAGAATGTCATTG
CGCTGCCATTCTCCAAATTGCAGTTCGCGCTTAGCTGGATAACGCCACGGAATGATGTCGTCGT
GCACAACAATGGTGACTTCTACAGCGCGGAGAATCTCGCTCTCTCCAGGGGAAGCCGAAGTTTC
CAAAAGGTCGTTGATCAAAGCTCGCCGCGTTGTTTCATCAAGCCTTACGGTCACCGTAACCAGC
AAATCAATATCACTGTGTGGCTTCAGGCCGCCATCCACTGCGGAGCCGTACAAATGTACGGCCA
GCAACGTCGGTTCGAGATGGCGCTCGATGACGCCAACTACCTCTGATAGTTGAGTCGATACTTC
GGCGATCACCGCTTCCCTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTAT
TGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGCTAGCTCACTCG
TCGCTACGCTCCGGGCGTGAGACTGCGGCGG

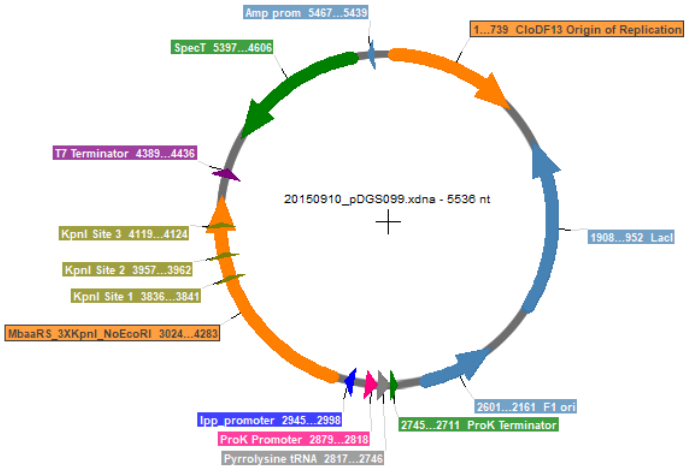

Figure A.1. Plasmid map of pDGS_Ultra-lpp-Inact-CUA (5536 nt). Plasmid map was generated using Serial Cloner [8].

M. barkeri Inactive Starting aaRS, 1260 nt [14].

ATGGATAAAAAACCGCTGGATGTGCTGATTAGCGCGACCGGCCTGTGGATGAGCCGTACCGGCA
CCCTGCATAAAATCAAACATCATGAAGTGAGCCGCAGCAAAATCTATATTGAAATGGCGTGCGG
CGATCATCTGGTGGTGAACAACAGCCGTAGCTGCCGTACCGCGCGTGCGTTTCGTCATCATAAA
TACCGCAAAACCTGCAAACGTTGCCGTGTGAGCGATGAAGATATCAACAACTTTCTGACCCGTA
GCACCGAAAGCAAAAACAGCGTGAAAGTGCGTGTGGTGAGCGCGCCGAAAGTGAAAAAAGCGAT
GCCGAAAAGCGTGAGCCGTGCGCCGAAACCGCTGGAAAATAGCGTGAGCGCGAAAGCGAGCACC
AACACCAGCCGTAGCGTTCCGAGCCCGGCGAAAAGCACCCCGAACAGCAGCGTTCCGGCGTCTG
CGCCGGCACCGAGCCTGACCCGCAGCCAGCTGGATCGTGTGGAAGCGCTGCTGTCTCCGGAAGA
TAAAATTAGCCTGAACATGGCGAAACCGTTTCGTGAACTGGAACCGGAACTGGTGACCCGTCGT
AAAAACGATTTTCAGCGCCTGTATACCAACGATCGTGAAGATTATCTGGGCAAACTGGAACGTG
ATATCACCAAATTTTTTGTGGATCGCGGCTTTCTGGAAATTAAAAGCCCGATTCTGATTCCGGC
GGAATATGTGGAACGTATGGGCATTAACAACGACACCGAACTGAGCAAACAAATTTTCCGCGTG
GATAAAAACCTGTGCCTGCGTCCGATGCTGGCCCCGACCCTGTAGGTACCTCTGCGTAAACTGG
ATCGTATTCTGCCGGGTCCGATCAAAATTTTTGAAGTGGGCCCGTGCTATCGCAAAGAAAGCGA
TGGCAAAGAACACCTGGAAGAATTTACCATGGTTAACGGTACCCAAATGGGCAGCGGCTGCACC
CGTGAAAACCTGGAAGCGCTGATCAAAGAATTTCTGGATTATCTGGAAATCGACTTCGAAATTG
TGGGCGATAGCTGCATGGTGTTTGGCGATACCCTGGATATTATGCATGGCGATCTGGAACTGAG
CAGCGCGGGTACCGGTCCGGTTAGCCTGGATCGTGAATGGGGCATTGATAAACCGTGGATTGGC
GCGGGTTTTGGCCTGGAACGTCTGCTGAAAGTGATGCATGGCTTCAAAAACATTAAACGTGCGA
GCCGTAGCGAAAGCTACTATAACGGCATTAGCACGAACCTGTAA

M. barkeri tRNA (CUA Anticodon), 72 nt [14].

GGGAACCTGATCATGTAGATCGAATGGACTCTAAATCCGTTCAGCCGGGTTAGATTCCCGGGGT
TTCCGCCA

M. barkeri tRNA^{Opt} (CUA Anticodon), 72 nt [15].

GGAAACGTGATCATGTAGATCGAATGGACTCTAAATCCGTTCAGTGGGGTTAGATTCCCCACGT
TTCCGCCA

DNA sequences of vectors containing the M. jannaschii translational components

pWB_Ultra-tac-TyrRS-XhoI, 5190 nt [8].

GCGCTGCGGACACATACAAAGTTACCCACAGATTCCGTGGATAAGCAGGGGACTAACATGTGAG
GCAAAACAGCAGGGCCGCGCCGGTGGCGTTTTTCCATAGGCTCCGCCCTCCTGCCAGAGTTCAC
ATAAACAGACGCTTTTCCGGTGCATCTGTGGGAGCCGTGAGGCTCAACCATGAATCTGACAGTA
CGGGCGAAACCCGACAGGACTTAAAGATCCCCACCGTTTCCGGCGGGTCGCTCCCTCTTGCGCT
CTCCTGTTCCGACCCTGCCGTTTACCGGATACCTGTTCCGCCTTTCTCCCTTACGGGAAGTGTG
GCGCTTTCTCATAGCTCACACACTGGTATCTCGGCTCGGTGTAGGTCGTTCGCTCCAAGCTGGG
CTGTAAGCAAGAACTCCCCGTTCAGCCCGACTGCTGCGCCTTATCCGGTAACTGTTCACTTGAG
TCCAACCCGGAAAAGCACGGTAAAACGCCACTGGCAGCAGCCATTGGTAACTGGGAGTTCGCAG
AGGATTTGTTTAGCTAAACACGCGGTTGCTCTTGAAGTGTGCGCCAAAGTCCGGCTACACTGGA
AGGACAGATTTGGTTGCTGTGCTCTGCGAAAGCCAGTTACCACGGTTAAGCAGTTCCCCAACTG
ACTTAACCTTCGATCAAACCACCTCCCCAGGTGGTTTTTTCGTTTACAGGGCAAAAGATTACGC
GCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACTGAACCGCTCTAGATTTCAGT
GCAATTTATCTCTTCAAATGTAGCACCTGAAGTCAGCCCCATACGATATAAGTTGTAATTCTCA
TGTTAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGG
TCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCT
TTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCG
GTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGC
CCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCG
AAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTAT
CCCACTACCGAGATGTCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCA
GCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCAT
GGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGA
TTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGC
CCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACC
GTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCC
GGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGA
TCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCT
TCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCG
ACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTT
TGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCAC
TTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAA
GAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATT
GACTCTCTTCCGGGCGCTATCATGCCAAGCTTTGAGCGGAAGAGCGCCTGATGCGGTATTTTCT
CCTTACGCATCTGTGCGGTATTTCACACCGCATACAGATCCTGACGCGCCCTGTAGCGGCGCAT
TAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCC
CGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTA
AATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTG
ATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTT

GGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCG
GTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGA
TTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAGGATCTGTATGGTGCAC
TCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTG
ACTGCCTCGACCTGCAGCTAGCAAAAAAGCCTGCTCGTTGAGCAGGCTTTTCGAATTTGGTCCG
GCGGAGGGGATTTGAACCCCTGCCATGCGGCTCGAGAGTCCGCCGTTCTGCCCTGCTGAACTAC
CGCCGGAATGCGGGGCGCATCTTACTGCGCAGATACGCCCTCGTCAATCCCTTAATAGCAAAAT
GCCTTTTGATCGGCGAGAAAGTCAGCGGATCCATCTGCAGATCTCCCGGGTAGGCTACTAGTTG
ACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAA
CAGAATTCATGGACGAGTTCGAAATGATTAAACGCAACACCAGCGAAATTATCTCTGAAGAAGA
GCTGCGCGAGGTGCTGAAGAAAGACGAGAAGAGCGCGTATATTGGCTTTGAGCCGTCCGGTAAA
ATTCACCTGGGTCACTACCTGCAAATCAAGAAGATGATTGATCTGCAAAACGCTGGTTTTGACA
TCATTATCCTGCTGGCGGACCTGCACGCCTACCTGAATCAAAAGGGCGAGCTGGATGAGATTCG
CAAGATCGGCGACTACAATAAGAAAGTCTTCGAAGCCATGGGTTTGAAGGCTAAATACGTCTAC
GGTAGCGAATTTCAGCTGGATAAGGATTACACGTTGAATGTGTACCGTCTGGCGCTGAAAACCA
CGCTGAAACGCGCCCGTCGTTCCATGGAGCTGATTGCGCGCGAGGATGAGAATCCAAAAGTTGC
TGAGGTTATTTACCCTATTATGCAAGTTAATGATATTCACTACCTGGGTGTTGATGTTGCAGTC
GGTGGTATGGAGCAACGCAAAATTCACATGCTGGCACGTGAACTGCTGCCGAAAAAGGTTGTCT
GTATTCACAATCCGGTCCTGACCGGCCTGGATGGCGAGGGTAAAATGAGCAGCAGCAAGGGTAA
CTTTATTGCAGTTGACGATAGCCCGGAAGAAATCCGTGCGAAGATCAAGAAAGCGTACTGCCCG
GCAGGCGTGGTTGAGGGTAACCCGATCATGGAAATCGCCAAGTATTTTCTGGAATACCCACTGA
CGATTAAGCGCCCGGAGAAATTTGGCGGCGACCTGACCGTCAACAGCTACGAGGAGCTGGAAAG
CTTGTTTAAGAACAAAGAACTGCACCCGATGGATCTGAAAAACGCCGTGGCGGAAGAGCTGATT
AAGATTCTGGAACCAATTCGCAAACGTCTGTAAGCGGCCGCGTCTGATAAAACAGAATTTGCCT
GGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGTTAATTAACCTAGGCTGCTGCCACCGCT
GAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAC
CTCAGGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA
ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCATCGTGGCCGGAT
CTTGCGGCCCCTCGGCTTGAACGAATTGTTAGACATTATTTGCCGACTACCTTGGTGATCTCGC
CTTTCACGTAGTGGACAAATTCTTCCAACTGATCTGCGCGCGAGGCCAAGCGATCTTCTTCTTG
TCCAAGATAAGCCTGTCTAGCTTCAAGTATGACGGGCTGATACTGGGCCGGCAGGCGCTCCATT
GCCCAGTCGGCAGCGACATCCTTCGGCGCGATTTTGCCGGTTACTGCGCTGTACCAAATGCGGG
ACAACGTAAGCACTACATTTCGCTCATCGCCAGCCCAGTCGGGCGGCGAGTTCCATAGCGTTAA
GGTTTCATTTAGCGCCTCAAATAGATCCTGTTCAGGAACCGGATCAAAGAGTTCCTCCGCCGCT
GGACCTACCAAGGCAACGCTATGTTCTCTTGCTTTTGTCAGCAAGATAGCCAGATCAATGTCGA
TCGTGGCTGGCTCGAAGATACCTGCAAGAATGTCATTGCGCTGCCATTCTCCAAATTGCAGTTC
GCGCTTAGCTGGATAACGCCACGGAATGATGTCGTCGTGCACAACAATGGTGACTTCTACAGCG
CGGAGAATCTCGCTCTCTCCAGGGGAAGCCGAAGTTTCCAAAAGGTCGTTGATCAAAGCTCGCC
GCGTTGTTTCATCAAGCCTTACGGTCACCGTAACCAGCAAATCAATATCACTGTGTGGCTTCAG
GCCGCCATCCACTGCGGAGCCGTACAAATGTACGGCCAGCAACGTCGGTTCGAGATGGCGCTCG
ATGACGCCAACTACCTCTGATAGTTGAGTCGATACTTCGGCGATCACCGCTTCCCTCATACTCT
TCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGA
ATGTATTTAGAAAAATAAACAAATAGCTAGCTCACTCGGTCGCTACGCTCCGGGCGTGAGACTG
CGGCGG

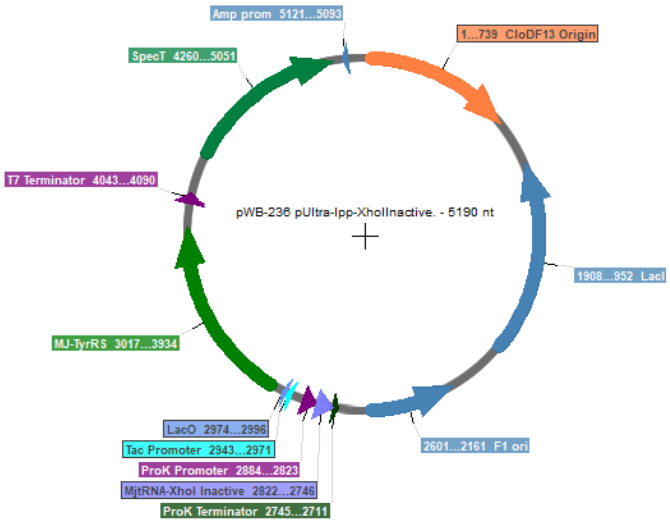

Figure A.2. Plasmid map of pWB_Ultra-tac-TyRS-XhoI (5190 nt). Plasmid map was generated using Serial Cloner [8].

M. jannaschii TyrRS aaRS, 918 nt [8].

```
ATGGACGAGTTCGAAATGATTAAACGCAACACCAGCGAAATTATCTCTGAAGAAGAGCTGCGCG
AGGTGCTGAAGAAAGACGAGAAGAGCGCGTATATTGGCTTTGAGCCGTCCGGTAAAATTCACCT
GGGTCACTACCTGCAAATCAAGAAGATGATTGATCTGCAAAACGCTGGTTTTGACATCATTATC
CTGCTGGCGGACCTGCACGCCTACCTGAATCAAAAGGGCGAGCTGGATGAGATTCGCAAGATCG
GCGACTACAATAAGAAAGTCTTCGAAGCCATGGGTTTGAAGGCTAAATACGTCTACGGTAGCGA
ATTTCAGCTGGATAAGGATTACACGTTGAATGTGTACCGTCTGGCGCTGAAAACCACGCTGAAA
CGCGCCCGTCGTTCCATGGAGCTGATTGCGCGCGAGGATGAGAATCCAAAAGTTGCTGAGGTTA
TTTACCCTATTATGCAAGTTAATGATATTCACTACCTGGGTGTTGATGTTGCAGTCGGTGGTAT
GGAGCAACGCAAAATTCACATGCTGGCACGTGAACTGCTGCCGAAAAAGGTTGTCTGTATTCAC
AATCCGGTCCTGACCGGCCTGGATGGCGAGGGTAAAATGAGCAGCAGCAAGGGTAACTTTATTG
CAGTTGACGATAGCCCGGAAGAAATCCGTGCGAAGATCAAGAAAGCGTACTGCCCGGCAGGCGT
GGTTGAGGGTAACCCGATCATGGAAATCGCCAAGTATTTTCTGGAATACCCACTGACGATTAAG
CGCCCGGAGAAATTTGGCGGCGACCTGACCGTCAACAGCTACGAGGAGCTGGAAAGCTTGTTTA
AGAACAAAGAACTGCACCCGATGGATCTGAAAAACGCCGTGGCGGAAGAGCTGATTAAGATTCT
GGAACCAATTCGCAAACGTCTG
```

M. jannaschii $tRNA^{Opt}$ (CUA Anticodon) 72 nt [8].

```
CCGGCGGTAGTTCAGCAGGGCAGAACGGCGGACTCTAAATCCGCATGGCAGGGGTTCAAATCCC
CTCCGCCGGACCA
```

DNA sequences of vectors containing the GFP reporter proteins

pGFPy66tat, 5223 nt [8].

GGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGC
TCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCT
CCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTC
GGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGC
TCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACT
ATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAG
GATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGC
TACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAG
TTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCA
GCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGAC
GCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCA
CCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTG
GTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCA
TCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCC
CCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCA
GCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATT
AATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCA
TTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCA
ACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCT
CCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATA
ATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTC
ATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACC
GCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCT
CAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTC
AGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAA
AAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAA
GCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACA
AATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATC
ATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCACCTCGAGAAATCAT
AAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGATTCAATTGTGAGCGGATAACA
ATTTCACACAGAAAACATTAAAGAGGAGAAAGAATTCTATGGCACTCGAAGGAGAAGAACTCTT
CCACGGAGTCGTCCCAATCCTCGTCGAATTAGACGGCGACGTCAACGGGCACAAATTCTCCGTC
CGTGGAGAAGGCGAAGGCGACGCCACAAACGGAAAACTCACCCTCAAATTCATCTGCACCACCG
GAAAACTCCCCGTCCCATGGCCAACACTCGTCACCACCCTGACCTACGGCGTCCAGTGCTTCTC
CCGTTACCCGGACCACATGAAACGGCACGACTTCTTCAAAAGCGCCATGCCCGAAGGCTACGTA
CAGGAACGTACCATCTCCTTCAAAGACGACGGGACCTACAAAACCCGTGCCGAAGTCAAATTCG
AAGGCGACACCCTCGTCAACCGTATCGAATTAAAAGGCATCGACTTCAAAGAAGACGGAAACAT
CCTCGGACACAAACTCGAATACAACTTCAACTCACACAACGTATACATCACGGCAGACAAACAG
AAAAACGGAATCAAAGCCAACTTCAAAATCCGTCACAACGTCGAAGACGGCTCCGTCCAGCTCG
CAGACCACTACCAGCAGAACACCCCAATCGGCGACGGCCCCGTCCTCTTACCAGACAACCACTA
CCTGTCCACACAGTCCGTCCTCTCGAAAGACCCCAACGAAAAACGTGACCACATGGTCCTCCTC
GAATTTGTAACCGCCGCCGGGATCACACACGGCATGGACGAACTCTACAAACACCACCACCACC
ACCACTAAGCGGCCGCGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCAC

CTGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCA
TGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTT
TCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGAT
TTGAACGTTGCGAAGCAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGC
ATCAAATTAAGCAGAAGGCCATCCTGACGGATGGCCTTTTTGCGTTTCTACAAACTCTTTTGTT
TATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCGATGATCCCGGGTCG
ACCTGCAGCCAAGCTTAATTAGCTGAGCTTGGACTCCTGTTGATAGATCCAGTAATGACCTCAG
AACTCCATCTGGATTTGTTCAGAACGCTCGGTTGCCGCCGGGCGTTTTTTATTGGTGAGAATCC
AAGCTAGACCCCGAAAAGTGCCACCTGACGTCGACACCATCGAATGGTGCAAAACCTTTCGCGG
TATGGCATGATAGCGCCCGGAAGAGAGTCAATTCAGGGTGGTGAATGTGAAACCAGTAACGTTA
TACGATGTCGCAGAGTATGCCGGTGTCTCTTATCAGACCGTTTCCCGCGTGGTGAACCAGGCCA
GCCACGTTTCTGCGAAAACGCGGGAAAAAGTGGAAGCGGCGATGGCGGAGCTGAATTACATTCC
CAACCGCGTGGCACAACAACTGGCGGGCAAACAGTCGTTGCTGATTGGCGTTGCCACCTCCAGT
CTGGCCCTGCACGCGCCGTCGCAAATTGTCGCGGCGATTAAATCTCGCGCCGATCAACTGGGTG
CCAGCGTGGTGGTGTCGATGGTAGAACGAAGCGGCGTCGAAGCCTGTAAAGCGGCGGTGCACAA
TCTTCTCGCGCAACGCGTCAGTGGGCTGATCATTAACTATCCGCTGGATGACCAGGATGCCATT
GCTGTGGAAGCTGCCTGCACTAATGTTCCGGCGTTATTTCTTGATGTCTCTGACCAGACACCCG
TCAACAGTATTATTTTCTCCCATGAAGACGGTACGCGACTGGGCGTGGAGCATCTGGTCGCATT
GGGTCACCAGCAAATCGCGCTGTTAGCGGGCCCATTAAGTTCTGTCTCGGCGCGTCTGCGTCTG
GCTGGCTGGCATAAATATCTCACTCGCAATCAAATTCAGCCGATAGCGGAACGGGAAGGCGACT
GGAGTGCCATGTCCGGTTTTCAACAAACCATGCAAATGCTGAATGAGGGCATCGTTCCCACTGC
GATGCTGGTTGCCAACGATCAGATGGCGCTGGGCGCAATGCGCGCCATTACCGAGTCCGGGCTG
CGCGTTGGTGCGGATATCTCGGTAGTGGGATACGACGATACCGAAGACAGCTCATGTTATATCC
CGCCGTTAACCACCATCAAACAGGATTTTCGCCTGCTGGGGCAAACCAGCGTGGACCGCTTGCT
GCAACTCTCTCAGGGCCAGGCGGTGAAGGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGA
AAAACCACCCTGGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGC
AGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTAAGTT
AGCGCGAATTGATCTGGTTTGACTTAATTAAGCGGCATCAGAGCAGATTGTACTGAGAGTGCAC
CATATGTGTAAACGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTT
TTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGT
TGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGG
GCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTG
GGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGAC
GGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGC
GCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTA
CAGGGCGCGTCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGG
CGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTAT
CAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACAT
GTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAA

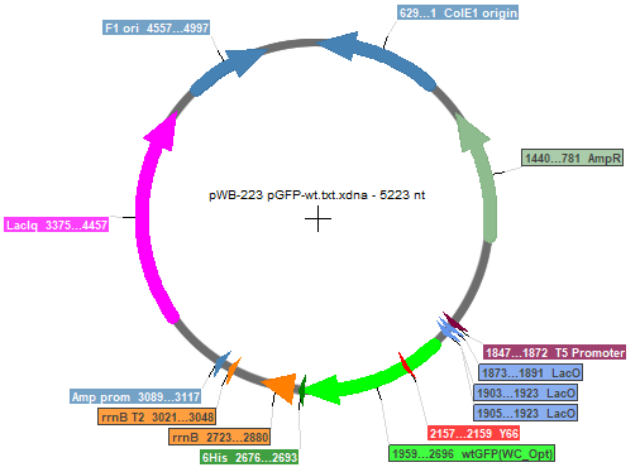

Figure A.3. Plasmid map of pGFPy66tat (5223 nt). Plasmid map was generated using Serial Cloner [8].

GFP gene, wild-type, 738 nt [8, 16].

ATGGCACTCGAAGGAGAAGAACTCTTCCACGGAGTCGTCCCAATCCTCGTCGAATTAGACGGCG
ACGTCAACGGGCACAAATTCTCCGTCCGTGGAGAAGGCGAAGGCGACGCCACAAACGGAAAACT
CACCCTCAAATTCATCTGCACCACCGGAAAACTCCCCGTCCCATGGCCAACACTCGTCACCACC
CTGACCTACGGCGTCCAGTGCTTCTCCCGTTACCCGGACCACATGAAACGGCACGACTTCTTCA
AAAGCGCCATGCCCGAAGGCTACGTACAGGAACGTACCATCTCCTTCAAAGACGACGGGACCTA
CAAAACCCGTGCCGAAGTCAAATTCGAAGGCGACACCCTCGTCAACCGTATCGAATTAAAAGGC
ATCGACTTCAAAGAAGACGGAAACATCCTCGGACACAAACTCGAATACAACTTCAACTCACACA
ACGTATACATCACGGCAGACAAACAGAAAAACGGAATCAAAGCCAACTTCAAAATCCGTCACAA
CGTCGAAGACGGCTCCGTCCAGCTCGCAGACCACTACCAGCAGAACACCCCAATCGGCGACGGC
CCCGTCCTCTTACCAGACAACCACTACCTGTCCACACAGTCCGTCCTCTCGAAAGACCCCAACG
AAAAACGTGACCACATGGTCCTCCTCGAATTTGTAACCGCCGCCGGGATCACACACGGCATGGA
CGAACTCTACAAACACCACCACCACCACTAA

DNA sequences of vectors containing the Z-domain of Protein A reporter proteins

pSPEL253-F5AGG, 4898 nt [10].

CTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGATTCAATTG
TGAGCGGATAACAATTTCACACAGAATTCATTAAAGAGGAGAAATTAACTATGAGAGGATCGCA
TCACCATCACCATCACGGATCCATGGCCGTAGACAACAAAAGGAACAAAGAACAACAAAACGCG
TTCTATGAGATCTTACATTTACCTAACTTAAACGAAGAACAACGAAACGCCTTCATCCAAAGTT
TAAAAGATGACCCAAGCCAAAGCGCTAACCTTTTAGCAGAAGCTAAAAAGCTAAATGATGCTCA
GGCGCCGTGGTAAAAGCTTAATTAGCTGAGCTTGGACTCCTGTTGATAGATCCAGTAATGACCT
CAGAACTCCATCTGGATTTGTTCAGAACGCTCGGTTGCCGCCGGGCGTTTTTTATTGGTGAGAA
TCCAAGCTAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAAATCACTGG
ATATACCACCGTTGATATATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTT

GCTCAATGTACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGA
AAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATCC
GGAATTTCGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTCACCCTTGTTAC
ACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCC
GGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCC
TAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTT
GATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTTCACCATGGGCAAATATTATA
CGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTCAGGTTCATCATGCCGTTTGTGATGGCTT
CCATGTCGGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAA
TTTTTTTAAGGCAGTTATTGGTGCCCTTAAACGCCTGGGGTAATGACTCTCTAGCTTGAGGCAT
CAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGA
ACGCTCTCCTGAGTAGGACAAATCCGCCCTCTAGATTACGTGCAGTCGATGATAAGCTGTCAAA
CATGAGAATTGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTT
TCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG
TTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCC
CTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGA
AAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATC
CCACTACCGAGATATCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAG
CGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATG
GTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGAT
TGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCC
CGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCG
TCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCG
GAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGAT
CAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTT
CGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGA
CAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTT
GCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACT
TTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAG
AGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTG
ACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCACCATTCGATGGTGTCGGAA
TTTCGGGCAGCGTTGGGTCCTGGCCACGGGTGCGCATGATCTAGAGCTGCCTCGCGCGTTTCGG
TGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCG
GATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCGCAG
CCATGACCCAGTCACGTAGCGATAGCGGAGTGTATACTGGCTTAACTATGCGGCATCAGAGCAG
ATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACC
GCATCAGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCG
AGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGA
AAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGT
TTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCG
AAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCT
GTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTT
CTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGT
GCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAAC
CCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGT
ATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGT
ATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCC

GGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAA
AAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAA
CTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAAT
TAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAAT
GCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACT
CCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATA
CCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCG
AGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGC
TAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTG
GTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTA
CATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAG
TAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATG
CCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTA
TGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAAC
TTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTG
TTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCA
CCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGAC
ACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTAT
TGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCA
CATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAA
AAATAGGCGTATCACGAGGCCCTTTCGTCTTCAC

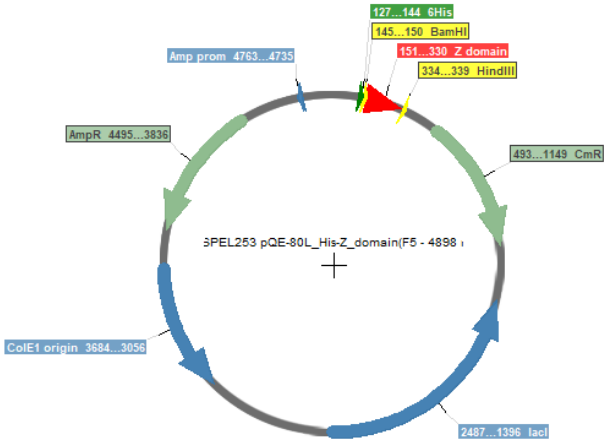

Figure A.4. Plasmid map of pSPEL253-F5AGG, (4898 nt). Plasmid map was generated using Serial Cloner [10].

Z-domain F5AGG, 216 nt [10].

ATGAGAGGATCGCATCACCATCACCATCACGGATCCATGGCCGTAGACAACAAAAGGAACAAAG
AACAACAAAACGCGTTCTATGAGATCTTACATTTACCTAACTTAAACGAAGAACAACGAAACGC
CTTCATCCAAAGTTTAAAAGATGACCCAAGCCAAAGCGCTAACCTTTTAGCAGAAGCTAAAAAG
CTAAATGATGCTCAGGCGCCGTGG

A.18 TABLES OF PLASMID DESIGNATIONS

Table A.5. pDGS_Ultra aaRS/tRNA Expression Vectors (*M. jannaschii*)

aaRS/tRNA Plasmid	tRNA Anticodon	Codon Read via Watson-Crick Interaction	Use of Vector
pWB_Ultra-tac-MjTyRS-XhoI	N/A	N/A	Inactive starting template for doing site-directed mutagenesis to change the anticodon of the *jannaschii* tRNA. Also used as Neg. Control (Ch. 6)
pWB_Ultra-lpp-MjTyr-CUA	CUA	Amber UAG	Vector that contains *M. jannaschii* Tyrosine aaRS. aaRS is driven by an *lpp* promoter. (Ch.6) Used with amber stop codon incorporation.
pDGS_ tac-MjTyRS-UAG	UAG	Leu CUA	Sense codon reassignment.
pDGS_ tac-MjTyRS-UAU	UAU	Ile AUA	Sense codon reassignment
pDGS_ tac-MjTyRS-UGA	UGA	Ser UCA	Sense codon reassignment
pDGS_ tac-MjTyRS-GCA	GCA	Cys UGC	Sense codon reassignment
pDGS_ tac-MjTyRS-CCG	CCG	Arg CGG	Sense codon reassignment
pDGS_ tac-MjTyRS-UCU	UCU	Arg AGA	Sense codon reassignment
pDGS_ tac-MjTyRS-CCU	CCU	Arg AGG	Sense codon reassignment
pDGS_ tac-MjTyRS-UCC	UCC	Gly GGA	Sense codon reassignment
pDGS_ tac-MjTyRS-GGG	GGG	Pro CCC	Sense codon reassignment, unable to construct
pDGS_ tac-MjTyRS-CCC	CCC	Gly GGG	Sense codon reassignment, unable to construct
The sequence of the plasmid backbone is described in above and in [8], with the only differences being the tRNA anticodon sequences and where indicated in the plasmid name the aaRS being driven by an *lpp* promoter instead of a *tac* promoter.			

Table A.6. pDGS_Ultra aaRS/tRNA Expression Vectors (*M. barkeri*)

aaRS/tRNA Plasmid	tRNA Anticodon	Codon Read via Watson-Crick Interaction	Use of Vector
pDGS_Ultra-lpp-Inact-CUA	CUA	UAG	Starting template for *M. barkeri* aaRS library. aaRS is driven by an *lpp* promoter..
pDGS_Ultra-lpp-Tyr-CUA	CUA	UAG	Vector that contains *M. barkeri* Tyrosine aaRS library hit. aaRS is driven by an *lpp* promoter.
pDGS_Ultra-Inact-CUA	CUA	UAG	Negative control "no aaRS" vector. Used in combination with the pGFP66uag reporter.
pDGS_Ultra-Tyr-notRNA	N/A	N/A	Negative control "no tRNA" vector. Used in combination with the pGFP66uag reporter.
pDGS_Ultra-Tyr-XhoI	N/A	N/A	Inactive starting template for doing site-directed mutagenesis to change the anticodon of the *M. barkeri* tRNA.
pDGS_Ultra-Tyr-CUA-Opt	CUA	UAG	Vector containing the *M. barkeri* $tRNA^{Opt}$ instead of *M. barker* tRNA
pDGS_Ultra-Tyr-CUA	CUA	UAG	Amber stop codon reassignment.
pDGS_Ultra-Tyr-CUU	CUU	AAG	Sense codon reassignment.
pDGS_Ultra-Tyr-CUC	CUC	GAG	Sense codon reassignment
pDGS_Ultra-Tyr-AAA	AAA	UUU	Sense codon reassignment
pDGS_Ultra-Tyr-AUC	AUC	GAU	Sense codon reassignment
pDGS_Ultra-Tyr-AGA	AGA	UCU	Sense codon reassignment
pDGS_Ultra-Tyr-CCU	CCU	AGG	Sense codon reassignment
pDGS_Ultra-Tyr-AGC	AGC	GCU	Sense codon reassignment
pDGS_Ultra-Tyr-AUU	AUU	AAU	Sense codon reassignment
pDGS_Ultra-ImprTyr-CUA	CUA	UAG	Vector that contains improved tyrosine-incorporating Pyl *M. barkeri* aaRS variant.
pDGS_Ultra-ImprTyr-AAA	AAA	Phe UUU	Sense codon reassignment.
pDGS_Ultra-ImprTyr-AAG	AAG	Leu CUU	Sense codon reassignment
pDGS_Ultra-ImprTyr-AAU	AAU	Ile AUU	Sense codon reassignment
pDGS_Ultra-ImprTyr-AAC	AAC	Val GUU	Sense codon reassignment
pDGS_Ultra-ImprTyr-AGA	AGA	Ser UCU	Sense codon reassignment
pDGS_Ultra-ImprTyr-AGG	AGG	Pro CCU	Sense codon reassignment
pDGS_Ultra-ImprTyr-AGU	AGU	Thr ACU	Sense codon reassignment
pDGS_Ultra-ImprTyr-AGC	AGC	Ala GCU	Sense codon reassignment
pDGS_Ultra-ImprTyr-AUG	AUG	His CAU	Sense codon reassignment.
pDGS_Ultra-ImprTyr-AUU	AUU	Asn AAU	Sense codon reassignment
pDGS_Ultra-ImprTyr-AUC	AUC	Asp GAU	Sense codon reassignment
pDGS_Ultra-ImprTyr-ACA	ACA	Cys UGU	Sense codon reassignment
pDGS_Ultra-ImprTyr-ACU	ACU	Ser AGU	Sense codon reassignment
pDGS_Ultra-ImprTyr-ACC	ACC	Gly GGU	Sense codon reassignment
pDGS_Ultra-ImprTyr-CAC	CAC	Val GUG	Sense codon reassignment
pDGS_Ultra-ImprTyr-CGC	CGC	Ala GCG	Sense codon reassignment
pDGS_Ultra-ImprTyr-CUU	CUU	Lys AAG	Sense codon reassignment.
pDGS_Ultra-ImprTyr-CUC	CUC	Glu GAG	Sense codon reassignment
pDGS_Ultra-ImprTyr-GCG	GCG	Arg CGC	Sense codon reassignment
pDGS_Ultra-ImprTyr-UCG	UCG	Arg CGA	Sense codon reassignment
pDGS_Ultra-ImprTyr-UAG	UAG	Leu CUA	Sense codon reassignment.
pDGS_Ultra-ImprTyr-UAU	UAU	Ile AUA	Sense codon reassignment
pDGS_Ultra-ImprTyr-UGA	UGA	Ser UCA	Sense codon reassignment
pDGS_Ultra-ImprTyr-GCA	GCA	Cys UGC	Sense codon reassignment
pDGS_Ultra-ImprTyr-CCG	CCG	Arg CGG	Sense codon reassignment
pDGS_Ultra-ImprTyr-UCU	UCU	Arg AGA	Sense codon reassignment
pDGS_Ultra-ImprTyr-CCU	CCU	Arg AGG	Sense codon reassignment
pDGS_Ultra-ImprTyr-UCC	UCC	Gly GGA	Sense codon reassignment
The sequence of the plasmid backbone is described in above and in [8], with the only differences being the aaRS, tRNA sequences and where indicated in the plasmid name and the aaRS being driven by an *lpp* promoter instead of a *tac* promoter where indicated.			

Table A.7. pGFP Reporter Protein Plasmid Designations and Usage for Sense Codon Reassignment Efficiency Measurements (Chapters 2-5) [8, 9].

Reporter Plasmid	Codon at Position 66	Canonically Encoded Amino Acid	Use
pGFP66tat	UAU	Tyr	Wild type GFP vector for 100% fluorescence reference
pGFP66uag	UAG	Amber Stop Codon	Vector for evaluating tyrosine incorporation at an amber stop codon
pGFP66aag	AAG	Lys	Vector for evaluating reassignment of AAG and used as a negative control vector
pGFP66uca	UCA	Ser	Vector for evaluating reassignment of UCA
pGFP66gga	GGA	Gly	Vector for evaluating reassignment of GGA
pGFP66ggg	GGG	Gly	Vector for evaluating reassignment of GGG
pGFP66ucg	UCG	Ser	Vector for evaluating reassignment of UCG
pGFP66cug	CUG	Leu	Vector for evaluating reassignment of CUG
pGFP66acg	ACG	Thr	Vector for evaluating reassignment of ACG
pGFP66aca	ACA	Thr	Vector for evaluating reassignment of ACA
pGFP66uua	UUA	Leu	Vector for evaluating reassignment of UUA
pGFP66uug	UUG	Leu	Vector for evaluating reassignment of UUG
pGFP66uaa	UAA	Ochre Stop	Vector for evaluating reassignment of Ochre
pGFP66uga	UGA	Opal Stop	Vector for evaluating reassignment of Opal
pGFP66uau	UAU	Tyr	Vector for evaluating reassignment of Tyr
pGFP66ugg	UGG	Trp	Vector for evaluating reassignment of UGG
pGFP66uuu	UUU	Phe	Vector for evaluating reassignment of UUU
pGFP66uuc	UUC	Phe	Vector for evaluating reassignment of UUC
pGFP66cuu	CUU	Leu	Vector for evaluating reassignment of CUU
pGFP66cuc	CUC	Leu	Vector for evaluating reassignment of CUC
pGFP66cua	CUA	Leu	Vector for evaluating reassignment of CUA
pGFP66auu	AUU	Ile	Vector for evaluating reassignment of AUU
pGFP66auc	AUC	Ile	Vector for evaluating reassignment of AUC
pGFP66aua	AUA	Ile	Vector for evaluating reassignment of AUA
pGFP66aug	AUG	Ile	Vector for evaluating reassignment of AUG
pGFP66guu	GUU	Val	Vector for evaluating reassignment of GUU
pGFP66guc	GUC	Val	Vector for evaluating reassignment of GUC
pGFP66gua	GUA	Val	Vector for evaluating reassignment of GUA
pGFP66gug	GUG	Val	Vector for evaluating reassignment of GUG
pGFP66ucu	UCU	Ser	Vector for evaluating reassignment of UCU
pGFP66ucc	UCC	Ser	Vector for evaluating reassignment of UCC
pGFP66ccu	CCU	Pro	Vector for evaluating reassignment of CCU
pGFP66ccc	CCC	Pro	Vector for evaluating reassignment of CCC
pGFP66cca	CCA	Pro	Vector for evaluating reassignment of CCA
pGFP66ccg	CCG	Pro	Vector for evaluating reassignment of CCG
pGFP66acu	ACU	Thr	Vector for evaluating reassignment of ACU
pGFP66acc	ACC	Thr	Vector for evaluating reassignment of ACC
pGFP66gcu	GCU	Ala	Vector for evaluating reassignment of GCU
pGFP66gcc	GCC	Ala	Vector for evaluating reassignment of GCC
pGFP66gca	GCA	Ala	Vector for evaluating reassignment of GCA
pGFP66gcg	GCG	Ala	Vector for evaluating reassignment of GCG
pGFP66cau	CAU	His	Vector for evaluating reassignment of CAU
pGFP66cac	CAC	His	Vector for evaluating reassignment of CAC
pGFP66caa	CAA	Gln	Vector for evaluating reassignment of CAA
pGFP66cag	CAG	Gln	Vector for evaluating reassignment of CAG
pGFP66aau	AAU	Asn	Vector for evaluating reassignment of AAU
pGFP66aac	AAC	Asn	Vector for evaluating reassignment of AAC
pGFP66aaa	AAA	Lys	Vector for evaluating reassignment of AAA
pGFP66gau	GAU	Asp	Vector for evaluating reassignment of GAU
pGFP66gac	GAC	Asp	Vector for evaluating reassignment of GAC
pGFP66gaa	GAA	Glu	Vector for evaluating reassignment of GAA
pGFP66gag	GAG	Glu	Vector for evaluating reassignment of GAG

Table A.7 continued.

Reporter Plasmid	Codon at Position 66	Canonically Encoded Amino Acid	Use
pGFP66ugu	UGU	Cys	Vector for evaluating reassignment of UGU
pGFP66ugc	UGC	Cys	Vector for evaluating reassignment of UGC
pGFP66cgu	CGU	Arg	Vector for evaluating reassignment of CGU
pGFP66cgc	CGC	Arg	Vector for evaluating reassignment of CGC
pGFP66cga	CGA	Arg	Vector for evaluating reassignment of CGA
pGFP66cgg	CGG	Arg	Vector for evaluating reassignment of CGG
pGFP66agu	AGU	Ser	Vector for evaluating reassignment of AGU
pGFP66agc	AGC	Ser	Vector for evaluating reassignment of AGC
pGFP66aga	AGA	Arg	Vector for evaluating reassignment of AGA
pGFP66agg	AGG	Arg	Vector for evaluating reassignment of AGG
pGFP66agg	AGG	Arg	Vector for evaluating reassignment of AGG
pGFP66ggu	GGU	Gly	Vector for evaluating reassignment of GGU
pGFP66ggc	GGC	Gly	Vector for evaluating reassignment of GGC
Vector design, construction, and backbone sequence reported in [8]. Sequences 4-16 were generated during the course of the research in this thesis. The remaining variants were constructed previously by other members of the Fisk lab [[8, 9]].			

Table A.8. pGFP Reporter Protein Plasmid Designations and Usage for Amber Stop Codon Incorporation Efficiency Measurements (Chapter 6).

Reporter Plasmid	Position(s) of UAG Codon	Canonically Encoded Amino Acid(s)	Use
pGFP66tat	None	Tyr	Wild type GFP vector for 100% fluorescence reference
pGFP66uag	66	Tyr	Vector for evaluating amber stop codon incorporation
pGFP73uag	74	Tyr	Vector for evaluating amber stop codon incorporation
pGFP92uag	92	Tyr	Vector for evaluating amber stop codon incorporation
pGFP106uag	106	Tyr	Vector for evaluating amber stop codon incorporation
pGFP143uag	143	Tyr	Vector for evaluating amber stop codon incorporation
pGFP151uag	151	Tyr	Vector for evaluating amber stop codon incorporation
pGFP182uag	182	Tyr	Vector for evaluating amber stop codon incorporation
pGFP200uag	200	Tyr	Vector for evaluating amber stop codon incorporation
pGFP237uag	237	Tyr	Vector for evaluating amber stop codon incorporation
pGFP74,182uag	74,182	Tyr	Vector for evaluating amber stop codon incorporation
pGFP66,143uag	66,143	Tyr	Vector for evaluating amber stop codon incorporation
pGFP66,74uag	66,74	Tyr	Vector for evaluating amber stop codon incorporation
pGFP74,143uag	74,143	Tyr	Vector for evaluating amber stop codon incorporation
pGFP74,200uag	74,200	Tyr	Vector for evaluating amber stop codon incorporation
pGFP74,92uag	74,92	Tyr	Vector for evaluating amber stop codon incorporation
pGFP143,200uag	143,200	Tyr	Vector for evaluating amber stop codon incorporation
pGFP143,182uag	143,182	Tyr	Vector for evaluating amber stop codon incorporation
pGFP106,143uag	106,143	Tyr	Vector for evaluating amber stop codon incorporation
pGFP92,143uag	92,143	Tyr	Vector for evaluating amber stop codon incorporation
pGFP74,92,200uag	74,92,200	Tyr	Vector for evaluating amber stop codon incorporation
pGFP74,182,200uag	74,182,200	Tyr	Vector for evaluating amber stop codon incorporation
pGFP74,92,106,200uag	74,92,106,200uag	Tyr	Vector for evaluating amber stop codon incorporation
pGFP74,92,106,182,200uag	74,92,106,182,200	Tyr	Vector for evaluating amber stop codon incorporation
Vector design, construction, and backbone sequence reported in [8].			

Table A.9. pSPEL253 (Z-Domain of Protein A) Plasmid Designations and Usage

Reporter Plasmid	Codon at Position 66	Canonically Encoded Amino Acid	Use
pSPEL253-F5Y	UAC	Tyr	Wild type Z-domain vector for 100% tyrosine reference
pSPEL253-F5Amber	UAG	Amber Stop Codon	Vector for evaluating tyrosine incorporation at an amber stop codon
Vector design and construction reported in [17]. Vectors described above were constructed by using Quikchange-style site directed mutagenesis to change position 5 to the indicated codons.			

A.19 TABLES OF PRIMERS USED

Table A.10 Primers for Generating the *M. barkeri* aaRS by TBIO [11, 13].

Primer	Primer sequence	Primer name
1	ATGGATAAAAAACCGCTGGATGTGCTGATTAGCG	MbaaRS TBIO 1A
2	CGCTGGATGTGCTGATTAGCGCGACCGGCCTGTGGATGAGCCGTACCGGCACCC	MbaaRS TBIO 1B
3	GAGCCGTACCGGCACCCTGCATAAAATCAAACATCATGAAGTGAGCCGCAGCAA	MbaaRS TBIO 1C
4	CATGAAGTGAGCCGCAGCAAAATCTATATTGAAATGGCGTGCGGCGATCATCTG	MbaaRS TBIO 1D
5	GCGTGCGGCGATCATCTGGTGGTGAACAACAGCCGTAGCTGCCGTACCGCGCGT	MbaaRS TBIO 1E
6	CGTTTGCAGGTTTTGCGGTATTTATGATGACGAAACGCACGCGCGGTACGGCAG	MbaaRS TBIO 1F
7	TGTTGATATCTTCATCGCTCACACGGCAACGTTTGCAGGTTTTGCGGTATTTAT	MbaaRS TBIO 1G
8	TTTGCTTTCGGTGCTACGGGTCAGAAAGTTGTTGATATCTTCATCGCTCACACG	MbaaRS TBIO 1H
9	TCGGCGCGCTCACCACACGCACTTTCACGCTGTTTTTGCTTTCGGTGCTACGGG	MbaaRS TBIO 1I
10	GCACGGCTCACGCTTTTCGGCATCGCTTTTTTCACTTTCGGCGCGCTCACCAC	MbaaRS TBIO 1J
11	GCCGAAAAGCGTGAGCCGTGCGCCGAAACCGCTGGAAAATA	MbaaRS TBIO 2K
12	CGCCGAAACCGCTGGAAAATAGCGTGAGCGCGAAAGCGAGCACCAACACCAGCC	MbaaRS TBIO 2L
13	CGAGCACCAACACCAGCCGTAGCGTTCCGAGCCCGGCGAAAAGCACCCCGAACA	MbaaRS TBIO 2M
14	GCGAAAAGCACCCCGAACAGCAGCGTTCCGGCGTCTGCGCCGGCACCGAGCCTG	MbaaRS TBIO 2N
15	GCCGGCACCGAGCCTGACCCGCAGCCAGCTGGATCGTGTGGAAGCGCTGCTGTC	MbaaRS TBIO 2O
16	GTTTCGCCATGTTCAGGCTAATTTTATCTTCCGGAGACAGCAGCGCTTCCACAC	MbaaRS TBIO 2P
17	CACCAGTTCCGGTTCCAGTTCACGAAACGGTTTCGCCATGTTCAGGCTAATTTT	MbaaRS TBIO 2Q
18	ATACAGGCGCTGAAAATCGTTTTTACGACGGGTCACCAGTTCCGGTTCCAGTTC	MbaaRS TBIO 2R
19	TTGCCCAGATAATCTTCACGATCGTTGGTATACAGGCGCTGAAAATCGTTTTTA	MbaaRS TBIO 2S
20	CACGTTCCAGTTTGCCCAGATAATCTTCACGATCG	MbaaRS TBIO 2T
21	ATCTGGGCAAACTGGAACGTGATATCACCA	MbaaRS TBIO 3U
22	GCAAACTGGAACGTGATATCACCAAATTTTTTGTGGATCGCGGCTTTCTGGAAA	MbaaRS TBIO 3V
23	GGATCGCGGCTTTCTGGAAATTAAAAGCCCGATTCTGATTCCGGCGGAATATGT	MbaaRS TBIO 3W
24	GATTCTGATTCCGGCGGAATATGTGGAACGTATGGGCATTAACAACGACACCGA	MbaaRS TBIO 3X
25	GGGCATTAACAACGACACCGAACTGAGCAAACAAATTTTCCGCGTGGATAAAAA	MbaaRS TBIO 3Y
26	CAAACAAATTTTCCGCGTGGATAAAAACCTGTGCCTGCGTCCGATGCTGGCCCC	MbaaRS TBIO 3Z
27	AATACGATCCAGTTTACGCAGATAGTTATACAGGGTCGGGGCCAGCATCGGACG	MbaaRS TBIO 3AA
28	CAAAAATTTTGATCGGACCCGGCAGAATACGATCCAGTTTACGCAGATAGTTAT	MbaaRS TBIO 3AB
29	TCGCTTTCTTTGCGATAGCACGGGCCCACTTCAAAAATTTTGATCGGACCCGGC	MbaaRS TBIO 3AC
30	CCATGGTGAATTCTTCCAGGTGTTCTTTGCCATCGCTTTCTTTGCGATAGCACG	MbaaRS TBIO 3AD
31	GTGCAGCCGCTGCCCATTTGGCAAAAGTTAACCATGGTGAATTCTTCCAGGTGT	MbaaRS TBIO 4AE
32	GCGCTTCCAGGTTTTCACGGGTGCAGCCGCTGCCCA	MbaaRS TBIO 4AF
33	CGTGAAAACCTGGAAGCGCTGATCAAAGA	MbaaRS TBIO 4AG
34	CCTGGAAGCGCTGATCAAAGAATTCCTGGATTATCTGGAAATCGACTTCGAAAT	MbaaRS TBIO 4AH
35	GGATTATCTGGAAATCGACTTCGAAATTGTGGGCGATAGCTGCATGGTGTATGG	MbaaRS TBIO 4AI
36	GCGATAGCTGCATGGTGTATGGCGATACCCTGGATATTATGCATGGCGATCTGG	MbaaRS TBIO 4AJ
37	TGGATATTATGCATGGCGATCTGGAACTGAGCAGCGCGGTGGTGGGTCCGGTTA	MbaaRS TBIO 4AK
38	TCCACGGTTTATCAATGCCCCATTCACGATCCAGGCTAACCGGACCCACCACCG	MbaaRS TBIO 4AL
39	TCAGCAGACGTTCCAGGCCAAAACCCGCGCCAATCCACGGTTTATCAATGCCCC	MbaaRS TBIO 4AM
40	CGCACGTTTAATGTTTTTGAAGCCATGCATCACTTTCAGCAGACGTTCCAGGCC	MbaaRS TBIO 4AN
41	AATGCCGTTATAGTAGCTTTCGCTACGGCTCGCACGTTTAATGTTTTTGAAGCC	MbaaRS TBIO 4AO
42	TTACAGGTTCGTGCTAATGCCGTTATAGTAGCTTTCGCTA	MbaaRS TBIO 4AP
The *Mb* aaRS was constructed by TBIO in 4 quarters. Quarters 1-2 and 3-4 were then overlap-extended with each other. Due to difficulties in cloning, each half was ligated into the destination vector.		

Table A.11. Primers for Generating the *M. barkeri* Pyrrolysyl-tRNA and the *M. barkeri* Pyrrolysyl-tRNA^{Opt} by TBIO [11, 15].

Primer	Primer sequence	Primer name
43	CCAGGTCTCGAGCATGCAAAAAAGCCTGCTCGTTGAG	MbtRNA TBIO A
44	AAAAAAGCCTGCTCGTTGAGCAGGCTTTTCGAATTTGGCGGAAACCCCGGG	MbtRNA TBIO B
45	GGCGGAAACCCCGGGAATCTAACCCGGCTGAACGGATTTAGAGTCCATTCG	MbtRNA TBIO C
46	GAACGGATTTAGAGTCCATTCGATCTACATGATCAGGTTCCCAATGCGGGG	MbtRNA TBIO D
47	GATTGACGAGGGCGTATCTGCGCAGTAAGATGCGCCCCGCATTGGGAACCT	MbtRNA TBIO E
48	TCGCCGATCAAAAGGCATTTTGCTATTAAGGGATTGACGAGGGCGTATCTG	MbtRNA TBIO F
49	TGCACGGCTAACTAAGCGGCCTGCTGACTTTCTCGCCGATCAAAAGGCATT	MbtRNA TBIO G
50	ACTAGAGTGCACGGCTAACTAAGCG	MbtRNA TBIO H
51	GATTGACGAGGGCGTATCTGCGCAGTAAGATGCGCCCCGCATTGGAAACGT	MbtRNA^{Opt} TBIO E
52	GAACGGATTTAGAGTCCATTCGATCTACATGATCACGTTTCCAATGCGGGG	MbtRNA^{Opt} TBIO D
53	GGCGGAAACGTGGGGAATCTAACCCCACTGAACGGATTTAGAGTCCATTCG	MbtRNA^{Opt} TBIO C
54	AAAAAAGCCTGCTCGTTGAGCAGGCTTTTCGAATTTGGCGGAAACGTGGGG	MbtRNA^{Opt} TBIO B
The *Mb* tRNA^{Opt} was constructed in the same manner as the *Mb* tRNA, with the Mb tRNA^{Opt} primers indicated above replacing the corresponding primers for the *Mb* tRNA.		

Table A.12. Primers for Mutagenesis and Cloning of the Pyrrolysyl *M. barkeri* aaRS.

Primer	Primer sequence	Primer name
55	TTAGTTACTGCAGATGGATCCGCTGACTTTCTCGCCG	tRNA Amplify Fwd
56	AAAAAAGCCTGCTCGTTGAG	tRNA Amplify Rev
57	AACGATCGTGAAGATTATCTGGGCAAACTGGAAC	MbaaRS SpeI_Out
58	TGGGTACCGTTAACCATGGTAAATTCTTCCAGGTGTTCTTTGCC	MbaaRS EcoRI_Out1
59	GAAGTCGATTTCCAGATAATCCAGAAATTCTTTGATCAGCGCTTCCAG	MbaaRS EcoRI_Out2
60	AGTCTGTGGAATTCATGGATAAAAAACCGCTGGATGTG	MbaaRS Fwd Amplify
61	ATGATTGCGGCCGCTTACAGGTTCGTGCTAATGC	MbaaRS Rev Amplify
62	CATAATATCCAGGGTATCGCCAAACACCATGCAGCTATCGCC	MbaaRS Y384F
63	CAGAATACGATCCAGTTTACGCAGAGGTACCTACAGGGTCGGGGCC	MbaaRS Inactive 1
64	GCCGCTGCCCATTTGGGTACCGTTAACCATGGTGAATTCTTCCAGG	MbaaRS Inactive 2
65	CCAGGCTAACCGGACCGGTACCCGCGCTGCTCAGTTC	MbaaRS Inactive 3
66	GGCAGAATACGATCCAGTTTACGCAGATAGTTATACAGGGTCGGGGCCAGCATCGGACGCAGG	MbaaRS Library 1
67	GCCGCTGCCCATTTGGCAAAAGTTAACCATGGTAAATTCTTCCAGGTGTTCTTTGCC	MbaaRS Library 2
68	CCAGGCTAACCGGACCCACCACCGCGCTGCTCAGTTC	MbaaRS Library 3

Table A.13. Primers for Mutagenesis of the Pyrrolysyl *M. barkeri* tRNA Anticodon.

Primer	Primer sequence	Primer name
69	ACCTGATCATGTAGATCGAATGGACTCGAGATCCGTTCAGCCGGGTTAG	MbtRNA-XhoI
70	ACCTGATCATGTAGATCGAATGGACTAAAAATCCGTTCAGCCGGGTTAG	MbtRNA-PheUUU
71	ACCTGATCATGTAGATCGAATGGACTAAGAATCCGTTCAGCCGGGTTAG	MbtRNA-LeuCUU
72	ACCTGATCATGTAGATCGAATGGACTAAUAATCCGTTCAGCCGGGTTAG	MbtRNA-IleAUU
73	ACCTGATCATGTAGATCGAATGGACTAACAATCCGTTCAGCCGGGTTAG	MbtRNA-ValGUU
74	ACCTGATCATGTAGATCGAATGGACTAGAAATCCGTTCAGCCGGGTTAG	MbtRNA-SerUCU
75	ACCTGATCATGTAGATCGAATGGACTAGGAATCCGTTCAGCCGGGTTAG	MbtRNA-ProCCU
76	ACCTGATCATGTAGATCGAATGGACTAGUAATCCGTTCAGCCGGGTTAG	MbtRNA-ThrACU
77	ACCTGATCATGTAGATCGAATGGACTAGCAATCCGTTCAGCCGGGTTAG	MbtRNA-AlaGCU
78	ACCTGATCATGTAGATCGAATGGACTAUGAATCCGTTCAGCCGGGTTAG	MbtRNA-HisCAU
79	ACCTGATCATGTAGATCGAATGGACTAUUAATCCGTTCAGCCGGGTTAG	MbtRNA-AsnAAU
80	ACCTGATCATGTAGATCGAATGGACTAUCAATCCGTTCAGCCGGGTTAG	MbtRNA-AspGAU
81	ACCTGATCATGTAGATCGAATGGACTACAAATCCGTTCAGCCGGGTTAG	MbtRNA-CysUGU
82	ACCTGATCATGTAGATCGAATGGACTACUAATCCGTTCAGCCGGGTTAG	MbtRNA-SerAGU
83	ACCTGATCATGTAGATCGAATGGACTACCAATCCGTTCAGCCGGGTTAG	MbtRNA-GlyGGU
84	ACCTGATCATGTAGATCGAATGGACTCACAATCCGTTCAGCCGGGTTAG	MbtRNA-ValGUG
85	ACCTGATCATGTAGATCGAATGGACTCGCAATCCGTTCAGCCGGGTTAG	MbtRNA-AlaGCG
86	ACCTGATCATGTAGATCGAATGGACTCUUAATCCGTTCAGCCGGGTTAG	MbtRNA-LysAAG
87	ACCTGATCATGTAGATCGAATGGACTCUCAATCCGTTCAGCCGGGTTAG	MbtRNA-GluGAG
88	ACCTGATCATGTAGATCGAATGGACTGCGAATCCGTTCAGCCGGGTTAG	MbtRNA-ArgCGC
89	ACCTGATCATGTAGATCGAATGGACTUCGAATCCGTTCAGCCGGGTTAG	MbtRNA-ArgCGA
90	ACCTGATCATGTAGATCGAATGGACTUAGAATCCGTTCAGCCGGGTTAG	MbtRNA-LeuCUA
91	ACCTGATCATGTAGATCGAATGGACTUAUAATCCGTTCAGCCGGGTTAG	MbtRNA-IleAUA
92	ACCTGATCATGTAGATCGAATGGACTUGAAATCCGTTCAGCCGGGTTAG	MbtRNA-SerUCA
93	ACCTGATCATGTAGATCGAATGGACTUGUAATCCGTTCAGCCGGGTTAG	MbtRNA-ThrACA
94	ACCTGATCATGTAGATCGAATGGACTGCAAATCCGTTCAGCCGGGTTAG	MbtRNA-CysUGC
95	ACCTGATCATGTAGATCGAATGGACTCCGAATCCGTTCAGCCGGGTTAG	MbtRNA-ArgCGG
96	ACCTGATCATGTAGATCGAATGGACTUCUAATCCGTTCAGCCGGGTTAG	MbtRNA-ArgAGA
97	ACCTGATCATGTAGATCGAATGGACTCCUAATCCGTTCAGCCGGGTTAG	MbtRNA-ArgAGG
98	ACCTGATCATGTAGATCGAATGGACTUCCAATCCGTTCAGCCGGGTTAG	MbtRNA-GlyGGA
99	ACCTGATCATGTAGATCGAATGGACTGGGAATCCGTTCAGCCGGGTTAG	MbtRNA-ProCCC
100	ACCTGATCATGTAGATCGAATGGACTCCCAATCCGTTCAGCCGGGTTAG	MbtRNA-GlyGGG
101	ACCTGATCATGTAGATCGAATGGACTCUAAATCCGTTCAGCCGGGTTAG	MbtRNA-AmberUAG

Table A.14. Primers for Mutagenesis of the GFP Reporter Protein.

Primer	Primer sequence	Primer name
101	CACTCGTCACCACACTCACCTAGGGCGTCCAGTGCTTCTCCCG	pGFPy66uag
102	TCCAGTGCTTCTCCCGTTAGCCGGACCACATGAAACGG	pGFPy74uag
103	GCCATGCCCGAAGGCTAGGTACAGGAACGTACCATCTCCTTC	pGFPy92uag
104	CTTCAAAGACGACGGGACCTAGAAAACCCGTGCCGAAGTC	pGFPy106uag
105	CCTCGGACACAAACTCGAATAGAACTTCAACTCACACAACGTATACATC	pGFPy143uag
106	CAACTTCAACTCACACAACGTATAGATCACGGCAGACAAACAGAAAAAC	pGFPy151uag
107	CCAGCTCGCAGACCACTAGCAGCAGAACACCCCAATCG	pGFPy182uag
108	GTCCTCTTACCAGACAACCACTAGCTGTCCACACAGTCCGTC	pGFPy200uag
109	CGGCATGGACGAACTCTAGAAACACCACCACCACCAC	pGFPy237uag
110	CACTCGTCACCACACACCACTGGCGTCCAGTGCTTCTCCCG	pGFPy66ThrACA1
111	CACTCGTCACCACACTCACCACTGGCGTCCAGTGCTTCTCCCG	pGFPy66ThrACA2
112	CACTCGTCACCACACTCACCTCAGGCGTCCAGTGCTTCTCCCG	pGFPy66SerUCA
113	CACTCGTCACCACACTCACCGGAGGCGTCCAGTGCTTCTCCCG	pGFPy66GlyGGA
114	CACTCGTCACCACACTCACCGGGGGCGTCCAGTGCTTCTCCCG	pGFPy66GlyGGG
115	CACTCGTCACCACACTCACCTCGGGCGTCCAGTGCTTCTCCCG	pGFPy66SerUCG
116	CACTCGTCACCACACTCACCCTGGGCGTCCAGTGCTTCTCCCG	pGFPy66LeuCUG
117	CACTCGTCACCACACTCACCACGGGCGTCCAGTGCTTCTCCCG	pGFPy66ThrACG
118	CACTCGTCACCACACTCACCTTAGGCGTCCAGTGCTTCTCCCG	pGFPy66LeuUUA
119	CACTCGTCACCACACTCACCTTGGGCGTCCAGTGCTTCTCCCG	pGFPy66LeuUUG
120	CACTCGTCACCACACTCACCTAAGGCGTCCAGTGCTTCTCCCG	pGFPy66OchreUAA
121	CACTCGTCACCACACTCACCTGAGGCGTCCAGTGCTTCTCCCG	pGFPy66OpalUGA
122	CACTCGTCACCACACTCACCTATGGCGTCCAGTGCTTCTCCCG	pGFPy66TyrUAU
123	CACTCGTCACCACACTCACCTGGGGCGTCCAGTGCTTCTCCCG	pGFPy66TrpUGG
Note: GFP reporters containing all other sense codons at the fluorophore tyrosine position were previously constructed [8, 9].		

Table A.15. Primers for Mutating Position 5 in the Z-Domain of Protein A.

Primer	Primer sequence	Primer name
124	GATCCATGGCCGTAGACAACAAATACAACAAAGAACAACAAAACGCG	QC Z-Dom F5Y Fwd
125	CGCGTTTTGTTGTTCTTTGTTGTATTTGTTGTCTACGGCCATGGATC	QC Z-Dom F5Y Rev
126	GATCCATGGCCGTAGACAACAAATAGAACAAAGAACAACAAAACGCG	QC Z-Dom F5UAG Fwd
127	GCGTTTTGTTGTTCTTTGTTCTATTTGTTGTCTACGGCCATGG	QC Z-Dom F5UAG Rev

REFERENCES

1. Sambrook and Russell, *Molecular Cloning: a laboratory manual 3rd Edition*. CSHL Press.
2. Clackson, T., *Phage display : a practical approach*. 2007, Oxford [u.a.]: Oxford Univ. Press.
3. Barbas, C.F., *Phage display : a laboratory manual*. 2001, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
4. Miyazaki, K., et al., *Directed evolution study of temperature adaptation in a psychrophilic enzyme.* J Mol Biol, 2000. **297**(4): p. 1015-26.
5. Arnold, F.H. and J.C. Moore, *Optimizing industrial enzymes by directed evolution*, in *New Enzymes for Organic Synthesis*. 1997, Springer. p. 1-14.
6. Moore, J.C. and F.H. Arnold, *Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents.* Nature biotechnology, 1996. **14**(4): p. 458.
7. Hanson-Manful, P. and W.M. Patrick, *Construction and analysis of randomized protein-encoding libraries using error-prone PCR.* Methods Mol Biol, 2013. **996**: p. 251-67.
8. Biddle, W., M.A. Schmitt, and J.D. Fisk, *Evaluating Sense Codon Reassignment with a Simple Fluorescence Screen.* Biochemistry, 2015. **54**(50): p. 7355-64.
9. Schmitt, M.A., Biddle, Wil, and John D. Fisk, *Mapping the Plasticity of the E. Coli Genetic Code with Orthgonal Pair Directed Sense Codon Reassignment.* submitted, 2017.

10. Lee, B.S., et al., *Incorporation of Unnatural Amino Acids in Response to the AGG Codon.* ACS Chem Biol, 2015. **10**(7): p. 1648-53.
11. Gao, X., *Thermodynamically balanced inside-out (TBIO) PCR-based gene synthesis: a novel method of primer design for high-fidelity assembly of longer gene sequences.* Nucleic Acids Research, 2003. **31**(22): p. 143e-143.
12. Yanagisawa, T., et al., *Crystallographic studies on multiple conformational states of active-site loops in pyrrolysyl-tRNA synthetase.* J Mol Biol, 2008. **378**(3): p. 634-52.
13. Neumann, H., et al., *Genetically encoding protein oxidative damage.* Journal of the American Chemical Society, 2008. **130**(12): p. 4028-4033.
14. Neumann, H., S.Y. Peak-Chew, and J.W. Chin, *Genetically encoding N(epsilon)-acetyllysine in recombinant proteins.* Nat Chem Biol, 2008. **4**(4): p. 232-4.
15. Fan, C., et al., *Rationally evolving tRNAPyl for efficient incorporation of noncanonical amino acids.* Nucleic Acids Res, 2015. **43**(22): p. e156.
16. Pédelacq, J.-D., et al., *Engineering and characterization of a superfolder green fluorescent protein.* Nature biotechnology, 2006. **24**(1): p. 79.
17. Lee, B.S., et al., *Incorporation of unnatural amino acids in response to the AGG codon.* ACS chemical biology, 2015. **10**(7): p. 1648-1653.