DISSERTATION

EXPANDING AND EVALUATING SENSE CODON REASSIGNMENT FOR GENETIC CODE EXPANSION

Submitted by

C. William Biddle

Department of Chemistry

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2017

Doctoral Committee:

Advisor: John D. (Nick) Fisk

Chris Ackerson
Charles Henry
Tim Stasevich

ABSTRACT

EXPANDING AND EVALUATING SENSE CODON REASSIGNMENT

FOR GENETIC CODE EXPANSION

Genetic code expansion is a field of synthetic biology that aims to incorporate non-canonical amino acids (ncAAs) into proteins as though they were one of the 20 “natural” amino acids. The amino acids which naturally make up proteins are chemical limited, and ncAAs can carry new chemical functionality into proteins. Proteins are of interest because they are simple to produce with good consistency and have immense potential due to the diversity of structure and function. Incorporating ncAA into proteins expands the scope of function of proteins even further.

Two methods have been widely used for genetic code expansion, global amino acid replacement and amber stop codon suppression. Global amino acid replacement exchanges one of the natural amino acids for a ncAA, producing an altered 20 amino acid genetic code. Amber stop codon suppression incorporates ncAA in response to the UAG stop codon making a 21 amino acid genetic code, but is limited in incorporation efficiency and producing proteins with multiple instances of a ncAA is challenging. We wanted to use a third genetic code expansion system called sense codon reassignment which has not been widely employed at all but should enable multisite incorporation of ncAAs. When the work presented in this dissertation was started, a single report of sense codon reassignment existed in the literature. We set out to improve and expand sense codon reassignment for the incorporation of multiple copies of ncAAs into proteins. We quickly discovered disparities in what was known regarding the variables that could be used to manipulate genetic code expansion, and the focus of our work shifted to systems for improving sense codon reassignment using quantitative measurements.

The first chapter of this dissertation is an introduction to genetic code expansion and the processes of translation and gene expression that are likely involved or could be involved in genetic code expansion.

The three following chapters will build upon the fundamentals described in Chapter 1. The second chapter is a complete story about how a screen to quantify sense codon reassignment was developed. The fluorescence based screen was used in a high throughput fashion to screen a directed evolution library of variants for increased sense codon reassignment efficiency at the Lys AAG sense codon. While evaluating various sense codons for potential reassignment efficiency, the AUG anticodon was found to be incapable of discriminating between the CAU and CAC codons. This was anomalous relative to the other anticodons we tested. Chapter 3 describes how unintended modifications to an engineered tRNA were identified and then how the fluorescence based screen was used to engineer the tRNA further for increased sense codon reassignment efficiency and to avoid the unintentional modification. Most applications of genetic code expansion rely on modifications to tRNAs but few reports actually consider them, The final chapter of this dissertation is a manuscript in preparation describing the reassignment of a rare sense codon to incorporate ncAAs. The chapter focuses on how improvements made in a system specific for an amino acid can be transferred to systems specific for other ncAAs. Over 150 different ncAAs have been incorporated into proteins using genetic code expansion technologies, but the extent to which the various systems are combinable has barely been evaluated.

This dissertation is a story about developing sense codon reassignment to functional levels and quantifying the effects of different variables along the way.

ACKNOWLEDGMENTS

I am tremendously grateful to my advisor, Dr. Nick Fisk for his contagious enthusiasm for science. When I first sat in his office (and countless times since) his excitement about the possibilities in the world of science and engineering convinced me that he would be an excellent advisor for my first formal scientific pursuits. I was not wrong. I am grateful for Nick's patience, support, inspiration, and guidance throughout my PhD.

Thank you to Dr. Meg Schmitt who acted as an unofficial co-advisor, staff scientist and friend. Meg was always available for help with technical protocols or with dealing with the mundane challenges in graduate school.

A special thank you to my PhD committee, Dr. Chris Ackerson, Dr. Chuck Henry, and Dr. Tim Stasevich. I owe Dr. Ackerson for several occasions where he was available to discuss my research, or sometimes more importantly, the academic world outside my lab and how my research fit into a larger scientific frame. Dr. henry was always available to help, whether I needed supplies for an analytical procedure or logistics help getting to and from conferences. Working between two departments was sometimes a challenge and Chuck was there to help. And thank you to Dr. Stasevich, who agreed last minute to fill in on and increase the biological perspective on my committee.

I owe a debt of gratitude to all of the professors in the Chemistry Department for encouraging, inspiring and educating me, but I especially want to thank the late Dr. Michael Elliott. Mike inspired me to not only be a better chemist and think more carefully, but also to try and inspire others around me without worrying about what people will think. Dr. Elliott was his own person and left a lasting impact on my impression of what a scientist and a good human being should be. Thank you, Mike, and I miss you.

Special thanks go to my brother Hovis, with whom I had several late night conversations where he encouraged me to keep pursuing my research. His passion for science is tangible and I was able to hold on

to that passion when the tedium of research sometimes seemed too much. Thank you to the rest of my family as well who supported me emotionally and physically throughout not only my PhD, but also through my undergraduate experience. Thank you to my parents who nourished a natural curiosity in me from an early age.

And finally, thank you to all my friends who have become my family abroad. You have helped me continue exploring the world around us. My colleagues from grad school have become my most trusted of friends.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION: EXPANDING THE GENETIC CODE

CHEMICAL DIVERSITY OF NATURAL PROTEINS

Carbohydrates, nucleic acids, and proteins are the functional polymeric materials that perform all of the critical functions in living systems. Proteins are ubiquitous in industrial applications, medical diagnostics and therapeutics, and as tools for understanding biochemistry and cellular biology. The primary sequence of proteins is genetically encoded and DNA base pairs can be precisely mutated, making it possible to modify protein sequences with a high degree of precision. As the ability to manipulate DNA is continually becoming cheaper and easier, the ability to engineer novel proteins with unique and diverse sequences has improved.

While the *de novo* design of proteins is also continually improving, designing structure, stability, and function still present obstacles for the field of protein engineering. The space of possible protein sequences is inconceivably vast, and the rules that determine protein folding and stability are not known with the precision needed for design. Nature does not use rational design, but rather relies on evolution to navigate the space of functional proteins. The tools of DNA manipulation now available have enabled high throughput studies where millions or billions of protein candidates can be made and evaluated in parallel. Proteins combine 20 natural amino acids into primary sequences that fold into specific secondary and tertiary structures based largely on the hydrophobic packing interactions of the amino acids. The primary sequence of the amino acids control not only the structure of the protein, but positions chemical functionality precisely on protein surfaces and in pockets to define the interactions and enzymatic functions of the protein. Although proteins are sufficient for the functions of living systems and have found utility in solving problems in science, health, and engineering, the set of chemical functionalities found in natural proteins is very limited. The 20 natural amino acids, although sufficient for the proteins of life, are limited to three categories of chemical functionality: acid/base chemistry using primarily

amines and carboxylic acids, polar non-ionic residues which include hydroxyl and amide side chains, and nonpolar hydrophobic residues with aliphatic side chains of different steric shapes. Genetic code expansion (GCE) is one of the more promising techniques for incorporating non-natural chemical functionality into proteins. Genetic code expansion involves using the natural machinery that produces proteins to incorporate non-canonical amino acids (ncAAs) with diverse chemical side chains in response to specific nucleic acid sequences. The "expansion" of the genetic code involves modifying the process of protein translation such that additional or repurposed sequences of nucleic acids signal the incorporation of additional (beyond the 20 natural) amino acids into proteins.

An example of the utility of genetic code expansion is the expansion of the binding ability of antibodies. Antibodies are well known for their role in binding and identifying foreign materials in higher living organisms. Biotechnologically, the process of antibody development can be used to raise antibodies to bind a great diversity of targets. However, there are some targets that antibodies cannot bind well. A large and important class of molecules that typically produce weak antibody responses and generate weakly binding antibodies are glycan moieties.[1–3] The difficulty in binding glycans is in large part due to the limited set of the chemical functionality of the natural amino acids. When antibody libraries with amino acid diversity in their binding domains are challenged with glycans, the chemical functionality present in the side chains of the 20 natural amino acids is not sufficient to produce tight binding antibodies. The original problem that motivated the work described in this thesis, was to pursue directed evolution of peptides containing non-canonical amino acids specifically able to target glycan moieties. Boronic acids have been shown to form tight, reversible covalent bonds with germinal and vicinal diols, like those found in the carbohydrates.[4,5] By including boronic acid in the side chain of an ncAA (para-boronophenylalanine, pBF) in large libraries of antibody variants screened against glycans, anti-glycan antibodies could be evolved. The idea was to potentially employ multiple copies of boronic acids in antibodies to target and interact specifically and tightly with glycans. None of the methodologies for genetic code expansion available at the time would have enabled incorporation of multiple ncAAs into

genetically encoded libraries. We set out to combine already established methods for the incorporation of pBF into proteins in response to non-sense codons and sense codon reassignment to direct incorporation of pBF at sense codons *in vivo*.

As we started working on the project to develop boronic acid for directed evolution libraries, it became clear that there were large gaps in knowledge surrounding the engineering of orthogonal translation systems (OTS), and although we set out to build antibodies, we ended up spending much more time evaluating orthogonal translation systems and the biochemistry of translating proteins with orthogonal components. This dissertation describes efforts to modify and engineer critical components of translation to enable genetic code expansion using the relatively new technique termed sense codon reassignment. Additionally, this dissertation describes several examples of quantitative evaluations of factors that affect the efficiency of sense codon reassignment.

The first chapter of the thesis provides a detailed description of the critical steps in the *in vivo* process of producing proteins, and highlights the component parts and variables that can be manipulated to alter the translation mechanism to expand the genetic code. Current methods for genetic code expansion including nonsense codon suppression (NCS) and global amino acid replacement (GAAR) are briefly reviewed. The discussion of the GCE techniques is focused on the benefits, limitations, and details of NCS and GAAR relative to other techniques which can be used to add to the chemical diversity of proteins. Finally, a discussion regarding how genetic code expansion systems have typically been quantified and why quantification of GCE systems is important will be presented.

FUNDAMENTALS OF TRANSLATION

Translation is one of the most central processes in living organisms. In rapidly growing cells, such as bacteria, maintenance of the translation apparatus and the synthesis of proteins accounts for approximately 75 percent of the cells energy budget.[6] The amount of energy (measured as tri-phosphate bonds) devoted to gene expression is an indicator of how important the process of protein biosynthesis is.

The process of protein biosynthesis is remarkably similar across all domains of life, nearly identical mechanisms employing similar proteins and complex machinery are employed. There are some differences between translation in prokaryotes, archaea, and eukaryotes, particularly in the mechanism of initiation but similarities greatly outweigh differences. The work described in this thesis exclusively employed prokaryotic translation system, that of the model bacterium *Escherichia coli* (*E. coli*), and the following discussion of translation will be focused on the *E. coli* system.

The central dogma of information transfer in molecular biology states that genetic information stored in DNA is transcribed into RNA, which is then translated into proteins. Organisms across all domains of life use the same genetic code where specific triplet nucleic acid codons encode the incorporation of specific amino acids into polypeptides. Protein synthesis is the result of interactions between an mRNA transcript, and aminoacyl-tRNA molecules, coordinated by ribosomes and additional cellular factors involved in initiation, control, and termination of the process. Three protein factors are involved in assembling the mRNA, initiator tRNA and ribosome to initiate translation of an mRNA transcript. The process of elongation is mediated by two protein factors and involves cycles of tRNA selection, amino acid addition and ratcheting of the mRNA in the ribosome to introduce the next codon for translation. The process of termination is mediated by two protein factors that recognize stop codons and hydrolyze the completed peptide chain from the tRNA. Additional protein factors disassemble and recycle ribosomes after the process of translation is complete. Further protein factors monitor the process of translation to hydrolyze peptides from stalled ribosomes and assist in the forming of nascent peptide chains. Beyond the set of proteins that are directly involved in the process of translation, multiple protein factors are responsible for maturation and modification of tRNAs and ribosomes. A set of proteins, the amino acyl tRNA synthetases (aaRSs) are responsible for selecting and attaching the appropriate amino acid to the appropriate tRNA. Several of the critical components and processes of translation will be discussed in greater detail below, with a focus on how each of the processes may be involved in or targeted for genetic code expansion. The material presented in the following sections is an in-depth

description of the molecular biology of translation. The description is presented because an understanding of the complex and interconnected process of translation is critical to understanding how genetic code expansion systems function and can be engineered.

Transcription

To be expressed, genes—sequences of DNA that encode proteins or functional RNA molecules, which can be either in the genome of the cell or on extrachromosomal elements (e.g. plasmids)—must first be transcribed into mRNA by RNA polymerase (RNAP). The extent to which different genes are expressed is dependent on the amount of mRNA that is produced and the extent that mRNA can compete for ribosomes. Specific DNA sequences called promoters bind RNAP. Both the binding strength of the promoter for a given gene as well as the number of copies of that gene in the DNA of a cell influence the amount of mRNA transcribed. The relative abundance of a specific mRNA in the intracellular pool of mRNA molecules along with the relative strength of the ribosome binding site (RBS) on that mRNA, which determines the extent of interaction with ribosomes control the extent of protein production. One common approach to affecting the expression of the genetic code is the overexpression of altered translation components, by having them expressed from a high copy number plasmid with a strong promoter and ribosome binding site. Most reported systems for genetic code expansion involve the over expression of an aminoacyl-tRNA synthetase (aaRS) and its cognate tRNA, to increase the overall concentration of engineered components in the cell. The extent that commonly used systems increase concentrations relative to natural components has not been widely investigated. A few examples exist of plasmid design experiments that suggest that overexpression increases the effects of genetic code expansion systems.[7,8]

Additional factors in the sequence of an mRNA that can affect translation rates include secondary structure in the mRNA, particularly hairpin formation that inhibits ribosome binding, and mRNA structures which can stall or terminate translation. The particular choice of synonymous codons can also affect the rate of translation; the use of specific rare codons have been suggested to decrease translation

rate and highly expressed genes typically employ codons read by the most abundant tRNAs in the cell.[9–11] While not commonly considered, mRNA secondary structure and codon usage may provide handles to control the expression of genes using engineered genetic codes.

Although outside the scope of this thesis, one could imagine expanding the genetic code at the nucleic acid level by expanding the number of nucleic acid bases that are employed in the cell. The natural genetic code is in part limited by the fact that there are only 64 possible triplets of the 4 nucleic acid bases. By expanding the set of nucleotides used in DNA and RNA to 5 (involving an orthogonal self-pair) or 6 bases the number of available triplet codons increases to 125 or 216. The development of expanded sets of nucleic acids is an active area of research that has only recently reached the stage where altered nucleotides in DNA can be replicated in cells and transcribed into RNA.[12–14] Using increased nucleic acid diversity systems to expand the set of amino acids that can be encoded has not yet been demonstrated. Almost all reports of genetic code expansion to date have employed natural nucleic acids and modified other components of the translation system.

Translation

After a gene has been transcribed into mRNA, the ribosome binding site on the mRNA, in conjunction with initiating protein factors, binds to the small ribosomal subunit and assembles the full ribosome charged with the initiator tRNA on the initiating codon of the mRNA. The mRNA signal that initiates translation is usually an AUG codon which typically codes for methionine. A special N-formylated methionyl-tRNA is a component of the initiating complex that assembles the ribosome on an mRNA to begin the translation process. The process of amino acid addition is sequential and involves repeated cycles of aminoacyl-tRNA selection, peptide bond formation (that functionally transfers the growing peptide chain from one tRNA to the next), and translocation of the tRNAs and mRNA to sequentially introduce the next codon into the ribosome. Aminoacyl-tRNA molecules are composed of an amino acid covalently linked to a tRNA molecule which acts as a decoder between the mRNA codon and a specific amino acid. The tRNA molecule is the physical link in the genetic code, and the correct

coupling of amino acids to tRNA molecules is carried out by a suite of enzymes called aminoacyl-tRNA synthetases (aaRSs). After a tRNA is aminoacylated, it is transported into the ribosome near the active mRNA codon as a ternary complex EF-Tu and a molecule of guanosine triphosphate (GTP). The EF-Tu complex associates with the ribosome where the anticodon of the tRNA base-pairs with the mRNA codon. If the codon and anticodon base-pair with high enough affinity, the tRNA and the amino acid it carries are accepted into the ribosome. Once inside, the ribosome catalyzes the formation of a new peptide bond between the growing polypeptide chain and the amino acid just brought into the ribosome. After the formation of the peptide bond, the ribosome ratchets forward and the process repeats, with aminoacyl-tRNAs decoding codons and adding to the polypeptide chain until a termination signal (stop codon) is reached. Stop codons are recognized by release factors instead of aminoacyl-tRNAs, and release factors carry no amino acid. Upon termination of protein synthesis, the ribosome cleaves the protein from the last tRNA, dissociates from the protein and the mRNA, and is recycled into the cell to start protein synthesis on a new mRNA strand.

Ribosomes

Ribosomes are the factories which facilitate protein biosynthesis in cells and are the location where the peptide bond formation is catalyzed to produce proteins from individual amino acids. The ribosome is among the most complex molecular machines in the cell and composed of both ribonucleic acids and proteins. The small subunit of the ribosome (30S subunit) is a riboprotein complex composed of the 16S ribosomal RNA (rRNA) and 22 proteins. The role of the small subunit is to bind mRNA and read the interaction between codons and anticodons. The 16S rRNA contains a region of nucleotides that specifically recognize and pair with the mRNA and orient the initiating tRNA in the appropriate position to assemble complete ribosomes. The 16S RNA contains a specific sequence that is complementary to the Shine-Delgarno sequence contained in the ribosome binding site (RBS) of an mRNA. The pairing of the 30S subunit of the ribosome and the mRNA occurs due to the creation of effectively double stranded RNA from the 16S rRNA and the RBS in the mRNA. The alignment of the Shine-Delgarno sequence and

the complimentary region in the 16S rRNA helps to position the start codon in the A site of the completed ribosome complex. After the 30S subunit binds to the mRNA, the larger subunit of the ribosome (50S subunit) binds to the 30S subunit assembling the intact ribosome. The 50S subunit is a complex of two rRNA molecules (23S and 5S) and 31 proteins. The catalytic peptidyl transfer reaction of the ribosome is carried out by a specific domain of the 23S rRNA. The protein components of the 50S subunit play a critical role in protein folding both *in vivo* and *in vitro* as the protein components of the subunit provide a hydrophobic surface for the nascent polypeptide to interact with as it exits the ribosome.[15,16] This mode of protein folding is used especially for proteins that are translated at slower rates.[17] The production of the translational apparatus and particularly, ribosomes requires a large fraction of the cellular resources. A bacterial cell with a typical doubling time of 30 minutes will contain approximately 45,000 ribosomes.[18] With 53 proteins per ribosome, 45,000 ribosomes represents a huge proportion of the millions of proteins estimated in a typical *E. coli* cell.[19] In the process of expanding the genetic code and reassigning the meaning of codons, it becomes likely that mutations could be introduced in some of the proteins in the ribosome. This could lead to organisms with defective ribosomes, but previous studies on missense incorporation *in vivo* suggest that high levels of reassignment of codons nearing 25 % can still result in viable, although challenged cells.[20,21] Even with the abundance of protein components in the ribosome, when random missense mutations are introduced into ribosomes cells are surprisingly tolerant.[22] While specific missense mutations could present issues in the ribosome, in general codon reassignment should not dramatically affect the structure or function of the ribosome.

Although the ribosome is frequently considered the catalyst and central organizing component of translation, the overall fidelity of the translation process is the result of multiple components outside of pairing of codon to anticodon that occur in the ribosome. What has been termed the second genetic code involves the multiple competing interactions between amino acids, tRNAs and the aaRSs that result in the appropriate amino acids being acylated onto the correct tRNA molecules. Additional protein components selectively modify both the tRNA and ribosomal RNAs to modulate the translation process.

tRNAs

The molecules of RNA responsible for carrying amino acids into the ribosome are termed transfer RNAs (tRNAs). tRNA molecules are relatively short segments of RNA with between 77 nucleotides and as many as 90 nucleotides.[23] The variability in nucleotide count is centered on one arm of the clover leaf secondary structure of tRNA molecules. The overall tertiary structure of tRNA molecules is conserved across all domains of life. All tRNAs have five defined regions: acceptor stem, T arm, D arm, anticodon arm, and the variable loop region (Figure 1.1A). The nucleotides of tRNA molecules are arranged in such a way that three hairpin-like stem-loops are generated and tRNAs are frequently referred to as "cloverleaf" structures, even though the physiological structure of the tRNA is well defined as a closely packed "L" shape (Figure 1.1B). The anticodon stem loops of tRNAs contain the three nucleotides that make up the anticodon which decode the complimentary mRNA codons in the ribosome.

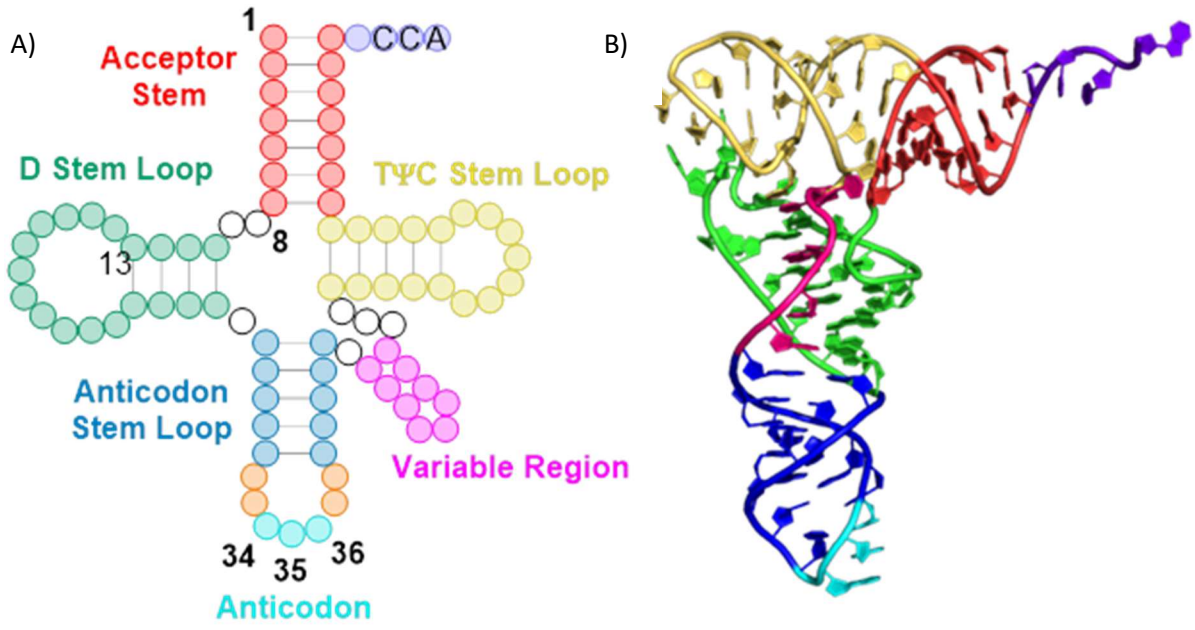


Figure 1.1 A) Secondary structure of tRNA molecule in the cloverleaf structure, showing the 5 characteristic regions of tRNA molecules. B) Cartoon representation of crystal structure of the Phe-tRNA$_{UUC}$ from Yeast (pdb: 1ehz).[109] Colors are coordinated between the two images to show the location of regions in the "L" structure. Figure adapted from reference [110]

On the opposite end of the tRNA from the anticodon is the acceptor stem which is a series of about 5-8 pairs of nucleosides and a conserved cytidine-cytidine-adenosine (CCA) tail. The CCA tail is

transcribed in the tRNA gene for some tRNA species and some tRNAs are modified by editing enzymes to have the CCA tail, but all tRNAs have the CCA tail which serves as an acceptor for the amino acid. Aminoacyl tRNA synthetases covalently link an amino acid to the tRNA, forming an ester linkage between either the 2' or 3' hydroxyl group on the ribose of the terminal adenosine and the carbonyl in the amino. Theoretically modifications to the acceptor stem of the tRNA could result in changes in the overall structure of the tRNA·AMP·ncAA in the aaRS which could lead to higher aminoacylation efficiency for a ncAA which would lead to a higher efficiency of genetic code expansion.

Specific examples of engineering approaches to modify the tRNA to increase genetic code expansion will be discussed later in the chapter, but most reports have focused primarily on changing the anticodon residues. However changes at other positions in the tRNA can result in changes in the recognition of the tRNA by the components of the translation machinery.

tRNAs from all domains of life contain modified nucleotides other than the canonical guanosine, cytidine, adenosine, and uracil. Modified nucleosides are found at positions throughout the tRNA but are especially common in the anticodon stem loop (Figure 1.2). Modifications of tRNA nucleotides allow for differentiation between the structurally similar tRNAs, some modified nucleotides are known to alter the structure of tRNAs and to enable specific interactions between modified tRNAs and amino acyl tRNA synthetases. tRNA modifications are also important in codon recognition in the ribosome. For example, the nucleotide at position 37 of most natural tRNAs is a conserved purine and is frequently modified to be methylated but can also include modifications with some small polar groups.[24] Even though position 37 is adjacent to the anticodon and does not participate directly in hydrogen bonding interactions with the codon, the modifications enforce a particular structure in the anticodon loop that orients the anticodon so that all 6 nucleotides of the codon and anticodon can interact.[24,25] Modifications of position 37 in tRNAs also prevent frameshifting mutations in translation[26–28], which come about when nucleotides in the tRNA

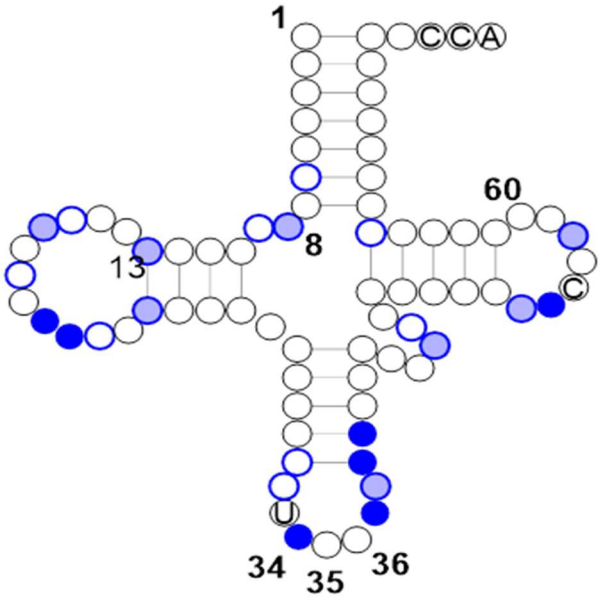


Figure 1.2 Location and frequency of modified nucleosides in prokaryotic tRNA molecules. Circles represent nucleosides in the standard cloverleaf formation. Universal tRNA numbering system from reference [109]. Positions which are known to have modified nucleosides are shown as blue circles. Blue shaded circles have are more frequently modified (at least 25% of known prokaryotic tRNA contain modified nucleosides), and filled blue circles represent positions where more than 75% of prokaryotic tRNAs have modified nucleosides. The 5 absolutely conserved prokaryotic residues, U37, C56, and the CCA tail (74-76) are highlighted. Figure adapted from reference [110].

anticodon stem loop outside of the anticodon interact with the mRNA transcript. If four nucleotides of the tRNA anticodon stem loop base-paired with four nucleotides of the mRNA, a +1 frameshift would result. Most nucleotide modifications are the result of post transcriptional modification by RNA editing enzymes which recognize specific features or structures of tRNAs. Modified nucleotides in the anticodon itself enable a single tRNA to decode more than one codon. For example, one of the *E. coli* arginyl-tRNAs is transcribed with an ACG anticodon corresponding to positions 34-36 in the tRNA. An ACG anticodon would be able to decode only a CGU codon employing normal Watson-Crick base pairing rules: adenosine and uracil base pair using 2 pairs of hydrogen bond donors and acceptors and cytidine and guanosine pair together with 3 hydrogen bonds. In *E. coli*, the adenosine at position 34 is deaminated by an RNA editing enzyme (tadA) to produce the modified base inosine[29]. The nucleoside inosine can base pair with uracil, adenosine, or cytidine, so the modified tRNA with an ICG anticodon is then capable of decoding the original CGU codon as well as GCC and GCA codons. Knocking out the tadA gene results in cells that are not viable, likely due to the inability to efficiently decode GCC and GCA codons.[29]

Controlling the type and extent of tRNA modification is a general strategy for tuning genetic code expansion. Unanticipated tRNA modification can also present challenges for orthogonality. Although no reports currently describe actively engineering modified nucleosides into tRNAs for the purpose of genetic code expansion, specific modifications could help to increase the affinity of a tRNA for its aaRS, or for other components of the translation machinery like EF-Tu. Further, modifications to tRNA nucleosides in the anticodon could be used to restrict wobble base decoding to more finely target specific codons. Although all tRNA molecules likely contain modified nucleosides, genetic code expansion experiments have largely failed to consider the presence or absence of modifications in introduced orthogonal tRNAs. Chapter 3 of this dissertation presents a description of how tRNA modifications can affect genetic code expansion, how modifications were identified, and ultimately how the efficiency and fidelity of orthogonal translation components can be modulated by controlling the extent of modifications.

Aminoacyl tRNA synthetases

Aminoacylation of an amino acid to a tRNA is a two-step reaction catalyzed by an aminoacyl tRNA synthetase (aaRSs). Most organisms have exactly 20 aaRSs, one for each amino acid. Although an aaRS is specific for a single amino acid, most aaRS recognize more than one subtype of tRNA molecule as substrates. There are 49 tRNA species in *E. coli*[30] and only 2 of the 20 aaRS recognize single tRNA species, meaning the other 18 aaRS have on average two tRNAs which are substrates. The aaRSs are responsible for correctly linking anticodons on tRNAs, with specific amino acids. Properly aminoacylated tRNAs are then selected in the ribosome complementary to mRNA codons. In the first step of the aminoacylation reaction, an amino acid and a molecule of adenosine triphosphate (ATP) bind to the aaRS in the amino acid binding pocket (Figure 1.3, reaction 1) to form aminoacyl-adenosine monophosphate (AMP) and a molecule of inorganic pyrophosphate. Subsequently, the tRNA binds to the aaRS, making nonspecific interactions between the D-arm and the surface of the aaRS, and specific interactions with the identity elements of the tRNA[31,32]. The anticodon and the acceptor stem of the tRNA are usually identity elements, although several aaRS do not use anticodons of substrate tRNAs as identity elements[33–35]. aaRS

which do not recognize the anticodon of a tRNA as an identity element for aminoacylation present an opportunity for codon reassignment to expand the genetic code. By altering the anticodon of a tRNA to decode a new codon, without significantly impacting the recognition of an aaRS for the tRNA, aminoacyl-tRNAs may be prepared with non-standard anticodons.

Proper binding of a tRNA by the aaRS positions either the 2' or 3' hydroxyl groups on the terminal adenosine in the C-C-A acceptor stem adjacent to the aminoacyl-AMP molecule and the aaRS catalyzes the substitution of the tRNA for the AMP to covalently link the amino acid to the tRNA (Figure 1.3, reaction 2). Misacylation occurs when an aaRS charges its amino acid onto a tRNA which is not a canonical substrate, or when an aaRS activates the incorrect amino acid and attaches it to the cognate tRNA substrate. Misacylation events that result in errors in translation where amino acids are covalently linked to noncognate tRNAs occur on the frequency of 1 in 10000, depending on the rate of translation.[36–38]

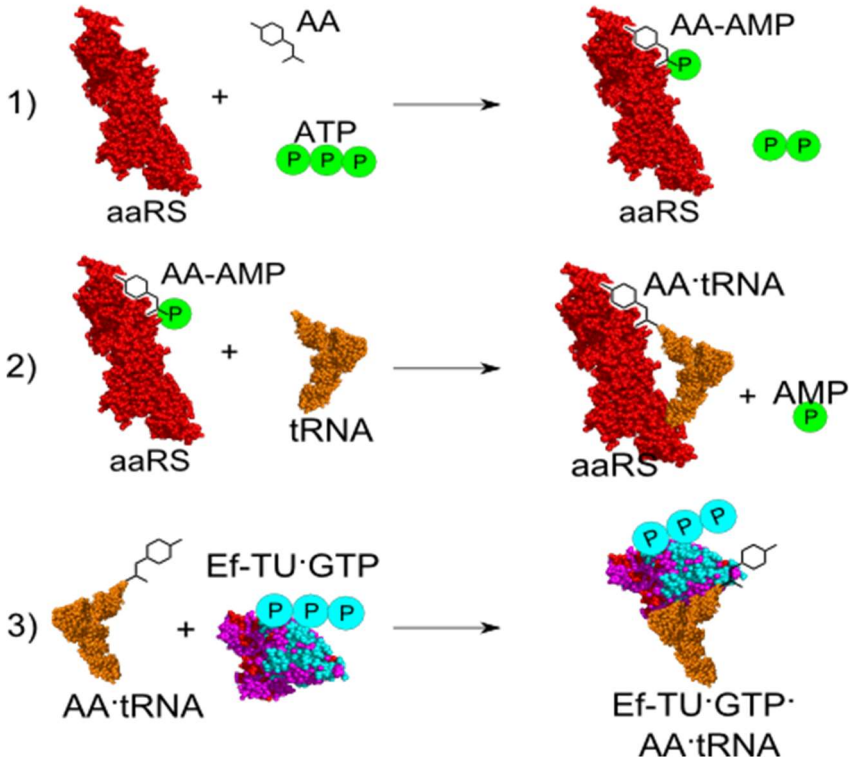


Figure 1.3 Reactions to generate the pool of EF-Tu·GTP·AA-tRNA ternary complex. Reactions 1 and 2 collectively represent aminoacylation. PDB codes: 1j1u, 5jbq.

aaRSs are multiple domain enzymes that typically contain a catalytic domain to bind ATP and activate an amino acid, a separate domain is typically present to facilitate binding to the appropriate tRNA and aaRSs often contain editing domains to hydrolyze misacylated tRNAs. It has been shown that aaRSs can have relatively promiscuous amino acid binding pockets. The Ile aaRS (IleRS) can readily activate valine (which differs by only a single methyl group) and the combination or the relative concentrations and of Ile and Val in the cell and the relatively high efficiency of valine activation can lead to very high levels of val-tRNA^{Ile} if the editing domain of IleRS is inactivated.[39] aaRSs have evolved two general editing strategies to minimize the extent of tRNA misacylation. Pretransfer editing occurs when the activation domain is able to hydrolyze a misacylated aa-AMP before transfer to the tRNA. Post-transfer editing employs a second domain that is able to hydrolyze misacylated tRNAs before dissociation from the aaRS. Post-transfer editing relies on a double sieve mechanism to select inappropriately acylated tRNAs.[40] The first steric sieve occurs in the activation, where amino acids smaller than the appropriate amino acid can enter the activation domain and become attached to tRNA (i.e. Val is activated by IleRS). The second sieve does not allow the appropriate amino acid to enter the binding pocket but does allow smaller amino acids to enter and get hydrolyzed. Pre and post transfer editing of certain aaRSs is augmented by non-aaRS editing enzymes that surveil the tRNA pool.[41,42]

aaRS have evolved to control the level of misacylation of natural amino acids (and natural metabolic amino acid precursors.[43,44] Many aaRS in nature are able to accept non-canonical amino acids (ncAAs) as substrates and aminoacylate their tRNAs with the alternate amino acid. For example, the *E. coli* Methionyl-tRNA synthetase (MetRS) is known to activate multiple structural analogues of methionine when methionine is unavailable.[45–47] Probably the first broad application of genetic code expansion involved the incorporation of the ncAA selenomethionine into proteins in place of methionine.[45]

Elongation factor EF-Tu

Once a tRNA has been aminoacylated, the aminoacyl-tRNA needs to be transferred to the acceptor site of the ribosome (A site) where the exposed codon on the mRNA is available for binding. To bring tRNAs into the ribosome, a highly-conserved elongation factor called EF-Tu binds in a ternary complex with a guanosine triphosphate (GTP) molecule and an aminoacyl-tRNA (Figure 1.3, Reaction 3). Ribosomes bind very quickly to the EF-Tu-GTP-aminoacyl-tRNA complexes (ternary complex) near the A site regardless of what the aminoacyl-tRNA or the mRNA codon in the A site is (Figure 1.4). Once the ternary complex is in the A site, the strength of hydrogen bonding interactions between the anticodon and codon are used as a screen for aminoacyl-tRNA selection. The more strongly associated the anticodon and codon are, the longer the tRNA will remain in the active site. Ternary complexes that do not bind to the codon in the A site dissociate away from the ribosome at rates approximately 85 times faster than those aminoacyl-tRNAs which bind to the codon, based on calculations with empirically determined equilibrium constants.[37] Many tRNAs are sampled before one complimentary to the codon in the A site is found. An aminoacyl-tRNA which binds to the codon with enough stabilization energy in the A site of the ribosome induces a structural change in the EF-Tu molecule which directs the GTPase domain towards the GTP molecule. Once a ternary complex is stabilized in the A site, indicating a match between codon and anticodon, GTP is hydrolyzed to guanosine diphosphate (GDP) which results in a structural rearrangement of EF-Tu leading to dissociation of EF-Tu·GDP and a phosphate molecule from the A site. The aminoacyl-tRNA bound to the codon in the A site has one final opportunity for proofreading. In the absence of EF-Tu stabilization, the free aminoacyl-tRNA can either dissociate from the A site or remain in the A site and be accommodated into the peptidyl transferase site (P site). The accommodation process involves conformational changes in the tRNA and ribosome. Once the aminoacyl-tRNA in the A site moves towards the peptidyl-tRNA in the P site, the transfer of the growing peptide chain to the aminoacyl-tRNA is catalyzed very quickly.

To complete a cycle of peptide elongation as the ratchet completes, the newly selected tRNA will now be covalently attached to the peptide chain and occupy the P site, while the recently released tRNA which was occupying the P site will be in the exit site (E site) and will dissociate from the ribosome to be recycled in the cytoplasm and re-charged with a new amino acid.

The fine details of the rates of all the steps in the process are largely unknown, it is clear that differences in tRNAs and amino acids affect the kinetics of the various steps but the extent of the differences and their relative importance are unclear. The process of genetic code expansion has typically focused on producing non-canonical aminoacyl-tRNA molecules. These non-canonical adapters between codons and the ncAA need to not only be successfully aminoacylated, they need to interact with all of the components of the translation machinery efficiently. EF-Tu serves as a transport molecule but is also involved in the proofreading of aminoacyl-tRNAs in that EF-Tu binding involves contributions from both

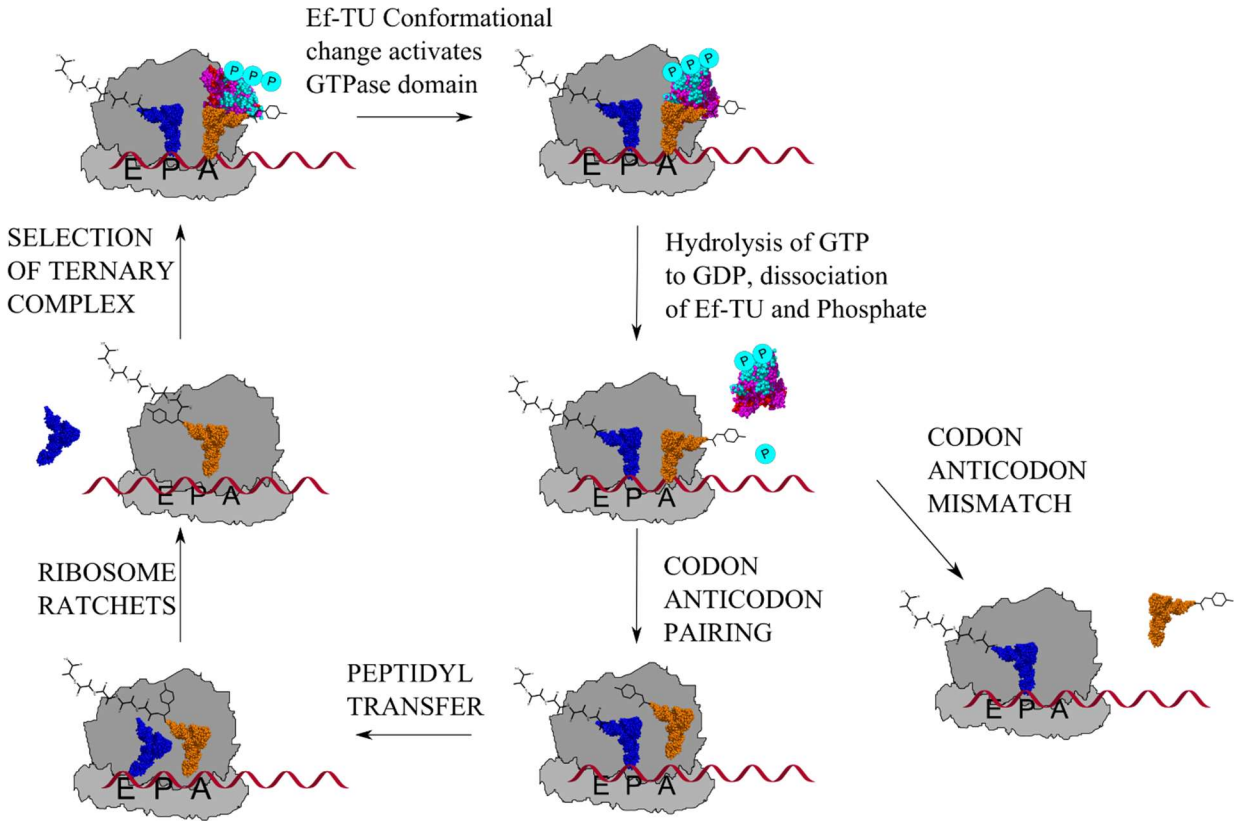


Figure 1.4 Schematic representation of selection of an aminoacyl-tRNA in the A site of the ribosome (as ternary complex) and subsequent steps resulting in catalysis of peptide bond formation and elongation of polypeptide chain.

the tRNA and amino acid. Bulky or charged amino acids decrease the affinity of EF-Tu for aminoacyl-tRNA complexes.[48,49]. Engineering EF-Tu may be necessary for genetic code expansion with some ncAAs and tRNAs. Alternately the tRNA molecule itself or potentially even the A site of the ribosome could be engineered to accommodate different ncAAs and anticodon codon interactions.

Engineering Orthogonal Translation Systems

Gene expression is the cyclic process of aminoacylation of tRNAs, transport of aa-tRNAs into the ribosome after forming the ternary EF-Tu-GTP-aa-tRNA complex, selection and rejection of aa-tRNAs based on structure and anticodon binding in response to the mRNA codon, and finally catalysis of the peptidyl transfer reaction and exit and folding of the protein from the ribosome. In bacterial systems protein synthesis occurs at a speed of about 20 amino acids per second, highlighting how finely tuned the entire gene expression system is.[50] Methods to expand the genetic code and try to incorporate additional amino acids into proteins using ribosome based protein synthesis machinery *in vivo* must engineer these systems to accept new amino acids, yet also remain compatible with all of the existing components. tRNA modifying enzymes, high energy molecules like ATP and GTP, and initiation, elongation, and termination factors are all involved in protein synthesis, but the subsystems of translation that have been engineered for genetic code expansion involve amino acids, tRNAs, aaRSs, EF-Tu and the ribosome.

Several approaches have been taken to engineer the components of translation in order to produce orthogonal translation systems (OTS). Translation components have been engineered, with rational approaches in concert with directed evolution approaches. To date it remains unclear which components of translation hold the greatest potential for improving genetic code expansion technologies. The field of genetic code expansion is largely pursuing codon reassignment as the dominant technology, but other groups are working on diverse projects which may prove to be more impactful.

COMMONLY USED METHODS FOR GENETIC CODE EXPANSION

Non canonical amino acids (ncAAs) are amino acids that are not naturally encoded in the canonical genetic code or incorporated into proteins in the process of ribosomal protein synthesis. The terms unnatural or non-natural amino acid are sometimes used to refer to ncAAs, but there are over 500 amino acids which occur in natural systems that are not ribosomally incorporated into proteins.[51] Genetic code expansion technologies seek to add ncAAs as though they were one of the 20 canonical genetically encoded amino acids. Genetic incorporation of ncAAs enables the production of proteins with non-natural chemical functionality. Some of the most widely used ncAAs include amino acids with side chains that contain azide and alkyne functionality, photocaged natural amino acids that produce a wildtype protein only after exposure to specific wavelengths of light, and analogues of natural amino acids with different structures in the side chain of the amino acid which enables studying how a particular residue interacts structurally. The current modern iteration of genetic code expansion technology is only 20 years old, but in this time has enabled the incorporation of over 150 different types of ncAA into proteins *in vivo.*

Although major improvements have been made in the last two decades, genetic code expansion is not a new technology. The longer history of genetic code expansion can be traced back to experiments that were performed in the process of elucidating the genetic code. Cohen and Cowie showed that auxotrophic cells could use structural analogues of natural amino acids, replacing methionine with selenomethionine.[45,52] Weisblum and Benzer used Raney nickel to convert Cys-tRNA^{Cys} into Ala-tRNA^{Cys} to rewrite the code and show that tRNAs sequence determined what amino acid was incorporated in response to a particular codon.[53] There was an extended period in the 1980s and 1990s when the dominant methods for genetic code expansion involved chemical amino acyclation of tRNAs *in vitro*.[54] The modern in vivo methods for genetic code expansion began with the development of orthogonal synthetases in 1998.[55,56] Several different methodologies have been developed to expand the genetic code, and three major classes of genetic expansion technologies will be discussed: global amino acid replacement, nonsense codon suppression, and sense codon reassignment.

Global Amino Acid Replacement

One of the first bioengineering technologies to expand the genetic code was global amino acid replacement (GAAR). In the 1950s Cohen and Cowie discovered that some endogenous aaRS were relatively promiscuous with regard to their amino acid substrate[45,52]. By either increasing the concentration of an aaRS, or by decreasing the relative concentration of its natural amino acid substrate, the aaRS will aminoacylate its cognate tRNA with structural analogues of the amino acid if they are available. For practical applications, GAAR usually requires using cell lines that are auxotrophic for the given amino acid and depleting the growth medium of the amino acid after cells have reached a reasonable cell density. Methionine is a frequently targeted residue for replacement. The methionyl-tRNA synthetase in *E. coli* (and other organisms) can aminoacylate different structural analogues of methionine including azidohomoalanine, homopropargylglycine, and selenomethionine. Efficiencies of up to 98% incorporation of methionine analogues with unsaturated side chains have been reported.[57]

As with all genetic code expansion technologies, the ncAA to be incorporated must be soluble and accessible to the cell, either by crossing the cell membrane or being produced inside the cell. GAAR exchanges all instances of a natural amino acid for a ncAA, resulting in an altered 20 amino acid genetic code. Methionine is a special case in *E. coli* in that there is only one codon which encodes methionine and only one tRNA which decodes it. Other aaRS are available and can be engineered for GAAR, including aaRS from exogenous sources and endogenous aaRS which have been engineered to have activity towards ncAAs. When using an aaRS which aminoacylates multiple tRNAs, this again results in all of the instances of a particular amino acid being replaced by the ncAA. Methionine is frequently targeted for GAAR because methionine is rare in the genome; only 2.6 out of every 100 codons in the *E. coli*.[58]

Global amino acid replacement was originally used to incorporate heavy selenium atoms (as selenomethionine) into proteins of interest to solve phasing problem in protein crystal structure determination.[59] Replacement of methionine with selenomethionine resulted in ribonuclease H that could

be purified with complete substitution at the three methionine codons, resulting in a unique signal for multiwavelength anomalous diffraction.[59]

The Tirrell lab recently reported using GAAR in a technique called BONCAT (Bio Orthogonal Non Canonical Amino acid Tagging) to label proteins that were produced as a result of heat shock to an organism. Working in vegetative *Aridopsis thaliana* the ncAA azidohomoalanine (Aha) was incorporated into proteins in response to the Met AUG codon[60]. By adding the Aha in timed pulses, Aha is only incorporated into newly synthesized proteins. Proteins can then be either labeled or purified using the azido functional group not present in any other proteins in the cell. Because the handle used to label or purify the proteins was pulsed into the cells. Temporal resolution from pulse-chase addition of the ncAA label can be combined with proteomic approaches to investigate the differential protein expression of an organism in response to physiological events. In the case of this study with Aridopsis, BONCAT was used to evaluate how protein expression levels changed as a function of heat shock, but BONCAT using GAAR has been used for studies of newly synthesized proteins in plants, bacteria, and mammals[60–63].

Methionine is not the only residue which can be replaced using GAAR. In attempt to push the limits of genetic code expansion using residue replacement, Merkel and co-workers simultaneously incorporated three analogues of natural amino acids (Pro, Phe, and Trp) into a protein of interest using a polyauxotrophic cell line. Using the endogenous aaRSs , 4(S)-fluoroproline, 4-fluorophenylalanine, and 6-fluorotryptophan were reassigned at six, sixteen, and two positions simultaneously in a single enzyme.[6465] Fluorinated ncAAs are useful for ^{19}F-nuclear magnetic resonance studies of protein structure[56] and studying biophysical properties of proteins[66], or for trying to create “Teflon-like” enzymes with enhanced solvent stability and specific activity.[65]

Global amino acid replacement traditionally uses endogenous aaRS in the host organism that have substrate promiscuity to aminoacylate ncAAs. Engineered aaRS can also be expressed in cells to aminoacylate tRNAs which expands the ncAAs which can be incorporated *in vivo*. Auxotrophic cells are still required in order to eliminate the competition with the natural amino acid to be replaced.

Nonsense Codon Suppression

Codon reassignment is a method for changing the way in which the 64 triplet codons are decoded in an organism. GAAR effectively recodes all the codons for a specific residue simultaneously by aminoacylating all of the tRNA isoacceptors for a natural amino acid with a ncAA. Selective reassignment experiments focus on engineering the anticodon of only specific tRNA molecules to have altered decoding relative to the canonical genetic code. Over the last two decades nonsense codon suppression, and specifically the incorporation of ncAAs in response to the UAG amber stop codon has been extensively used to expand the genetic code. Simultaneously developed in multiple labs and most extensively developed in the Schultz laboratory[56,67–69], amber stop codon suppression involves using a tRNA with CUA anticodon which decodes the UAG codon which typically codes for termination of protein synthesis. Typically, stop codons are recognized by release factors in the A site of the ribosome which recruit a water molecule into position to hydrolyze the polypeptide chain from the tRNA in the P site of the ribosome.[70] The amber stop codon is the least frequently used codon in the *E. coli* genome, making it a logical choice for codon reassignment.[58] Fewer reassigned codons in the proteome lead to fewer missense mutations that may affect cell health, and typically codons that are used less frequently will have fewer cellular factors to compete with.

Recoding of the stop codon allows for the incorporation of a 21st amino acid into proteins translationally. Natural stop codon suppressors are found in organisms from all domains of life. Mutations to the anticodon of a tRNA allow base pairing with a stop codon in the ribosome A site. Most suppressor tRNAs are those requiring only a single mutation in the anticodon to provide recognition of a stop codon. The resulting protein then contains the natural amino acid the suppressor tRNA was aminoacylated with. These natural systems for stop codon suppression along with the substrate permissivity of aaRSs served as inspiration for the first report of engineered genetic code expansion in response to stop codons *in vivo*. Yeast use a modified Phe $tRNA_{CUA}$ and the canonical phenylalanyl tRNA synthetase (PhrRS) to suppress amber stop codons (UAG) with Phe residues. Furter combined the PheRS and Phe-$tRNA_{CUA}$ from yeast

into a special *E. coli* strain that contained a mutant PheRS[56]. It was already known that the yeast PheRS would aminoacylate the ncAA para-fluoro-phenylalanine(pFPhe)[71], and that the mutant *E. coli* PheRS would not.[72] Using site directed mutagenesis, an amber stop codon was introduced in the reading frame of a reporter protein in place of a sense codon. When pFPhe was introduced into the growth media, the stop codon in the reporter protein was suppressed with approximately 65% efficiency based on protein yields[56]. Although yields of protein were low, and there were high levels of natural amino acid (Phe) contamination at the desired location, nonsense codon suppression (NCS)technology was realized as a promising technique and has become the predominant method for incorporating ncAA into proteins site-specifically. Since the first report, several other aaRS/tRNA pairs have been engineered to aminoacylate over 150 ncAAs.[73]. Aminoacyl-tRNA synthetases and their cognate tRNAs have also been engineered to increase their orthogonality to the endogenous aaRSs and tRNAs of model organisms.[74–76] The Schultz lab discovered that the Tyrosyl tRNA synthetase (TyrRS) and the cognate tyr-tRNA from the methanogenic archaea *Methanocaldococcus jannaschii* (MJ) is orthogonal to the set of aaRS and tRNA in *E. coli*.[77] The *M. jannaschii* TyrRS does not aminoacylate any of the tRNA species in *E. coli* and the tyr-tRNA from *M. jannaschii* was not aminoacylated by any of the 20 *E. coli* aaRSs. Additionally, it was discovered that unlike many aaRSs, the MJ TyrRS does not rely heavily on the anticodon of its cognate tRNA as an identity element.[55] When the anticodon of MJ tyr-tRNA is changed from GUA to CUA *in vitro* aminoacylation decreases 15 fold.[78] Changing the anticodon of the Tyr-tRNA to CUA *in vivo* resulted in aminoacyl-$tRNA_{CUA}$ that suppressed amber stop codons in *E. coli*.

In order to incorporate ncAA in response to the stop codon, the amino acid binding pocket of the Mj aaRS was engineered to aminoacylate the ncAA L-3-(2-napthyl)alanine (NapA) and not to accept tyrosine.[79] A library of aaRS variants with diversity at the residues lining the amino acid binding pocket in MJ TyrRS was screened using a selection system that depends on the suppression of a stop codon in an essential gene.[69] Orthogonal aaRS/tRNA pairs have now been engineered to suppress each of the three nonsense codons which encode stop signals as well as decoding quadruplet codons as sense codons

instead of resulting in frame shifts[73,80–82]. The Mj Tyr-tRNA has been engineered by to try to increase suppression efficiency, and decrease the interactions of the tRNA with endogenous translation machinery. There is some evidence of minor interactions between the MJ TyrRS/tRNA pair and *E. coli* proline tRNA and aaRS. Most work with the Mj Tyr-tRNA has focused on mutating the identity elements in the acceptor stem and the T arm to find optimized variants.[78,83] One particular variant($tRNA^{Opt}_{CUA}$) has been widely used for nonsense suppression. Over 100 ncAA have been incorporated using the Mj aaRS/tRNA pair. Several other aaRS/tRNA pairs are also being developed for ncAA incorporation, most notably the pyrolysyl pair from *Methanosarcina barkeri* (MB) or *Methanosarcina mazei*.[73,84] The MB pair is of particular interest as the anticodon of the tRNA does not interact with the aaRS, which means that engineering the anticodon stem loop should have only limited effects on aminoacylation efficiency.[85,86]

Suppression of nonsense codons allows for the incorporation of a ncAA at a particular location in a protein and has enabled the design of specific protein interactions and protein labeling schemes. Nonsense suppression is somewhat limited by the low relative efficiency of the orthogonal components, and the competition of suppressor tRNAs with release factors. Release factor 1 (RF1) binds in the A site of the ribosome complementary to the UAG and UAA stop codons.[87] Attempting to incorporate an ncAA at more than one stop codon in a protein typically results in truncated proteins and low yields of incorporated protein in general because of competition with release factors. Class 1 release factors bind to stop codons in the A site of the ribosome and lead to hydrolysis of the peptide chain and disassociation of the ribosome. Efficient nonsense suppression relies on an aminoacyl suppressor tRNA complexed with EF-Tu outcompeting release factors to decode the stop codon. The Chatterjee group measured the efficiency of suppressing multiple stop codons in *E.* coli strains and found that suppression of a second stop codon decreases protein yield by 60%, and four or more stop codons results in 99% decrease in protein yield.[88] Three different groups have recently reported different approaches to remove RF1 from bacterial cells to eliminate competition between an orthogonal suppressor tRNA and the release factor. The Wang lab reported mutations to release factor 2 (RF2) that increased its efficiency to terminate UAA

codons and restore fitness to cells when RF1 was knocked out.[89] Although the cells were able to grow with the modified RF2 in the absence of RF1, the suppression of UAG codons in the genome could lead to issues from proteins that are not terminated appropriately. Issacs *et al* reported a strategy to completely free the amber sense codon from any meaning in the genetic code by recoding the 314 naturally occurring UAG codons in the *E. coli* genome with alternate stop codons. After tour de force genome re-engineering project, RF1 was knocked out of cells to eliminate competition for the orthogonal suppressor tRNA.[90] However, even with massive undertakings to eliminate termination pressure at the amber stop codon—more than 5 research groups were collaboratively involved in the refactoring of the *E. coli* genome—there still exists termination pressure and recent reports suggest that when trying to suppress more than a few amber stop codons in the same gene leads to significant loss of efficiency.[88,91] In addition to the apparent termination pressure at nonsense codons, the expansion of the genetic code to more than 21 amino acids will require additional codons. With the removal of RF1, it is unlikely that suppression of the other two stop codons will be efficient. Recent work has focused on the reassignment of sense codons to expand the genetic code.

Sense Codon Reassignment: Breaking the Degeneracy of the Genetic Code

In a triplet codon translation system using 4 nucleobases, there are 64 possible codons. Three of these codons are recognized by *E. coli* as stop signals and do not typically code for amino acids. Some organisms, including some strains of *E. coli* have evolved suppressor mechanisms that decode one of the three stop codons as a sense codon, typically incorporating one of the 20 canonical amino acids. The remaining 61 sense codons are decoded by approximately 46 tRNA species in *E. coli*, with an average of two tRNA species per amino acid.[92] The majority of the tRNA species decode multiple mRNA codons which leads to one element of degeneracy in the genetic code; multiple codons encode the same amino acid. The deciphering of multiple codon sequences by a single tRNA species is accomplished using non-Watson-Crick base-pairing interactions in the third position of the mRNA codon, where non-typical hydrogen bond donors and acceptors interact.[93] The guanosine-uridine wobble pair is the most common

non-Watson-Crick base pair found in translation and has a Gibbs free energy of melting ($\Delta G°_{37}$) of 4.1 kcal/mol[94]. In contrast, canonical guanosine-cytosine base pairs and adenosine-uracil base pairs have $\Delta G°_{37}$ of 6.5 and 6.3 kcal/mol, respectively.[94]

The Tirrell group reported reassigning one of the two phenylalanine codons in *E. coli* to direct the incorporation of the ncAA NapA.[95] In *E. coli* Phe is incorporated into proteins by one tRNA with a GAA anticodon sequence which recognizes the mRNA codons UUC and UUU. The UUC codon is decoded with three canonical base pairs and the UUU codon is decoded with two Watson-Crick pairs and a G-U wobble pairing. By supplying an orthogonal aaRS engineered to aminoacylate NapA onto its cognate tRNA, and by mutating the tRNA to have an AAA anticodon which base pairs in a canonical fashion with the UUU Phe codon, the ncAA was incorporated at UUU codons while leaving the UUC codon to be decoded by the natural Phe-tRNA$_{GAA}$ (Figure 1.5). The canonical base pairing interactions between A-U and C-G bases have Gibbs free energies of melting ($\Delta G°_{37}$) of 6.3 and 6.5 kcal/mol respectively. The G-U wobble interaction has a $\Delta G°_{37}$ of only 4.1 kcal/mol[94]. Using the relation between Gibbs free energy and equilibrium constants for these base pairing reactions at 37°, changing from a G-U basepair to an A-U basepair produces an increase in binding affinity for this specific nucleotide position which translates to a 35-fold greater k_a. When Kwon and co-workers introduced an AAA anticodon to compete against the endogenous AAG anticodon, they reported near quantitative incorporation of a ncAA in response to multiple UUU codons in the same reporter protein.[95] Sense codon reassignment combines the suite of promiscuous and engineered aaRS/tRNA pairs that have been developed for global amino acid replacement and nonsense suppression techniques with the directed anticodon mutations of stop codon reassignment.

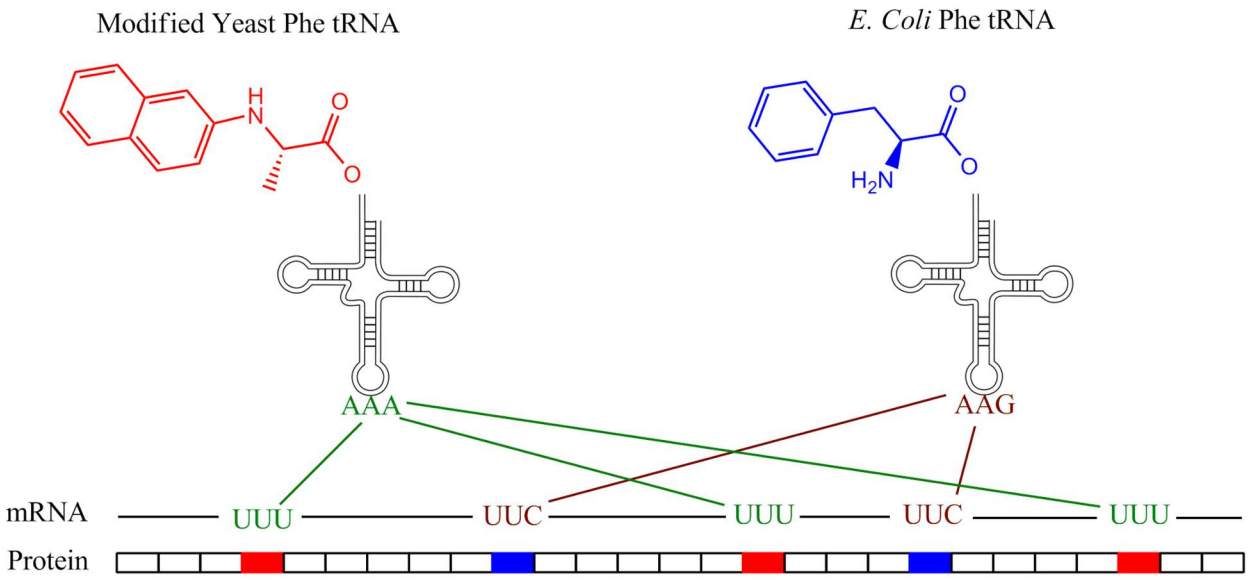


Figure 1.5 Breaking the Degeneracy of the Genetic Code. Supplying a ncAA-tRNA with an anticodon with greater binding affinity for a codon typically read through wobble interactions leads to incorporation of a ncAA in a site specific manner while maintaining the ability to code for all 20 proteinogenic interactions. ***Adapted from reference*** **[95].**

Breaking the degeneracy of the genetic code following the method of Tirrell avoids the problems of lower protein yields with suppression based methods and the elimination of natural amino acids as in the global amino acid replacement technique. Because a sense codon can be decoded by either an endogenous tRNA bearing a natural amino acid or by an engineered tRNA, the proteins produced in sense codon reassignment systems have some probability of mixed amino acids at reassignment positions; a single codon can direct incorporation of two amino acids. Improving the efficiency of aminoacylation of the engineered tRNA and increasing the overall concentration of the aminoacyl-tRNA for reassignment in the intercellular pool can bias competition with endogenous tRNAs to favor incorporation of the ncAA. Compared to either NCS or GAAR, little work has been done to evaluate sense codon reassignment using exogenous aaRS/tRNA pairs and still only a few reports describe reassigning sense codons *in vivo* to incorporate ncAAs.[96–98] Most of the reports of sense codon reassignment to date have focused on reassigning rare codons in the genome which will likely have aminoacyl-tRNA concentrations that are lower and also produce fewer mutations in the proteome that could lead to unhealthy cells.

This dissertation focuses on methods for sense codon reassignment that measure and increase the efficiency of reassignment of multiple sense codons. The variable which control to what extent a sense codon may be reassigned to an alternate amino acid are not well understood. We assumed that sense codon reassignment efficiency (the frequency with which a specific codon is successfully reassigned) would depend largely on the concentration of the competing endogenous tRNA and the efficiency of aminoacylation reaction for the engineered aaRS/tRNA pair.

OTHER METHODS FOR INCREASING CHEMICAL DIVERSITY OF PROTEINS

Solid Phase Synthesis

Several technologies that do not rely on *in vivo* translation have been used to introduce extra chemical functionality into proteins, and although less relevant to this dissertation the limits and benefits of some popular methods are described here for comparison and reference. Solid phase peptide synthesis (SPPS) has been used extensively to build peptides with defined primary sequences, and because of the nature of the reaction any amino acid that can be protected can be incorporated into peptides.[99] In a typical SPPS reaction, the first amino acid residue in the peptide is covalently linked to a solid support by making an ester linkage with the carboxylic acid of the amino acid. All of the reactive functional groups in the amino acids must be protected—amines are usually protected as Fmoc carbamates, carboxyl residues are usually protected as tert-butyl esters—except for the carboxylic acid of the residue to be coupled next. After covalently linking the first residue to the solid support, the Fmoc protecting group is removed from the first residue, and an Fmoc N-protected residue is added to the reaction. The carboxylic acid on the second residue is activated and reacts with the unprotected amine on the first residue to produce a natural peptide bond. Free unreacted amino acid and activating agents are washed away from the immobilized support, the N-terminus of the peptide is deprotected, and a third amino acid (with reactive carboxylic acid) is added to the solution. This process is repeated as many times as necessary to produce the desired peptide. Because of the reaction process, any amino acid, natural or non-canonical

can be incorporated into peptides. Due to less than 100% efficiency, SPPS can produce side products that require purification steps, and SPPS is limited in the number of amino acids that can be incorporated into peptides. Due to the cyclic nature of this reaction, although the efficiency of each coupling step is high, the overall efficiency is amplified exponentially across all the coupling steps. The necessary purification to get the final peptide product isolated from side products becomes logistically challenging for peptides with more than about 50 residues. The overall yield of peptides suffers greatly as well, assuming a 99% efficiency for each coupling, a 50-mer peptide would have an overall yield of 60%.

Techniques like native chemical ligation allow for connection of multiple smaller peptides, to make polypeptides the length of proteins[100], but the ligation must occur at cysteine residues. Native chemical ligation reacts a non-natural thioester from the C terminus of the first peptide with an N-terminal cysteine residue from the second peptide. A thioester exchange reaction initially links the two peptides covalently using a non-peptide bond. Near pH 7, the free amine from peptide second peptide attacks the carbonyl of the thioester from the c-terminus of the first peptide and an S,N acyl chain shift results, restoring the peptide backbone and leaving a cysteine residue at the position of the covalent link between the peptides. Some proteins will not have cysteine residues in convenient locations in the primary sequence for NCL, and alterations to the primary sequence may be necessary to use native chemical ligation. Some proteins may not be able to tolerate the addition of cysteine residues.

Cell Free Translation Systems

Cell free systems for *in vitro* translation can be used to expand the genetic code. Specific translation components can be included or omitted from *in vitro* systems which provides some control over translation. For example, removing an endogenous tRNA from the *in* vitro reaction may decrease the competition an exogenous tRNA may encounter when trying to decode a codon.[101] Unfortunately, the more controlled and specialized an *in vitro* translation experiment becomes, the more expensive, time consuming and fragile it becomes as well. The PURE (protein synthesis using recombinant elements) system reported by Shimuzu *et al* is a finely tunable system, where each of 32 necessary components for

translation are purified and then added separately to the translation reaction.[102] Most cell free translation reactions work by producing crude cell lysates, removing contaminating mRNA and then adding additional components (new mRNA, orthogonal translation systems) back into the lysate.[103] The PURE system produces complete control over the genetic code, but the cost to purify all the components is prohibitively expensive for most research laboratories.

While both SPPS and *in vitro* translation allow a wide variety of amino acids to be included in peptides, polypeptides, and proteins, there are additional drawbacks to these methods. Frequently *in vitro* produced proteins have folding and solubility issues[16], and *in vitro* synthesized proteins can only be used to study biochemistry *in vivo* or *in situ*. Additionally, the yield of *in vitro* methods for producing peptides and proteins is limited and cannot compete logistically with *in vivo* techniques. However, the expression of proteins that would be toxic to a host, or issues of undesired reassignment of codons in the proteome of the host become non-issues with *in vitro* systems.

QUANTIFYING GENETIC CODE EXPANSION

While interest in genetic code expansion technologies has burgeoned over the last two decades, methods to quantitate the levels of efficiency of codon reassignment systems have not been well developed. Most reports of GCE in the literature do not provide quantitative measurement of efficiency and present qualitative evidence of the incorporation of the ncAA as the only measurement. In the case of nonsense suppression, most reports assume that there are two options for a GCE experiment: either the desired ncAA is incorporated and a full-length protein will be produced or the ncAA will not be incorporated, resulting in a truncated or frameshifted (and likely truncated) protein. However, a likely third option is frequently observed, the suppression of a nonsense codon with a natural amino acid (several reports have shown that engineered aaRS have residual activity towards the natural amino acid they were derived from as well as other amino acids usually at lower levels).[7] Engineering orthogonal translation system components results in multiple cellular interactions being affected. In an ideal experiment, the magnitude of the changes of aminoacylation efficiency, expression levels of the

orthogonal aaRS and tRNA, the altered interaction between the aminoacyl-tRNA and the EF-Tu, and the change in affinity for the altered anticodon and the mRNA codon would all be measured. The determination of all of the changes made in an *in vivo* system simultaneously are not practical, and even when working in *in vitro* systems interactions between translation components must effectively be measured one at a time. More holistic approaches to the quantitation of orthogonal translation will enable comparisons of systems and provide insight towards the reactions which may provide the most control and opportunity for increasing genetic code expansion efficiencies. For nonsense suppression experiments, comparing the protein yield of a reporter protein produced in both the presence and absence of a ncAA gives an estimate of how efficient the entire suppression system is.[7,88] The relative efficiency of a system when changes to the amino acid binding pocket in the aaRS, or mutations in the tRNA molecule are made could be measured *in vitro*, but rarely are. When new ncAA specificity is engineered into aaRS, rarely is the activity of the new aaRS towards the new ncAA compared to other quantified systems. When efficiency is evaluated in reports of GCE it is frequently qualitative (suppression of a nonsense codon in a gene for a resistance marker results in colonies that are capable of growing on a given antibiotic).[104]

The problem of quantitation becomes even more difficult when trying to measure sense codon reassignment systems. Unlike nonsense suppression, the null result of a sense codon reassignment experiment is a full-length protein meaning that production of a protein cannot be used as an indicator of codon reassignment. In the limited reports of sense codon reassignment in the literature, efficiency of reassigning a sense codon has been accomplished either by mass spectral characterization or by comparative Western Blot responses. When the ncAA to be incorporated installs a reactive handle in the protein, sense codon reassignment efficiency can be measured by comparing the amount of a reactive label (a fluorescent probe, a radioisotope, or a reactive antibody coupled to an enzymatic reporter) in a protein sample to the total amount of protein present.[98] Using antibodies as reporters for sense codon reassignment presents challenges in comparing identical concentrations of proteins, and having linear

signals from western blot probes. When using mass spectrometry to identify and attempt to quantify the incorporation of ncAAs in peptides, electrospray ionization mass spectrometry (ESI-MS) or matrix assisted laser desorption ionization mass spectrometry (MALDI-MS) are typically used as these methods have high mass resolution and are relatively cheap. However, as changes in peptide identity can interfere with ionization efficiency, these methods are not quantitative unless heavy isotope internal standards are used as well. Further, the sensitivity of these methods can be low enough that small amounts of protein or peptide with masses reasonably close to the target protein can be "hidden" under the curve of the target protein. Many reports rely on mass spectral results to quantitate the level of reassignment in a GCE experiment, but without expensive internal standards for the proteins being produced, or isotopically labeled standards, mass spectrometry by itself is barely quantitative.[105,106] Coupling the separation of digested peptides using liquid chromatography followed by tandem mass spectrometry (LC-MS/MS) can provide high level resolution of incorporated amino acids in peptide fragments, including primary sequence of the peptides. With enough of a protein sample, quantitation of peptide fragments could be achieved using UV spectroscopy after LC separation, but this requires complex equipment setups and an impractically large concentration of peptides. Instead, quantitation using LC-MS/MS uses radioisotopically labeled standards of a protein as internal standards from which response curves can be generated.[107,108]

The second chapter of this dissertation will describe in depth a fluorescence based screen that enabled quantitation of sense codon reassignment at any non-tyrosine codon using the frequently used MJ TyrRS/tRNA pair.

DISSERTATION OUTLINE

In this dissertation three complete stories regarding sense codon reassignment will be described, all of which revolve around a quantitative screen described in detail in Chapter 2. The green fluorescent protein based screen that was developed enabled quantitative measurement of the reassignment of several codons *in* vivo as well as how well a given anticodon discriminates between two codons. The screen was

then used to improve sense codon reassignment at a specific lysine codon using directed evolution and fluorescence activated cell sorting.

In attempting to reassign histidine codons, anomalies in codon discrimination were discovered. Chapter 3 of this dissertation deals with how an tRNA that may be orthogonal in regard to a set of endogenous aaRS can still be a substrate for other host enzymes. Specifically in this case, engineering the anticodon of the MJ Tyr-tRNA to be AUG resulted in the modification of the tRNA by the enzyme tadA. The GFP based screen was used to measure the effects of this modification, as well as to screen directed evolution libraries for tRNAs with altered sequences that would not be modified by tadA but still be good substrates for the aaRS.

The third story of sense codon reassignment evaluates how directed evolution can be used to improve and enable ncAA incorporation in response to sense codons, particularly the AGG arginine codon in *E. coli*. One important question for GCE that has received insufficient attention, is to what extent can improvements made in one system for GCE be transferred to other orthogonal translation systems. For example, mutations in an aaRS sequence that increase sense codon reassignment for a specific amino acid may or may not improve sense codon reassignment for other amino acids based on that same aaRS. Over 100 different ncAAs have been incorporated into proteins using variants of the *M. jannaschii* TyrRS/tRNA pair[73], and the mutations to allow this diverse substrate scope revolve around changes in only a few residues in the binding pocket for the specific ncAA. To increase sense codon reassignment efficiency, modifications were focused on regions outside of the amino acid binding pocket that interact with the anticodon. It is unclear how or if mutations that increase sense codon reassignment for one synthetase variant will transfer to other variants of the aaRS that activate other ncAAs. Using the fluorescence screen for TyrRS sense codon reassignment activity and also by quantifying the incorporation of the ncAA para-azidophenylalanine (pAzF) into proteins the transferability of orthogonal translation system modifications was evaluated.

Sense codon reassignment has great potential as a method for genetic code expansion, both in industrial improvements to proteins with unique chemical capabilities, but also as a tool for understanding the biochemistry of the translation process.

REFERENCES

(1) Lundquist, J. J.; Toone, E. J. The Cluster Glycoside Effect. *Chemical reviews* **2002**, *102*, 555–578.

(2) Mammen, M.; Choi, S.-K.; Whitesides, G. M. Polyvalent Interactions in Biological Systems: Implications for Design and Use of Multivalent Ligands and Inhibitors. *Angewandte Chemie International Edition* **1998**, *37*, 2754–2794.

(3) Dimick, S. M.; Powell, S. C.; McMahon, S. A.; Moothoo, D. N.; Naismith, J. H.; Toone, E. J. On the Meaning of Affinity: Cluster Glycoside Effects and Concanavalin A. *Journal of the American Chemical Society* **1999**, *121*, 10286–10296.

(4) Springsteen, G.; Wang, B. A Detailed Examination of Boronic Acid-Diol Complexation. *Tetrahedron* **2002**, *58*, 5291–5300.

(5) Jin, S.; Cheng, Y.; Reid, S.; Li, M.; Wang, B. Carbohydrate Recognition by Boronolectins, Small Molecules, and Lectins. *Medicinal research reviews* **2010**, *30*, 171–257.

(6) Princiotta, M. F.; Finzi, D.; Qian, S.-B.; Gibbs, J.; Schuchmann, S.; Buttgereit, F.; Bennink, J. R.; Yewdell, J. W. Quantitating Protein Synthesis, Degradation, and Endogenous Antigen Processing. *Immunity* **2003**, *18*, 343–354.

(7) Young, T. S.; Ahmad, I.; Yin, J. A.; Schultz, P. G. An Enhanced System for Unnatural Amino Acid Mutagenesis in E. Coli. *Journal of molecular biology* **2010**, *395*, 361–374.

(8) Ryu, Y.; Schultz, P. G. Efficient Incorporation of Unnatural Amino Acids into Proteins in Escherichia Coli. *Nature methods* **2006**, *3*, 263–265.

(9) Plotkin, J. B.; Kudla, G. Synonymous but Not the Same: The Causes and Consequences of Codon Bias. *Nature Reviews Genetics* **2010**, *12*, 32–42.

(10) Sharp, P. M.; Li, W.-H. Codon Usage in Regulatory Genes in Escherichia Coli Does Not Reflect Selection for “rare”codons. *Nucleic Acids Research* **1986**, *14*, 7737–7749.

(11) Elf, J.; Nilsson, D.; Tenson, T.; Ehrenberg, M. Selective Charging of tRNA Isoacceptors Explains Patterns of Codon Usage. *Science* **2003**, *300*, 1718–1722.

(12) Benner, S. A. Redesigning Genetics. *Science* **2004**, *306*, 625–626.

(13) Henry, A. A.; Romesberg, F. E. Beyond A, C, G and T: Augmenting Nature’s Alphabet. *Current opinion in chemical biology* **2003**, *7*, 727–733.

(14) Kool, E. T. Synthetically Modified DNAs as Substrates for Polymerases. *Current opinion in chemical biology* **2000**, *4*, 602–608.

(15) Gloge, F.; Becker, A. H.; Kramer, G.; Bukau, B. Co-Translational Mechanisms of Protein Maturation. *Current opinion in structural biology* **2014**, *24*, 24–33.

(16) Hingorani, K. S.; Gierasch, L. M. Comparing Protein Folding in Vitro and in Vivo: Foldability Meets the Fitness Challenge. *Current opinion in structural biology* **2014**, *24*, 81–90.

(17) Sivertsson, E. M.; Itzhaki, L. S. Protein Folding: When Ribosomes Pick the Structure. *Nat Chem* **2014**, *6*, 378–379.

(18) Bremer, H.; Dennis, P. P.; others. Modulation of Chemical Composition and Other Parameters of the Cell by Growth Rate. *Escherichia coli and Salmonella: cellular and molecular biology* **1996**, *2*, 1553–1569.

(19) Milo, R. What Is the Total Number of Protein Molecules per Cell Volume? A Call to Rethink Some Published Values. *Bioessays* **2013**, *35*, 1050–1055.

(20) Ruan, B.; Palioura, S.; Sabina, J.; Marvin-Guy, L.; Kochhar, S.; LaRossa, R. A.; Söll, D. Quality Control despite Mistranslation Caused by an Ambiguous Genetic Code. *Proceedings of the National Academy of Sciences* **2008**, *105*, 16502–16507.

(21) Min, B.; Kitabatake, M.; Polycarpo, C.; Pelaschier, J.; Raczniak, G.; Ruan, B.; Kobayashi, H.; Namgoong, S.; Söll, D. Protein Synthesis in Escherichia Coli with Mischarged tRNA. *Journal of bacteriology* **2003**, *185*, 3524–3526.

(22) Lind, P. A.; Berg, O. G.; Andersson, D. I. Mutational Robustness of Ribosomal Protein Genes. *Science* **2010**, *330*, 825–827.

(23) Chan, P. P.; Lowe, T. M. GtRNAdb: A Database of Transfer RNA Genes Detected in Genomic Sequence. *Nucleic acids research* **2009**, *37*, D93–D97.

(24) Agris, P. F. The Importance of Being Modified: Roles of Modified Nucleosides and Mg 2+ in RNA Structure and Function. *Progress in nucleic acid research and molecular biology* **1996**, *53*, 79–129.

(25) Konevega, A. L.; Soboleva, N. G.; Makhno, V. I.; Semenkov, Y. P.; Wintermeyer, W.; Rodnina, M. V.; Katunin, V. I. Purine Bases at Position 37 of tRNA Stabilize Codon-Anticodon Interaction in the Ribosomal A Site by Stacking and Mg2+-Dependent Interactions. *Rna* **2004**, *10*, 90–101.

(26) Björk, G. R.; Durand, J.; Hagervall, T. G.; Lundgren, H. K.; Nilsson, K.; Chen, P.; Qian, Q.; Urbonavi\vcius, J.; others. Transfer RNA Modification: Influence on Translational

Frameshifting and Metabolism. *FEBS letters* **1999**, *452*, 47–51.

(27) Urbonavi\vcius, J.; Qian, Q.; Durand, J. M. B.; Hagervall, T. G.; Björk, G. R. Improvement of Reading Frame Maintenance Is a Common Function for Several tRNA Modifications. *The EMBO journal* **2001**, *20*, 4863–4873.

(28) Lim, V. I. Analysis of Interactions between the Codon-Anticodon Duplexes within the Ribosome: Their Role in Translation. *J. Mol. Biol.* **1997**, *266*, 877–890.

(29) Wolf, J.; Gerber, A. P.; Keller, W. tadA, an Essential tRNA-Specific Adenosine Deaminase from Escherichia Coli. *The EMBO journal* **2002**, *21*, 3841–3851.

(30) Söll, D.; RajBhandary, U. L. *tRNA: Structure, Biosynthesis, and Function*; Zondervan, 1995.

(31) Senger, B.; Aphasizhev, R.; Walter, P.; Fasiolo, F. The Presence of a D-Stem but Not a T-Stem Is Essential for Triggering Aminoacylation upon Anticodon Binding in Yeast Methionine tRNA. *Journal of molecular biology* **1995**, *249*, 45–58.

(32) Romby, P.; Moras, D.; Bergdoll, M.; Dumas, P.; Vlassov, V. V.; Westhof, E.; Ebel, J. P.; Giegé, R. Yeast tRNAAsp Tertiary Structure in Solution and Areas of Interaction of the tRNA with Aspartyl-tRNA Synthetase: A Comparative Study of the Yeast Phenylalanine System by Phosphate Alkylation Experiments with Ethylnitrosourea. *Journal of molecular biology* **1985**, *184*, 455–471.

(33) McClain, W. H. Rules That Govern tRNA Identity in Protein Synthesis. *Journal of molecular biology* **1993**, *234*, 257–280.

(34) Yarus, M. tRNA Identity: A Hair of the Dogma That Bit Us. *Cell* **1988**, *55*, 739–741.

(35) Söll, D.; RajBhandary, U. L. *tRNA: Structure, Biosynthesis, and Function*; Zondervan, 1995.

(36) Bouadloun, F.; Donner, D.; Kurland, C. G. Codon-Specific Missense Errors in Vivo. *The EMBO journal* **1983**, *2*, 1351–1356.

(37) Rodnina, M. V.; Wintermeyer, W. Fidelity of Aminoacyl-tRNA Selection on the Ribosome: Kinetic and Structural Mechanisms. *Annual review of biochemistry* **2001**, *70*, 415–435.

(38) Kramer, E. B.; Farabaugh, P. J. The Frequency of Translational Misreading Errors in E. Coli Is Largely Determined by tRNA Competition. *RNA* **2007**, *13*, 87–96.

(39) Pezo, V.; Metzgar, D.; Hendrickson, T.; Waas, W.; Hazebrouck, S.; Döring, V.; Marliere, P.; Schimmel, P.; De Crécy-Lagard, V. Artificially Ambiguous Genetic Code Confers Growth Yield Advantage. *Proceedings of the National Academy of Sciences of*

the United States of America **2004**, *101*, 8593.

(40) Nureki, O.; Vassylyev, D. G.; Tateno, M.; Shimada, A.; Nakama, T.; Fukai, S.; Konno, M.; Hendrickson, T. L.; Schimmel, P.; Yokoyama, S. Enzyme Structure with Two Catalytic Sites for Double-Sieve Selection of Substrate. *Science* **1998**, *280*, 578–582.

(41) Hendrickson, T. L.; Nomanbhoy, T. K.; de Crécy-Lagard, V.; Fukai, S.; Nureki, O.; Yokoyama, S.; Schimmel, P. Mutational Separation of Two Pathways for Editing by a Class I tRNA Synthetase. *Molecular cell* **2002**, *9*, 353–362.

(42) Lin, L.; Schimmel, P. Mutational Analysis Suggests the Same Design for Editing Activities of Two tRNA Synthetases. *Biochemistry* **1996**, *35*, 5596–5601.

(43) Bullwinkle, T. J.; Ibba, M. Translation Quality Control Is Critical for Bacterial Responses to Amino Acid Stress. *Proceedings of the National Academy of Sciences* **2016**, *113*, 2252–2257.

(44) Shepherd, J.; Ibba, M. Bacterial Transfer RNAs. *FEMS microbiology reviews* **2015**, *39*, 280–300.

(45) Cohen, G.; Cowie, D.; others. [Total Replacement of Methionine by Selenomethionine in the Proteins of Escherichia Coli.]. *Comptes rendus hebdomadaires des seances de l'Academie des sciences* **1957**, *244*, 680.

(46) Kiick, K.; Weberskirch, R.; Tirrell, D. Identification of an Expanded Set of Translationally Active Methionine Analogues in Escherichia Coli. *Febs Letters* **2001**, *502*, 25–30.

(47) Van Hest, J. C. M.; Kiick, K. L.; Tirrell, D. A. Efficient Incorporation of Unsaturated Methionine Analogues into Proteins in Vivo. *Journal of the American Chemical Society* **2000**, *122*, 1282–1288.

(48) Park, H.-S.; Hohn, M. J.; Umehara, T.; Guo, L.-T.; Osborne, E. M.; Benner, J.; Noren, C. J.; Rinehart, J.; Söll, D. Expanding the Genetic Code of Escherichia Coli with Phosphoserine. *Science* **2011**, *333*, 1151–1154.

(49) Fan, C.; Ip, K.; Söll, D. Expanding the Genetic Code of Escherichia Coli with Phosphotyrosine. *FEBS letters* **2016**, *590*, 3040–3047.

(50) Young, R.; Bremer, H. Polypeptide-Chain-Elongation Rate in Escherichia Coli B/r as a Function of Growth Rate. *Biochemical Journal* **1976**, *160*, 185–194.

(51) Wagner, I.; Musso, H. New Naturally Occurring Amino Acids. *Angewandte Chemie International Edition in English* **1983**, *22*, 816–828.

(52) Cowie, D. B.; Cohen, G. N.; Bolton, E. T.; De Robichon-Szulmajster, H. Amino Acid Analog Incorporation into Bacterial Proteins. *Biochimica et biophysica acta* **1959**, *34*, 39–46.

(53) Danenberg, P. V.; Heidelberger, C. The Effect of Raney Nickel on the Covalent Thymidylate Synthetase-5-Fluoro-2'-Deoxyuridylate-5, 10-Methylenetetrahydrofolate Complex. *Biochemistry* **1976**, *15*, 1331–1337.

(54) Robertson, S. A.; Ellman, J. A.; Schultz, P. G. A General and Efficient Route for Chemical Aminoacylation of Transfer RNAs. *Journal of the American Chemical Society* **1991**, *113*, 2722–2729.

(55) Steer, B. A.; Schimmel, P. Major Anticodon-Binding Region Missing from an Archaebacterial tRNA Synthetase. *Journal of Biological Chemistry* **1999**, *274*, 35601–35606.

(56) Furter, R. Expansion of the Genetic Code: Site-Directed P-Fluoro-Phenylalanine Incorporation in Escherichia Coli. *Protein science* **1998**, *7*, 419–426.

(57) Kiick, K. L.; Tirrell, D. A. Protein Engineering by in Vivo Incorporation of Non-Natural Amino Acids: Control of Incorporation of Methionine Analogues by Methionyl-tRNA Synthetase. *Tetrahedron* **2000**, *56*, 9487–9493.

(58) Maloy, S. R.; Stewart, V. J.; Taylor, R. K.; others. *Genetic Analysis of Pathogenic Bacteria: A Laboratory Manual.*; Cold Spring Harbor Laboratory Press., 1996.

(59) Yang, S.; Hendrickson, W. A.; Crouch, R. J.; Satow, Y. Structure of Ribonuclease H Phased at 2 Angstrom Resolution by MAD Analysis of the Selenomethionyl Protein. *Science* **1990**, *249*, 1398–1406.

(60) Glenn, W. S.; Stone, S. E.; Ho, S. H.; Sweredoski, M. J.; Moradian, A.; Hess, S.; Bailey-Serres, J.; Tirrell, D. A. Bioorthogonal Noncanonical Amino Acid Tagging (BONCAT) Enables Time-Resolved Analysis of Protein Synthesis in Native Plant Tissue. *Plant physiology* **2017**, *173*, 1543–1553.

(61) Link, A. J.; Vink, M. K. S.; Tirrell, D. A. Preparation of the Functionalizable Methionine Surrogate Azidohomoalanine via Copper-Catalyzed Diazo Transfer. *Nat Protoc* **2007**, *2*, 1879–1883.

(62) Tanrikulu, I. C.; Schmitt, E.; Mechulam, Y.; Goddard, W. A.; Tirrell, D. A. Discovery of Escherichia Coli Methionyl-tRNA Synthetase Mutants for Efficient Labeling of Proteins with Azidonorleucine in Vivo. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 15285–15290.

(63) Dieterich, D. C.; Link, A. J.; Graumann, J.; Tirrell, D. A.; Schuman, E. M. Selective Identification of Newly Synthesized Proteins in Mammalian Cells Using Bioorthogonal Noncanonical Amino Acid Tagging (BONCAT). *Proceedings of the National Academy*

of Sciences **2006**, *103*, 9482.

(64) MISSING:pmid:20572253. MISSING:pmid:20572253. **2017**.

(65) Merkel, L.; Schauer, M.; Antranikian, G.; Budisa, N. Parallel Incorporation of Different Fluorinated Amino Acids: On the Way to "teflon" Proteins. *ChemBioChem* **2010**, *11*, 1505–1507.

(66) Salwiczek, M.; Nyakatura, E. K.; Gerling, U. I.; Ye, S.; Koksch, B. Fluorinated Amino Acids: Compatibility with Native Protein Structures and Effects on Protein-Protein Interactions. *Chemical Society Reviews* **2012**, *41*, 2135–2171.

(67) Giegé, R. Genetic Code Expansion. *Nature Structural \& Molecular Biology* **2003**, *10*, 414–416.

(68) Liu, D. R.; Schultz, P. G. Progress toward the Evolution of an Organism with an Expanded Genetic Code. *Proceedings of the National Academy of Sciences* **1999**, *96*, 4780.

(69) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. Expanding the Genetic Code of Escherichia Coli. *Science* **2001**, *292*, 498.

(70) Zavialov, A. V.; Mora, L.; Buckingham, R. H.; Ehrenberg, M. Release of Peptide Promoted by the GGQ Motif of Class 1 Release Factors Regulates the GTPase Activity of RF3. *Molecular cell* **2002**, *10*, 789–798.

(71) Cohen, G.; Halvorson, H.; Spiegelman, S. Effects of P-Fluorophenylalanine on the Growth and Physiology of Yeast. *Microsomal Particles and Protein Synthesis* **1958**, *1*, 100–108.

(72) Kast, P.; Hennecke, H. Amino Acid Substrate Specificity of Escherichia Coli Phenylalanyl-tRNA Synthetase Altered by Distinct Mutations. *J. Mol. Biol.* **1991**, *222*, 99–124.

(73) Dumas, A.; Lercher, L.; Spicer, C. D.; Davis, B. G. Designing Logical Codon Reassignment-Expanding the Chemistry in Biology. *Chemical Science* **2015**, *6*, 50–69.

(74) Chatterjee, A.; Xiao, H.; Schultz, P. G. 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*, 14841–14846.

(75) Neumann, H.; Slusarczyk, A. L.; Chin, J. W. De Novo Generation of Mutually Orthogonal Aminoacyl-tRNA synthetase/tRNA Pairs. *J. Am. Chem. Soc.* **2010**, *132*, 2142–2144.

(76) Pastrnak, M.; Magliery, T. J.; Schultz, P. G. A New Orthogonal Suppressor tRNA/aminoacyl-tRNA Synthetase Pair for Evolving an Organism with an Expanded Genetic Code. *Helvetica Chimica Acta* **2000**, *83*, 2277–2286.

(77) Wang, L.; Magliery, T. J.; Liu, D. R.; Schultz, P. G. 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*, 5010–5011.

(78) Fechter, P.; Rudinger-Thirion, J.; Tukalo, M.; Giegé, R. Major Tyrosine Identity Determinants in Methanococcus Jannaschii and Saccharomyces Cerevisiae tRNATyr Are Conserved but Expressed Differently. *European Journal of Biochemistry* **2001**, *268*, 761–767.

(79) Wang, L.; Brock, A.; Schultz, P. G. Adding L-3-(2-Naphthyl) Alanine to the Genetic Code of E. Coli. *Journal of the American Chemical Society* **2002**, *124*, 1836–1837.

(80) Wang, K.; Schmied, W. H.; Chin, J. W. Reprogramming the Genetic Code: From Triplet to Quadruplet Codes. *Angewandte Chemie International Edition* **2012**, *51*, 2288–2297.

(81) O'Donoghue, P.; Prat, L.; Heinemann, I. U.; Ling, J.; Odoi, K.; Liu, W. R.; Söll, D. Near-Cognate Suppression of Amber, Opal and Quadruplet Codons Competes with Aminoacyl-tRNA< Sup> Pyl</sup> for Genetic Code Expansion. *FEBS letters* **2012**.

(82) Anderson, J. C.; Wu, N.; Santoro, S. W.; Lakshman, V.; King, D. S.; Schultz, P. G. An Expanded Genetic Code with a Functional Quadruplet Codon. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 7566–7571.

(83) Guo, J.; Melançon, C. E.; Lee, H. S.; Groff, D.; Schultz, P. G. Evolution of Amber Suppressor tRNAs for Efficient Bacterial Production of Proteins Containing Nonnatural Amino Acids. *Angewandte Chemie International Edition* **2009**, *48*, 9148–9151.

(84) Wan, W.; Tharp, J. M.; Liu, W. R. Pyrrolysyl-tRNA Synthetase: An Ordinary Enzyme but an Outstanding Genetic Code Expansion Tool. *Biochim. Biophys. Acta* **2014**, *1844*, 1059–1070.

(85) Yanagisawa, T.; Ishii, R.; Fukunaga, R.; Kobayashi, T.; Sakamoto, K.; Yokoyama, S. Multistep Engineering of Pyrrolysyl-tRNA Synthetase to Genetically Encode N Ɛ-(O-Azidobenzyloxycarbonyl) Lysine for Site-Specific Protein Modification. *Chemistry \& biology* **2008**, *15*, 1187–1197.

(86) Ambrogelly, A.; Palioura, S.; Söll, D. Natural Expansion of the Genetic Code. *Nature chemical biology* **2007**, *3*, 29–35.

(87) Scolnick, E.; Tompkins, R.; Caskey, T.; Nirenberg, M. Release Factors Differing in Specificity for Terminator Codons. *Proceedings of the National Academy of Sciences*

1968, *61*, 768–774.

(88) Zheng, Y.; Lajoie, M. J.; Italia, J. S.; Chin, M. A.; Church, G. M.; Chatterjee, A. Performance of Optimized Noncanonical Amino Acid Mutagenesis Systems in the Absence of Release Factor 1. *Molecular BioSystems* **2016**, *12*, 1746–1749.

(89) Johnson, D. B. F.; Xu, J.; Shen, Z.; Takimoto, J. K.; Schultz, M. D.; Schmitz, R. J.; Xiang, Z.; Ecker, J. R.; Briggs, S. P.; Wang, L. RF1 Knockout Allows Ribosomal Incorporation of Unnatural Amino Acids at Multiple Sites. *Nature Chemical Biology* **2011**.

(90) Isaacs, F. J.; Carr, P. A.; Wang, H. H.; Lajoie, M. J.; Sterling, B.; Kraal, L.; Tolonen, A. C.; Gianoulis, T. A.; Goodman, D. B.; Reppas, N. B.; Emig, C. J.; Bang, D.; Hwang, S. J.; Jewett, M. C.; Jacobson, J. M.; Church, G. M. Precise Manipulation of Chromosomes in Vivo Enables Genome-Wide Codon Replacement. *Science* **2011**, *333*, 348–353.

(91) Gan, R.; Perez, J. G.; Carlson, E. D.; Ntai, I.; Isaacs, F. J.; Kelleher, N. L.; Jewett, M. C. Translation System Engineering in Escherichia Coli Enhances Non-Canonical Amino Acid Incorporation into Proteins. *Biotechnology and bioengineering* **2017**, *114*, 1074–1086.

(92) Söll, D.; RajBhandary, U. L. *tRNA: Structure, Biosynthesis, and Function*; Zondervan, 1995.

(93) Crick, F. H. Codon—anticodon Pairing: The Wobble Hypothesis. *Journal of molecular biology* **1966**, *19*, 548–555.

(94) Meroueh, M.; Chow, C. S. Thermodynamics of RNA Hairpins Containing Single Internal Mismatches. *Nucleic acids research* **1999**, *27*, 1118–1125.

(95) Kwon, I.; Kirshenbaum, K.; Tirrell, D. A. Breaking the Degeneracy of the Genetic Code. *Journal of the American Chemical Society* **2003**, *125*, 7512–7513.

(96) Mukai, T.; Yamaguchi, A.; Ohtake, K.; Takahashi, M.; Hayashi, A.; Iraha, F.; Kira, S.; Yanagisawa, T.; Yokoyama, S.; Hoshi, H.; others. Reassignment of a Rare Sense Codon to a Non-Canonical Amino Acid in Escherichia Coli. *Nucleic acids research* **2015**, *43*, 8111–8122.

(97) Ho, J. M.; Reynolds, N. M.; Rivera, K.; Connolly, M.; Guo, L.-T.; Ling, J.; Pappin, D. J.; Church, G. M.; Söll, D. Efficient Reassignment of a Frequent Serine Codon in Wild-Type Escherichia Coli. *ACS synthetic biology* **2015**, *5*, 163–171.

(98) Lee, B. S.; Shin, S.; Jeon, J. Y.; Jang, K.; Lee, B. Y.; Choi, S.; Yoo, T. H. Incorporation of Unnatural Amino Acids in Response to the AGG Codon. *ACS chemical biology* **2015**.

(99) Merrifield, R. Solid-Phase Peptide Synthesis. *Advances in Enzymology and Related Areas of Molecular Biology, Volume 32* **2006**, 221–296.

(100) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. Synthesis of Proteins by Native Chemical Ligation. *Science* **1994**, *266*, 776–779.

(101) Hong, S. H.; Kwon, Y.-C.; Jewett, M. C. Non-Standard Amino Acid Incorporation into Proteins Using Escherichia Coli Cell-Free Protein Synthesis. *Front Chem* **2014**, *2*, 34.

(102) Shimizu, Y.; Inoue, A.; Tomari, Y.; Suzuki, T.; Yokogawa, T.; Nishikawa, K.; Ueda, T. Cell-Free Translation Reconstituted with Purified Components. *Nat. Biotechnol.* **2001**, *19*, 751–755.

(103) Albayrak, C.; Swartz, J. R. Cell-Free Co-Production of an Orthogonal Transfer RNA Activates Efficient Site-Specific Non-Natural Amino Acid Incorporation. *Nucleic acids research* **2013**, *41*, 5949–5963.

(104) Wang, L.; Schultz, P. G. A General Approach for the Generation of Orthogonal tRNAs. *Chemistry \& biology* **2001**, *8*, 883–890.

(105) Reinders, J.; Lewandrowski, U.; Moebius, J.; Wagner, Y.; Sickmann, A. Challenges in Mass Spectrometry-Based Proteomics. *Proteomics* **2004**, *4*, 3686–3703.

(106) Lai, X.; Wang, L.; Witzmann, F. A. Issues and Applications in Label-Free Quantitative Mass Spectrometry. *International journal of proteomics* **2013**, *2013*.

(107) Barnidge, D. R.; Dratz, E. A.; Martin, T.; Bonilla, L. E.; Moran, L. B.; Lindall, A. Absolute Quantification of the G Protein-Coupled Receptor Rhodopsin by LC/MS/MS Using Proteolysis Product Peptides and Synthetic Peptide Standards. *Analytical chemistry* **2003**, *75*, 445–451.

(108) Arrivault, S.; Guenther, M.; Fry, S. C.; Fuenfgeld, M. M.; Veyel, D.; Mettler-Altmann, T.; Stitt, M.; Lunn, J. E. Synthesis and Use of Stable-Isotope-Labeled Internal Standards for Quantification of Phosphorylated Metabolites by LC-MS/MS. *Analytical chemistry* **2015**, *87*, 6896–6904.

(109) Shi, H.; Moore, P. B. The Crystal Structure of Yeast Phenylalanine tRNA at 1.93 A Resolution: A Classic Structure Revisited. *Rna* **2000**, *6*, 1091–1105.

(110) Biddle, W.; Schmitt, M. A.; Fisk, J. D. Modification of Orthogonal tRNAs: Unexpected Consequences for Sense Codon Reassignment. *Nucleic Acids Research* **2016**, *44*, 10042–10050.

CHAPTER 2

EVALUATING SENSE CODON REASSIGNMENT WITH A SIMPLE FLUORESCENCE SCREEN[1]

OVERVIEW

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, particularly those with chemistries not provided by Nature. Because 61 sense codons encode 20 amino acids, reassigning the meaning of sense codons provides an avenue for biosynthetic modification of proteins, furthering both fundamental and applied biochemical research. We developed a simple screen that exploits the absolute requirement for fluorescence of an active site tyrosine in green fluorescent protein (GFP) to probe the pliability of the degeneracy of the genetic code. Our screen monitors the restoration of the fluorophore of GFP by incorporation of a tyrosine in response to a sense codon typically assigned another meaning in the genetic code. We evaluated sense codon reassignment at 4 of the 21 sense codons read through wobble interactions in *E. coli* using the *Methanocaldococcus jannaschii (M. jannaschii)* orthogonal tRNA/aminoacyl tRNA synthetase pair originally developed and commonly used for amber stop codon suppression. By changing only the anticodon of the orthogonal tRNA, sense codon reassignment efficiencies between 1% (Phe UUU) and 6% (Lys AAG) were achieved. Each of the orthogonal tRNAs preferentially decoded the codon traditionally read via a wobble interaction in *E. coli* with the exception of the orthogonal tRNA with an AUG anticodon, which incorporated tyrosine in response to both the His CAU and His CAC codons with approximately equal frequency. We applied our screen in a high throughput manner to evaluate a 10^9 member combined tRNA/aminoacyl tRNA

1. The text in this chapter is reprinted with permission from Biddle, W.; Schmitt, M.A.; Fisk, J.D.; *Biochemistry* **2015**, 54, 7355-7364. Copyright 2015 American Chemical Society

synthetase library to identify improved sense codon reassigning variants for the Lys AAG codon. A single, rapid screen with the ability to broadly evaluate reassignable codons will facilitate identification and improvement of the combinations of sense codons and orthogonal pairs that display efficient reassignment.

INTRODUCTION

One of the driving motivations of both chemical and synthetic biology is the expansion of the set of building blocks that can be employed in the templated synthesis of biopolymers. The expansion of the set of amino acids that can be utilized in translation is particularly challenging. 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.[1-3] Genetic code expansion is further hampered by the fact that all 64 codon triplets have an assigned function. However, the genetic code is degenerate: 61 sense codons specify 20 canonical amino acids. While the amino acid specified by each triplet of nucleotides has not changed during evolution (with some rare exceptions), codon usage frequencies and the complements of adapter tRNA molecules used to translate the code have diverged considerably.[4, 5] The fact that codon usage varies widely across organisms and that different species employ different sets of tRNAs to decode their genomes implies that a fair degree of plasticity exists in the machinery that specifies the genetic code.[6, 7]

Breaking the degeneracy of the genetic code and 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. Genetic code expansion via nonsense (i.e. amber stop (UAG) or four-base codon) suppression has been limited to incorporation of non-canonical amino acids (ncAAs) at a single site because competition with termination signals or frameshift mutations curtail the amount of protein produced when multiple suppressions are attempted in a single protein.[8-11] Recently, the Sakamoto, Wang, and Church laboratories each engineered genomic changes in *Escherichia coli* (*E. coli)* that mitigate the

usual cytotoxic effect of deletion of the release factor that competes for decoding the amber stop signal.[12-14] These efforts have produced cells where the meaning of the amber codon is "free" and can be utilized to encode multiple copies of a non-canonical amino acid, generating a 21 amino acid genetic code. However, none of the methods is easily extendable to additional codons.

An alternative approach to genetic code expansion, residue specific reassignment, allows multisite incorporation of ncAAs, but requires replacement of every occurrence of one natural amino acid with the ncAA. Reassignment is controlled by switching the growth medium such that the targeted natural amino acid is removed and replaced by a ncAA that is a close structural analog.[15-17] The ncAA is utilized in translation in place of the removed natural amino acid, resulting in proteins with high levels of non-canonical amino acid incorporation at multiple sites without appreciable reductions in yield. The resulting genetic code is an altered 20 amino acid genetic code because one natural amino acid must be removed entirely.

Using a strategy that combines aspects of the amber suppression and residue specific methods, Kwon, Kirshenbaum, and Tirrell broke the degeneracy of the genetic code and demonstrated that one of two codons specifying phenylalanine (Phe) in *E. coli* could be reassigned to allow multisite incorporation of ncAAs.[18, 19] *E. coli* utilize one tRNA species with a GAA anticodon to decode two Phe codons: UUU and UUC. Naphthylalanine was incorporated in response to the UUU codon (usually read by a wobble interaction) by supplying an orthogonal yeast tRNA modified with an AAA anticodon and an orthogonal yeast aminoacyl tRNA synthetase. The UUC codon continued to direct incorporation of Phe, resulting in a 21 amino acid genetic code.

Expanding the genetic code by reassigning the meaning of sense codons should be broadly generalizable. Unfortunately, predicting which sense codons are amenable to reassignment and which orthogonal machinery is best suited for the task is made challenging by the largely unknown and idiosyncratic recognition and discrimination features of each organism's complement of tRNAs and

aminoacyl tRNA synthetases (aaRSs). The interactions between the 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. Understanding the limits to which modifications can be made without breaking the translational system has practical implications beyond the expansion of the set of encodable amino acids. Measurements of the global effects of translational system modifications where the reading of specific sense codons is altered, including the extent to which a given sense codon is naturally subject to errors, contribute to understanding the molecular mechanisms of translation-related diseases and certain cancers.[20-22]

Five key conditions must be met in *E. coli* in order to break the degeneracy of the genetic code. 1) A natural amino acid must be specified by multiple codons such that one of the codons can be reassigned to the ncAA. In the standard genetic code, 18 of the 20 natural amino acids are encoded by more than one codon. 2) The codon targeted for sense codon reassignment should be read by the set of natural *E. coli* tRNAs via a wobble interaction in which the tRNA anticodon is able to decode a codon triplet through recognition of two positions by Watson-Crick base pairing.[23] The third codon position is recognized via an energetically less favourable wobble pairing of nucleotide bases other than G with C or A with U. In *E. coli*, 40 tRNA species decode 61 sense codons.[24] The subset of 21 codons that are read through wobble interactions are potential targets for sense codon reassignment.[18] 3) The orthogonal tRNA must be able to outcompete the natural translational system for the codon to be reassigned. A combination of increased concentration, preferential energetic interactions of the Watson-Crick vs wobble base pairing, and more subtle effects of precise geometry of the decoding hairpin in the codon-anticodon interaction contribute to the ability of tRNAs to compete for decoding the same sense codon.[4, 25] 4) The nucleotide changes in the anticodon of the orthogonal tRNA required to read the selected sense codon must not abrogate the interaction with the cognate orthogonal aminoacyl tRNA synthetase. The anticodon is often an important identity element that allows a specific aaRS to recognize its appropriate tRNA. Changing the anticodon can affect the efficiency with which the ncAA is attached to the sense codon reassigning tRNA.[26] 5)

Orthogonality of the tRNA/aaRS pair must be maintained as the tRNA anticodon is modified and the aaRS is evolved to recognize the new tRNA. The complete set of tRNA/aaRS interactions is not known even for *E. coli,* and how each modified heterologous pair will interact with the natural system is extremely difficult to predict.

In this report we describe a simple screen to evaluate the reassignment of sense codons in *E. coli* through the introduction of modified forms of an orthogonal tRNA/aminoacyl tRNA synthetase pair. Our system employs the orthogonal tyrosine-incorporating tRNA and aaRS pair from *Methanocaldococcus jannaschii* (*M. jannaschii*), and a GFP reporter protein with modifications to the codon specifying the fluorophore tyrosine residue. We use the fluorescence-based screen to evaluate the reassignment of four *E. coli* sense codons and improve the extent of reassignment of one codon through the directed evolution of tRNA and aaRS variants. Because our intent was to improve the interactions between an orthogonal tRNA with an altered anticodon and the orthogonal aminoacyl tRNA synthetase, rather than change the identity of the amino acid charged to the tRNA, libraries were designed to explore the space of the anticodon loop of the tRNA and the anticodon binding domain of the aaRS. The expansion of the genetic code via sense codon reassignment can be viewed as a two part problem. The simple screen we developed addresses the first facet: identification of the particular sense codons that can be most productively targeted by a particular orthogonal pair. The second facet of the problem involves modulating the levels of aminoacyl tRNA synthetase function (through adjustments in expression levels or activity) and controlling protein expression conditions to effectively compete with the natural translational apparatus. Our screen allows improvement of the aminoacylation efficiency for a specific tRNA by providing a high throughput measurable for incorporation.

The selection of initial codons to evaluate was guided by the eventual goal of using variants of the *M. jannaschii* tRNA/aaRS pair to incorporate non-canonical amino acids in response to sense codons. The *M. jannaschii* pair has been subjected to repeated directed evolution experiments, leading to identification of variants capable of incorporating nearly 100 ncAAs into proteins in response to the amber stop codon.[8]

The choice of codons for this initial evaluation also considered maintenance of orthogonal pair function with anticodon modification, cellular codon usage, tRNA levels, and the cellular tolerance toward substitution of the amino acid targeted for reassignment. Beyond the utility of the fluorescence screen for evaluating sense codon reassignment, the screen is sufficiently sensitive to measure natural background levels of missense incorporation of tyrosine.

EXPERIMENTAL PROCEDURES:

General Methods

Reagent details and sources, DNA manipulation and mutagenesis details, and vector and oligonucleotide primer sequences are described in the Supporting Information.

Antibiotics for Vector Maintenance

Spectinomycin was used at 50 μg/mL to maintain the vectors harbouring the tRNA and aaRS genes. Carbenicillin was used at 50 μg/mL to maintain the vectors harbouring the GFP reporter gene. Unless otherwise noted, these antibiotics were present in all liquid and solid media.

GFP Fluorescence-Based Sense Codon Reassignment Efficiency Assays Reassignment Efficiency *In Vivo*

Superfolder green fluorescent protein (GFP) reporter plasmids (pGFP66xxx, where xxx specifies the codon at position 66) were co-transformed with vectors expressing the modified orthogonal translational components (pWB_Ultra-Tyr-yyy, where yyy indicates the anticodon on the tRNA) into SB3930 E. coli cells (*λ*-, *ΔhisB463*). After overnight growth, colonies were picked into 200 μL LB media in a 96 well plate. Cells were grown to saturation (usually 12 hours) with shaking at 37 °C. Cells were diluted 10-fold into LB media with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for induction of aaRS and GFP reporter. Assays were performed in a Fluorotrac 200 clear bottom 96 well plate (Greiner 655096) and monitored in a BioTek Synergy H1 plate reader at 37 °C with continuous double orbital

shaking. The optical density (OD600) and fluorescence of each well was measured every 20 minutes; optical density was measured at 600 nm, and fluorescence was measured with an excitation at 485 nm and detection at 515 nm with an 8 nm band pass. The relative fluorescence of each 200 µL culture was calculated by dividing the LB media-blanked fluorescence by the OD600. The 100% relative fluorescence unit (reported as fluorescence units per unit optical density at 600 nm, RFU) value for sense codon reassignment efficiency was defined by taking an average of three cultures expressing wild type sfGFP (pGFP66tat) in cells harboring a negative control translational machinery plasmid (pWB_Ultra-Tyr-no_tRNA) to maintain a similar metabolic burden on the cell. Sense codon reassignment efficiency for each tRNA variant was calculated by dividing the individual RFU values from each of three colonies by the average 100% reference RFU value and then averaging the results.

Protein Purification and Quantification for Reassignment Efficiency

Colonies of SB3930 cells harboring vectors expressing both the orthogonal translational components and a GFP reporter gene were picked into 200 µL of LB and grown to saturation with shaking at 37 °C. An 80 µL aliquot from each culture was diluted with 3 mL of LB media containing 1 mM IPTG. The aaRS and reporter protein were expressed with shaking at 37 °C for 12 hours. 2.6 mL of each expression culture was pelleted by centrifugation at 17,000 x*g* for 2 min, and the supernatant was discarded. Cell pellets were frozen overnight at -80 °C. The pellets were thawed at room temperature and 200 µL of B-PER lysis reagent (Thermo Scientific) was added to each cell pellet and vortexed to mix. Lysis reactions were incubated at room temperature for 30 minutes, and whole cell lysates were clarified by pelleting at 17,000 x*g* for 5 minutes. Ni-NTA spin columns (Qiagen) were equilibrated with NPI-10 buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole pH: 8.0). 200 µL of cleared lysate was loaded onto the columns and eluted following the recommendations from Qiagen. Proteins were washed twice with NPI-20 buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole pH: 8.0), and then eluted twice with 350 µL of NPI-500 (50 mM NaH_2PO_4, 300 mM NaCl, 500 mM imidazole pH: 8.0) for a total elution volume of 700 µL. Imidazole and other salts were removed from the eluted protein samples using

Amicon® 3,000 molecular weight cut-off spin columns (Millipore, 0.5 mL volume). Briefly, 150 µL of each protein elution was combined with 350 µL of phosphate buffered saline (PBS) and spun through the filter at 17,000 x*g* at 4 °C for 30 minutes. After the first spin, approximately 50µL of condensate remained, to which 450 µL of PBS was added. After the second spin identical to the first, 150 µL of PBS was added to the 50 µL of condensate and the filter was inverted. The dilute condensate was spun out of the filter for 2 minutes at 1,000 x*g*, resulting in a final imidazole concentration of 3.8 mM. Two samples of 200 µg/mL bovine serum albumin (BSA) were diluted in NPI-500 elution buffer and desalted using the filtration process described above to prepare a standard curve for protein quantification using the Pierce Micro BCA Protein Assay kit.

4 and 8-fold dilutions of each protein sample (150 µL each) were placed into a Fluorotrac 200 clear bottom 96 well plate, and the fluorescence of each sample was measured (excitation at 485 nm, emission read at 515 nm). Following quantification of fluorescence, 150 µL of BCA working reagent (25:24:1) was added to each well to determine the protein concentration. The plate was shaken for 30 seconds to mix and was then incubated at 37 °C for 2 hours before reading the absorbance of each well at 562 nm. Protein concentrations were determined based on a BSA standard curve (see supporting information).

Construction of 256 Member tRNA Anticodon Loop Library

Primer GX (see supporting information) was used to construct a small (256 member) library of tRNA anticodon loop variants via Kunkel mutagenesis.[27, 28] The mutagenic primer was designed to simultaneously randomize tRNA positions 33, 34, 38 and 39 flanking the anticodon (Figure 2.2a) and change the anticodon from AUG to CUU. We created the library using vector pWB_Ultra-Tyr-AUG as a template in order to reduce background fluorescence from non-library members; the AUG anticodon in the starting material tRNA would not recognize the AAG codon in the GFP reporter fluorophore position and therefore Lys would be incorporated into a non-fluorescent protein. The library was transformed into NEB 5-alpha cells harboring the pGFP66aag reporter plasmid and plated onto LB agar with 0.5 mM

IPTG to induce expression of the aaRS and reporter proteins. Out of 295 colonies on a 10,000 fold dilute plate (2.95×10^6 total unique transformants), 7 were visibly fluorescent when viewed on a dark reader (DR46B Transilluminator, Clare Chemical Research). Colonies from this plate, including the seven visibly fluorescent colonies, were evaluated using the *in vivo* assay described above. Clones from the library with fluorescence on the same order as pWB_Ultra-Tyr-CUU with the same pGFP66aag reporter were colony purified on LB agar plates. Colonies from these plates were then analyzed to evaluate colony-to-colony variation.

Construction of Combined tRNA Anticodon Loop and aaRS tRNA Binding Domain Library

The starting material for the large combined tRNA/aaRS library was pWB_Ultra-XhoI-Inactive, a variant of pWB_Ultra-Tyr-AUG that contains two XhoI restriction sites within the aaRS gene and an additional XhoI site in the anticodon position of the tRNA. Restriction sites were introduced using primers GZ, HA, and HB (see supporting information). The primers used to construct the library remove the XhoI restriction sites, providing a facile method to determine the efficiency of mutagenesis. Most library members will not be green because most mutations will be deleterious to tRNA/aaRS or tRNA/mRNA recognition. The library was created using primer GX (anticodon loop randomization) and primers HE, HF, and HG, which target nine amino acid residues in three regions in the aaRS. The amino acids chosen for randomization (Y230, C231, P232, A233, F261, H283, P284, M285, and D286) are within 5Å of the tRNA anticodon in the crystal structure of the *M. jannaschii* TyrRS complexed with the *M. jannaschii* tyrosyl-tRNA (pdb 1J1U) (Figure 2.2b).[29] In designing the library, we used PyMol software to approximate the tRNA anticodon as CUU (instead of the Tyr GUA).[30] Degenerate codons for each position of the library were chosen to include the original wild type residue for both the aaRS and the tRNA anticodon loop. The theoretical diversity of the combined library is 1.77×10^{10}. The library was generated using Kunkel mutagenesis and then transformed into SS320 cells in order to remove the ss dU DNA template. The transformation was allowed to recover in 50 mL of SOC without antibiotics for 60

min at 37 °C. A small aliquot was plated on LB agar (50 μg/mL spectinomycin) to determine mutation efficiency (1×10^9 unique transformants). The remainder of the transformation recovery was diluted up to 150 mL of 2xYT media with 0.2% glucose (to mimic SOC recovery media) and 50 μg/mL spectinomycin and allowed to grow for another 60 min at 37 °C. 18.4 μg DNA was isolated using a Qiagen maxi prep kit (Product 12162). 16 μg of this DNA were digested with XhoI to remove unmutated template DNA. The resulting DNA was transformed into NEB 5-alpha cells harboring the pGFP66aag reporter plasmid and recovered in 50 mL of SOC media at 37 °C for 60 min. A small aliquot was plated on LB agar with 1 mM IPTG to determine transformation efficiency (2.5×10^8 unique transformants). The remainder of the transformation recovery was transferred into 600 mL of LB media with 1 mM IPTG and grown overnight to a final OD_{600} of 1.7. The LB agar plates from each transformation were analyzed and PCR fragments amplified from individual colonies were restricted with XhoI to determine mutation efficiency. After transformation into SS320 cells (to remove uridine-containing template DNA), only 3 out of 11 colonies analyzed were library members (3/3 XhoI restriction sites in the starting material had been removed). The maxiprep-restrict-retransform step eliminated the remaining starting material: 10/10 colonies analyzed had all 3 XhoI restriction sites removed.

Flow Cytometry and Cell Sorting.

Approximately 1.6×10^9 cells from a 600 mL library culture (final OD_{600} of 1.7) were pelleted in a benchtop centrifuge at 8,000 x*g* for 3 minutes and then resuspended in sterile 0.9% NaCl solution. These cells were sorted via fluorescence-activated cell sorting (FACS) on a Dako-Cytomation MoFlo Legacy using the 488 nm laser line. To remove non-cell particles and cell clumps, cells were initially gated on forward and side scatter. Accepted cells were then gated on fluorescence using a 530/40 nm band pass filter in front of the GFP emission detector. Cells were sorted using the Purify 1-2 mode and either a 70 μm or a 100 μm flow cell tip. As a control, we used NEB 5-alpha cells harboring the most efficient clone from the 256-member anticodon loop library and the unmodified Tyr aaRS, pWB_Ultra-Tyr-CUU^{AU} and the pGFP66aag reporter expressed in an identical fashion to the combined library. Approximately 6×10^7

library cells were analyzed in the first sort, and 3×10^5 cells with fluorescence greater than 100 fluorescence units were collected in PBS. The collected cells were diluted directly into LB media and incubated with shaking overnight at 37 °C. The saturated culture of FACS-enriched cells was plated on LB agar with 1 mM IPTG to yield isolated colonies. The remainder of the library culture was frozen in 35% glycerol at -80 °C. Isolated colonies from the first round of FACS screening were evaluated using the *in vivo* fluorescence assay described above.

For the second round of FACS, 1 mL of the cells frozen after the first sort (representing approximately 10% of the enriched clones) was diluted into 20 mL of LB with 1 mM IPTG and grown with shaking at 37 °C for 10 hours. 200 μL of the cell culture (approximately 1.4×10^9 cells) was pelleted at 8000 x*g* for 3 min and resuspended in 0.9% NaCl. 3.7×10^7 of these cells were sorted using a narrowed collection gate, and 1.0×10^6 cells with fluorescence greater than the top 5% of our control cells were collected. These cells were diluted into LB, amplified, and plated on LB agar with 1 mM IPTG. Individual clones were screened using the *in vivo* fluorescence assay as described above.

RESULTS AND DISCUSSION

Principle of the GFP-Based Screen

Our screen evaluates the ability of anticodon-modified tRNAs to incorporate tyrosine (Tyr) in response to a sense codon that is assigned another identity in the standard *E. coli* genetic code. Residues 65-67 of superfolder GFP specify the Thr-Tyr-Gly sequence that autocatalytically folds into the tripeptide fluorophore. Replacement of Tyr at position 66 with any other natural amino acid effectively abolishes the fluorescence of the protein. (Phenylalanine or histidine substitution lead to proteins with greatly reduced and shifted fluorescence.[31] Tryptophan incorporation without additional mutations leads to unfolded protein and no fluorophore formation.) We generated a series of reporter plasmids in which the fluorophore tyrosine position in GFP was mutated to each sense codon under evaluation for reassignment. Amino acids incorporated into the fluorophore tyrosine position by the *E. coli* translational machinery

lead to proteins that are not fluorescent; when the orthogonal translational machinery incorporates tyrosine, fluorescent proteins are produced (Table of contents image). The amount of fluorescence observed for each reporter is a measure of the extent to which the orthogonal aaRS is able to recognize and aminoacylate its cognate tRNA with a modified anticodon and the ability of the orthogonal tRNA to compete with an *E. coli* tRNA to decode the sense codon specifying the fluorophore Tyr. Similar gain of function via missense mutation enzyme reporter systems have been employed to measure natural levels of missense incorporation.[32-35] The levels of sense codon reassignment measured under the conditions of this screen (e.g. rich media) represent the lower limit of reassignment that should be achievable under the controlled conditions utilized for typical protein expression employing the residue specific reassignment strategy. Even low level sense codon reassignment is useful because missed incorporations do not lead to termination, and multiple non-canonical amino acids can be incorporated into a single protein, producing a statistical mixture, but in high yield.

Selection of Sense Codons for Initial Evaluation

We selected 4 of the 21 sense codons that are read through wobble interactions by endogenous *E. coli* tRNAs: lysine (Lys) AAG, histidine (His) CAU, asparagine (Asn) AAU, and phenylalanine (Phe) UUU for our initial evaluation of sense codon reassignment. Our choice of codons considered cellular codon usage, tRNA levels, and the cellular tolerance toward substitution of the amino acid targeted for reassignment as well as the predicted efficiency of aminoacylation of anticodon-modified *M. jannaschii* tRNAs. The recognition elements that define the *M. jannaschii* tRNA have been mapped through kinetic measurements of tRNA variants.[36] Based on an assumption of additivity in the effects of measured changes in the anticodon sequence on aminoacylation, we calculated expected aminoacylation efficiency reductions for tRNA species with altered anticodons.[36] Predicted losses in aminoacylation efficiency across the anticodon sequences directed at the 61 sense codons range between 1 and 3,600-fold relative to the natural tyrosine anticodon. The nucleotide changes required to convert a Tyr (GUA anticodon) to an amber suppressor (CUA anticodon) tRNA result in a 97-fold decrease in the recognition of the tRNA

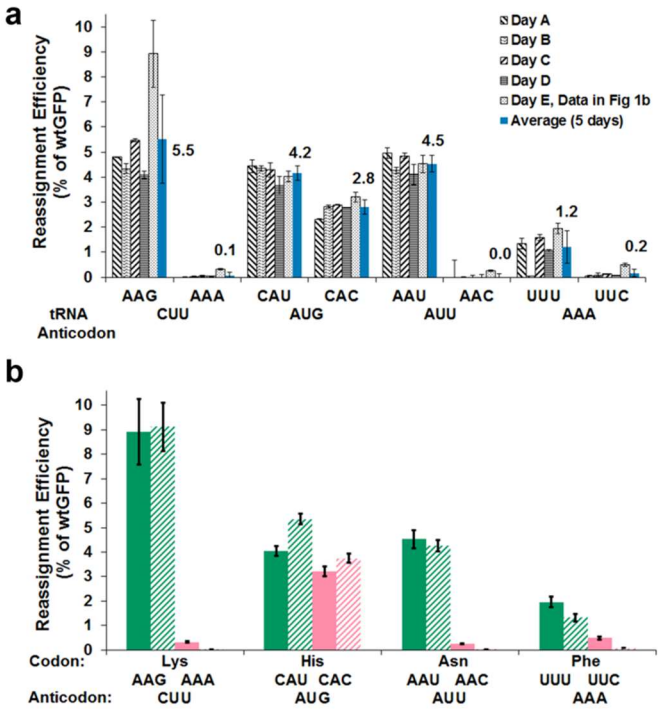
species by the aaRS. The sense codons we selected for our initial evaluation have predicted losses in aminoacylation efficiency of 116-fold (Lys AAG), 108-fold (Asn AAU), 360-fold (His CAU), and 864-fold (Phe UUU). These reductions in predicted aminoacylation efficiency reflect at most a 9-fold decrease in expected tRNA charging efficiency relative to the amber suppressor (predicted 97-fold reduction), suggesting that the altered tRNA species should be charged by the aaRS.[36]

Evaluation of Sense Codon Reassignment

We prepared variants of a form of the *M. jannaschii* tRNA gene (optimized for amber suppression)[37, 38] in which the anticodon was modified to either CUU, AUG, AUU, or AAA in order to base pair via Watson-Crick interactions with the codons for Lys (AAG), His (CAU), Asn (AAU), or Phe (UUU) which are read through wobble interactions by endogenous *E. coli* tRNAs. In all cases, both the GFP reporter and orthogonal aaRS genes were under the control of an inducible promoter.[38, 39] A single plasmid containing the *M. jannaschii* tRNA and aaRS genes was cotransformed into SB3930 *E. coli* cells along with a GFP reporter plasmid.

The observed relative fluorescence (reported as fluorescence units per unit optical density at 600 nm, RFU) for the four sense codons ranges from 1% to 6% of the relative fluorescence of wild type GFP (Figure 2.1a). The trend in sense codon reassignment efficiencies, reported as the relative fluorescence observed in each of the test cases divided by the relative fluorescence of wild type GFP, loosely correlates with the predicted reductions in aminoacylation of anticodon-modified *M. jannaschii* tRNAs by the *M. jannaschii* aaRS. Reassignment of the Lys AAG codon has an average effeciency of 6%, the Asn and His codons AAU and CAU were reassigned with 4% efficiency, and the Phe UUU codon with 1% efficiency. The 1-6% levels of sense codon reassignment we observe are similar to the levels seen in other systems designed to force missense incorporation.[32-34, 40] The variation in sense codon reassignment levels between colonies on a given day is small, typically less than 5% of the measured value. Variation in observed reassignment efficiency for a given codon between tests performed on different days is up to 16% of the

measured values, with the exception of a high outlier for Lys AAG and a low outlier for Phe UUC. The day-to-day variability is primarily due to variations in the amount of wild type GFP produced.



c

tRNA Anticodon	Sense Codon	Reassignment Percentage (In Cell Screen)	Reassignment Percentage (Purified Protein)	Codon Discrimination[a]
CUU	AAG	8.9 ± 1.3	9.1 ± 1.0	2275
	AAA	0.3 ± 0.03	0.004 ± 0.002	
AUG	CAU	4.0 ± 0.2	5.4 ± 0.2	1.5
	CAC	3.2 ± 0.2	3.7 ± 0.2	
AUU	AAU	4.5 ± 0.4	4.3 ± 0.2	614
	AAC	0.3 ± 0.02	0.007 ± 0.004	
AAA	UUU	2.0 ± 0.2	1.3 ± 0.2	16.3
	UUC	0.5 ± 0.06	0.08 ± 0.01	

[a] Codon discrimination is the ratio of reassignment percentage for the sense codon read by an *E. coli* tRNA through wobble interactions relative to the reassignment percentage for the sense codon read by an *E. coli* tRNA through Watson-Crick base pairing.

Figure 2.1. Reassignment efficiencies reported as the observed optical density-corrected fluorescence divided by the optical density-corrected fluorescence of wild type GFP for each combination of orthogonal pair and GFP reporter. a) Representation of day-to-day variation in observed reassignment efficiencies. Data are grouped by the anticodon of the tRNA utilized in each evaluation. Labeled blue bars represent the average reassignment efficiency observed for each codon. Data for each of the eight codons from five different experiments are shown. Data for Day E are those for which protein samples were also isolated and quantified (see Figure 2.1b). b) Data are grouped by the anticodon of the tRNA utilized in each evaluation. Green bars represent data for the reassigned codon (i.e. Watson-Crick base pairing to the orthogonal tRNA). Rose bars represent data for the alternative sense codon specifying the replaced amino acid (i.e. wobble base pairing to the orthogonal tRNA). Solid bars represent reassignment efficiencies determined from the in cell screen. Hashed bars represent reasignment efficiencies determined from purified protein samples. c) Values for reassignment efficiencies of each codon evaluated. Codon discrimination for each orthogonal tRNA is reported based on an analysis of purified protein. Both in cell and purified protein data were collected for the same set of cell cultures. Error bars represent variation across 3 individual colonies for each of the eight sense codons and wild type GFP.

Comparison of Fluorescence-Based Screen with Purified Protein

In our screen, the fluorescence observed in a bacterial culture is an absolute measurement of the production of GFP with a functional tyrosine-containing fluorophore. The reassignment efficiency calculation assumes that the normalization of fluorescence per cell approximates fluorescence per protein (i.e. the amount of all forms of GFP produced in all cases are comparable). In order to confirm that cell density relates to protein production level, we isolated and quantified the GFP produced in each reassignment system. Sense codon reassignment efficiencies determined from protein concentration-corrected fluorescence measurements parallel those determined from the in cell assay (Figure 2.1b). The reassignment efficiencies determined from purified protein samples are based not on the amount of protein obtained, but on the relative fluorescence of approximately equal amounts of protein obtained in all cases. The sensitivity of the in cell assay is hampered by the background fluorescence of cells and growth medium. The limit of detection in cells allows the facile identification of sense codon reassignments of about 1 part in 1,000. The isolated protein assay has an expanded detection limit due to decreased background. Sense codon substitutions can be detected at 1 part in 25,000, corresponding to a substitution frequency of 4×10^{-5}. The background missense incorporation rate is generally estimated at between 10^{-3} and 10^{-4} per codon.[35, 41] Using isolated proteins, this fluorescense assay should be sensitive enough to measure background incorporation of tyrosine at sense codons in *E. coli*. The sensitivity of the purified protein assay is based on the absolute detection of fluorescence in a sample and could be further improved by evaluating larger protein samples.

Evaluation of Codon Discrimination

The guiding principle of breaking the degeneracy of the genetic code is that wobble vs Watson-Crick base pairing can be used to specifically reassign one of two (or more) codons decoded by a single tRNA species. To examine the ability of the anticodon-modified orthogonal tRNAs to discriminate between decoding the reassigned codon and the alternative sense codon specifying the replaced amino acid, we evaluated the same set of anticodon-modified tRNAs in GFP gene reporter systems in which the

fluorophore Tyr is specified by the alternative Lys (AAA), His (CAC), Asn (AAC), and Phe (UUC) codons which are read via Watson-Crick interactions by the endogenous *E. coli* tRNAs. For the sense codons in our initial evaluation, codon discrimination measured using purified proteins ranges across 3 orders of magnitude from 1.5-fold for the CAU vs CAC histidine codons to 2,275-fold between the lysine AAG and AAA codons (Figure 2.1c). Codon discrimination between Lys AAG and AAA requires discrimination between a G and an A in the codon by a C at position 34 (see tRNA numbering in Figure 2.2a) of the anticodon-modified *M. jannaschii* tRNA. For His, Asn and Phe codons, discrimination between the Watson-Crick and wobble pairings involves an A at position 34 of the anticodon recognizing a U or C in the codon. A-C pairs are expected to be energtically unfavorable.[18, 25] Despite the A-C pairing, and in contrast to the levels of discrimination observed for Asn and Phe, the AUG anticodon-modified *M. jannaschii* tRNA does not discriminate well between the two His codons. The lack of discrimination may be due to modification of A34 in the tRNA to inosine. An IUG anticodon would recogize the CAU (His), CAC (His), and CAA (Glu) codons with similar efficiencies.[23] The only *E. coli* tRNA with an A to I modification is the Arg2 tRNA which contains an ACG anticodon.[42] The exact substrate determinants of the A to I deaminase tadA responsible for tRNA Arg2 modification have not been mapped, but the anticodon and anticodon loop sequences appear to be important recognition elements.[42] With the exception of the central anticodon position (35), 6 of 7 nucleobases in the *E. coli* Arg2 and AUG anticodon-modifed *M. jannaschii* tRNA anticodon and anticodon loop sequences are identical.

High Throughput in Cell Fluorescence Evaluations

In the *M. jannaschii* system, the tRNA anticodon changes required for sense codon reassignment are expected to adversely affect aminoacylation and may also change the anticodon stem loop geometry, impacting decoding on the ribosome. In order to achieve useful levels of sense codon reassignment for orthogonal aaRSs that incorporate non-canonical amino acids, we expect that the aminoacylation efficiencies of the anticodon-modified tRNAs will need to be improved. We applied our fluorescence-based screen in a high throughput manner for the directed evolution of improved sense codon reassigning

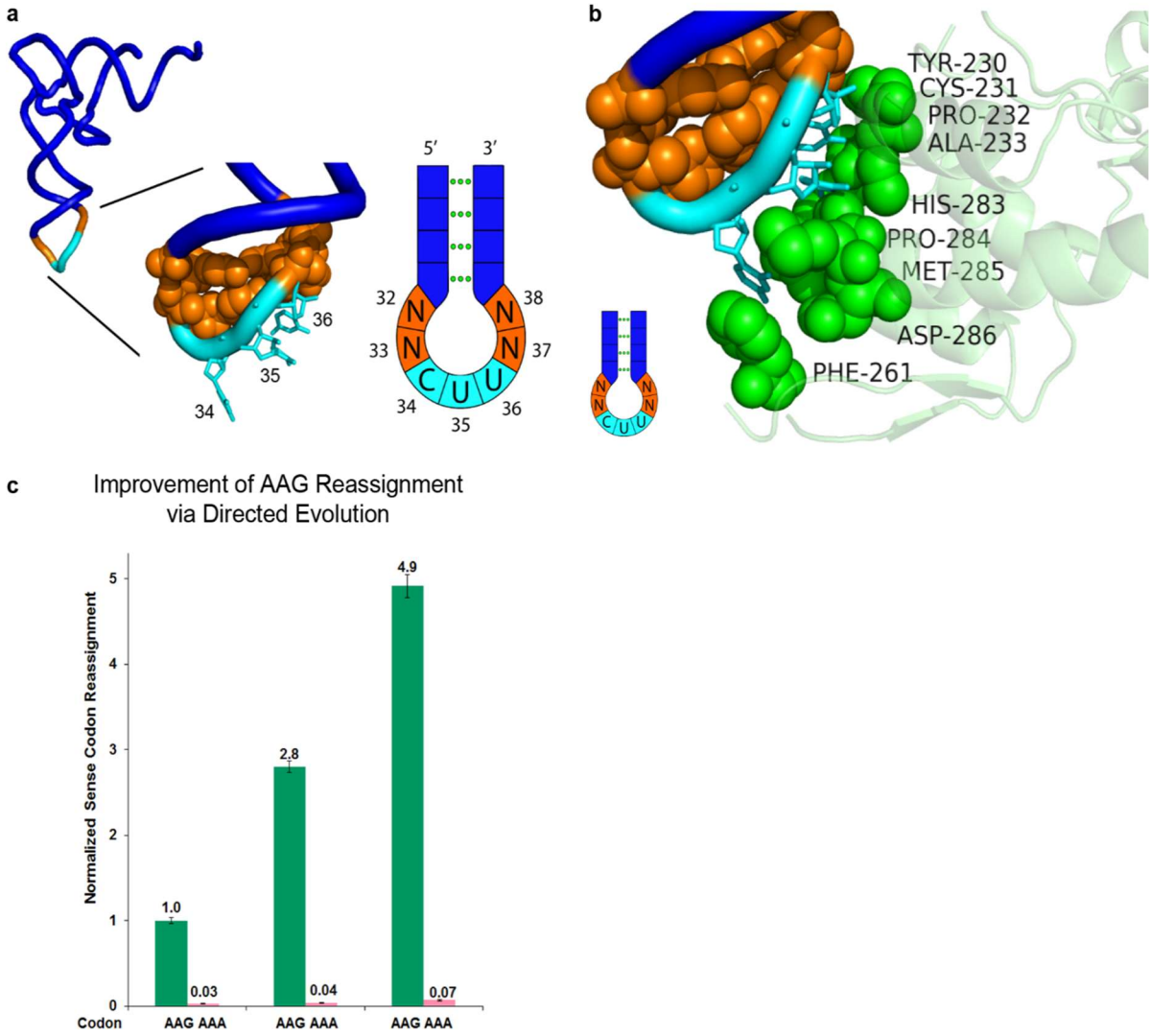


Figure 2.2 Library design and improvement of sense codon reassignment at the Lys AAG codon. a) Enlargement of the *M. jannaschii* tyrosyl-tRNA anticodon stem loop with anticodon nucleotides shown in stick representation and remaining nucleotides in the anticodon loop shown in space filling representation (pdb 1J1U). The two-dimensional cartoon representation is color coded to match the backbone trace of the tRNA, and nucleotides in the anticodon loop are numbered. The four anticodon loop positions that flank the anticodon nucleotides were randomized in a 256 member tRNA anticodon loop library. b) Crystal structure of *M. jannaschii* tyrosyl-aaRS complexed with the *M. jannaschii* tyrosyl-tRNA (pdb 1J1U) with positions targeted for diversity in the combined tRNA/aaRS library shown in space filling representation. We have altered the anticodon of the tRNA to CUU. Our combined tRNA/aaRS library for Lys reassignment includes diversity at the four positions in the tRNA anticodon loop as well as the amino acids on the aaRS that are within 5Å of the anticodon. c) Increase in efficiency of sense codon reassignment for the best tRNA identified from our initial 256 member tRNA library (5' CU CUU AU 3') as well as an improved orthogonal pair variant identified via a FACS sort of the combined tRNA/aaRS library (clone F7 in Figure 3b, tRNA 5' CU CUU AC and aaRS H283L, P284C, M285C, D286P). Fluorescence data for the alternative sense codon specifying the replaced amino acid (i.e. wobble base pairing to the orthogonal tRNA) is also shown to demonstrate retention of codon discrimination by improved variants.

tRNA/aaRS pairs. We first prepared a small library of tRNA anticodon loop variants of the lysine AAG codon reassigning *M. jannaschii* tRNA (CUU anticodon). All of the nucleotides in the anticodon loop flanking the anticodon (nucleotides 32, 33, 37, and 38) were varied to produce a library of 256 variants (Figure 2.2a). Similar tRNA libraries have been employed to improve the efficiency and orthogonality of amber codon reassigning variants with different levels of success.[43-45] The tRNA library was transformed into cells containing the reporter plasmid with a lysine AAG codon specifying the fluorophore Tyr position. Fluorescent clones from the transformed mini-library were grown up in 96 well plates and their fluorescence was measured. The most fluorescent clone analyzed from this library showed a surprisingly large improvement in sense codon reassignment efficiency of 2.8-fold over the starting tRNA (Figure 2.2c). The improved tRNA maintained the high level of discrimination between the AAG and AAA codons. The selected variant had a single modification at position 38 from U to A and maintained the G 37 A modification previously identified when optimizing the *M. jannaschii* tRNA for amber suppression.[37, 44]

In order to evaluate co-evolution of the orthogonal tRNA and aminoacyl tRNA synthetase, we generated a combined *M. jannaschii* tyrosyl-aaRS tRNA binding domain/tRNA library (Figure 2.2b). The combined library included the 256-member tRNA anticodon loop library and diversified 9 amino acid residues that are within 5Å of the tRNA anticodon in the co-crystal structure of the *M. jannaschii* tRNA/aaRS complex (pdb 1J1U).[29] The library employed restricted codon sets at the 9 varied positions to keep the total library theoretical diversity to 10^{10}. The library was prepared via Kunkel mutagenesis, unmutated template was removed by restriction enzyme digestion, and the library was transformed into cells harboring the GFP reporter plasmid with AAG specifying the fluorophore tyrosine position. Restriction analysis of randomly picked clones after digestion and transformation revealed minimal carry-through of unmutated starting template DNA: 10/10 clones were full library members. The library was sorted via FACS; the distribution of fluorescence observed for the initial sort of 7×10^7 cells included a long tail of many highly fluorescent variants (Figure 2.3a). The top 0.5% of cells were collected,

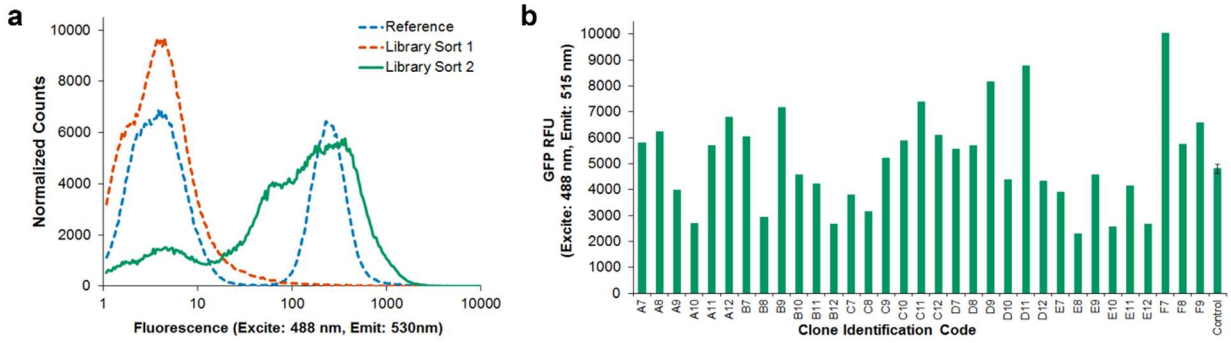


Figure 2.3 a) Distribution of fluorescence of sorted cells with counts normalized to 500,000 to show multiple sorts on the same axes. Reference cells express the tRNA variant identified the first 256 member library. In our first sort of 6.5 x 10^7 cells from the combined tRNA/aaRS library, we collected all cells that showed fluorescence greater than 100 fluorescence units (library sort 1 trace). We amplified the fraction collected in sort 1 and resorted (library sort 2 trace). Cells in the second sort with fluorescence greater than that displayed by our reference cells were collected for further analysis. b) Relative fluorescence (RFU) for 33 randomly picked clones after plating FACS sort 2. Data for the tRNA variant identified from the mini-library (control) are shown for comparison. Clone F7 was selected for further characterization (see also, Figure 2.2c).

amplified, and resorted. The cell population for the second sort showed a large increase in average fluorescence. The top 3% of these cells were collected. Randomly picked isolated clones were grown in 96 well plates and their fluorescence was quantified (Figure 2.3b). The majority of the analysed clones (17/33, 52%) were more fluorescent than the positive control cells from the mini-library selected modified tRNA with unmodified aaRS. The most fluorescent clone from the analysed variants (clone F7) appeared to be nearly twice as fluorescent (Figure 2.3b). The liquid culture of clone F7 was replated for further analysis. Seven replicate colonies of clone F7 were analysed, and reassignment efficiency improved 4.9-fold relative to the starting tRNA/aaRS pair, which included only an anticodon modification (Figure 2.2c). In both the tRNA mini-library hit and in clone F7, position 38 in the tRNA changed from the original A, but the two variants did not share the same mutation; position 38 of the tRNA from the mini-library is U and from clone F7 is C. The aaRS from the clone F7 pair included changes to 4 out of 9 amino acids. Amino acids at positions 230-233 in the aaRS are closest to position 36 of the tRNA. This region of the library was represented by minimal diversity and contained only 256 protein sequence variants. That none of the amino acids at these positions changed from the wild type *M. jannaschii* aaRS is not surprising. Greater diversity was allowed at position 261 (15 amino acids + 1 stop), but a mutation from the wild type Phe was not present in clone F7. The majority of diversity in the library was available

for amino acids 283-286, and clone F7 does include mutations at these positions. The His-Pro-Met-Asp in the wild type aaRS changed to Leu-Cys-Cys-Pro. These mutations suggest backbone rearrangements in the structure. Movement of the proline from position 284 to 286 suggests a shortening of an α-helix.

The fact that the tRNA anticodon loop and aaRS evolve together to improve sense codon reassignment highlights the fact that the effects of changing the aaRS and the tRNA are not independent. Using the GFP-based screen for directed evolution of the orthogonal tRNA and aminoacyl tRNA synthetase together allows identification of improved pairs that balance many factors, including discrimination, efficiency, and orthogonality. The extent to which tRNA and aaRS modifications that affect the efficiency of sense codon reassignment are transferable between aaRSs evolved for different non-canonical amino acids is difficult to assess. Attempts to identify tRNA modifications that improve amber suppression efficiency by modulating interactions with Ef-Tu identified tRNAs that produced generally improved suppression for multiple aaRSs evolved to incorporate different ncAAs. However, some tRNA modifications appeared to be beneficial to only one aaRS.[44] Additional studies looking at the transferability of modifications to the aaRS anticodon-binding domain suggest that transferability is related to the similarity between the ncAA sizes.[45] Regardless of the transferability, measurement and improvement of tyrosine-incorporating orthogonal aaRS variants provides a probable high end estimate for the efficiency of ncAA incorporation attainable at a particular sense codon and will be useful as starting points for future studies employing non-canonical amino acid incorporating aaRSs. Our expectation is that each orthogonal tRNA/aaRS pair will have a different set of idiosyncratic cross reactions with the translational system into which it is transplanted.

Comparison to Reported Levels of Missense Incorporation

The extent of sense codon reassignment obtained with the evolved lysine AAG-decoding tRNA/aaRS pair (clone F7) is among the highest reported to date in rich medium. Döring and Marlière described an editing-deficient mutant of the valine aaRS that allows incorporation of aminobutyric acid in

response to valine codons at a level of 24%.[32] Min and Söll reported the introduction of a non-discriminating archeal aspartyl tRNA synthetase into *E. coli*. This change results in the aminoacylation of the aspartyl tRNA with both Asp and Asn.[33] The Asp for Asn substitutions of between 16% and 38% are detected using a screen that monitors restoration of enzyme function as a result of missense incorporation. Ruan and Söll expanded on the work of Min and evaluated four different missense substitutions (Cys for Pro, Glu for Gln, Ser for Thr, and Asp for Asn).[34] For these systems, levels of replacement are between 1.8% and 28%. Together, these reports describe the highest levels of missense incorporation observed in cells without controlling the medium composition or employing auxotrophic strains.

Effect of Sense Codon Reassignment on Cell Health

A possible complicating factor in our screening and selection strategy is the effect of sense codon reassignment on cell health. Of the four codons for which data are presented in this manuscript, cell growth is significantly negatively impacted only when either the directed evolution-improved machinery for reassignment of lysine AAG codons (clone F7) or the machinery to reassign His codons is expressed (supplementary Figures S4-S5). Only minor effects on the growth of cells are observed for all other reassignment systems. In general, cell growth rates are slowed slightly relative to control systems in which no reassignment takes place. Cell growth rate decreases are within the ranges of the wild type GFP producing control systems for all reassignment systems, except His and clone F7 for Lys AAG. The standard deviation in the calculated exponential growth rates is approximately 5 to 7.5% of the measured values. The final optical densities observed in culture are slightly reduced for the majority of systems in which reassignment takes place. Final optical densities for cells expressing either the directed evolution-improved machinery for Lys AAG codons (clone F7) or for histidine reassignment decreased further. This observation may suggest cumulative effects of incorporation of tyrosine at histidine or lysine codons throughout the *E. coli* genome.

Our results are consistent with several reports that measure the tolerance of bacterial cells toward missense incorporation. In the series of reports describing directed attempts to induce high levels of missense incorporation in *E. coli*, only minor effects on growth were noted. Ruan and Söll specifically evaluated growth effects on *E. coli* cells with normal protein quality control machinery and missense incorporation systems that led to as much as 28% missense incorporation. No effects on cell growth were observed. However, upon evaluation of the missense incorporation systems in *E. coli* cells with compromised protein quality control mechanisms, Ruan and Söll report pronounced growth inhibition effects when high levels of missense mutations are induced.

The lysine AAG codon encodes 22% of the lysine residues in the *E. coli* genome. Our observed levels of sense codon reassignment, and even nearly complete reassignment of the AAG codon, would be within the range of misssense levels generally tolerated by bacteria. Approximately half of the codons read via wobble interactions are used in fewer than a third of the instances in which a given amino acid is specified in the *E. coli* genome. This subset of wobble codons represents target codons for ready reassignment. We expect that the cell growth effects will correlate with the relative use of the particular codon targeted for reassignment and the catalytic importance of the substituted amino acid (e.g. histidine). The exact extent to which either of these factors plays a role in observed effects on cellular health has not been established. Our screen could be readily applied to evaluate the relative importance of particular amino acids in living systems.

CONCLUSIONS

We developed a simple, high throughput screen to measure and improve sense codon reassignment by the most commonly employed orthogonal tRNA/aaRS pair for non-canonical amino acid incorporation at amber stop codons. By changing only the anticodon of the orthogonal tRNA, sense codon reassignment efficiencies between 1% (Phe UUU) and 6% (Lys AAG) were achieved. Each of the orthogonal tRNAs preferentially decoded the codon traditionally read via a wobble interaction in *E. coli* with the exception

of the orthogonal tRNA with an AUG anticodon, which incorporated tyrosine in response to both the His CAU and His CAC codons with approximately equal frequency. We applied our screen in a high throughput manner to evaluate a 10^9 member combined tRNA/aminoacyl tRNA synthetase library to identify improved sense codon reassigning variants for the Lys AAG codon. We believe that a single screen with the ability to broadly evaluate sense codons and the available set of orthogonal tRNA/aaRS pairs will facilitate quick identification of the combinations that can be employed to efficiently reassign the meaning of sense codons.

Producing proteins using expanded genetic codes combines exquisite positional control over sites of modification with a large toolbox of encodable functionalities. Unlike amber stop codon suppression methodology, sense codon reassignment has the potential to expand the genetic code further by allowing the reassignment of more than one codon in the same system. With the present suite of nearly 100 aaRSs evolved to incorporate ncAAs, there are 100 possible 21 amino acid genetic codes. We are interested in employing the set of previously-evolved aminoacyl tRNA synthetase variants to incorporate multiple copies of non-canonical amino acids in response to sense codons in *E. coli*. As a first step toward this goal, we developed a simple, high throughput screen to measure and improve sense codon reassignment by the most commonly employed orthogonal tRNA/aaRS pair for non-canonical amino acid incorporation at amber stop codons. We began our evaluation of sense codon reassignment in *E. coli* using the *M. jannaschii* tyrosyl-tRNA/aaRS pair, but the fluorescence-based screen will be applicable to other organisms and to other orthogonal pairs that can be evolved to incorporate tyrosine. Differences in the complements of tRNAs, codon usage and vagrancies in determinants of aminoacylation between organisms may require targeting alternative sets of sense codons across different organisms. At present, three well-studied, mutually orthogonal sets of tRNA/aaRSs can be employed for stop and four-base codon-directed ncAA incorporation (two orthogonal pairs derived from the *M. jannaschii* pair and the *Methanosarcina barkeri* pyrrolysyl-tRNA/aaRS pair).[46-48] The *Methanosarcina barkeri* pyrrolysyl-tRNA/aaRS pair is evolutionarily closely related to phenylalanine aaRSs and has been evolved to

introduce para-substituted phenyl ring containing amino acids. Several other pairs have been described, but not widely employed.[11, 49-52] Extending the sense codon reassignment strategy to exploit these ncAA-incorporating orthogonal pairs is a step towards generating expanded genetic codes where multiple copies of multiple ncAAs can be incorporated into proteins simultaneously. The ability to incorporate more than one type of ncAA in the same protein expands the genetic code exponentially to 10,000 possible 22 amino acid genetic codes. Beyond the utility towards expanded genetic codes, the screening system is sensitive enough to measure natural rates of tyrosine missense incorporation, and evolved tyrosine-incorporating aaRSs will allow the evaluation of the biological effects of high level directed amino acid substitutions at various codons of interest.

ASSOCIATED CONTENT

Supporting Information. Methods and results for the following: General methods, reagent details and sources, DNA manipulation and mutagenesis details, vector and oligonucleotide primer sequences. This material is available free of charge via the Internet at http://pubs.acs.org as well as in Appendix 1 in this dissertation.

REFERENCES

(1) Harold, F. M. (1986) *The vital force a study of bioenergetics*, W.H. Freeman, New York.

(2) Russell, J. B., and Cook, G. M. (1995) Energetics of bacterial-growth - balance of anabolic and catabolic reactions, *Microbiol. Rev. 59*, 48-62.

(3) Stouthamer, A. H. (1973) A theoretical study on the amount of ATP required for synthesis of microbial cell material, *Antonie Van Leeuwenhoek Journal of Microbiology 39*, 545-565.

(4) Agris, P. F. (2004) Decoding the genome: a modified view, *Nucleic Acids Res. 32*, 223-238.

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

(6) Plotkin, J. B., and Kudla, G. (2011) Synonymous but not the same: the causes and consequences of codon bias, *Nat. Rev. Genet. 12*, 32-42.

(7) Knight, R. D., Freeland, S. J., and Landweber, L. F. (2001) Rewiring the keyboard evolvability of the genetic code, *Nat. Rev. Genet. 2*, 49-58.

(8) Liu, C. C., and Schultz, P. G. (2010) Adding new chemistries to the genetic code, In *Annual Review of Biochemistry, Vol 79* (Kornberg, R. D., Raetz, C. R. H., Rothman, J. E., and Thorner, J. W., Eds.), pp 413-444, Annual Reviews, Palo Alto.

(9) Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C., and Schultz, P. G. (1989) A general method for site-specific incorporation of unnatural amino-acids into proteins, *Science 244*, 182-188.

(10) Bain, J. D., Glabe, C. G., Dix, T. A., Chamberlin, A. R., and Diala, E. S. (1989) Biosynthetic site-specific incorporation of a non-natural amino-acid into a polypeptide, *J. Am. Chem. Soc. 111*, 8013-8014.

(11) Furter, R. (1998) Expansion of the genetic code: Site-directed p-fluoro-phenylalanine incorporation in *Escherichia coli*, *Protein Sci. 7*, 419-426.

(12) Mukai, T., Hayashi, A., Iraha, F., Sato, A., Ohtake, K., Yokoyama, S., and Sakamoto, K. (2010) Codon reassignment in the *Escherichia coli* genetic code, *Nucleic Acids Res. 38*, 8188-8195.

(13) Johnson, D. B. F., Xu, J. F., Shen, Z. X., Takimoto, J. K., Schultz, M. D., Schmitz, R. J., Xiang, Z., Ecker, J. R., Briggs, S. P., and Wang, L. (2011) RF1 knockout allows

ribosomal incorporation of unnatural amino acids at multiple sites, *Nat. Chem. Biol. 7*, 779-786.

(14) Isaacs, F. J., Carr, P. A., Wang, H. H., Lajoie, M. J., Sterling, B., Kraal, L., Tolonen, A. C., Gianoulis, T. A., Goodman, D. B., Reppas, N. B., Emig, C. J., Bang, D., Hwang, S. J., Jewett, M. C., Jacobson, J. M., and Church, G. M. (2011) Precise manipulation of chromosomes in vivo enables genome-wide codon replacement, *Science 333*, 348-353.

(15) Dougherty, M. J., Kothakota, S., Mason, T. L., Tirrell, D. A., and Fournier, M. J. (1993) Synthesis of a genetically engineered repetitive polypeptide containing periodic selenomethionine residues, *Macromolecules 26*, 1779-1781.

(16) Budisa, N., Steipe, B., Demange, P., Eckerskorn, C., Kellermann, J., and Huber, R. (1995) High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in *Escherichia coli*, *Eur. J. Biochem. 230*, 788-796.

(17) Budisa, N. (2006) *Engineering the genetic code expanding the amino acid repertoire for the design of novel proteins*, Wiley-VCH; John Wiley [distributor], Weinheim.

(18) Kwon, I., Kirshenbaum, K., and Tirrell, D. A. (2003) Breaking the degeneracy of the genetic code, *J. Am. Chem. Soc. 125*, 7512-7513.

(19) Kwon, I., and Lim, S. I. (2015) Tailoring the substrate specificity of yeast phenylalanyl-trna synthetase toward a phenylalanine analog using multiple-site-specific incorporation, *ACS Synth. Biol*. *4*, 634-643.

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

(21) Schimmel, P., and Guo, M. (2009) A tipping point for mistranslation and disease, *Nat. Struct. Mol. Biol. 16*, 348-349.

(22) Kirchner, S., and Ignatova, Z. (2015) Emerging roles of tRNA in adaptive translation, signalling dynamics and disease, *Nat. Rev. Genet. 16*, 98-112.

(23) Crick, F. H. C. (1966) Codon-Anticodon Pairing - Wobble Hypothesis, *J. Mol. Biol. 19*, 548-555.

(24) Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A., and Steinberg, S. (1998) Compilation of tRNA sequences and sequences of tRNA genes, *Nucleic Acids Res. 26*, 148-153.

(25) Meroueh, M., and Chow, C. S. (1999) Thermodynamics of RNA hairpins containing single internal mismatches, *Nucleic Acids Res. 27*, 1118-1125.

(26) Giege, R., Sissler, M., and Florentz, C. (1998) Universal rules and idiosyncratic features in tRNA identity, *Nucleic Acids Res. 26*, 5017-5035.

(27) Sidhu, S. S., and Weiss, G. A. (2004) Phage display a practical approach, In *Practical approach series* (Clackson, T., and Lowman, H. B., Eds.), Oxford University Press, Oxford.

(28) Kunkel, T. A. (1985) Rapid and efficient site-specific mutagenesis without phenotypic selection, *Proc. Natl. Acad. Sci. U. S. A. 82*, 488-492.

(29) Kobayashi, T., Nureki, O., Ishitani, R., Yaremchuk, A., Tukalo, M., Cusack, S., Sakamoto, K., and Yokoyama, S. (2003) Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion, *Nat. Struct. Biol. 10*, 425-432.

(30) Schrodinger, LLC. (2010) The PyMOL Molecular Graphics System, Version 1.3r1.

(31) Hein, R., and Tsien, R. Y. (1996) Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer, *Curr. Biol. 6*, 178-182.

(32) Döring, V., Mootz, H. D., Nangle, L. A., Hendrickson, T. L., de Crécy-Lagard, V., Schimmel, P., and Marlière, P. (2001) Enlarging the amino acid set of *Escherichia coli* by infiltration of the valine coding pathway, *Science 292*, 501-504.

(33) Min, B., Kitabatake, M., Polycarpo, C., Pelaschier, J., Rraczniak, G., Ruan, B., Kobayashi, H., Namgoong, S., and Söll, D. (2003) Protein synthesis in *Escherichia coli* with mischarged tRNA, *J. Bacteriol. 185*, 3524-3526.

(34) Ruan, B. F., Palioura, S., Sabina, J., Marvin-Guy, L., Kochhar, S., LaRossa, R. A., and Söll, D. (2008) Quality control despite mistranslation caused by an ambiguous genetic code, *Proc. Natl. Acad. Sci. U. S. A. 105*, 16502-16507.

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

(36) Fechter, P., Rudinger-Thirion, J., Tukalo, M., and Giege, R. (2001) Major tyrosine identity determinants in *Methanococcus jannaschii* and *Saccharomyces cerevisiae* tRNA(Tyr) conserved but expressed differently, *Eur. J. Biochem 268*, 761-767.

(37) Young, T. S., Ahmad, I., Yin, J. A., and Schultz, P. G. (2010) An enhanced system for unnatural amino acid mutagenesis in *E. coli*, *J. Mol. Biol. 395*, 361-374.

(38) Chatterjee, A., Sun, S. B., Furman, J. L., Xiao, H., and Schultz, P. G. (2013) A versatile platform for single- and multiple-unnatural amino acid mutagenesis in *Escherichia coli*, *Biochemistry 52*, 1828-1837.

(39) Ryu, Y. H., and Schultz, P. G. (2006) Efficient incorporation of unnatural amino acids into proteins in *Escherichia coli*, *Nat. Methods 3*, 263-265.

(40) Doring, V., and Marliere, P. (1998) Reassigning cysteine in the genetic code of *Escherichia coli*, *Genetics 150*, 543-551.

(41) Parker, J. (1989) Errors and alternatives in reading the universal genetic-code, *Microbiol. Rev. 53*, 273-298.

(42) Wolf, J., Gerber, A. P., and Keller, W. (2002) tadA, an essential tRNA-specific adenosine deaminase from *Escherichia coli*, *Embo J. 21*, 3841-3851.

(43) Wang, L., and Schultz, P. G. (2001) A general approach for the generation of orthogonal tRNAs, *Chem. Biol. 8*, 883-890.

(44) Guo, J. T., Melancon, C. E., Lee, H. S., Groff, D., and Schultz, P. G. (2009) Evolution of amber suppressor trnas for efficient bacterial production of proteins containing nonnatural amino acids, *Angew. Chem. Int. Edit. 48*, 9148-9151.

(45) Wang, N. X., Ju, T., Niu, W., and Guo, J. T. (2015) Fine-tuning interaction between aminoacyl-trna synthetase and trna for efficient synthesis of proteins containing unnatural amino acids, *ACS Synth. Biol. 4*, 207-212.

(46) Neumann, H., Slusarczyk, A. L., and Chin, J. W. (2010) De novo neneration of mutually orthogonal aminoacyl-tRNA synthetase/tRNA pairs, *J. Am. Chem. Soc. 132*, 2142-2144.

(47) Blight, S. K., Larue, R. C., Mahapatra, A., Longstaff, D. G., Chang, E., Zhao, G., Kang, P. T., Green-Church, K. B., Chan, M. K., and Krzycki, J. A. (2004) Direct charging of tRNA(CUA) with pyrrolysine in vitro and in vivo, *Nature 431*, 333-335.

(48) Wang, L., Brock, A., Herberich, B., and Schultz, P. G. (2001) Expanding the genetic code of *Escherichia coli*, *Science 292*, 498-500.

(49) Anderson, J. C., and Schultz, P. G. (2003) Adaptation of an orthogonal archaeal leucyl-tRNA and synthetase pair for four-base, amber, and opal suppression, *Biochemistry 42*, 9598-9608.

(50) Santoro, S. W., Anderson, J. C., Lakshman, V., and Schultz, P. G. (2003) An archaebacteria-derived glutamyl-tRNA synthetase and tRNA pair for unnatural amino acid mutagenesis of proteins in *Escherichia coli*, *Nucleic Acids Res. 31*, 6700-6709.

(51) Hughes, R. A., and Ellington, A. D. (2010) Rational design of an orthogonal tryptophanyl nonsense suppressor tRNA, *Nucleic Acids Res. 38*, 6813-6830.

(52) Chatterjee, A., Xiao, H., and Schultz, P. G. (2012) Evolution of multiple, mutually orthogonal prolyl-tRNA synthetase/tRNA pairs for unnatural amino acid mutagenesis in *Escherichia coli*, *Proc. Natl. Acad. Sci. U. S. A. 109*, 14841-14846.

CHAPTER 3

MODIFICATION OF ORTHOGONAL tRNAs: UNEXPECTED CONSEQUENCES FOR SENSE CODON REASSIGNMENT[2]

OVERVIEW

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

INTRODUCTION

Genetic code engineering, expanding the set of non-canonical amino acids (ncAAs) that can be biosynthetically incorporated into proteins, is of increasing interest as a means of precisely endowing proteins with chemical functionalities that are not present in the natural set of 20 amino acids [1-3]. The genetic code is often thought of as the straightforward pairing of mRNA codon with tRNA anticodon that

2.

specifies the amino acid sequence of a protein. A broader view of the genetic code includes the complete set of interactions that are involved in the conversion of information contained in the genome of an organism into the functional protein molecules that ultimately produce phenotypic behaviors.

The genetic code operates on at least 3 distinct levels where different sets of interactions affect the information transfer that is described in the code table. The pairing of mRNA codon with tRNA anticodon represents the primary genetic code. The idiosyncratic set of interactions between the complement of tRNA species and the suite of aminoacyl tRNA synthetase (aaRS) enzymes in an organism represents a second level of the genetic code[4, 5]. The combined interactions between tRNA species and aaRSs specify that the appropriate amino acid is attached to the appropriate tRNA molecule. The set of protein-tRNA interactions that determine the extent and positioning of modified nucleobases in tRNA molecules represents a third level of the genetic code. tRNA modification influences each level of the genetic code, including the interactions of mRNA codons with tRNA anticodons and the interactions of tRNAs with their aaRSs[6]. In addition, modified tRNAs serve regulatory roles modulating the interactions between the process of translation and cellular stress and metabolism[7, 8]. The three levels of the genetic code are tightly connected: the pairing of tRNA anticodon to mRNA codon as a means of information transfer is limited by the efficiency of the joining of an amino acid to the appropriate tRNA, and the effectiveness of both processes are modulated by the modification of tRNA species.

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

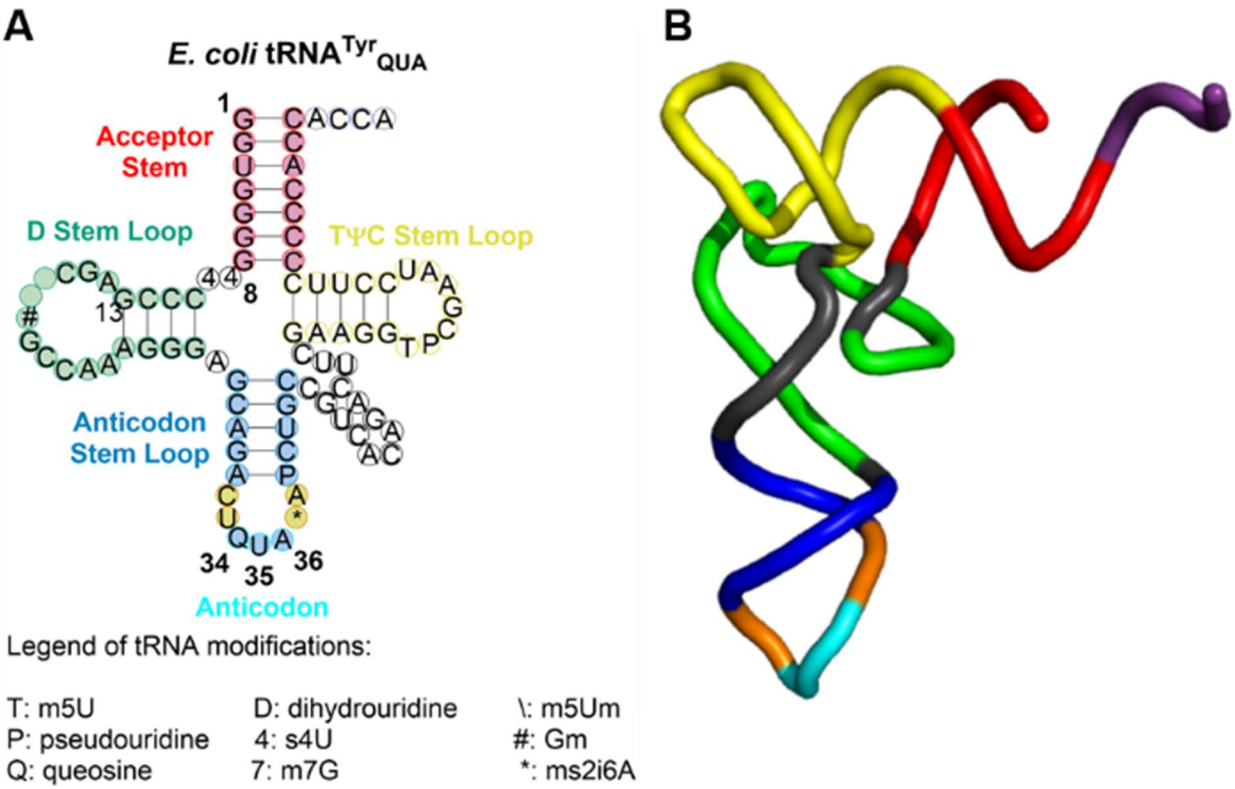


Figure 3.1 (A) Sequence and secondary structure of E. coli tyrosyl tRNA molecule. Modified nucleobases are indicated. (B) Backbone trace of tRNA structure (pdb 1EHZ), color coded to match secondary structural elements in (A).

Modification of tRNA is widespread and essential for the proper functioning of the translational apparatus in all kingdoms of life. In *Escherichia coli* (*E. coli*), one of the three model systems for which the full extent of modifications has been mapped, the 46 tRNA species contain an average of 7.5 modified bases, approximately 10% of the entire tRNA molecule[6]. All *E. coli* tRNAs contain the modified bases ribothymidine and pseudouridine in the T-Psi-C loop, and 90% of *E. coli* tRNAs contain at least one dihydrouridine in the D stem loop (Figure 3.1). Although the most prevalent modifications occur in the T-Psi-C and D stem loops, the greatest diversity of modifications occurs in the anticodon loop. *E. coli* anticodon loops, tRNA positions 32-38, contain 21 different nucleobase modifications. In the complement of *E. coli* tRNAs, position 34, which recognizes the wobble position of a codon, may be modified to 1 of 14 non-AUGC bases to allow a single tRNA to read multiple codons. A single *E. coli* tRNA, $tRNA^{Arg2}$, includes inosine modification of position 34, which allows one tRNA to decode three arginine codons: CGU, CGC, and CGA. Position 37, directly 3' of the anticodon, may be modified to 1 of 7 non-AUGC

bases[14-16]. The large diversity of modifications that occur in the anticodon loop often have direct effects on the fidelity of translation, particularly modulating the reading of codons at the wobble position and helping to maintain the reading frame.

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

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

of deletion of the release factor that competes for decoding the amber stop signal[21-23]. These efforts have produced cells where the meaning of the amber codon is "free" and can be utilized to encode multiple copies of a non-canonical amino acid, generating a 21 amino acid genetic code. However, none of the methods is easily extendable to additional codons.

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

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

We recently reported sense codon reassignment at lysine AAG, histidine CAU, asparagine AAU and phenylalanine UUU codons[35]. Our evaluation utilized the orthogonal *Methanocaldococcus jannaschii* (*M. jannaschii*) tyrosyl tRNA and aaRS pair, most commonly employed for the introduction of ncAAs. The *M. jannaschii* pair has been evolved to recognize and activate over 100 different ncAAs[3, 19]. We found that the degeneracy of the genetic code could be broken at each of the four sense selected codons with between 1% and 6% efficiency simply by introducing a variant of an engineered *M. jannaschii* tyrosyl tRNA ($tRNA^{Opt}$, [36]) with an altered anticodon designed to recognize one of the selected codons through Watson-Crick base pairing. The levels of sense codon reassignment can be improved through directed evolution. The finding that many sense codons could be reassigned with the *M. jannaschii* system is significant because the anticodon is often an important identity element that allows a specific aaRS to recognize its appropriate tRNA. Changing the anticodon may affect the efficiency with which an amino acid is attached to the tRNA. In order for sense codon reassignment to be possible, the nucleotide changes in the anticodon of the orthogonal tRNA required to read the selected sense codon must not abrogate the interaction with the cognate orthogonal aminoacyl tRNA synthetase.

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

Relative to efforts to improve the behavior of orthogonal translational components through engineering mRNA codon/orthogonal tRNA anticodon interactions and orthogonal tRNA/aaRS interactions, little effort has been devoted to consideration of the third level of the genetic code:

interactions between the orthogonal tRNA and host tRNA modifying enzymes. In this report, we describe experiments that identify a modification of $tRNA^{Opt}_{AUG}$ as the cause of the aberrant lack of discrimination between the two histidine codons. $tRNA^{Opt}_{AUG}$ is a substrate for the A to I deaminase TadA that is essential for maturation of the *E. coli* $tRNA^{Arg2}$. The resulting $tRNA^{Opt}_{IUG}$ has inosine at the first anticodon position and is capable of incorporating the amino acid it carries in response to both the CAU and CAC codons. We utilized directed evolution and our fluorescence-based screen to identify orthogonal tRNA variants that are not substrates for TadA and allow increased reassignment of the histidine CAU codon with high discrimination for the CAC codon. The unexpected modification of an orthogonal tRNA is, to our knowledge, the first described instance of a clash between genetic code engineering and tRNA modifications. These results highlight the fact that tRNA modifications should not be disregarded, as they may be both an obstacle to and target for genetic code engineering.

MATERIAL AND METHODS

Cell strains and vector composition and construction have been described in detail previously[35]; brief summaries are provided as supporting information (S.1-S.3). pWB_Ultra plasmids used in this study differ from those reported only in the nucleotides corresponding to positions 32 through 38 of the $tRNA^{Opt}$ genes (Table S1). Variants of pGFP reporters used in this study differ from those reported only in the sense codon that specifies the fluorophore position 66 (Table S2)[37]. Each histidine residue in GFP is encoded by a CAC codon in our reporter constructs; the only CAU codon in the reporter vectors occurs at position 66 to specify the critical fluorophore position. The GFP-based screen utilized to quantify sense codon reassignment has been described[35]; details are provided as supporting information (S.4).

Identification of Inosine in tRNA Anticodons

The complement of tRNA molecules was extracted from NEB 5-alpha cells harboring a pWB_Ultra plasmid[38]. Briefly, RNA was precipitated by phenol:chloroform extraction from cell lysates; a detailed protocol is provided in the supporting information (S.5). 2.5 μL of extracted RNA solution was used as

the template for reverse transcription with primers specific for either *E. coli* $tRNA^{Arg2}$ (primer LW, Table S3) or *M. jannaschii* $tRNA^{Opt}$ (primer LZ, Table S3). Reverse transcription was performed using SuperScript IV reverse transcriptase (Thermo Scientific) according to the manufacturer's instructions. Duplicate reactions to which no reverse transcriptase was added were included as negative controls to ensure subsequent PCR amplification resulted from tRNA as opposed to contaminating DNA. The cDNA products from the reverse transcription reactions were amplified with Taq polymerase (New England Biolabs) using primer LX and either primer LY or primer CC (Table S3). PCR products were purified using the GeneJET PCR Cleanup kit (Thermo Scientific) and analyzed by agarose gel electrophoresis. Amplified cDNA products were sequenced with primer LX (*M. jannaschii* $tRNA^{Opt}$ products) or primer LY (*E. coli* $tRNA^{Arg2}$) (Genewiz LLC). Sanger sequencing.ab1 files were converted to .xml files using abi2xml.exe (freely available at www.abi2xml.sourceforge.net. Figures were created by importing data lines from the .xml file into Microsoft Excel.

Construction of the tRNA Anticodon Loop Library

A 64 member library of nucleotide variants at positions 32, 37, and 38 of the *M. jannaschii* $tRNA^{Opt}_{AUG}$ (Figure 3.1a) was generated via Kunkel mutagenesis using primer JQ (Table S3). Plasmid pWB_Ultra-Tyr-(XhoI) has an XhoI restriction site in the anticodon position of the tRNA and was used as an inactive template for library construction. Primer JQ allows A, C, G, or U to be incorporated at positions 32, 37 and 38 and restores the AUG anticodon. Nearly all domains of life feature a conserved uridine at position 33 [14-16]; this position was not varied. 5 µg of pWB_Ultra-Tyr-(XhoI) ss dU DNA were used as template for library construction. Library DNA was electroporated into SS320 cells. 10 µL of the 1000 µL transformation recovery were plated to determine transformation efficiency. The electroporation yielded 4 $x10^9$ unique transformants. PCR products from 8 of 13 of these colonies were not able to be digested by XhoI, suggesting a mutagenic efficiency of approximately 60%.

tRNA molecules are inherently highly structured, and mutation of these polynucleotides is challenging. The introduction of an XhoI restriction site into the template DNA provides a handle for improving the efficiency of mutation in addition to utilization of uridine-enriched template DNA[39]. The remaining 990 µl of the library transformation recovery media was diluted 5-fold into LB/spectinomycin 50 µg/mL media and grown for 5 hours at 37 °C with shaking. Isolation of DNA from 5 mL of culture yielded 7.4 µg of double stranded DNA, 2 µg of which were restricted using XhoI (New England Biolabs) at 37 °C for 1 hour. Following PCR spin kit cleanup and elution in 35 µL ultrapure water, one-half of the restricted DNA was transformed into SB3930 cells harboring the pGFP66cau reporter plasmid. The electroporation yielded 4 x10^6 unique transformants. PCR products from 13 of 13 evaluated colonies were susceptible to XhoI digestion, suggesting that all transformants were library members.

RESULTS AND DISCUSSION

Motivated by a general interest in reassigning the meaning of sense codons for incorporation of ncAAs into proteins *in vivo*, we recently reported a screen to evaluate the sense codon reassignment potential of the *M. jannaschii* tyrosyl tRNA/aaRS orthogonal pair[35]. The screen exploits the absolute requirement for an active site tyrosine in green fluorescent protein (GFP). The screen monitors the restoration of GFP fluorescence by incorporation of tyrosine at position 66 in response to a sense codon typically assigned another meaning in the genetic code (Figure 3.2). When the sense codon at position 66 is read by *M. jannaschii* tyrosine-charged $tRNA^{Opt}$, tyrosine is incorporated, and the resulting protein is fluorescent. When the sense codon is decoded by an endogenous *E.coli* tRNA, another canonical amino acid is incorporated, and the resulting protein is not fluorescent. The screen provides a quantitative measurement of sense codon reassignment by providing a combined measurement of the extent to which the orthogonal *M. jannaschii* tyrosyl aaRS recognizes and aminoacylates an *M. jannaschii* tRNA with an alternative anticodon and the extent to which the altered orthogonal tRNA competes against *E. coli* tRNA species to decode the codon specifying the essential tyrosine position of GFP.

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

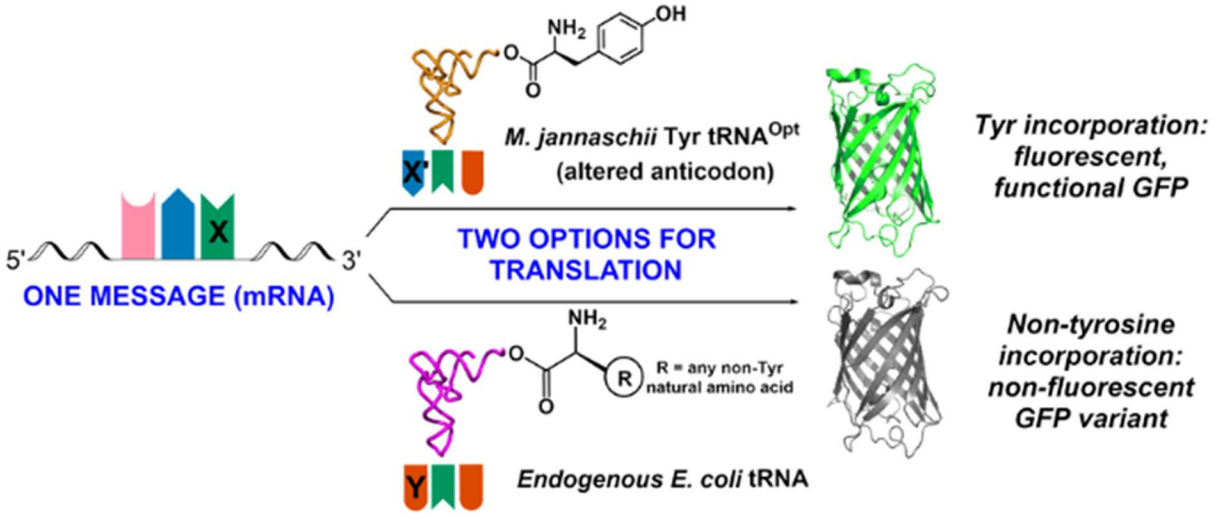


Figure 3.2 Depiction of the fluorescence-based screen utilized to evaluate sense codon reassignment.

The supplied sense codon suppressing $tRNA^{Opt}_{AUG}$, $tRNA^{Opt}_{AAA}$, and $tRNA^{Opt}_{AUU}$ each include an adenosine at position 34, a feature uncommon in *E. coli* tRNAs. Only a single *E. coli* tRNA, $tRNA^{Arg2}$, is genetically-encoded with an adenosine at position 34. Although the anticodon of $tRNA^{Arg2}$ is genetically

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

Evaluation of Inosine Modification of *M. jannaschii* $tRNA^{Opt}_{AUG}$

In comparing the sequences of the *E. coli* $tRNA^{Arg2}$ and the orthogonal *M. jannaschii* tRNA with an AUG anticodon, marked similarities and differences were apparent. The anticodon stems of the two tRNA species were largely different in sequence, sharing only one of five base pairs. The anticodon loops of

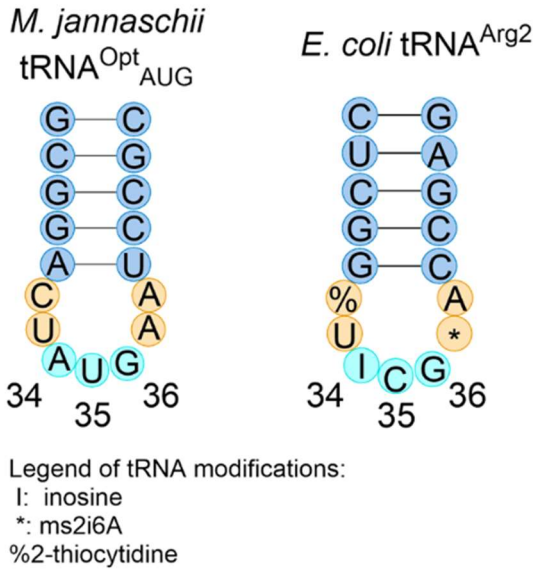


Figure 3.3 Comparison of anticodon stem loop sequences of *M. jannaschii* $tRNA^{Opt}_{AUG}$ and *E. coli* $tRNA^{Arg2}_{IUG}$. Modified bases in the *E. coli* tRNA are labelled. The inosine modification of A34 confirmed in this work is the only modification noted in *M. jannaschii* tRNA $tRNA^{Opt}_{AUG}$. Other modifications that have not yet been identified may exist.

precursor $tRNA^{Arg2}_{ACG}$ and $tRNA^{Opt}_{AUG}$ were remarkably similar in sequence, differing only by a single nucleotide, position 35, the middle of the anticodon (Figure 3.3). We hypothesized that $tRNA^{Opt}_{AUG}$ was a substrate for TadA, resulting in $tRNA^{Opt}_{IUG}$. An IUG anticodon would theoretically be able to decode CAC, CAU, and CAA codons and would explain the lack of discrimination observed for the 2 synonymous histidine codons. Phe and Asn codon reassigning $tRNA^{Opt}_{AAA}$ and $tRNA^{Opt}_{AUU}$ differ from the precursor $tRNA^{Arg2}_{ACG}$ at both positions 35 and 36. The combined differences make modification of $tRNA^{Opt}_{AAA}$ and $tRNA^{Opt}_{AUU}$ by TadA less likely.

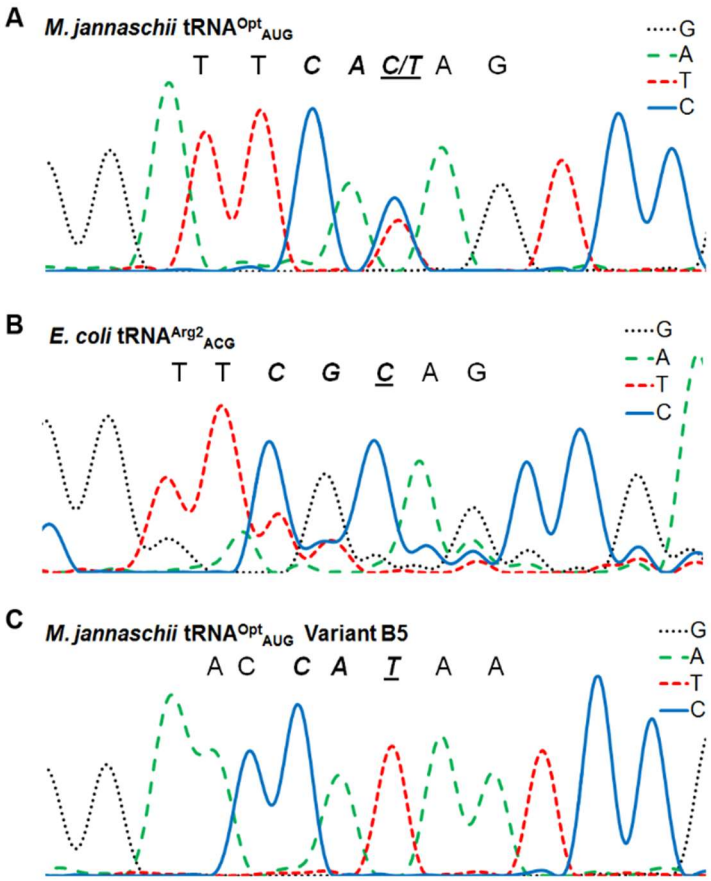


Figure 3.4 Sequence traces of reverse transcribed isolated tRNA species for (A) *M. jannaschii* $tRNA^{Opt}_{AUG}$, (B) E. coli $tRNA^{Arg2}$, and (C) *M. jannaschii* tRNAOptAUG variant B5 identified from the anticodon loop library (traces of additional isolated tRNA variants are shown in Figure S1). The sequenced stand corresponds to the reverse complement of the tRNA sequence. The portion of the chromatograms that correspond to the anticodon are italicized and bolded. The position corresponding to first anticodon position of the A to I modification is underlined. At position 34, observation of C is indicative of inosine modification; observation of T is indicative of unmodified adenosine.

Inosine modification of RNA may be detected by reverse transcription, amplification and sequencing of the resulting cDNA[43, 46]. Unmodified adenosine at position 34 of the tRNA is expected to trigger incorporation of only thymidine during reverse transcription. Inosine pairs most stably with cytidine and is expected to trigger significant incorporation of cytidine during reverse transcription. The complement of tRNAs was isolated from overnight cultures by phenol:chloroform extraction. Sanger sequencing of cDNA products replicated from the isolated tRNA fraction confirmed the incorporation of cytidine in response to the nucleotide at position 34 of both the *E. coli* $tRNA^{Arg2}_{ICG}$ and the *M. jannaschii* $tRNA^{Opt}_{AUG}$. The chromatogram for the *E. coli* $tRNA^{Arg2}$, which is presumably nearly quantitatively modified, shows almost exclusive incorporation of cytidine, indicating the presence of inosine (Figure 3.4B). The chromatogram for the original $tRNA^{Opt}_{AUG}$ shows an approximate 50:50 mixture of cytidine and thymidine (Figure 3.4A), suggesting that this tRNA is partially modified to inosine at position 34. The chromatograms for the $tRNA^{Opt}_{AUG}$ variants selected from the library show exclusive incorporation of thymidine (Figure 3.4C and S1), indicating that the tRNA is unmodified. The tRNA sequencing traces suggest that the lack of discrimination observed for the *M. jannaschii* $tRNA^{Opt}_{AUG}$ was the result of anticodon modification to IUG, leading to increased decoding of the CAC codon.

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

Directed Evolution of the *M. jannaschii* $tRNA^{Opt}_{AUG}$ Anticodon Loop Increases Sense Codon Reassignment and Prevents Inosine Modification of Adenosine 34

The TadA substrate recognition elements in $tRNA^{Arg2}$ have not been fully mapped, but the anticodon loop was shown to be an important factor[43]. We have reported that alterations to nucleotides flanking the anticodon are capable of increasing sense codon reassignment of Lys AAG codons by $tRNA^{Opt}_{CUU}$ [35]. We hypothesized that variation within the anticodon loop of $tRNA^{Opt}_{AUG}$ could simultaneously increase the efficiency of sense codon reassignment at CAU codons and decrease deamination of position 34 by TadA, leading to increased discrimination between CAU and CAC codons. We constructed a focused library of 64 $tRNA^{Opt}_{AUG}$ variants with diversity at nucleotide positions 32, 37, and 38 in the anticodon loop. The universal U33 was maintained.

The tRNA anticodon loop library was screened using a sfGFP reporter with a CAU codon at position 66. In a screen of 85 colonies from the library, 34 samples had detectable fluorescence and 17 colonies showed fluorescence equal to or greater than the original $tRNA^{Opt}_{AUG}$. The highly fluorescent colonies appeared to be clustered in groups of approximately equal fluorescence and 6 colonies representing the apparent groups were selected for further characterization. The vector DNA specifying the orthogonal translational machinery from each of these colonies was isolated and sequenced (Table 3.1). Colony G10 was identified as the original $tRNA^{Opt}_{AUG}$ with no mutations to the anticodon loop; this variant was not characterized further. Each of the other 5 variants had at least one mutation in the anticodon loop, and five of these clones had multiple mutations. Alterations at each position included in the library (32, 37, and 38) result in improved reassignment of the CAU codon.

Table 3.1 Sanger sequencing of reverse transcribed cDNA from isolated tRNA fractions

Variant	tRNA Nucleotide Number						
	32	33	34	35	36	37	38
$tRNA^{Opt}_{IUG}$	C	U	I	U	G	A	A
$tRNA^{Arg2}_{ICG}$	C[a]	U	I	C	G	A[b]	A
Colony B5	U	U	A	U	G	G	U
Colony C1	C	U	A	U	G	A	C
Colony E2	C	U	A	U	G	A	U
Colony G10	C	U	I	U	G	A	A
Colony G11	A	U	A	U	G	A	C
Colony H2	C	U	A	U	G	G	A

[a] Position is modified to thyocytidine
[b] Position is modified to 2-methylthio-N6-isopentenyladenosine

The five selected tRNA anticodon loop variants were evaluated for their ability to discriminate between the CAU and CAC codons. Each variant showed improved codon discrimination, suggesting that the mutations that improved the efficiency of sense codon reassignment of CAU also mitigated the A34I modification (Figure 3.5). The original $tRNA^{Opt}_{AUG}$ incorporated tyrosine in response to 5.1% of histidine codons, with a 3:2 preference for CAU over CAC codons. The incorporation of tyrosine in response to the CAC codon by the five $tRNA^{Opt}_{AUG}$ library variants was at or below the limit of detection for our in cell assay, 0.2% reassignment. The most improved systems expressing the tRNA variants selected from the library exhibit at least a 98:2 preference for CAU over CAC codons, a significant improvement compared

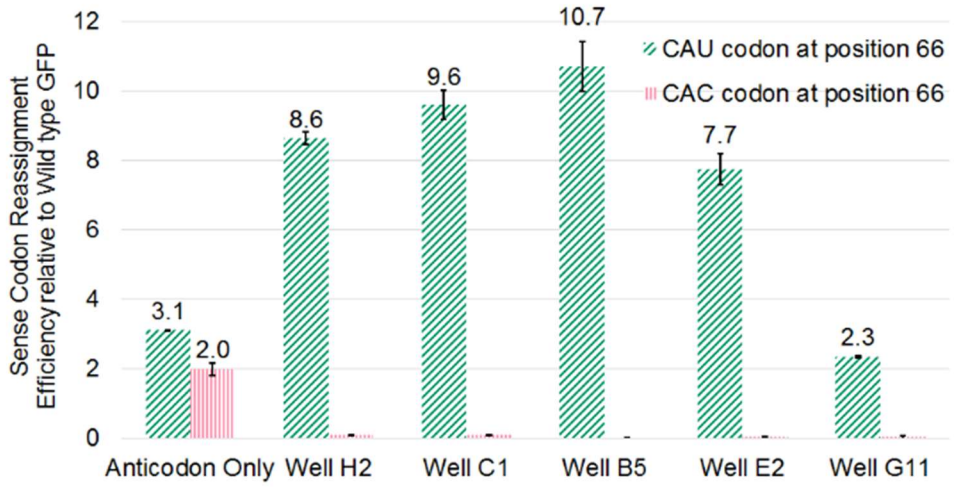


Figure 3.5 Codon discrimination for selected tRNA variants. Reassignment efficiencies reported as the observed optical density-corrected fluorescence divided by the optical density-corrected fluorescence of wild type GFP for each orthogonal pair and measured in response to CAU and CAC histidine codons specifying Y66 of GFP reporters.

to the 3:2 preference observed with the original $tRNA^{Opt}_{AUG}$. The tRNA fraction from cells expressing the $tRNA^{Opt}_{AUG}$ variants was isolated and sequenced. Sanger sequencing of reverse transcribed cDNA from the isolated tRNA fractions showed that all five clones had adenosine in position 34 rather than inosine (Figures 3.4C and S1). Although not all possible combinations of mutations were observed, mutations at positons 32, 37, or 38 of the $tRNA^{Opt}_{AUG}$ appear to eliminate recognition of $tRNA^{Opt}_{AUG}$ as a substrate for TadA and improve the reassignment efficiency of the targeted CAU codon.

Prevention of Inosine Modification Improves Health of Sense Codon Reassigning Cell Systems

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

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

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

CONCLUSIONS

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

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

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

ASSOCIATED CONTENT

Supporting Information is available in Appendix 2 in this dissertation.

REFERENCES

(1) Lang, K., and Chin, J. W. (2014) Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins, *Chem. Rev.* **114**, 4764-4806.

(2) Ngo, J. T., and Tirrell, D. A. (2011) Noncanonical amino acids in the interrogation of cellular protein synthesis, *Acc. Chem. Res.* **44**, 677-685.

(3) Dumas, A., Lercher, L., Spicer, C. D., and Davis, B. G. (2015) Designing logical codon reassignment - expanding the chemistry in biology, *Chem. Sci.* **6**, 50-69.

(4) de Duve, C. (1988) The second genetic code, *Nature* **333**, 117-118.

(5) Giege, R., Sissler, M., and Florentz, C. (1998) Universal rules and idiosyncratic features in tRNA identity, *Nucleic Acids Res.* **26**, 5017-5035.

(6) El Yacoubi, B., Bailly, M., and de Crecy-Lagard, V. (2012) Biosynthesis and function of posttranscriptional modifications of transfer RNAs, *Annu. Rev. Genet.* **46**, 69-95.

(7) Persson, B. C. (1993) Modification of transfer-RNA as a regulatory device, *Mol. Microbiol.* **8**, 1011-1016.

(8) Manickam, N., Joshi, K., Bhatt, M. J., and Farabaugh, P. J. (2016) Effects of tRNA modification on translational accuracy depend on intrinsic codon-anticodon strength, *Nucleic Acids Res.* **44**, 1871-1881.

(9) Limbach, P. A., Crain, P. F., and McCloskey, J. A. (1994) The modified nucleosides of RNA - summary, *Nucleic Acids Res.* **22**, 2183-2196.

(10) Gray, M. W. (2012) Evolutionary origin of RNA editing, *Biochemistry* **51**, 5235-5242.

(11) Alseth, I., Dalhus, B., and Bjoras, M. (2014) Inosine in DNA and RNA, *Curr. Opin. Genet. Dev.* **26**, 116-123.

(12) Bazak, L., Haviv, A., Barak, M., Jacob-Hirsch, J., Deng, P., Zhang, R., Isaacs, F. J., Rechavi, G., Li, J. B., Eisenberg, E., and Levanon, E. Y. (2014) A-to-i RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes, *Genome Res.* **24**, 365-376.

(13) Paul, M. S., and Bass, B. L. (1998) Inosine exists in mRNA at tissue-specific levels and is most abundant in brain mRNA, *Embo J.* **17**, 1120-1127.

(14) Dunin-Horkawicz, S., Czerwoniec, A., Gajda, M. J., Feder, M., Grosjean, H., and Bujnicki, J. M. (2006) Modomics: A database of RNA modification pathways, *Nucleic Acids Res.* **34**, D145-D149.

(15) Czerwoniec, A., Dunin-Horkawicz, S., Purta, E., Kaminska, K. H., Kasprzak, J. M., Bujnicki, J. M., Grosjean, H., and Rother, K. (2009) Modomics: A database of RNA modification pathways. 2008 update, *Nucleic Acids Res.* **37**, D118-D121.

(16) Machnicka, M. A., Milanowska, K., Oglou, O. O., Purta, E., Kurkowska, M., Olchowik, A., Januszewski, W., Kalinowski, S., Dunin-Horkawicz, S., Rother, K. M., Helm, M., Bujnicki, J. M., and Grosjean, H. (2013) Modomics: A database of RNA modification pathways-2013 update, *Nucleic Acids Res.* **41**, D262-D267.

(17) Furter, R. (1998) Expansion of the genetic code: Site-directed p-fluoro-phenylalanine incorporation in *Escherichia coli*, *Protein Sci.* **7**, 419-426.

(18) Wang, L., Brock, A., Herberich, B., and Schultz, P. G. (2001) Expanding the genetic code of *Escherichia coli*, *Science* **292**, 498-500.

(19) Liu, C. C., and Schultz, P. G. (2010) Adding new chemistries to the genetic code *Annu. Rev. Biochem.* **79**, 413-444.

(20) Wan, W., Tharp, J. M., and Liu, W. R. (2014) Pyrrolysyl-tRNA synthetase: An ordinary enzyme but an outstanding genetic code expansion tool, *Biochim. Biophys. Acta, Proteins Proteomics* **1844**, 1059-1070.

(21) Mukai, T., Hayashi, A., Iraha, F., Sato, A., Ohtake, K., Yokoyama, S., and Sakamoto, K. (2010) Codon reassignment in the *Escherichia coli* genetic code, *Nucleic Acids Res.* **38**, 8188-8195.

(22) Johnson, D. B. F., Xu, J. F., Shen, Z. X., Takimoto, J. K., Schultz, M. D., Schmitz, R. J., Xiang, Z., Ecker, J. R., Briggs, S. P., and Wang, L. (2011) RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites, *Nat. Chem. Biol.* **7**, 779-786.

(23) Isaacs, F. J., Carr, P. A., Wang, H. H., Lajoie, M. J., Sterling, B., Kraal, L., Tolonen, A. C., Gianoulis, T. A., Goodman, D. B., Reppas, N. B., Emig, C. J., Bang, D., Hwang, S. J., Jewett, M. C., Jacobson, J. M., and Church, G. M. (2011) Precise manipulation of chromosomes *in vivo* enables genome-wide codon replacement, *Science* **333**, 348-353.

(24) Dougherty, M. J., Kothakota, S., Mason, T. L., Tirrell, D. A., and Fournier, M. J. (1993) Synthesis of a genetically engineered repetitive polypeptide containing periodic selenomethionine residues, *Macromolecules* **26**, 1779-1781.

(25) Budisa, N., Steipe, B., Demange, P., Eckerskorn, C., Kellermann, J., and Huber, R. (1995) High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in *Escherichia coli*, *Eur. J. Biochem.* **230**, 788-796.

(26) Budisa, N. (2006) *Engineering the genetic code expanding the amino acid repertoire for the design of novel proteins*, Wiley-VCH; John Wiley [distributor], Weinheim.

(27) Bacher, J. M., and Ellington, A. D. (2001) Selection and characterization of *Escherichia coli* variants capable of growth on an otherwise toxic tryptophan analogue, *J. Bacteriol.* **183**, 5414-5425.

(28) Kwon, I., Kirshenbaum, K., and Tirrell, D. A. (2003) Breaking the degeneracy of the genetic code, *J. Am. Chem. Soc.* **125**, 7512-7513.

(29) Kwon, I., and Lim, S. I. (2014) Tailoring the substrate specificity of yeast phenylalanyl-tRNA synthetase toward a phenylalanine analog using multiple-site-specific incorporation, *ACS Synth. Biol*. **5**, 634-643.

(30) Krishnakumar, R., Prat, L., Aerni, H.-R., Ling, J., Merryman, C., Glass, J. I., Rinehart, J., and Söll, D. (2013) Transfer RNA misidentification scrambles sense codon recoding, *ChemBioChem* **14**, 1967-1972.

(31) Bohlke, N., and Budisa, N. (2014) Sense codon emancipation for proteome-wide incorporation of noncanonical amino acids: Rare isoleucine codon AUA as a target for genetic code expansion, *FEMS Microbiol. Lett.* **351**, 133-144.

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

(33) Lee, B. S., Shin, S., Jeon, J. Y., Jang, K.-S., Lee, B. Y., Choi, S., and Yoo, T. H. (2015) Incorporation of unnatural amino acids in response to the AGG codon, *ACS Chem. Biol.* **10**, 1648-1653.

(34) Ho, J. M., Reynolds, N. M., Rivera, K., Connolly, M., Guo, L.-T., Ling, J., Pappin, D. J., Church, G. M., and Söll, D. (2016) Efficient reassignment of a frequent serine codon in wild-type *Escherichia coli*, *ACS Synth. Biol.* **5**, 163-171.

(35) Biddle, W., Schmitt, M. A., and Fisk, J. D. (2015) Evaluating sense codon reassignment with a simple fluorescence screen, *Biochemistry* **54**, 7355-7364.

(36) Young, T. S., Ahmad, I., Yin, J. A., and Schultz, P. G. (2010) An enhanced system for unnatural amino acid mutagenesis in *E. coli*, *J. Mol. Biol.* **395**, 361-374.

(37) Pedelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C., and Waldo, G. S. (2006) Engineering and characterization of a superfolder green fluorescent protein, *Nat. Biotechnol.* **24**, 79-88.

(38) Varshney, U., Lee, C. P., and Rajbhandary, U. L. (1991) Direct analysis of aminoacylation levels of transfer-RNAs *in vivo* - application to studying recognition of *Escherichia coli* initiator transfer-RNA mutants by glutaminyl-transfer RNA-synthetase, *J. Biol. Chem.* **266**, 24712-24718.

(39) Holland, E. G., Acca, F. E., Belanger, K. M., Bylo, M. E., Kay, B. K., Weiner, M. P., and Kiss, M. M. (2015) *In vivo* elimination of parental clones in general and site-directed mutagenesis, *J. Immunol. Methods* **417**, 67-75.

(40) Meroueh, M., and Chow, C. S. (1999) Thermodynamics of RNA hairpins containing single internal mismatches, *Nucleic Acids Res.* **27**, 1118-1125.

(41) Agris, P. F. (2004) Decoding the genome: A modified view, *Nucleic Acids Res.* **32**, 223-238.

(42) Crick, F. H. C. (1966) Codon-anticodon pairing – the wobble hypothesis, *J. Mol. Biol.* **19**, 548-555.

(43) Wolf, J., Gerber, A. P., and Keller, W. (2002) TadA, an essential tRNA-specific adenosine deaminase from *Escherichia coli*, *Embo J.* **21**, 3841-3851.

(44) Auxilien, S., Crain, P. F., Trewyn, R. W., and Grosjean, H. (1996) Mechanism, specificity and general properties of the yeast enzyme catalysing the formation of inosine 34 in the anticodon of transfer RNA, *J. Mol. Biol.* **262**, 437-458.

(45) Gerber, A. P., and Keller, W. (1999) An adenosine deaminase that generates inosine at the wobble position of tRNAs, *Science* **286**, 1146-1149.

(46) Motorin, Y., Muller, S., Behm-Ansmant, I., and Brantant, C. (2007) Identification of modified residues in RNAs by reverse transcription-based methods, In *RNA modification* (Gott, J. M., Ed.), pp 21-53.

(47) Dong, H. J., Nilsson, L., and Kurland, C. G. (1996) Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates, *J. Mol. Biol.* **260**, 649-663.

CHAPTER 4

MEASURING TRANSFERABILITY OF MODIFICATIONS TO ORTHOGONAL TRANSLATION SYSTEMS FOR GENETIC CODE EXPANSION[3]

OVERVIEW

Methods to expand the genetic code to increase the chemical diversity of proteins have recently been widely developed. Many methods rely on genomic modifications to host cells and may be combinable with other methods to produce even higher efficiency genetic code expansion systems. We recently reported on an analytical method to quantitate the extent of reassignment of sense codons *in vivo* that enables screening of directed evolution libraries. Our screen relies on the incorporation of the natural amino acid Tyr from the commonly used orthogonal aminoacyl -tRNA synthetase (aaRS) and tRNA pair from *Methanocaldococcus jannaschii*. In order to incorporate non-canonical amino acids (ncAA) in response to sense codons, we attempted to transfer improvements to the TyrRS and $tRNA^{Opt}$ identified in a Tyr incorporating library to variants of the same aminoacyl tRNA synthetase and $tRNA^{Opt}$ specific for the ncAA para-azidophenylalanine (pAzF). Improvements identified in the Tyr library which increased sense codon reassignment (SCR) efficiency two-fold also increased reassignment efficiency for the pAzF specific variant almost two-fold. We combined our improvements from the directed evolution approach with a genomic engineering strategy to knockout the competing arg-$tRNA_{CCU}$ from *E. coli* encoded by the gene argW. When we attempted to quantify the effects of removing arg-$tRNA_{CCU}$ on SCR, we discovered that argW does not significantly affect SCR at Arg AGG codons. This was confirmed by performing experiments using materials provided by the original authors of the argW knockout paper. In the process

3. This chapter is a manuscript in preparation for publication. Contributing authors: Biddle,W., Schmitt, M.A., Schwark, D.G., and Fisk, J.D.

of evaluating their system, quantitation of SCR for the pEVOL1 plasmid backbone showed very high levels of reassignment efficiency: over 90%. This high level of reassignment efficiency could have contributed to the confusion about the effects of argW knockout. We attempted to combine our improved aaRS/tRNA pair from the directed evolution library with the high level of reassignment observed in the pEVOL1 plasmid system. Surprisingly, although unmodified aaRS specific for Tyr and pAzF reassign the AGG codon with high efficiency in this system, the library improvements which doubled incorporation in our pWB_Ultra plasmid system lead to a decrease in reassignment efficiency in the pEVOL1 backbone. This emphasizes that more work is needed to understand the variables that affect transferability of engineered improvements for orthogonal translation systems for genetic code expansion.

INTRODUCTION

Increasing the chemical diversity of proteins through the genetically encoded introduction of non-canonical amino acids (ncAAs) is a powerful tool for the precise modification of proteins. Two general approaches to expand the genetic code, nonsense suppression and global amino acid reassignment, have been widely employed. Advances in both technologies have improved the efficiency with which a 21st amino acid can be genetically encoded. Nonsense suppression technology has been improved through genomic modifications enabling the elimination of the translation factor that competes for decoding of the amber stop codon as a termination signal. The global amino acid reassignment strategy has been updated by breaking the degeneracy of the genetic code to enable the reassignment of individual sense codons.

Nonsense suppression targeting one of the codons typically read as a stop signal has been the most widely used method for introduction of a 21st amino acid into the genetic code. The amber stop codon is employed principally because it is the least commonly used codon in the *E. coli* genome. *In vivo* nonsense codon suppression involves the introduction of an orthogonal aaRS engineered to recognize and attach a ncAA onto its cognate, orthogonal tRNA. The orthogonal tRNA is not recognized by the set of *E. coli* aaRSs and the orthogonal aaRS does not recognize the complement of *E. coli* tRNAs. Orthogonal pairs are typically derived from phylogenetically distant organisms, and more recently, engineered orthogonal

aaRS/tRNA pairs have been identified.[1–4] Most amber stop codon suppression in *E. coli* is achieved using derivatives of only two orthogonal pairs: the tyrosyl tRNA synthetase (TyrRS) and its cognate tRNA ($tRNA^{Opt}$) from *M. jannaschii* and the pyrrolysine pair from the archaeal *Methanosarcina* genus. Variants of these two pairs that recognize, aminoacylate, and incorporate over 100 different ncAAs have been developed (reviewed by Dumas *et al.* [5]).

The efficiency of nonsense suppression has been limited as a result of competition with the natural termination machinery, which reads the codon as a stop signal, and also because engineered ncAA-activating orthogonal aaRSs are not as efficient as natural aaRSs. The introduction of an ncAA in response to a stop codon has the benefit of truncating any protein that fails to incorporate the ncAA, leading to high purity of the modified protein. The disadvantage of suppressing a stop codon is that competition with the normal termination functions leads to exponentially reduced protein yields when attempting to incorporate more than one copy of an ncAA. Typical efficiency of incorporation of ncAAs in response to stop codons averages about 25% per position. The low substitution frequency and high frequency of termination render incorporation of multiple ncAAs into a single protein difficult. Recent reports from the Wang, Church, and Sakamoto labs have each described different *E. coli* genome engineering strategies to improve the efficiency of nonsense suppression. All three methods involve compensating modifications to enable the removal of the ribosomal release factor 1 (RF1) that recognizes the amber stop codon.[6–10] All three methods essentially eliminate termination pressure and enable multiple copies of non-canonical amino acids to be introduced into the same protein in response to the amber stop codon. However, these methods are difficult to extend to other organisms or to expand to more than one "freed" codon. The performance of the various engineered cell lines in other applications suggest that some limitations on the efficiency of nonsense suppression are still present.[11,12]

Global amino acid reassignment involves growing cells under conditions in which one of the natural amino acids is replaced with a close structural analog. Global amino acid reassignment produces an altered 20 amino acid genetic code where every instance of one of the natural amino acids in a protein

is substituted with a ncAA. The most widely employed use of global amino acid reassignment is replacement of methionine by selenomethionine in proteins for X-ray crystallographic phasing.[13] Global amino acid reassignment typically requires an auxotrophic strain and relies on controlled minimal media to force substitution of the ncAA. The auxotrophic cell requirement and minimal media conditions can limit the expression of proteins relative to typical over expression systems. The approach has been expanded through the use of editing deficient, binding pocket modified and/or exogenous orthogonal aminoacyl tRNA synthetase pairs to increase the set of ncAAs that can be employed.

The use of orthogonal pairs to direct the incorporation of ncAAs in response to single sense codons allows the degeneracy of the genetic code to be broken. 18 of the 20 canonical amino acids are encoded by more than one codon. Reassigning a sense codon to an ncAA while leaving one (or more) other sense codon available to direct incorporation of the natural amino acid expands the genetic code to 21 amino acids. The original demonstration of breaking the degeneracy of the genetic code reassigned the Phe UUU codon, leaving the UUC codon available for incorporation of Phe.[14] Sense codon reassignment should be broadly generalizable. Unfortunately, predicting which sense codons are amenable to reassignment and which orthogonal machinery is best suited for the task is made challenging by the largely unknown and idiosyncratic recognition and discrimination features of each organism's complement of tRNAs and aaRSs.

Recent reports of breaking the degeneracy of the genetic code have targeted codons that are used infrequently in the *E. coli* genome. One of the factors affecting the efficiency of reassignment is the relative rate of tRNA sampling by the ribosome; rare codons are often read by low abundance tRNA species that should be easier to outcompete. Additionally, abundant substitution at rarely used codons is expected to limit the proteome-wide effects of reassignment. Incorporation of ncAAs in response to Ser AGU, Leu CUG, Ile AUA, Phe UUU, and Arg AGG and AGA codons has been investigated.[14–19]

In *E. coli*, the rarely used Arg AGG and AGA codons are decoded by two rare tRNAs with either a CCU or UCU anticodon. Both of the codons are expected to be read by the tRNA with the UCU

anticodon based on wobble pairing rules. The tRNA with the CCU anticodon should contribute to decoding only the AGG codon. As a result, the Arg AGG codon has been a target for sense codon reassignment in *E. coli*. Both of the most commonly used orthogonal pairs have been engineered to utilize a tRNA with a CCU anticodon. Two research groups have reported AGG reassignment using the orthogonal pyrrolysyl aaRS/tRNA pair. The Liu group reported reassignment levels of approximately 90% by using media where the Arg concentration was controlled.[20] The Sakamoto group reengineered a strain of *E. coli* to reassign the 38 AGG codons found in essential genes and subsequently removed the gene for $tRNA_{CCU}$ from the genome, allowing incorporation of a close structural analogue of Arg at very high levels.[16] The Yoo group described AGG codon reassignment using the *M. jannaschii* TyrRS/$tRNA^{Opt}$ pair.[17] To remove competition for the AGG codon, the *E. coli* genome was modified to remove the gene encoding arg-$tRNA_{CCU}$, argW, which resulted in high levels of incorporation.

That argW knockout cells were viable and exhibited only a slightly reduced growth effect, suggesting that the remaining Arg tRNA with a UCU anticodon can read AGG codons to some extent; double knockouts of both rare Arg tRNAs are not viable. [16] Lee and co-workers reported that by removing the argW gene encoding the endogenous *E. coli* Arg-$tRNA_{CCU}$, the AGG codon could be nearly quantitatively reassigned using the para-azidophenylanine (pAzF) aaRS.[17] The quantitative reassignment efficiency assessment was based on Western Blots and electrospray ionization mass spectrometry (ESI-MS) analysis.

Although the technologies for genome modification are continuously improving, genetic code expansion as a result of genomic changes limits the selection of targetable sense codons and requires reengineering and optimization in order to expand the genetic code of other organisms. Single gene knockouts tend to be of minimal difficulty, but removal of targeted codons from the genome or larger scale genome refactoring often require a prohibitive amount of work. Wholesale genome rewriting and adjustment of codon usage ratios may have unintended consequences on the life cycle of an organism. Changes to DNA sequence can result in changes to regulatory elements including binding sites and

secondary structural propensities may disrupt the delicate balance of replication, transcription and translation an organism has optimized through evolution.

Improving the interactions between an orthogonal aaRS/tRNA pair via directed evolution is another strategy for increasing the efficiency of codon reassignment. The anticodon is often an important identity element that allows a specific aaRS to recognize its appropriate tRNA. Changing the anticodon to target different codons can affect the efficiency with which an amino acid is attached to the tRNA. Increasing the recognition between the altered anticodon of the tRNA and the anticodon binding domain of the aaRS is expected to lead to a higher concentration of aminoacylated tRNA, which is expected to lead to better kinetic competition against endogenous tRNAs capable of decoding the targeted codon.

Here we discuss our efforts to improve reassignment of the AGG codon in *E. coli* by improving the interactions between the *M. jannaschii* $tRNA^{Opt}$ with a CCU anticodon and the orthogonal *M. jannaschii* tyrosyl-tRNA synthetase (TyrRS).[21] Utilization of a fluorescence-based screen that measures the efficiency of introduction of Tyr in response to non-Tyr sense codons allowed the screening of a large library of tRNA and aaRS variants via fluorescence activated cell sorting (FACS).[22] The fluorescence-based screen relies on the reassignment of an AGG codon in the critical fluorophore position of green fluorescent protein (GFP) to Tyr (instead of the naturally encoded Arg) in order to produce a mature fluorophore. In this system, GFP fluorescence serves as a direct measurement of reassignment of the AGG codon as proteins produced as the result of endogenous *E. coli* Arg-tRNA decoding will incorporate Arg into the fluorophore and will produce a non-fluorescent, full length protein. Although we are directly measuring only a single AGG codon in a reporter protein, we assume that sense codon reassignment occurs at AGG codons across the proteome and that the GFP screen approximates proteome-wide reassignment. The extent to which mutations to aaRS/tRNA pairs that improved reassignment of AGG codons to tyrosine were transferable to aminoacyl tRNA synthetases evolved for ncAAs was also examined. As aaRS/tRNA mutations and genomic modifications should be complementary, the ability to combine improved orthogonal pairs with tRNA-deleted cell strains was evaluated.

Our attempts to combine selected improved synthetase and tRNA variants with genome modifications suggest the surprising conclusion that competition with endogenous rare tRNAs is not an important contributor to the efficiency of sense codon reassignment.

RESULTS AND DISCUSSION

Directed Evolution of *M. jannaschii* $tRNA_{CCU}$/TyrRS for Improved Reassignment of AGG Codons

A previously described fluorescence-based screen was used to evaluate the *in vivo* efficiency of reassigning the AGG codon from Arg to Tyr by monitoring the recovery of fluorescence in GFP reporter protein variants where the central fluorophore Tyr position is encoded by a non-Tyr codon.[22] The efficiency of AGG reassignment was quantified in cells expressing the *M. jannaschii* tyrosyl aaRS (TyrRS) and its cognate $tRNA^{Opt}_{CCU}$ in combination with a GFP variant including an AGG codon at position 66. Simply altering the anticodon of the tRNA to Watson-Crick base pair with the AGG codon results in highly efficient reassignment of 56.6% (Figure 1B). The high efficiency of sense codon reassignment is evident in rich media without any modifications to the genome or orthogonal TyrRS or $tRNA^{Opt}$ genes beyond the anticodon change.

To improve the efficiency of AGG reassignment, a 10^{10} member library combining *M. jannaschii* tRNA anticodon loop and aaRS anticodon binding domain variants was constructed and screened. The library included diversity at nine amino acid positions in the aaRS and three of four nucleotides in the tRNA anticodon loop outside of the anticodon (positions 32, 37, and 38; Figure 4.1A). The varied amino acid positions were chosen based on proximity to the nucleotides of the anticodon in the co-crystal structure of the *M. jannaschii* tyrosyl aaRS/tRNA orthogonal pair (Figure 4.1B, pdb reference 1j1u)[23]. Amino acid diversity at each of the positions was limited in order to keep the library size close to the number of variants that could be readily screened. Degenerate codons for each position were chosen to include the wild type *M. jannaschii* aaRS sequence. A similar library strategy was previously used to

select *M. jannaschii* tyrosyl aaRS/tRNA variants with improved reassignment of the Lys AAG codon.[22] Four of the positions were allowed to vary between four amino acids, and greater diversity was included at the remaining five positions in the aaRS where between 9 and 15 amino acids were available (Table 4.1). Three of the four nucleotides flanking the anticodon in the tRNA were allowed to vary; the universally conserved U33 was preserved. The library had a theoretical diversity of 1.7×10^{10}. The constructed library contained 3.0×10^{8} unique transformants, which was amplified approximately 25-fold to 7.2×10^{9} cells in order to express GFP. Approximately 10^{8} cells were screened in the first round of fluorescence activated cell sorting (FACS).

The combined tRNA anticodon loop/aaRS anticodon binding domain library was subjected to three rounds of FACS screening through which the brightest 1.4%, 0.6% and 1.76% (approximately 1%) of cells were collected, amplified and rescreened. After the third round of FACS screening, collected cells were plated on IPTG, and 84 of the most visibly green colonies were selected for single colony analysis in the 96 well plate fluorescence-based screen. 75 of the 84 clones showed increased fluorescence relative to the starting variant. Overall, the clones showed a wide range of reassignment efficiencies, with the vast majority reassigning the AGG codon at greater than 75% efficiency.

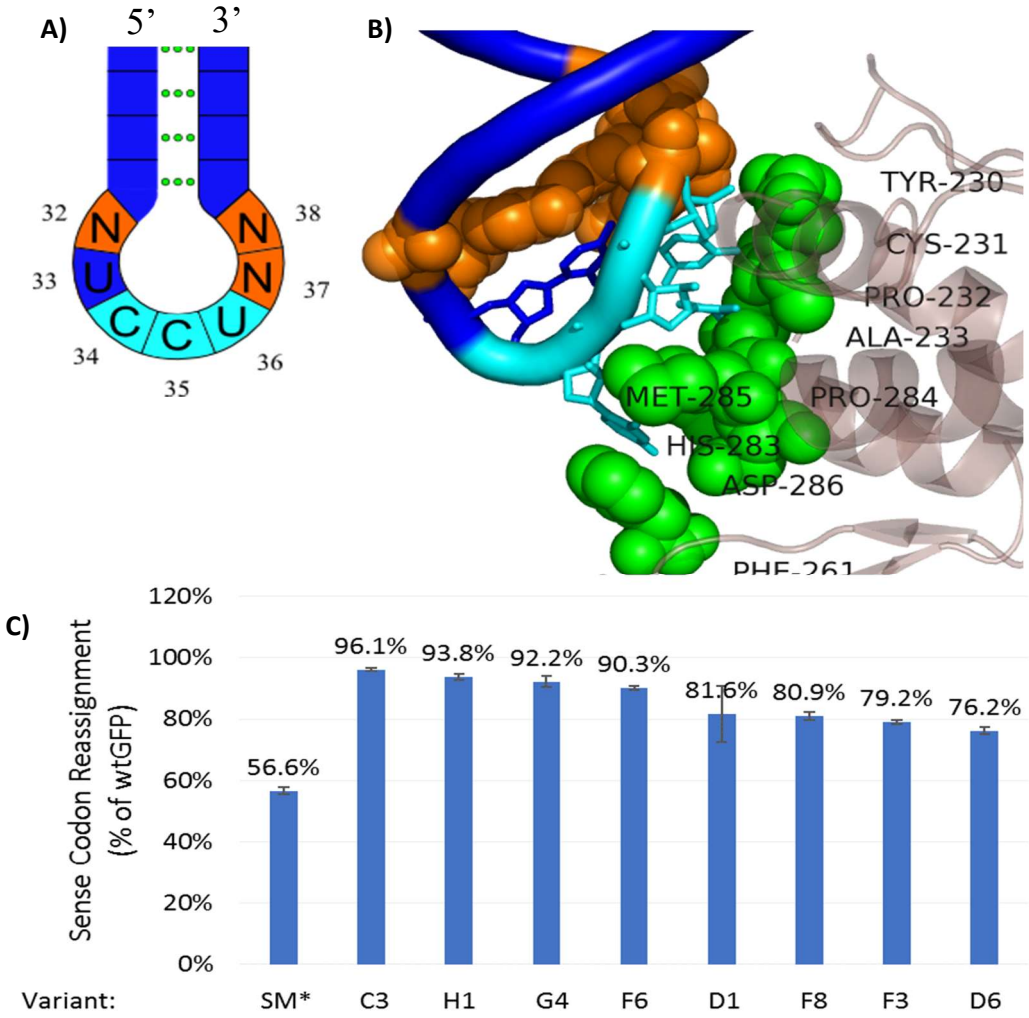


Figure 4.1. A) Cartoon schematic showing the anticodon-stem-loop of the orthogonal $tRNA^{Opt}$ used in this study. The anticodon nucleotides are shown in cyan, and the orange nucleotides indicate diversity positions in the library. U37 was conserved as this position is a uridine in all prokaryotes, and previous libraries did not show diversity at this position. B) A portion of the crystal structure of the MJ TyrRS complexed with its tRNA. The anticodon has been modified using PyMol software to represent the CCU anticodon. Residues represented as spheres show positions of diversity in the library; orange in the tRNA and green in the TyrRS. C) Sense codon reassignment efficiencies for eight clones from a combined TyrRS/tRNA library. SM represent the starting material for the library (TyrRS and $tRNA^{Opt}_{CCU}$). Variant titles represent clones selected from the directed evolution library and correspond to variants in Table 1. Values represent averages of three individual colonies of each variant. All values are given relative to the fluorescence level of a wtGFP reporter expressed in the same cells with the same pWB_Ultra reassignment vector.

Table 4.1 Library diversity and sequences of aaRS/ tRNA library residues from eight selected clones

	aaRS anticodon binding domain									tRNA anticodon loop						
TyrRS [a]	**Y**	**C**	**P**	**A**	**F**	**H**	**P**	**M**	**D**	**c**	**u**	**C**	**C**	**U**	**a**	**a**
Position [b]	**230**	**231**	**232**	**233**	**261**	**283**	**284**	**285**	**286**	**32**	**33**	**34**	**35**	**36**	**37**	**38**
Diversity in Library (Number of Residues/Nucleotides)	4	4	4	4	15	13	12	9	12	4	1	1	1	1	4	4
Diversity [c] (Available Residues/Nucleotides)	YADS	CYSF	PTKQ	AFSV	ADEFH IKLNP QSTVY	RNDQE GHILK MSV	ARDCG HLFPS YV	MLRW CFIT S	RNDQE GHKPS TA	n					n	n
Consensus [d]		+		++	+++	+++	+++	+++	+							
Clone C3 **	Y	F	Q	A	F	L	R	S	G	a	u	C	C	U	a	g
Clone F3 **	S	F	T	A	F	L	R	S	A	a	u	C	C	U	a	g
Clone F8	D	F	T	A	F	L	R	T	A	a	u	C	C	U	a	g
Clone C8/H1	A	Y	T	A	F	L	R	S	A	a	u	C	C	U	a	g
Clone D1	D	Y	T	A	Y	L	R	S	H	a	u	C	C	U	a	g
Clone E3/F6	S	Y	K	A	Y	L	R	S	N	a	u	C	C	U	a	g
Clone D6	A	Y	Q	A	F	Y	R	S	H	a	u	C	C	U	a	g
Clone G4 **	D	C	K	F	I	R	S	W	H	g	u	C	C	U	a	g

[a] Residues and nucleotides in the wild type TyrRS and $tRNA^{Opt}_{CCU}$ that was the starting point for the library
[b] Numbers correspond to either residue number for aaRS sequences or nucleotide number for tRNA sequences
[c] One letter abbreviations for amino acid residues or nucleotides available in the library based on codons used
[d] Similarity of clones selected in library based on properties of amino acid residues and available diversity at each position.
** indicates clones with additional mutations outside of residue positions varied in the library, see text.

In order to verify that the variants were functionally improved, and to further characterize their behavior, the aaRS/tRNA library plasmid was isolated from the cells for 10 of the AGG reassigning systems. DNA from the five most efficient apparent reassigning systems and five other representative highly efficient systems was isolated. Preparation of DNA from these cells resulted in a mixture of both the aaRS/tRNA vector and the GFP reporter plasmid with an AGG codon specifying the fluorophore Tyr residue. Unique DNA sequences within the regions flanking the aaRS/tRNA allowed complete sequencing of this region of the plasmid from the mixed preparation. Of the 10 clones selected, eight unique combined sequences for the *M. jannaschii* aaRS/tRNA were discovered (Table 1). DNA specifying the orthogonal translational machinery for the eight unique reassignment systems was separated from the plasmid encoding the GFP reporter via restriction digest. Clone H1 was selected to represent the sequence identified in both H1 and C8; clone F6 was selected to represent the sequence identified in both F6 and E3. Restriction sites unique to the GFP reporter were used to functionally eliminate that vector, and the restriction digests were transformed into cells. Transformations were plated onto antibiotics selective for one or both vectors to verify that the GFP reporter had been digested. Plating the transformed cells on spectinomycin, the antibiotic for which the orthogonal translation machinery plasmid confers resistance, yielded thousands of viable colonies. No viable colonies grew on the plates containing both spectinomycin and carbenicillin; the latter is an antibiotic selective for the GFP reporter, confirming successful digestion of the reporter.

The process of isolating the aaRS/tRNA vector from the GFP reporter vector with which it was selected in FACS, followed by another round of co-transformation with the GFP reporter and examination of reassignment efficiency, allowed analysis of multiple colonies of the same reassignment system. The process also served as verification that improvements in observed sense codon reassignment efficiency are related to the modifications in the aaRS/tRNA pair as opposed to a mutation or bias in the reporter plasmid or in the cell in which screening originally occurred. Analysis of three colonies of each of the eight selected aaRS/tRNA variants suggested a successful library selection process, as each variant

reassigned the AGG codon to Tyr with high efficiency (Figure 4.1C). Additionally, the colony to colony variation observed for each system was typical of that observed for these reassignment evaluations, suggesting that the aaRS/tRNA variants themselves were responsible for the increased reassignment efficiency. The variant with the highest reassignment efficiency, clone C3, reassigned the AGG codon to Tyr with 96% efficiency. This is a 1.6-fold improvement over the starting material, which reassigns the AGG codon to Tyr with 56% efficiency. When the same library strategy was used to evolve the TyrRS/$tRNA^{Opt}$ pair to reassign the Lys AAG codon, a five-fold improvement was observed over the starting material, but this is likely due to the low efficiency of the Lys starting material (6%).[22] The high level of reassignment of the AGG codon by Clone C3 suggests efficiencies comparable to natural *E. coli* aaRS/tRNA pairs, and to the levels of reassignment described by Lee and Mukai.[16,17]

Strong consensus was apparent across the unique clones identified. In the case of the tRNA anticodon loop, the nucleotides identified at position 32, 37, and 38 were identical for seven of the eight aaRS variants: 5'-auCCUag-3' (loop positions shown as lower case letters and upper case letters corresponding to the anticodon). The starting tRNA sequence was 5'-cuCCUaa-3'. None of the variants included a change at position 37, which is a conserved purine residue in prokaryotes.[24,25] The single clone (G4) that showed a different anticodon loop sequence was 5'-guCCUag-3'. Only position 32 differed from the other seven variants, and this change was accompanied by unique residues in the selected aaRS. The amino acids selected at six of nine varied positions in clone G4 are not identified in any of the other variants.

Three out of the eight unique clones we selected contained mutations in the aaRS sequence outside of the nine library positions we varied. Clones G4 and F3 had K228N mutations, and clone C3 showed E221K and R223G. All three of these mutations exist in an α-helix of the aaRS near the anticodon binding domain. The co-crystal structure of the *M. jannaschii* TyrRS complexed with its cognate tRNA with a GUA anticodon shows the ε- amine of K228 is within 2.9Å of a carbonyl oxygen in the tRNA phosphate backbone at position 29. In its most extended conformation, the Asn side chain is 2Å shorter

than a Lys side chain. Mutation of position 228 to Asn from Lys could be bringing the tRNA closer to the aaRS. This spontaneous mutation was also discovered in three of ten selected clones of the same library with a CUU Lys anticodon.[22] The E221K and R223G mutations in clone C3 occur near the distal end of the α-helix relative to the tRNA. These mutations may make structural changes to the α-helix, but likely do not interact directly with the tRNA molecule.

Effect on Cell Health/Growth

Although the high levels of sense codon reassignment could have resulted in reduced doubling times of the selected clones from the library due to damaged cell regulatory elements, this was not observed. In fact, the instantaneous doubling times of cells in their exponential phase of growth are faster than the doubling times for the starting material. Bacterial cells are largely tolerant of induced missense mutations, and growth defects only appear when the protein quality control machinery is overwhelmed.[26,27] The tolerance toward almost complete reassignment of AGG codons is likely due to the low usage of the AGG codon in the *E. coli* genome. Additionally, the conditions of the screen used to select clones with high reassignment efficiencies, involving repeated selections and amplifications, are also selecting against slow-growing cells.

Transferability of Mutations to the aaRS that Improved Reassignment Between aaRSs for Different Amino Acids

The primary goal for genetic code expansion is the incorporation of ncAAs. Just two orthogonal aaRS/tRNA pairs have been employed to incorporate the vast majority of ncAAs in response to amber stop codons: the *M. jannaschii* tyrosyl aaRS/tRNA pair and the pyrrolysine aaRS/tRNA pair from either *M. barkeri* or *M. mazei*. Our fluorescence based screen utilizes the *M. jannaschii* Tyr-incorporating aaRS as a stand-in for *M. jannaschii* aaRSs variants that activate ncAAs. The extent to which the anticodon binding domain (varied based upon the codon targeted for reassignment) and amino acid activation domain (varied based upon the amino acid being incorporated) of the aaRS interact in aminoacyl tRNA synthetases has not been intensely studied. The extent that modifications improving incorporation via

recognition between the tRNA and aaRS can be transferred between synthetase variants has not been investigated. The extent that tRNA mutations that improve selected ncAA incorporation in response to amber codons transfer between synthetases specific for different ncAAs suggest that the extent of transfer is idiosyncratic.[2,21]

The *M. jannaschii* tyrosyl aaRS evolved for incorporation of para-azidophenylalanine (pAzFRS) has been used to incorporate pAzF in response to both amber stop and Arg AGG codons.[17,28] pAzF has been widely used for protein crosslinking and derivatization via copper catalyzed Huisgen cyclization with alkynes. Additionally, pAzF produces an altered but observable fluorescence when incorporated into the fluorophore of GFP[29,30,31]. Fluorescence from pAzF incorporation into the fluorophore results in a slightly blue shifted fluorescence relative to the maximum wavelengths of sfGFP.[31]

The extent to which improvements in reassignment efficiency selected using the TyrRS can be transferred to the pAzFRS was evaluated both using the both the fluorescence-based screen and mass spectrometry. The mutations found to improve Tyr incorporation in response to AGG codons in clone C3 were transferred to a pAzF specific synthetase.[28] The fluorescence of cells expressing GFP with an AGG codon specifying the fluorophore position 66 in combination with either pAzFRS/$tRNA_{CCU}$ or the clone C3 pAzFRS/$tRNA_{CCU}$ mutations was monitored in the presence and absence of pAzF. Meaningful *in vivo* fluorescence was not observed for these cultures, possibly as a result of the azide not reducing to the amine *in vivo* as had been noted previously, or possibly because measurements were being made in cells in rich media, both of which have a level of background fluorescence. The modified fluorophore of sfGFP with an amine group replacing the hydroxyl side chain of the Tyr results in a blue shifted and decreased fluorescence. The quantum yield of pAzF containing sfGFP that had been reduced to the amine side chain was measured to be approximately 27% of the wild type sfGFP when quantum yield is measured at the maximum wavelengths for both protein species.[31] A slight increase in fluorescence was observed after irradiation of the cell cultures with UV light as described by Reddington, et al., however it was difficult to control this method for uniform irradiation of each culture. Instead, proteins from cell cultures where

pAzF was incorporated into the fluorophore of sfGFP either at an AGG codon or at a UAG stop codon were purified via an N-terminal hexahistidine tag. The fluorescence of the purified proteins was evaluated, and the concentration of protein in each sample was estimated from PAGE gels using GFP standards as a calibration curve. The observed fluorescence of each purified protein sample was normalized using the estimated protein concentration to derive a value for fluorescence per protein. The assumption that every protein produced from a control system in which the GFP reporter includes an amber stop UAG codon at position 66 contains the pAzF in the fluorophore is required to identify a value for 100% pAzF incorporation. This assumption is somewhat invalidated by clear evidence of Tyr incorporation in systems to which no pAzF was added. The amount of incorporation is low, but detectable.

The pAzFRS/$tRNA_{CCU}$ pair was able to reassign the AGG codon in the fluorophore of sfGFP to pAzF at 29% in rich media relative to the fluorescence / protein of stop codon incorporated pAzF-sfGFP expressed under the same conditions (Figure 4.2b). In the absence of the ncAA in the growth medium, fluorescence corresponding to 6% reassignment was detected, which likely represents misaminoacylation of the $tRNA_{CCU}$ by the pAzFRS. The pAzFRS/tRNA variants with the improvement mutations from clone C3 were able to reassign the AGG codon at 50%, a 1.7-fold improvement over the original pAzFRS/$tRNA_{CCU}$ pair. The clone C3 variant pair also had two-fold higher background in the absence of ncAA. This improvement is consistent with the increase in reassignment efficiency we observed for the TyrRS and tRNA pair (Figure 4.2a), suggesting that improvements made to the aaRS and tRNA for reassigning Tyr at AGG codons also increased the efficiency of reassigning AGG codons to the ncAA pAzF.

In order to confirm the increase in efficiency of pAzFRS and $tRNA_{CCU}$ variants with mutations from clone C3, electrospray ionization mass spectrometry (ESI-MS) was used to approximate the levels of incorporation. The Z domain of proteinA is a small 8.3 kDa peptide and can be expressed

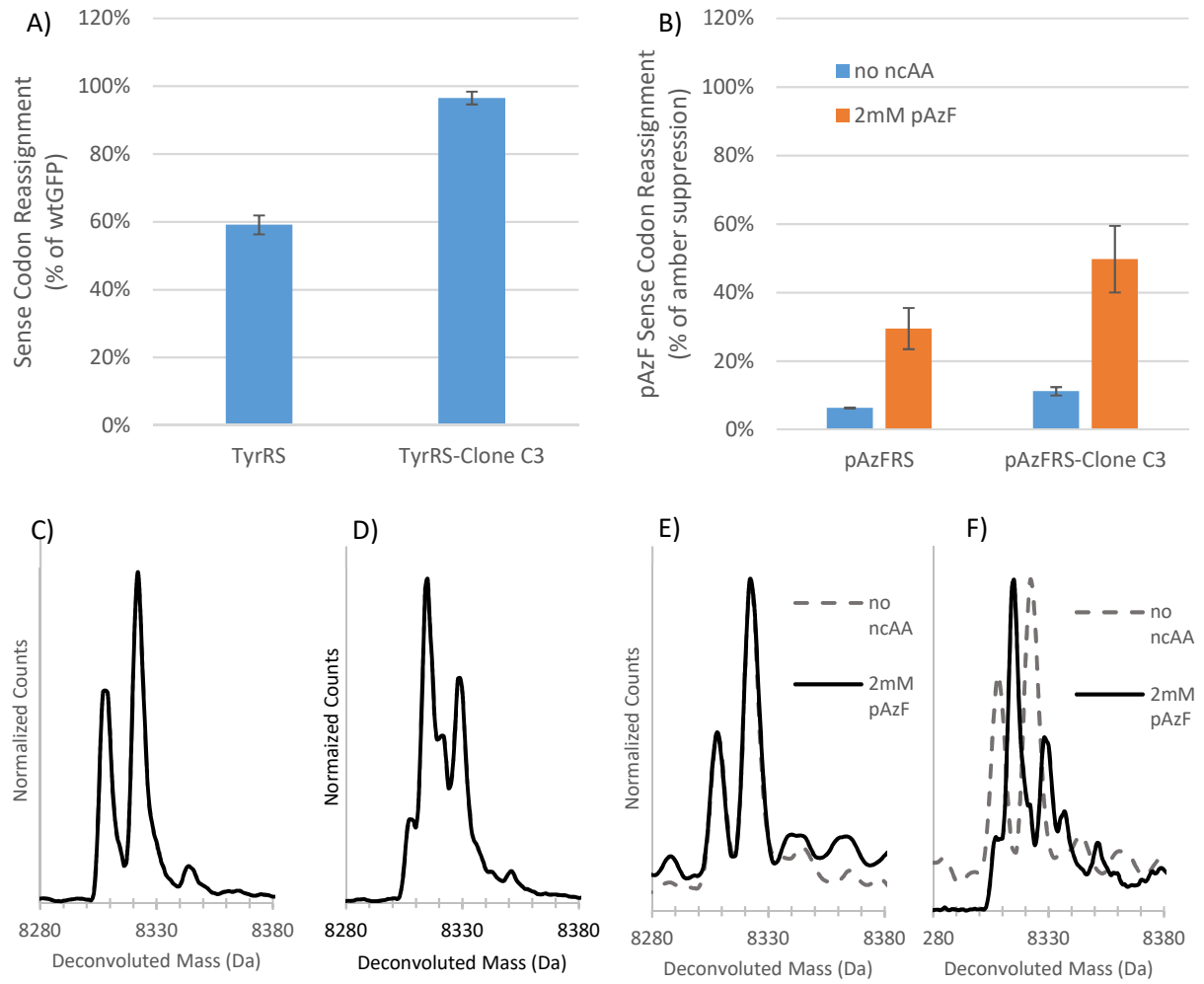


Figure 4.2 Transferability of modifications for a TyrRS/tRNA pair to a pAzFRS/tRNA pair. A) Sense codon reassignment of TyrRS/tRNA$^{Opt}_{CCU}$ and the clone C3 improved variants in pWB_Ultra backbone relative to levels of wtGFP expressed from the same system. B) Sense codon reassignment of pAzF incorporation at AGG codons expressing pAzFRS and tRNA$^{Opt}_{CCU}$ and the clone C3 improved variants from pWB_Ultra. Reassignment percent is determined from fluorescence per protein estimates from clarified lysates, relative to suppression of an amber stop codon using tRNA$^{Opt}_{CUA}$. C-F) ESI MS of the Z-domain peptide with incorporation of either Tyr (+7Da) or pAzF (+32Da) at the single AGG codon. Calculated mass of the Z domain peptide with arginine at the single AGG codon is 8308 Da. The +14Da shift seen on all peaks is likely due to a methylation of the Z-domain. Dotted lines in E) and F) represent proteins expressed in the absence of ncAA and solid lines in these figures were expressed in the presence of 2mM pAzF. C) TyrRS/tRNA$^{Opt}_{CCU}$ reassigning AGG codons to Tyr. D) Clone C3 improved variant of TyrRS/tRNA$_{CCU}$ reassigning Tyr at the AGG codon. E) pAzFRS/tRNA$_{CCU}$ reassigning AGG codons to pAzF. F) pAzFRS/tRNA$_{CCU}$ with improvement mutations from clone C3 reassigning AGG to pAzF.

independently as a soluble tri helix peptide.[32] This reporter protein had already been shown to work well with ESI-MS techniques.[17] The Z domain peptide with an F5R mutation to the AGG codon (see sup info for protein sequence) was expressed in cells reassigning the AGG codon to tyr or pAzF using both the

unimproved and clone C3 improvements to the aaRSs and tRNAs. Z domain proteins were purified and analyzed using ESI-MS of the intact peptide followed by deconvolution of the mass spectra using the Maximum Entropy algorithm (MassHunter Software, Agilent Technologies). Figures 4.2c-f show the deconvoluted mas spectra for Z domain proteins produced from various codon reassignment systems. In both of the systems with unimproved aaRS/$tRNA_{CCU}$ pairs, reassignment of the AGG codon is not detectable in the mass spectrum (Figures 4.2c and 4.2e). However, we see approximately the same level of reassignment of for the TyrRS and pAzFRS with clone C3 mutations in the aaRS and the $tRNA_{CCU}$ (Figures 4.2d and 4.2f). Although we cannot reconcile the apparent difference in reassignment efficiencies in the Z domain peptide as measured by ESI-MS and the sfGFP fluorescence screen, the trends in both reporter systems suggest that improvements selected from a library for the reassignment of Tyr at AGG codons also increase the reassignment of pAzF at AGG codons.

Combining Genome Modifications with Directed Evolution Improvements (argW knockout)

The improvements to sense codon reassignment efficiency made via modifying tRNA and aaRS interactions should be separate from and compatible with improvements that result from removing competition through genomic engineering approaches. The improved AGG-reassigning aaRS/tRNA clone C3 pair was evaluated in a cell strain with one of the simplest of the genome modifications: removal of the rare Arg $tRNA_{CCU}$ that Watson—Crick base pairs with the AGG codon and competes against the *M. jannaschii* tRNA for decoding the codon. The ArgW knockout of DH10B cells was prepared via the lambda red recombinase method to mimic the BS01 knockout cells reported by Lee and coworkers.[17] This cell strain has been designated BS01-Fisk. An additional knockout of argA renders the cells auxotrophic for Arg. The argW/argA dual knockout mimics the previously reported BS02 cells; this cell strain has been designated BS02-Fisk. Sequencing of the PCR amplifications of relevant segments of the chromosome in knockout cells and as well as tRNA amplification tests from cellular tRNAs indicated that the genomic knockouts were successful in removing the arg-$tRNA_{CCU}$ from DH10B cells (Supp Info).

Surprisingly, the reassignment of AGG to Tyr was less effective in the BS02-Fisk cells than in the parent strain DH10B (Figure 4.3). Additional argW knockouts were constructed from different starting strains. In all cases examined, *E. coli* DH10B, Top10, and C321.ΔA.exp argW tRNA knockout strains appeared to be less effective at AGG sense codon reassignment than $argW^{+}$ parent strains.

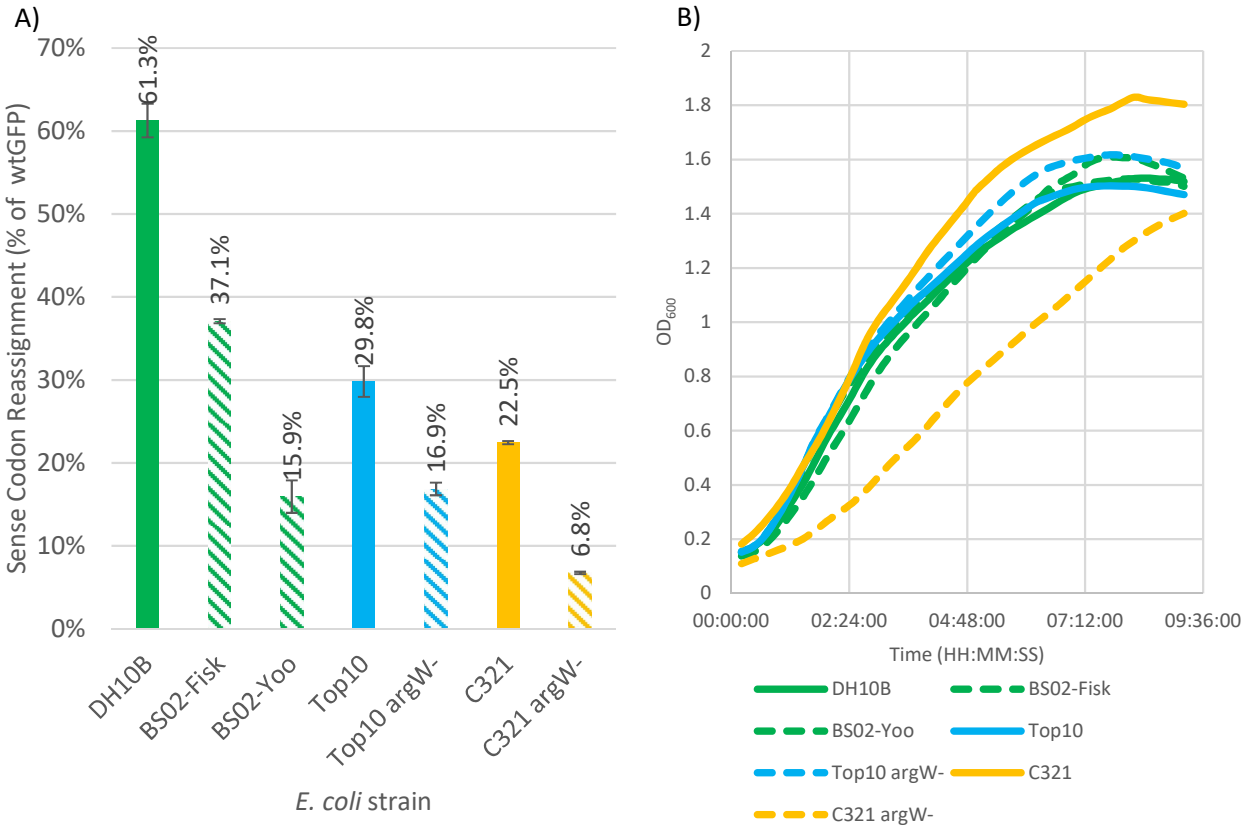


Figure 4.3 A)Effect of competing knocking out competing endogenous tRNA on sense codon reassigment. Sense codon reassignment efficiency of reassigning AGG codons to tyr using the TyrRS/$tRNA_{CCU}$ in strains of *E. coli* (Solid bars) and in the same strains with argW gene knocked out (hashed bars). BS02 cells are derivative of DH10B cells, and BS02-Yoo were obtained from the Yoo lab, BS02-Fisk were engineered in our lab. B) Growth curves of a representative expression experiment for the 7 strains of *E. coli* in Figure 3a. All strains are harboring the same plasmids to reassign the AGG codon in the fluorophore of sfGFP to Tyr.

We obtained an original sample of BS02 cells (designated BS02-Yoo) and the (pEVOL1/pQE-80) plasmid system employed for para-azidophenylalanine (pAzF) incorporation (generous gifts from Tae Hyeon Yoo). Sequencing and tRNA amplification tests suggested that both BS02-Yoo and BS02-Fisk cells were indeed ArgW knockouts. Multiple AGG codon reassignment experiments were performed using both Tyr and pAzF incorporating aaRSs and evaluated using the GFP based screen and by ESI-MS

of the Z domain peptide. In none of the tests did we find that the removal of argW from the genome of *E. coli* led to an increase in sense codon reassignment levels outside of standard deviations of the experiments. In most incorporation scenarios, incorporation efficiency was decreased in the absence of argW. Unexplainably we saw no significant difference in the growth rate of cells when we compared the strains with argW and their argW knockout counterparts (Figure 4.3b). Reports from the Yoo lab suggested that knocking out the argW gene from DH10B cells resulted in severe growth defects, which we did not observe.

We concluded that knocking argW out of cells was not a major contributing factor to sense codon reassignment efficiency. The lack of growth rate defect in argW knockout cells suggests that the Arg-$tRNA_{UCU}$ is largely capable of decoding the AGG codons and that it compensates for Arg-$tRNA_{CCU}$ removal. Zeng and co-workers (Liu 2014) attempted to reduce the effective concentration of $tRNA_{CCU}$ by introducing an antisense RNA and found that it had an insignificant effect on sense codon reassignment efficiency. Mukai et al. reported cells that cells were viable after knocking out argW, but were unable to survive after knocking out the genes for both the Arg-$tRNA_{CCU}$ and Arg-$tRNA_{UCU}$. Supplemental information in Lee et al. suggests that very high levels of AGG codon reassignment are present in $ArgW^{+}$ parent cells consistent with the AGG reassignment efficiencies measured by our GFP screen.

Effect of Plasmid Backbone on Reassignment Efficiency

All of our attempts to reassign the AGG codon in both $argW^{+}$ and $argW^{-}$ cells was done using our pWB_Ultra plasmid backbone to express the TyrRS and $tRNA_{CCU}$. We wanted to make sure that the difference in argW results we observed relative to the results reported by Lee et al. was not due to some unknown difference between our plasmid systems. We cloned our TyrRS gene into the pEVOL1 plasmid and measured the reassignment efficiency of AGG codons to Tyr in DH10B, BS02-Yoo, and BS02-Fisk cells (Figure 4.4a). We also ruled out any differences from the plasmid backbone for the reporter sfGFP plasmid by cloning our sfGFP reporter protein into the pQE-80 plasmid used by the Yoo lab to express the Z-domain reporter in their paper.[17] Results of Tyr incorporation using the pEVOL1 backbone

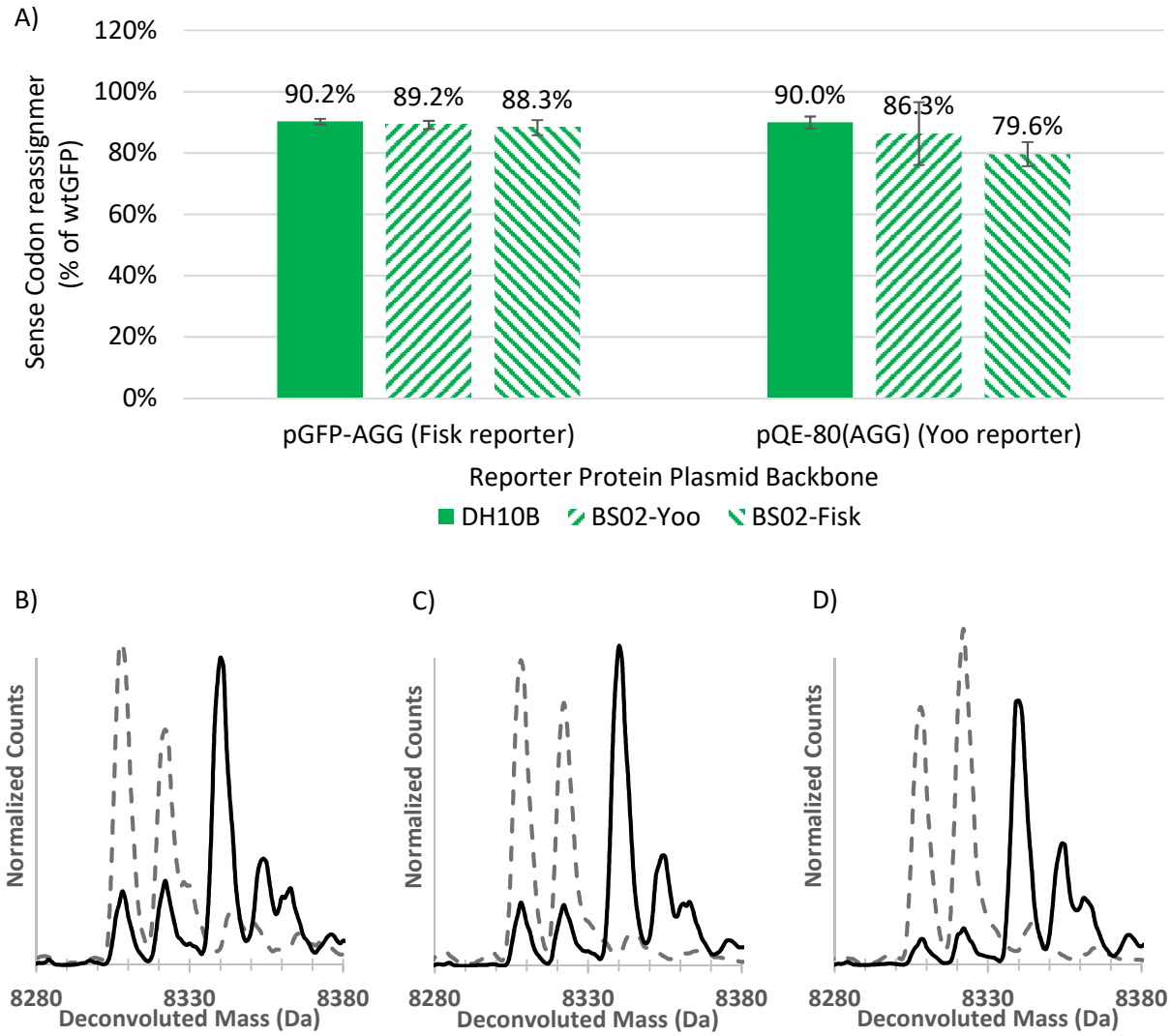


Figure 4.4 Reassignment of AGG codons using the pEVOL1 backbone to express orthogonal translation components. A) Reassignment of an AGG codon in the fluorophore of sfGFP to Tyr in three cell lines. Hashed boxes are argW^{-}. The sfGFP gene is the same in both reporter plasmid backbones. B) – D) Deconvoluted ESI-MS spectra for Z domains incorporating pAzF in response to a single AGG codon from three different cell lines. Dashed lines represent proteins expressed without ncAA, solid lines were expressed with 2mM pAzF. B) DH10B cells. C) BS02-Yoo Cells. D) BS02-Fisk Cells.

supported our previous findings that removal of argW did not lead to increased sense codon reassignment at AGG codons, but surprisingly sense codon reassignment with the pEVOL1 backbone was significantly higher than sense codon reassignment from the pWB_Ultra plasmid backbone. Regardless of the argW phenotype of the cell or the plasmid backbone from which the reporter was expressed, the pEVOL1 backbone expressing the TyrRS and the $tRNA^{Opt}_{CCU}$ reassigned AGG codons at approximately 90% efficiency relative to a wtGFP protein under the same expression conditions. We also compared

incorporation of pAzF into the AGG codon in the Z domain peptide using ESI-MS analysis (Figure 4.4b-d). The mass spectra of pAzF incorporated at the AGG codon supports high levels of reassignment form the pEVOL1 backbone, showing small peaks corresponding to Arg and predominantly incorporation of pAzF in all three cell types.

Although argW did not appear to be a significant contributor to the efficiency of AGG codon, the efficiencies of reassignment were dependent on plasmids that expressed the TyrRS and tRNA. Using the same aaRS and tRNA, reassignment efficiencies were consistently higher when expressed from the pEVOL1 as opposed to the pWB Ultra system. The high efficiency as a starting point would be difficult to improve, and is likely partially responsible for the confusing results that made it appear as though argW had an effect.

We tried to combine the high starting efficiency of the pEVOL1 plasmid system with the improvements discovered in our directed evolution library to reach even higher levels of incorporation. In the pEVOL1 bases system, the selected improved variant performed worse than the starting aaRS. The selected variant showed a sense codon reassignment efficiency of only 76% relative to the unmodified TyrRS and $tRNA^{Opt}_{CCU}$ which incorporate at 90% (Figure 4.5a). Although ESI-MS is not quantitative the trends it reveals agree with the decrease seen in the sfGFP screen for TyrRS expressed from pEVOL1. the pEVOL1 system, the starting material with the unmodified TyrRS incorporates at such a high level that no arg incorporation is detectable (Figure 4.5c). The TyrRS/tRNA pair with clone C3 modifications selected from the library, show a decrease in sense codon reassignment efficiency as incorporation of Arg as well as Tyr is plainly detected (Figure 4.5d). Arg incorporation results in the 8308 Da peak (as does the 8322 which has a +14Da methylation in the z-domain). Substituting Tyr for Arg results in a +7Da mass shift (8315 and 8329)

We evaluated the unmodified pAzFRS/$tRNA_{CCU}$ and clone C3 library improvements to pAzFRS and $tRNA_{CCU}$ expressed from the pEVOL1 plasmid. As we saw with the TyrRS screen however, the pAzFRS/tRNA pair with the library improvements led to a decrease in sense codon reassignment

efficiency when expressed from the pEVOL1 plasmid (Figure 4.5b). This is likely due to the same variables that limited reassignment in the pEVOL1-TyrRS system, and could also stem from the already very high levels of reassignment. Evaluation of the pAzFRS and $tRNA_{CCU}$ pair with and without library

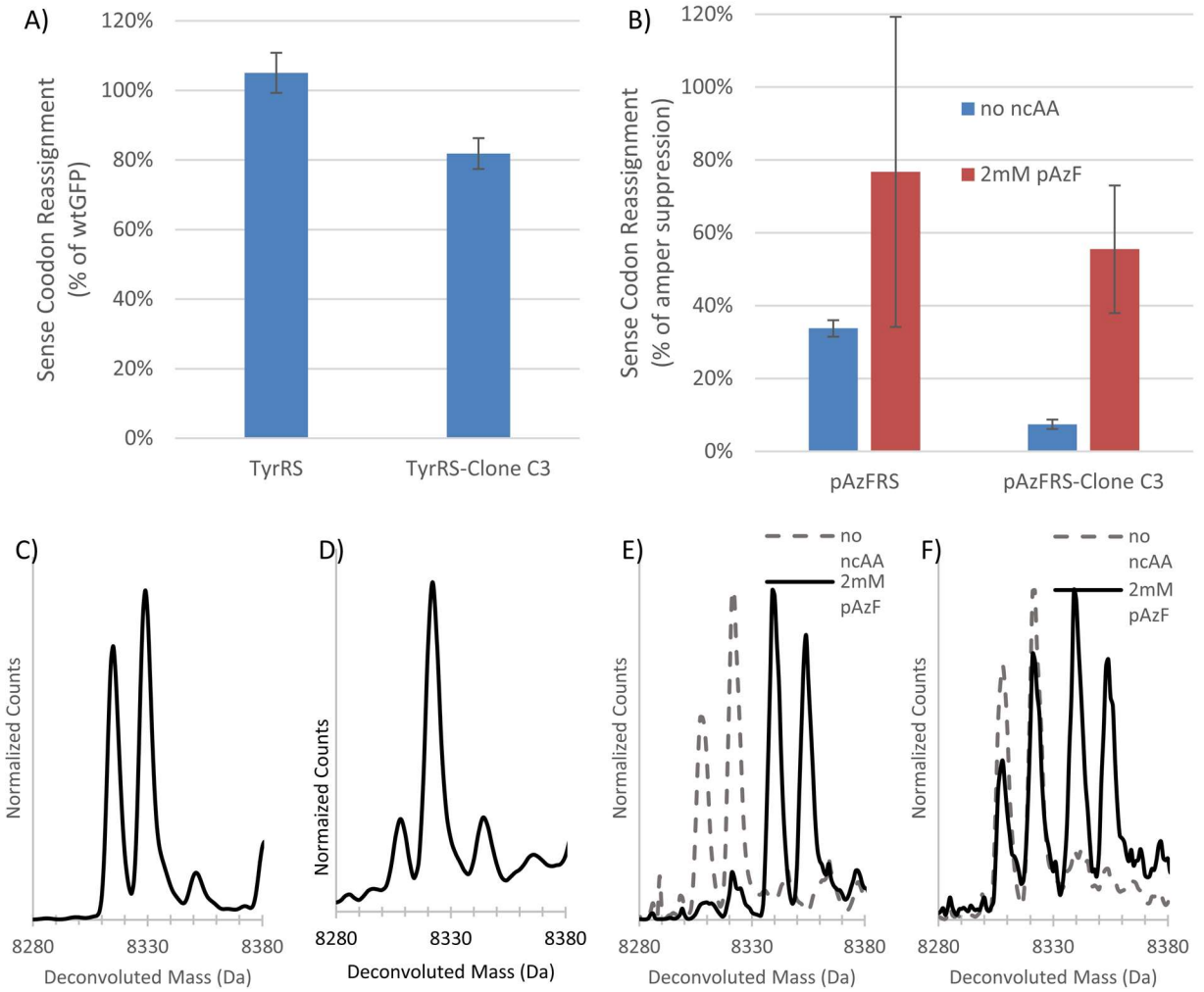


Figure 4.5 Transferability of improvements for a TyrRS / tRNACCU library evolved in pWB_Ultra, expressed in pEVOL1. A) Sense codon reassignment of TyrRS in pEVOL backbone relative to levels of wtGFP expressed from the same system. B) Sense codon reassignment of pAzF incorporation at AGG codons expressing pAzFRS and the tRNA from pEVOL. Reassignment percent is determined from fluorescence per protein estimates from clarified lysates, relative to suppression of an amber stop codon using $tRNA^{Opt}_{CUA}$. C) ESI MS of the Z-domain peptide with Tyr incorporated in response to an AGG codon. Left, the $tRNA^{Opt}_{CCU}$ and the TyrRS, Right, Clone C3 variants of the TyrRS and $tRNA^{Opt}_{CCU}$. Calculated mass for the Z domain with Arg is 8308 Da. Substitution of Tyr for Arg results in a +7Da mass shift. D) ESI MS of purified Z domain peptides in the presence (Solid line) or absence (dashed line) of 2mM pAzF. On the left is the pAzFRS and the $tRNA^{Opt}_{CCU}$ Substitution of pAzF for an Arg residue results in a +32Da mass shift. The +14Da shift is likely due to a methylation of the Z-domain.

modifications using ESI-MS analysis of Z domain peptides confirms a decrease in sense codon reassignment in systems with mutations originally identified in PWB-Ultra libraries (Figure 4.5e and f).

We assume that the loss in reassignment efficiency for aaRS/tRNA pairs with clone C3 modifications in the pEVOL1 backbone is because of the kinetic differences which come about from the different concentrations of translation components. The genes for the aaRS, the tRNA, and the reporter protein are identical in each scenario, while the plasmid copy number and the promoters driving those genes is variable across the systems. It remains unclear which factors are the most important for the transferability of improvements to aaRS and tRNAs, and expression systems clearly play some role in efficiency. While mutations which increased sense codon reassignment efficiency in the pWB_Ultra plasmid backbone were transferrable between aaRS with different amino acid specificities, the same mutations did not lead to an increase in sense codon reassignment in the pEVOL1 backbone. The maintenance of the trend of decreasing reassignment did however transfer between the TyrRS/$tRNA_{CCU}$ pair and the pAzFRS/$tRNA_{CCU}$ pair in the pEVOL1 backbone. This suggests that when using the same plasmid backbones, improvements to and decreases in sense codon reassignment efficiency may be transferable.

Although the library improvements did not increase pAzF sense codon reassignment in pEVOL1 plasmid system, the levels of pAzF reassignment in both the starting material and the library improvements were quite high. This once again supports the argument that different plasmid backbones can have different efficiencies for sense codon reassignment. One interesting result from the comparisons of these two different aaRS/tRNA plasmid expression systems is that although the sfGFP reporter gene was expressed from the same plasmid in all of the expression tests (pGFP-Y66aag), the expression levels of the reporter protein were dramatically different between the two plasmid systems, and also between the reassignment of an AGG codon and the suppression of an amber stop codon at position 66 (Figure 4.6).

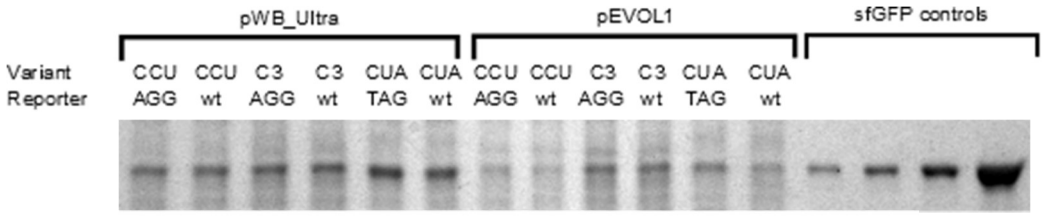


Figure 4.6. Coomassie stained PAGE Gel of clarified lysates of sfGFP expressing cultures with orthogonal translation systems expressed from two different backbones. Far right samples are purified protein samples representing 110ng, 225ng, 550ng, or 1100ng of sfGFP

Although our library modifications we identified from the pWB_Ultra plasmid system did not lead to improvements in the pEVOL1 for AGG sense codon reassignment future work with directed evolution libraries in alternate plasmid systems should lead to increased incorporation efficiencies at sense codons. The fact that library improvements do not transfer well between plasmid systems suggests that expression of orthogonal translation components is quite sensitive and that care should be taken to optimize expression systems for a given aaRS and tRNA in each scenario, including libraries for directed evolution. We are particularly interested in working with the pDULE plasmid system and autoinducing media, as comparative testing has shown this system to have very high levels of ncAA incorporation in response to stop codons as well as very high levels of protein yield. Although we have started working with pDULE, the plasmid construction is not amenable to high diversity DNA library construction, and so far we have not produced a library of mutants in pDULE based plasmids.

CONCLUSIONS

These results suggest that although directed evolution is in fact a valuable tool for improving GCE, and that modifying the interaction between the aaRS and tRNA can have positive effects on sense codon reassignment efficiency, there is no guarantee that those improvements will transfer outside of the *in vivo* reassignment system used for directed evolution. Further, there is large variability in reassignment values between different incorporation systems (as seen in this work, as well as in reports from several other groups). Not many reports directly compare plasmid or cellular systems for incorporation in the way we have (they don't have a good enough screen for it?)

More work to understand which individual components are important variables in these systems needs to be performed, not only to optimize genetic code expansion for a given system, but also to have better understanding *a priori* about moving GCE systems between organisms and ncAA and aaRS systems.

While genome engineering can lead to high levels of incorporation at sense codons, it is a challenging method that requires specific modification of a given strain of an organism. There is also some question as to the extent to which removing competition from endogenous tRNAs increases or changes sense codon reassignment, particularly at the AGG codon. Directed evolution can be used to realize high level incorporation efficiency of ncAA into proteins in response to sense codons. Reassigning the AGG Arg codon does not lead to high levels of toxicity in cells even when high levels of reporter protein are expressed with ncAA incorporations in rich LB media. This produces a simple method for the expansion of the genetic code that can be combined with other more complex methods if necessary, but can also be used in a standalone fashion.

REFERENCES

(1) Neumann, H.; Slusarczyk, A. L.; Chin, J. W. De Novo Generation of Mutually Orthogonal Aminoacyl-tRNA synthetase/tRNA Pairs. *J. Am. Chem. Soc.* **2010**, *132*, 2142–2144.

(2) Chatterjee, A.; Xiao, H.; Schultz, P. G. 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*, 14841–14846.

(3) Santoro, S. W.; Anderson, J. C.; Lakshman, V.; Schultz, P. G. An Archaebacteria-Derived Glutamyl-tRNA Synthetase and tRNA Pair for Unnatural Amino Acid Mutagenesis of Proteins in Escherichia Coli. *Nucleic acids research* **2003**, *31*, 6700–6709.

(4) Pastrnak, M.; Magliery, T. J.; Schultz, P. G. A New Orthogonal Suppressor tRNA/aminoacyl-tRNA Synthetase Pair for Evolving an Organism with an Expanded Genetic Code. *Helvetica Chimica Acta* **2000**, *83*, 2277–2286.

(5) Dumas, A.; Lercher, L.; Spicer, C. D.; Davis, B. G. Designing Logical Codon Reassignment-Expanding the Chemistry in Biology. *Chemical Science* **2015**, *6*, 50–69.

(6) Mukai, T.; Hayashi, A.; Iraha, F.; Sato, A.; Ohtake, K.; Yokoyama, S.; Sakamoto, K. Codon Reassignment in the Escherichia Coli Genetic Code. *Nucleic acids research* **2010**, *38*, 8188–8195.

(7) Ohtake, K.; Sato, A.; Mukai, T.; Hino, N.; Yokoyama, S.; Sakamoto, K. Efficient Decoding of the UAG Triplet as a Full-Fledged Sense Codon Enhances the Growth of a prfA-Deficient Strain of Escherichia Coli. *Journal of bacteriology* **2012**, *194*, 2606–2613.

(8) Johnson, D. B. F.; Wang, C.; Xu, J.; Schultz, M. D.; Schmitz, R. J.; Ecker, J. R.; Wang, L. Release Factor One Is Nonessential in Escherichia Coli. *ACS Chemical Biology* **2012**, *7*, 1337–1344.

(9) Johnson, D. B. F.; Xu, J.; Shen, Z.; Takimoto, J. K.; Schultz, M. D.; Schmitz, R. J.; Xiang, Z.; Ecker, J. R.; Briggs, S. P.; Wang, L. RF1 Knockout Allows Ribosomal Incorporation of Unnatural Amino Acids at Multiple Sites. *Nature Chemical Biology* **2011**.

(10) Lajoie, M. J.; Rovner, A. J.; Goodman, D. B.; Aerni, H.-R.; Haimovich, A. D.; Kuznetsov, G.; Mercer, J. A.; Wang, H. H.; Carr, P. A.; Mosberg, J. A.; others. Genomically Recoded Organisms Expand Biological Functions. *Science* **2013**, *342*, 357–360.

(11) Gan, R.; Perez, J. G.; Carlson, E. D.; Ntai, I.; Isaacs, F. J.; Kelleher, N. L.; Jewett, M. C. Translation System Engineering in Escherichia Coli Enhances Non-Canonical Amino Acid Incorporation into Proteins. *Biotechnology and bioengineering* **2017**, *114*, 1074–

1086.

(12) Zheng, Y.; Lajoie, M. J.; Italia, J. S.; Chin, M. A.; Church, G. M.; Chatterjee, A. Performance of Optimized Noncanonical Amino Acid Mutagenesis Systems in the Absence of Release Factor 1. *Molecular BioSystems* **2016**, *12*, 1746–1749.

(13) Yang, S.; Hendrickson, W. A.; Crouch, R. J.; Satow, Y. Structure of Ribonuclease H Phased at 2 Angstrom Resolution by MAD Analysis of the Selenomethionyl Protein. *Science* **1990**, *249*, 1398–1406.

(14) Kwon, I.; Kirshenbaum, K.; Tirrell, D. A. Breaking the Degeneracy of the Genetic Code. *Journal of the American Chemical Society* **2003**, *125*, 7512–7513.

(15) Wang, Y.; Tsao, M.-L. Reassigning Sense Codon AGA to Encode Noncanonical Amino Acids in Escherichia Coli. *ChemBioChem* **2016**, *17*, 2234–2239.

(16) Mukai, T.; Yamaguchi, A.; Ohtake, K.; Takahashi, M.; Hayashi, A.; Iraha, F.; Kira, S.; Yanagisawa, T.; Yokoyama, S.; Hoshi, H.; others. Reassignment of a Rare Sense Codon to a Non-Canonical Amino Acid in Escherichia Coli. *Nucleic acids research* **2015**, *43*, 8111–8122.

(17) Lee, B. S.; Shin, S.; Jeon, J. Y.; Jang, K.; Lee, B. Y.; Choi, S.; Yoo, T. H. Incorporation of Unnatural Amino Acids in Response to the AGG Codon. *ACS chemical biology* **2015**.

(18) Bohlke, N.; Budisa, N. 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*, 133–144.

(19) Ho, J. M.; Reynolds, N. M.; Rivera, K.; Connolly, M.; Guo, L.-T.; Ling, J.; Pappin, D. J.; Church, G. M.; Söll, D. Efficient Reassignment of a Frequent Serine Codon in Wild-Type Escherichia Coli. *ACS synthetic biology* **2015**, *5*, 163–171.

(20) Zeng, Y.; Wang, W.; Liu, W. R. Towards Reassigning the Rare AGG Codon in Escherichia Coli. *ChemBioChem* **2014**, *15*, 1750–1754.

(21) Young, T. S.; Ahmad, I.; Yin, J. A.; Schultz, P. G. An Enhanced System for Unnatural Amino Acid Mutagenesis in E. Coli. *Journal of molecular biology* **2010**, *395*, 361–374.

(22) Biddle, W.; Schmitt, M. A.; Fisk, J. D. Evaluating Sense Codon Reassignment with a Simple Fluorescence Screen. *Biochemistry* **2015**, *54*, 7355–7364.

(23) Kobayashi, T.; Nureki, O.; Ishitani, R.; Yaremchuk, A.; Tukalo, M.; Cusack, S.; Sakamoto, K.; Yokoyama, S. Structural Basis for Orthogonal tRNA Specificities of Tyrosyl-tRNA Synthetases for Genetic Code Expansion. *Nature Structural Biology* **2003**, *10*, 425–432.

(24) Konevega, A. L.; Soboleva, N. G.; Makhno, V. I.; Semenkov, Y. P.; Wintermeyer, W.; Rodnina, M. V.; Katunin, V. I. Purine Bases at Position 37 of tRNA Stabilize Codon-Anticodon Interaction in the Ribosomal A Site by Stacking and Mg2+-Dependent Interactions. *Rna* **2004**, *10*, 90–101.

(25) Agris, P. F. Decoding the Genome: A Modified View. *Nucleic acids research* **2004**, *32*, 223–238.

(26) Ruan, B.; Palioura, S.; Sabina, J.; Marvin-Guy, L.; Kochhar, S.; LaRossa, R. A.; Söll, D. Quality Control despite Mistranslation Caused by an Ambiguous Genetic Code. *Proceedings of the National Academy of Sciences* **2008**, *105*, 16502–16507.

(27) Min, B.; Kitabatake, M.; Polycarpo, C.; Pelaschier, J.; Raczniak, G.; Ruan, B.; Kobayashi, H.; Namgoong, S.; Söll, D. Protein Synthesis in Escherichia Coli with Mischarged tRNA. *Journal of bacteriology* **2003**, *185*, 3524–3526.

(28) Chin, J. W.; Santoro, S. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. Addition of P-Azido-L-Phenylalanine to the Genetic Code of Escherichia c Oli. *Journal of the American Chemical Society* **2002**, *124*, 9026–9027.

(29) Wang, L.; Xie, J.; Deniz, A. A.; Schultz, P. G. Unnatural Amino Acid Mutagenesis of Green Fluorescent Protein. *The Journal of Organic Chemistry* **2003**, *68*, 174–176.

(30) Maza, J. C.; Villa, J. K.; Landino, L. M.; Young, D. D. Utilizing Unnatural Amino Acids To Illustrate Protein Structure-Function Relationships: An Experiment Designed for an Undergraduate Biochemistry Laboratory. *Journal of Chemical Education* **2016**, *93*, 767–771.

(31) Reddington, S. C.; Rizkallah, P. J.; Watson, P. D.; Pearson, R.; Tippmann, E. M.; Jones, D. D. Different Photochemical Events of a Genetically Encoded Phenyl Azide Define and Modulate GFP Fluorescence. *Angewandte Chemie International Edition* **2013**.

(32) Braisted, A. C.; Wells, J. A. Minimizing a Binding Domain from Protein A. *Proceedings of the National Academy of Sciences* **1996**, *93*, 5688–5692.

APPENDIX A

Supporting Information for Chapter 2 — Evaluating Sense Codon Reassignment with a Simple Fluorescence Screen

TABLE OF CONTENTS

Cell Strains

NEB 5-alpha (New England Biolabs): *fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*

SB3930 (Yale CGSC): λ^-, *ΔhisB463*

SS320 (MC1061 F', Lucigen): [F'*proAB+lacIqlacZ*ΔM15 *Tn10* (tetr)] *hsdR mcrB araD139* Δ(*araABC-leu*)7679 *lac*X74 *galUgalK rpsL thi*

CJ236 (New England Biolabs): FΔ*(HindIII)::cat* (Tra^+ Pil^+ Cam^R)/ *ung-1 relA1 dut-1 thi-1 spoT1 mcrA*

S.1 General Materials and Reagents

All restriction enzymes, DNA polymerases, and T4 kinase were purchased from New England Biolabs and used according to the manufacturer's instructions. ATP was purchased from Fisher (BP413-25) and dNTPs were purchased form New England Biolabs (N0447S). DNA isolation was performed using either a Thermo Scientific GeneJET plasmid miniprep kit (K0503) or Sigma Aldrich GenElute plasmid miniprep kit (PLN350) according to the manufacturer's protocols. Purification of PCR and mutagenesis reactions were performed using either a Thermo Scientific GeneJET PCR purification kit (K0702) or Sigma Aldrich GenElute PCR clean-up kit (NA1020) according to the manufacturer's instructions.

LB liquid media (per liter: 10 g tryptone, 5 g yeast extract, 5 g NaCl) and LB agar plates with 15 g/L agar (TEKNova, A7777) were used unless otherwise noted. Isopropyl-beta-D-thiogalactoside (IPTG) was purchased from Gold Bio (I2481C5). Spectinomycin (Enzo Life Science, BML-A281) was used at 50 μg/mL to maintain the pUltra-based vectors harboring the tRNA and aaRS genes. Carbenicillin (PlantMedia, 40310000-2) was used at 50 μg/mL to maintain the vectors harboring the GFP reporter gene. All bacterial cultures were grown at 37 °C unless otherwise noted.

Electrocompetent stocks of all strains were prepared in-house according to the method of Sambrook and Russell[1]. Typical transformation efficiencies for electrocompetent cells produced in this way are 10^9 cfu/μg of supercoiled DNA. All electroporation transformations were recovered in SOC (20 g/L tryptone,

5 g/L yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 20 mM Glucose) for 1 hour at 37 °C with shaking prior to transfer to media containing appropriate antibiotics.

All oligonucleotides were purchased from Integrated DNA Technologies (Iowa, USA) in 25 nmol scales with standard desalting.

S.2 DNA Mutagenesis Protocols

QuikChange Style Site-Directed Mutagenesis

Non-library site directed mutagenesis for routine cloning and mutation of tRNA anticodons and GFP fluorophore codons was accomplished using a method adapted from the Stratagene QuikChange Multi-Site Directed Mutagenesis protocol. For each 25 µl reaction, 125 ng of mutagenic oligonucleotide was combined with 100 ng of plasmid DNA isolated from NEB 5-alpha cells in 1x Q5 High Fidelity Polymerase with 1x High GC Buffer, 200 µM dNTPs, and 1 unit of Q5 High-Fidelity Polymerase. Reactions were cycled in a PeqSTAR thermocycler (peqLab) for 18 cycles: 98 °C for 20 seconds, and 72 °C for 15 seconds/kb of template plasmid. A final extension at 72 °C was carried out for 5 minutes. 20 units of DpnI enzyme was added directly to the PCR reaction and incubated at 37 °C for 2 hours. Reactions were transformed into electrocompetent NEB 5-alpha without cleanup.

Isolation of single stranded deoxy-uridine containing DNA

DNA libraries were produced with Kunkel Mutagenesis using a method adapted from Sidhu and Weiss[2]. Briefly, cultures of CJ236 cells harboring the phagemid to be mutated were grown to an OD_{600} of 0.5 and infected with M13K07 helper phage at a multiplicity of infection of 10:1. The infected culture was transferred into 50-100 mL of LB media with 5 µg/mL chloramphenicol, appropriate antibiotic to maintain the phagemid, and 0.25 µg/ml uridine. Cultures were grown overnight at 30 °C or 37 °C. Cells were pelleted at 17,000 x*g* at 4 °C for 12 minutes in a Sorvall RC 6+ with a Thermo FIBERLite F13-14x50cy rotor. Phage particles were isolated by decanting the supernatant from the pelleted cells into 1/5th volume of 20% 8,000 molecular weight polyethylene glycol and 2.5M NaCl in water. Solutions were incubated on ice or at 4 °C for 2 hours or overnight (respectively). Phage particles were isolated by

pelleting at 17,000 x*g* at 4 °C for 20 minutes in a Sorvall RC 6+. The supernatant was decanted and the phage pellet was spun for an additional minute to collect all the remaining supernatant, which was then removed using a pipette. The phage pellet was resuspended in approximately 1 mL of Phosphate or Tris buffered saline, pH=7.4. Insoluble material was pelleted out of the phage solution at 17,000 x*g* for 5 minutes. Single stranded DNA was isolated from phage particles using a Qiagen M13 spin kit.

Kunkel Style Site-Directed Mutagenesis

Mutagenic primers were phosphorylated for 1 hour at 37 °C using T4 polynucleotide kinase. Phosphorylated primers were annealed to 1 μg of single stranded template DNA at a 10:1 molar ratio by incubating at 90 °C for 2 minutes, 50 °C for 3 minutes, and then 25 °C for 5 minutes. The annealed mixture was extended using T7 DNA polymerase in the presence of T4 DNA Ligase and 670 μm ATP and 330 μm dNTPs (each) at room temperature overnight. Reactions were cleaned up using a PCR spin kit and transformed into electrocompetent NEB-5alpha cells. This procedure typically yields 10^5 transformants with 40-80% mutation efficiency.

Construction of GFP Reporter Vectors

Reporter plasmids containing sfGFP gene variants (S30R, Y39N, F64L, S65T, F99S, N105T, Y145F, M153T, V163A, I171V, and A206V relative to the wtGFP sequence[3]) with codon modifications at position 66 (fluorophore Tyr) were constructed using standard DNA manipulation techniques. The sfGFP gene in our reporter vectors differs from the protein reported by Pédelacq *et al.* at the N-terminus of the protein; our sfGFP reporter has an N-terminal hexa-histidine tag followed by Ala-Leu-Glu in place of the first three residues of the wild type sequence Met-Ser-Lys. In addition in our sfGFP reporter variant, the Thr at position 9 is a His residue and glutamine 80 is an arginine. All other reporter plasmids used in this study vary only by the DNA codon corresponding to amino acid 66 (Tyr in sfGFP).

We used a modified pQE-40 protein expression plasmid (Qiagen) as the backbone for the reporter plasmids. Briefly, the unique EcoRI site in pQE-40 was moved upstream of the hexa-histidine tag, and a unique PacI site was inserted between the XbaI site and the M13 origin. The sfGFP gene was amplified

out of Np3-TAT-GFP (Gift from A. Bradbury[4]) and assembled into a highly repressed, IPTG-inducible cassette with a T5 promoter, 2 LacO operator sites, and an engineered strong ribosome binding site from the Registry of Standard Biological Parts (**BBa_B0030**). The sfGFP gene cassette was inserted into the vector backbone using NotI and the repositioned EcoRI site. Strong repression of the sfGFP reporter was achieved by incorporating a LacI gene cassette amplified from pBbB1a-GFP (Addgene plasmid #35340)[5] and cloned into the backbone using the XbaI and PacI sites. The camR gene cassette was removed from the reporter vector by digesting the DNA with NheI and XbaI and ligating the overhanging ends together. The entire DNA sequence of the sfGFP expressing reporter plasmid is given in S.3.

Construction of tRNA/aaRS Translational Components Vectors

The pWB_Ultra-Tyr-AUG vector expressing the orthogonal translational components was assembled from several vectors and modeled after pUltra plasmids reported by Chatterjee *et al.*[6]. The *M. jannaschii* $tRNA^{Opt}$ expression cassette was amplified out of pEVOL-pBpF (Addgene plasmid #311990)[7]. We assembled a vector backbone containing a medium copy number ClodF13 origin of replication, spectinomycin resistance gene, and LacI gene cassette from pCDFDuet-ADAT1-ADAT3 (Addgene plasmid #15917)[8]. The *E. coli* codon optimized *M. jannaschii* tyrosyl- tRNA synthetase (TyrRS) gene was obtained by changing the specificity of the p-azidophenylalanine incorporating aaRS (from pEVOL-pAzF, Addgene plasmid #31186)[9] to tyrosine by site directed mutagenesis. Quikchange-style mutagenesis was used to mutate the T32Y, N107E, P157D, L158I, Q162L and R286D in the pAzF-RS gene to restore TyrRS activity (based on GenBank accession no. Q57834). The TyrRS gene was cloned into our backbone under the control of an IPTG inducible tac promoter. The entire DNA sequence is provided for clarity in S.6

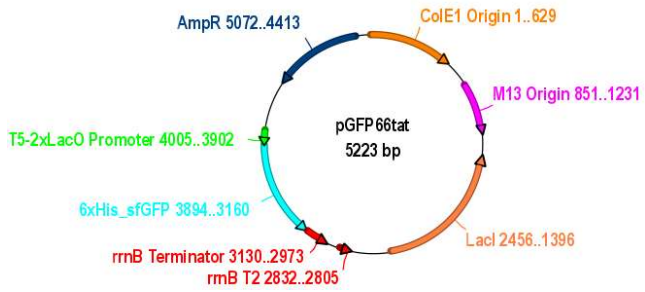


Figure S1. Graphic map for Reporter Phagemid pGFP66tat

S.3 DNA Sequence of pGFP66tat

> [pGFP66tat wtGFP reporter - 5223 bp]

CTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGATTCAATTGTGAGCGGATAACA
ATTTCACACAGAAAACATTAAAGAGGAGAAAGAATTCTATGCACCACCACCACCACCACGCACTCGAGGGAGAAGAA
CTTTTCCACGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCCGTGGAGA
GGGTGAAGGTGATGCTACAAACGGAAAACTCACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGC
CAACACTTGTCACTACTCTGACCTATGGTGTTCAATGCTTTTCCCGTTATCCGGATCACATGAAACGGCACGACTTT
TTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAACGCACTATATCTTTCAAAGATGACGGGACCTACAAGACGCG
TGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATCGTATCGAGTTAAAGGGTATTGATTTTAAAGAAGATGGAA
ACATTCTTGGACACAAACTCGAGTACAACTTTAACTCACACAATGTATACATCACGGCAGACAAACAAAAGAATGGA
ATCAAAGCTAACTTCAAAATTCGCCACAACGTTGAAGATGGTTCCGTTCAACTAGCAGACCACTATCAACAAAATAC
TCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCACTACCTGTCGACACAATCTGTCCTTTCGAAAGATCCCA
ACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTGGGATTACACACGGCATGGATGAGCTCTAC
AAATAAGCGGCCGCGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGA
ACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGAACTGCCAGGCATCA
AATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTA
GGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAGCAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATAA
ACTGCCAGGCATCAAATTAAGCAGAAGGCCATCCTGACGGATGGCCTTTTTGCGTTTCTACAAACTCTTTTGTTTAT
TTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCGATGATCCCGGGTCGACCTGCAGCCAAGCTT
AATTAGCTGAGCTTGGACTCCTGTTGATAGATCCAGTAATGACCTCAGAACTCCATCTGGATTTGTTCAGAACGCTC
GGTTGCCGCCGGGCGTTTTTTATTGGTGAGAATCCAAGCTAGACCCCGAAAAGTGCCACCTGACGTCGACACCATCG
AATGGTGCAAAACCTTTCGCGGTATGGCATGATAGCGCCCGGAAGAGAGTCAATTCAGGGTGGTGAATGTGAAACCA
GTAACGTTATACGATGTCGCAGAGTATGCCGGTGTCTCTTATCAGACCGTTTCCCGCGTGGTGAACCAGGCCAGCCA
CGTTTCTGCGAAAACGCGGGAAAAAGTGGAAGCGGCGATGGCGGAGCTGAATTACATTCCCAACCGCGTGGCACAAC
AACTGGCGGGCAAACAGTCGTTGCTGATTGGCGTTGCCACCTCCAGTCTGGCCCTGCACGCGCCGTCGCAAATTGTC
GCGGCGATTAAATCTCGCGCCGATCAACTGGGTGCCAGCGTGGTGGTGTCGATGGTAGAACGAAGCGGCGTCGAAGC
CTGTAAAGCGGCGGTGCACAATCTTCTCGCGCAACGCGTCAGTGGGCTGATCATTAACTATCCGCTGGATGACCAGG
ATGCCATTGCTGTGGAAGCTGCCTGCACTAATGTTCCGGCGTTATTTCTTGATGTCTCTGACCAGACACCCGTCAAC
AGTATTATTTTCTCCCATGAAGACGGTACGCGACTGGGCGTGGAGCATCTGGTCGCATTGGGTCACCAGCAAATCGC
GCTGTTAGCGGGCCCATTAAGTTCTGTCTCGGCGCGTCTGCGTCTGGCTGGCTGGCATAAATATCTCACTCGCAATC
AAATTCAGCCGATAGCGGAACGGGAAGGCGACTGGAGTGCCATGTCCGGTTTTCAACAAACCATGCAAATGCTGAAT
GAGGGCATCGTTCCCACTGCGATGCTGGTTGCCAACGATCAGATGGCGCTGGGCGCAATGCGCGCCATTACCGAGTC
CGGGCTGCGCGTTGGTGCGGATATCTCGGTAGTGGGATACGACGATACCGAAGACAGCTCATGTTATATCCCGCCGT
TAACCACCATCAAACAGGATTTTCGCCTGCTGGGGCAAACCAGCGTGGACCGCTTGCTGCAACTCTCTCAGGGCCAG
GCGGTGAAGGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACCACCCTGGCGCCCAATACGCAAACCGC

CTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGC
AACGCAATTAATGTAAGTTAGCGCGAATTGATCTGGTTTGACTTAATTAAGCGGCATCAGAGCAGATTGTACTGAGA
GTGCACCATATGTGTAAACGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACC
AATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGG
AACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACT
ACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCC
CCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCT
AGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGC
GTCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCTCTTCCGCTTCCTCGCTC
ACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCAC
AGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGT
TGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAAC
CCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCT
TACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTT
CGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGT
AACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAG
AGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTG
GTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCT
GGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTT
TTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCT
TCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGT
TACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGT
CGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCAC
CGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCC
TCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGC
CATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGC
GAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTG
GCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTC
TGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAA
TACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTC
TCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTAC
TTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAAT
GTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATA
TTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGA
AACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAC

Table S1: Reporter Protein Plasmid Designations and Usage

Reporter Plasmid	Position 66 Codon	Naturally Encoded Amino Acid	Usage
pGFP66tat	UAU	Tyr	Wild type GFP vector
pGFP66cac	CAC	His	Codon discrimination vector
pGFP66cat	CAU	His	Sense codon reassignment codon
pGFP66aaa	AAA	Lys	Codon discrimination vector
pGFP66aag	AAG	Lys	Sense codon reassignment codon
pGFP66aac	AAC	Asn	Codon discrimination vector
pGFP66aat	AAU	Asn	Sense codon reassignment codon
pGFP66ttc	UUC	Phe	Codon discrimination vector
pGFP66ttt	UUU	Phe	Sense codon reassignment codon

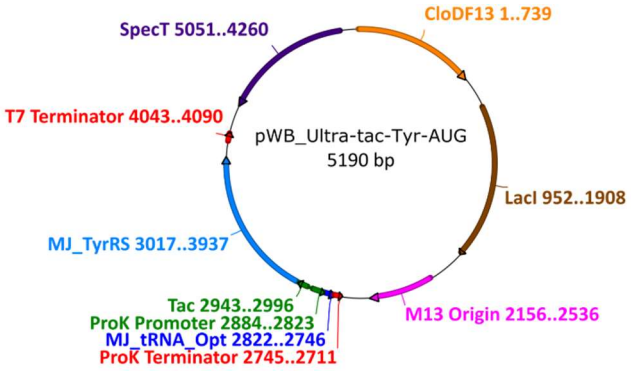


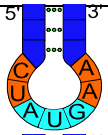
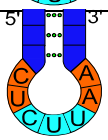
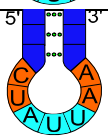
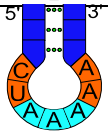
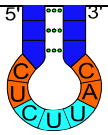
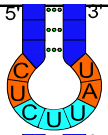
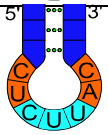
Figure S2. Graphic Map of pWB_Ultra-Tyr-AUG aaRS and tRNA expression vector

S.4 DNA Sequence of pWB_Ultra-Tyr-AUG (5190 bp)

GCGCTGCGGACACATACAAAGTTACCCACAGATTCCGTGGATAAGCAGGGGACTAACATGTGAGGCAAAACAGCAGG
GCCGCGCCGGTGGCGTTTTTCCATAGGCTCCGCCCTCCTGCCAGAGTTCACATAAACAGACGCTTTTCCGGTGCATC
TGTGGGAGCCGTGAGGCTCAACCATGAATCTGACAGTACGGGCGAAACCCGACAGGACTTAAAGATCCCCACCGTTT
CCGGCGGGTCGCTCCCTCTTGCGCTCTCCTGTTCCGACCCTGCCGTTTACCGGATACCTGTTCCGCCTTTCTCCCTT
ACGGGAAGTGTGGCGCTTTCTCATAGCTCACACACTGGTATCTCGGCTCGGTGTAGGTCGTTCGCTCCAAGCTGGGC
TGTAAGCAAGAACTCCCCGTTCAGCCCGACTGCTGCGCCTTATCCGGTAACTGTTCACTTGAGTCCAACCCGGAAAA
GCACGGTAAAACGCCACTGGCAGCAGCCATTGGTAACTGGGAGTTCGCAGAGGATTTGTTTAGCTAAACACGCGGTT
GCTCTTGAAGTGTGCGCCAAAGTCCGGCTACACTGGAAGGACAGATTTGGTTGCTGTGCTCTGCGAAAGCCAGTTAC
CACGGTTAAGCAGTTCCCCAACTGACTTAACCTTCGATCAAACCACCTCCCCAGGTGGTTTTTTCGTTTACAGGGCA
AAAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACTGAACCGCTCTAGATTTCAGTGC
AATTTATCTCTTCAAATGTAGCACCTGAAGTCAGCCCCATACGATATAAGTTGTAATTCTCATGTTAGTCATGCCCC
GCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGC
TAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAAT
CGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAAC
AGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAA
ATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATGT
CCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATC
GCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCG
TTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAAC
TTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCT
TCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGC
AGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGAT
TGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCG
GCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAA
CGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTT
CCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCATACTCT
GCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCAAGCT
TTGAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATACAGATCCTGA
CGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCC
TAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGG
GGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACG
TAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGT
TCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTAT
TGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAGGATCTGTATG

GTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACTGCCT
CGACCTGCAGCTAGCAAAAAAGCCTGCTCGTTGAGCAGGCTTTTCGAATTTGGTCCGGCGGAGGGGATTTGAACCCC
TGCCATGCGGATTCATAGTCCGCCGTTCTGCCCTGCTGAACTACCGCCGGAATGCGGGGCGCATCTTACTGCGCAGA
TACGCCCTCGTCAATCCCTTAATAGCAAAATGCCTTTTGATCGGCGAGAAAGTCAGCGGATCCATCTGCAGATCTCC
CGGGTAGGCTACTAGTTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTCACACA
GGAAACAGAATTCATGGACGAGTTCGAAATGATTAAACGCAACACCAGCGAAATTATCTCTGAAGAAGAGCTGCGCG
AGGTGCTGAAGAAAGACGAGAAGAGCGCGTATATTGGCTTTGAGCCGTCCGGTAAAATTCACCTGGGTCACTACCTG
CAAATCAAGAAGATGATTGATCTGCAAAACGCTGGTTTTGACATCATTATCCTGCTGGCGGACCTGCACGCCTACCT
GAATCAAAAGGGCGAGCTGGATGAGATTCGCAAGATCGGCGACTACAATAAGAAAGTCTTCGAAGCCATGGGTTTGA
AGGCTAAATACGTCTACGGTAGCGAATTTCAGCTGGATAAGGATTACACGTTGAATGTGTACCGTCTGGCGCTGAAA
ACCACGCTGAAACGCGCCCGTCGTTCCATGGAGCTGATTGCGCGCGAGGATGAGAATCCAAAAGTTGCTGAGGTTAT
TTACCCTATTATGCAAGTTAATGATATTCACTACCTGGGTGTTGATGTTGCAGTCGGTGGTATGGAGCAACGCAAAA
TTCACATGCTGGCACGTGAACTGCTGCCGAAAAAGGTTGTCTGTATTCACAATCCGGTCCTGACCGGCCTGGATGGC
GAGGGTAAAATGAGCAGCAGCAAGGGTAACTTTATTGCAGTTGACGATAGCCCGGAAGAAATCCGTGCGAAGATCAA
GAAAGCGTACTGCCCGGCAGGCGTGGTTGAGGGTAACCCGATCATGGAAATCGCCAAGTATTTTCTGGAATACCCAC
TGACGATTAAGCGCCCGGAGAAATTTGGCGGCGACCTGACCGTCAACAGCTACGAGGAGCTGGAAAGCTTGTTTAAG
AACAAAGAACTGCACCCGATGGATCTGAAAAACGCCGTGGCGGAAGAGCTGATTAAGATTCTGGAACCAATTCGCAA
ACGTCTGTAAGCGGCCGCGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATG
TTAATTAACCTAGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGG
GTTTTTTGCTGAAACCTCAGGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAG
CAATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCATCGTGGCCGGATCTTGCGGCCCC
TCGGCTTGAACGAATTGTTAGACATTATTTGCCGACTACCTTGGTGATCTCGCCTTTCACGTAGTGGACAAATTCTT
CCAACTGATCTGCGCGCGAGGCCAAGCGATCTTCTTCTTGTCCAAGATAAGCCTGTCTAGCTTCAAGTATGACGGGC
TGATACTGGGCCGGCAGGCGCTCCATTGCCCAGTCGGCAGCGACATCCTTCGGCGCGATTTTGCCGGTTACTGCGCT
GTACCAAATGCGGGACAACGTAAGCACTACATTTCGCTCATCGCCAGCCCAGTCGGGCGGCGAGTTCCATAGCGTTA
AGGTTTCATTTAGCGCCTCAAATAGATCCTGTTCAGGAACCGGATCAAAGAGTTCCTCCGCCGCTGGACCTACCAAG
GCAACGCTATGTTCTCTTGCTTTTGTCAGCAAGATAGCCAGATCAATGTCGATCGTGGCTGGCTCGAAGATACCTGC
AAGAATGTCATTGCGCTGCCATTCTCCAAATTGCAGTTCGCGCTTAGCTGGATAACGCCACGGAATGATGTCGTCGT
GCACAACAATGGTGACTTCTACAGCGCGGAGAATCTCGCTCTCTCCAGGGGAAGCCGAAGTTTCCAAAAGGTCGTTG
ATCAAAGCTCGCCGCGTTGTTTCATCAAGCCTTACGGTCACCGTAACCAGCAAATCAATATCACTGTGTGGCTTCAG
GCCGCCATCCACTGCGGAGCCGTACAAATGTACGGCCAGCAACGTCGGTTCGAGATGGCGCTCGATGACGCCAACTA
CCTCTGATAGTTGAGTCGATACTTCGGCGATCACCGCTTCCCTCATACTCTTCCTTTTTCAATATTATTGAAGCATT
TATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGCTAGCTCACTCGG
TCGCTACGCTCCGGGCGTGAGACTGCGGCGG

Table S2: pWB_Ultra aaRS/tRNA expression vectors

Phagemid	aaRS gene	Anticodon	Anticodon Loop	Recognized Codon	Coded Amino Acid	Comments	Primer used to mutate anticodon
pWB_Ultra-Tyr-AUG	TyrRS	AUG		CAU	His	His Reassignment Vector	CA, CB, CC, CG, CE, and CF [1]
pWB_Ultra-Tyr-CUU	TyrRS	CUU		AAG	Lys	Lys Reassignment Vector	FX
pWB_Ultra-Tyr-AUU	TyrRS	AUU		AAU	Asn	Asn Reassignment Vector	FW
pWB_Ultra-Tyr-AAA	TyrRS	AAA		UUU	Phe	Phe Reassignment Vector	FP
pWB_Ultra-Tyr-no_tRNA	TyrRS	no tRNA		--	--	Negative Control Vector	--[2]
pWB_Ultra-XhoI-Inactive	2x XhoI Inactive TyrRS	XhoI restriction site		--	--	Inactive Library Template	GZ
pWB_Ultra-Tyr-CUU^{AU}	TyrRS	CUU		AAG	Lys	Sequenced clone from 256 member tRNA library	GX
pWB_Ultra-Tyr(LCCP)-CUU^{AC}	TyrRS Library Hit	CUU		AAG	Lys	Sequenced clone from aaRS/tRNA combined Library	GX

1) The amber suppression optimized *M. jannaschii* tRNA with AUG anticodon was assembled using thermally balanced inside out PCR synthesis[10].

2) The tRNA cassette was removed using restriction and ligation to leave a single PstI site in the backbone.

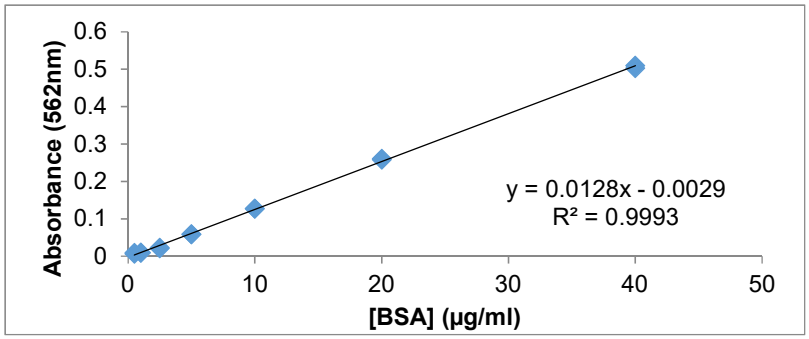


Figure S3. Standard Curve of Bovine Serum Albumin for BCA MicroAssay

Table S3: Design for combined aaRS/tRNA Library for CUU anticodon

	aaRS Amino Acid Positions									tRNA nt Positions				Diversity
Position	230	231	232	233	261	283	284	285	286	33	34	38	39	
wild type seq.	Y	C	P	A	F	H	P	M	D	C	U	A	A	
degenerate nucleotide	kmt	tnt	mmg	kyc	nhh	vds	bnt	wbb	vvw	N	N	N	N	
Codon Diversity	4	4	4	4	36	18	12	18	18	4	4	4	4	1.65E+11
Residue Diversity	4	4	4	4	16	13	12	9	12	4	4	4	4	1.77E+10
Possible Amino Acids	Y	C	P	A	*	R	A	M	A					
	A	Y	T	F	A	N	R	L	R					
	D	S	K	S	D	D	D	R	N					
	S	F	Q	V	E	Q	C	W	D					
					F	E	G	C	Q					
					H	G	H	F	E					
					I	H	L	I	G					
					K	I	F	T	H					
					L	L	P	S	K					
					N	K	S		P					
					P	M	Y		S					
					Q	S	V		T					
					S	V								
					T									
					V									
					Y									

Table S4: Primers Used for Mutagenesis

S4.1 Mutagenic primers for mutating GFP fluorophore codons

Primer code	Primer sequence	Primer name
DL	AAAGCATTGAACACCaTgGGTCAGAGTAGTGAC	QC GFP Y66H (cat) DL
FI	AAAGCATTGAACACCgTgGGTCAGAGTAGTGAC	QC GFP Y66H (cac) FI
FT	GGAAAAGCATTGAACACCaaAGGTCAGAGTAGTGAC	QC GFP Y73F (ttt) FT
FU	GGAAAAGCATTGAACACCattGGTCAGAGTAGTGAC	QC GFP Y73N (aat) FU
FV	GGAAAAGCATTGAACACCcttGGTCAGAGTAGTGAC	QC GFP Y73K (aag) FV
FY	GGAAAAGCATTGAACACCgaaGGTCAGAGTAGTGAC	QC GFP Y73F (uuc) FY
FZ	GGAAAAGCATTGAACACCgttGGTCAGAGTAGTGAC	QC GFP Y73N (aac) FZ
GA	GGAAAAGCATTGAACACCtttGGTCAGAGTAGTGAC	QC GFP Y73K (aaa) GA

S4.2 Mutagenic primers for mutating the anticodon of the MJ tRNA^{Opt}

Primer code	Primer sequence	Primer name
FX	GGCAGAACGGCGGACTcttAATCCGCATGGCAGG	QC tRNA Opt cuu anticodon FX
FW	GGCAGAACGGCGGACTattAATCCGCATGGCAGG	QC tRNA Opt auu anticodon FW
FP	GGCAGAACGGCGGACTaaaAATCCGCATGGCAGG	QC tRNA Opt aaa anticodon FP
CA	TATGATCAGTGCACGGCTAACTAAGCGGCCTGCTGACTTTCTCGCCGATCAAAAGGCA	tRNA TBIO CA
CB	CTCGCCGATCAAAAGGCATTTTGCTATTAAGGGATTGACGAGGGCGTATCTGCGCAGTAA	tRNA TBIO CB
CC	GGCGTATCTGCGCAGTAAGATGCGCCCCGCATTCCGGCGGTAGTTCAGCAGGGCAGAACG	tRNA TBIO CC
CG	CGGAGGGGATTTGAACCCCTGCCATGCGGATTCATAGTCCGCCGTTCTGCCCTGCTGAAC	tRNA TBIO CG
CE	AAAAAGCCTGCTCGTTGAGCAGGCTTTTCGAATTTGGTCCGGCGGAGGGGATTTGAACCC	tRNA TBIO CE
CF	CTGAGCTGCTCGAGCATGCAAAAAAGCCTGCTCGTTGAG	tRNA TBIO CF
GZ	GCAGAACGGCGGACTctcgAgCCGCATGGCAGG	tRNA_XhoI_Inact GZ
GX	GGCAGAACGGCGGANNCTTNNTCCGCATGGCAGG	tRNA_CUU Lib GX

S4.3 Primers for mutating aaRS genes

Primer code	Primer sequence	Primer name
HA	CCTGCCGGGCAGctCGagTTCTTGATCTTCGCACGG	aaRS_1 XhoI_Inact HA
HB	CCACGGCGTTTTTCAGcTCgAgcgggtgCAGTTC	aaRS_2 XhoI_Inact HB
HE	CCCTCAACCACGCCGRMCKKANAAKMCGCTTTCTTGATCTTCGCAC	aaRS AC Lib #1 HE
HF	CAGGTCGCCGCCDDNTTTCTCCGGGCGC	aaRS AC Lib #2 HF
HG	CCACGGCGTTTTTCAGWBBVVWANVSHRCAGTTCTTTGTTCTTAAACAAGCTTTC	aaRS AC Lib #3 HG

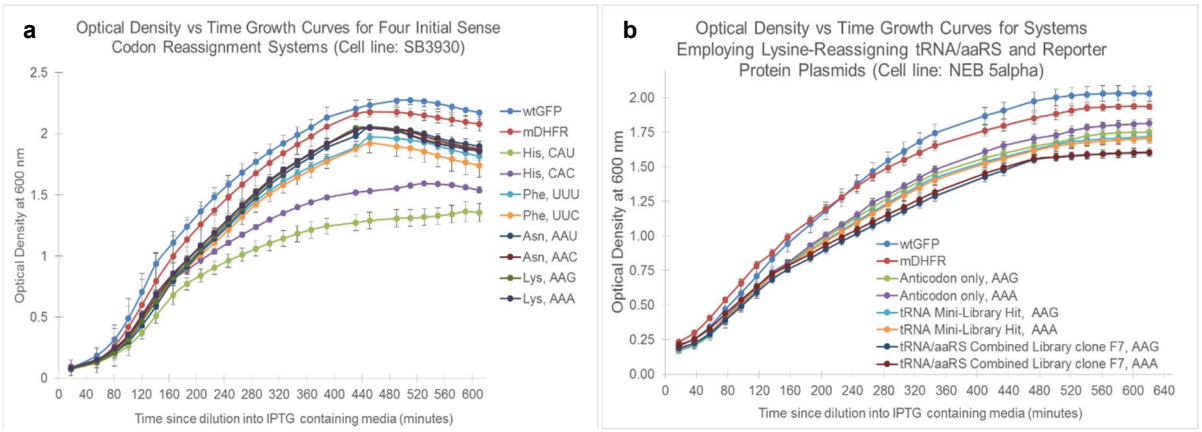


Figure S4. Representative optical density versus time plots for sense codon reassignment systems.

Data in each panel were collected on the same day. a) Growth profiles for *E. coli* SB3930 cells harboring anticodon-modified orthogonal translation machinery for the four codons targeted for reassignment and the related codon read via Watson-Crick interactions by an *E. coli* tRNA. Error bars represent variation across 3 colonies for each system. b) Growth profiles for *E. coli* NEB 5alpha cells harboring improved orthogonal translation machinery for reassigning the Lys AAG codon. Error bars represent variation across 7 colonies for each system. Control cells in which no reassignment can occur due to absence of the orthogonal tRNA and with reporter vectors containing either GFP with a Tyr codon in the fluorophore (wtGFP) or murine dihydrofolate reductase (mDHFR, a non-fluorescent control) are also shown. The vectors expressing the orthogonal translation machinery in the control cells lack the tRNA but do express the orthogonal aaRS in an effort to maintain a similar burden on the cells (e.g. required to produce a gene to confer antibiotic resistance).

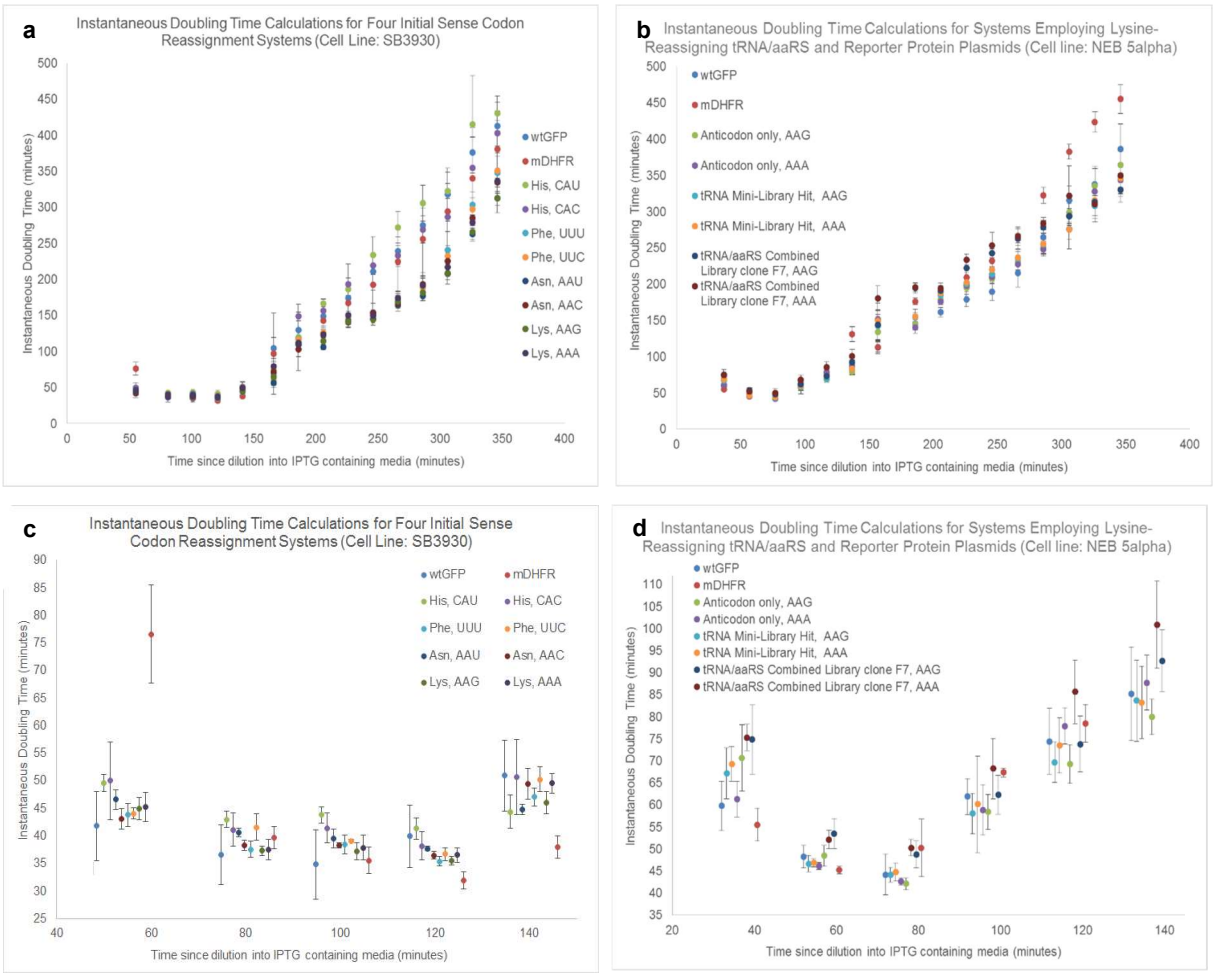


Figure S5. Conversion of optical density measurements plotted above into instantaneous doubling time versus time plots for sense codon reassignment systems. EQ S1 was used to convert the data and is provided below. Data in each panel were collected on the same day. a) Instantaneous doubling times for cells calculated between each measured optical density point for data collected over the first six hours of the experiment depicted in S4a. Error bars represent variation across 3 colonies for each system. b) Instantaneous doubling times calculated at each measured optical density point for data collected over the first 6 hours of the experiment depicted in S4b. Error bars represent variation across 7 colonies for each system. c) Enlarged view of the first approximately 2 hours of instantaneous doubling time values for data shown in Figure S5a. Data are spread across distinct time values to allow clear discernment of the error bars; however data for each grouping as shown in S5a were collected at the same time. d) Enlarged view of the first approximately 2 hours of instantaneous doubling time values for data shown in Figure S5b. Data are spread across distinct time values to allow clear discernment of the error bars; however data in each grouping as shown in S5b were collected at the same time. Control cells in which no reassignment can occur due to absence of the orthogonal tRNA and with reporter vectors containing either GFP with a Tyr codon in the fluorophore (wtGFP) or murine dihydrofolate reductase (mDHFR, a non-fluorescent control) are also shown. The vectors expressing the orthogonal translation machinery in the control cells lack the tRNA but do express the orthogonal aaRS in an effort to maintain a similar burden on the cells (e.g. required to produce a gene to confer antibiotic resistance).

EQ S1.Instantaneous Doubling Time Calculations:

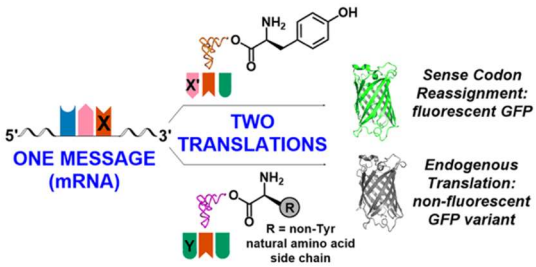
$$OD_{600,b} = OD_{600,a} \times e^{\frac{\ln 2}{T_d} t}$$

$$T_d = \frac{\ln 2 \times t}{\ln OD_{600,b} - \ln OD_{600,a}}$$

T_d is the doubling time in minutes

t is the amout of time elapsed between measuring $OD_{600,b}$ and $OD_{600,a}$

Figure S6. Table of Contents Image



Supplemental References:

(S1) Sambrook, J., and Russell, D. W. (2001) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory press.

(S2) Sidhu, S., and Weiss, G. (2004) Phage display: a practical approach (Clackson, T. and Lowman, H.B., Ed.). Oxford University Press, USA.

(S3) Pédelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C., and Waldo, G. S. (2005) Engineering and characterization of a superfolder green fluorescent protein, *Nat. Biotechnol. 24*, 79–88.

(S4) Velappan, N., Fisher, H. E., Pesavento, E., Chasteen, L., D'Angelo, S., Kiss, C., Longmire, M., Pavlik, P., and Bradbury, A. R. M. (2010) A comprehensive analysis of filamentous phage display vectors for cytoplasmic proteins: an analysis with different fluorescent proteins, *Nucleic Acids Res. 38*, e22.

(S5) Lee, T. S., Krupa, R. A., Zhang, F., Hajimorad, M., Holtz, W. J., Prasad, N., Kuk Lee, S., and Keasling, J. D. (2011) BglBrick vectors and datasheets: a synthetic biology platform for gene expression, *J. Biol. Eng. 5*, 12.

(S6) Chatterjee, A., Sun, S. B., Furman, J. L., Xiao, H., and Schultz, P. G. (2013) A versatile platform for single and multiple unnatural amino acid mutagenesis in *Escherichia coli*, *Biochemistry* 52, 1828-1837.

(S7) Chin, J. W., Martin, A. B., King, D. S., Wang, L., and Schultz, P. G. (2002) Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*, *Proc. Natl. Acad. Sci. U. S. A. 99*, 11020.

(S8) Rubio, M. A. T., Pastar, I., Gaston, K. W., Ragone, F. L., Janzen, C. J., Cross, G. A., Papavasiliou, F. N., and Alfonzo, J. D. (2007) An adenosine-to-inosine tRNA-editing enzyme that can perform C-to-U deamination of DNA, *Proc. Natl. Acad. Sci. U. S. A. 104*, 7821–7826.

(S9) Chin, J. W., Santoro, S. W., Martin, A. B., King, D. S., Wang, L., and Schultz, P. G. (2002) Addition of p-Azido-l-phenylalanine to the Genetic Code of *Escherichia coli*, *J. Am. Chem. Soc. 124*, 9026–9027.

(S10) Gao, X., Yo, P., Keith, A., Ragan, T. J., and Harris, T. K. (2003) 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 Res. 31*, e143.

APPENDIX B

Supporting Information for Chapter 3 — Modification of Orthogonal tRNAs: Unexpected Consequences for Sense Codon Reassignment

TABLE OF CONTENTS

S.1. Cell Strains

NEB 5-alpha (New England Biolabs): *fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*

SB3930 (Yale CGSC): *λ*$^{-}$, *ΔhisB463*

SS320 (MC1061 F', Lucigen): [F'*proAB+lacIqlacZ*ΔM15 *Tn10* (tetr)] *hsdR mcrB araD139* Δ(*araABC-leu*)7679 *lac*X74 *galUgalK rpsL thi*

CJ236 (New England Biolabs): FΔ*(HindIII)::cat* (Tra^{+} Pil^{+} Cam^{R})/ *ung-1 relA1 dut-1 thi-1 spoT1 mcrA*

S.2 General Materials and Reagents

All restriction enzymes, DNA polymerases, and T4 kinase were purchased from New England Biolabs and used according to the manufacturer's instructions. ATP was purchased from Fisher (BP413-25) and dNTPs were purchased form New England Biolabs (N0447S). DNA isolation was performed using a Thermo Scientific GeneJET plasmid miniprep kit (K0503) according to the manufacturer's protocols. Purification of PCR and mutagenesis reactions were performed using either a Thermo Scientific GeneJET PCR purification kit (K0702) according to the manufacturer's instructions.

LB liquid media (per liter: 10 g tryptone, 5 g yeast extract, 5 g NaCl) and LB agar plates with 15 g/L agar (TEKNova, A7777) were used unless otherwise noted. Isopropyl-beta-D-thiogalactoside (IPTG) was purchased from Gold Bio (I2481C5). Spectinomycin (Enzo Life Science, BML-A281) was used at 50 μg/mL to maintain the pUltra-based vectors harboring the tRNA and aaRS genes. Carbenicillin (PlantMedia, 40310000-2) was used at 50 μg/mL to maintain the vectors harboring the GFP reporter gene. All bacterial cultures were grown at 37 °C unless otherwise noted.

Electrocompetent stocks of all strains were prepared in-house according to the method of Sambrook and Russell (Sambrook, J., and Russell, D. W. (2001) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory press). Typical transformation efficiencies for electrocompetent cells produced in this way are 10^{9} cfu/μg of supercoiled DNA. All electroporation transformations were recovered in SOC (20 g/L tryptone, 5 g/L yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 20 mM glucose) for 1 hour at 37 °C with shaking prior to transfer to media containing appropriate antibiotics.

All oligonucleotides were purchased from Integrated DNA Technologies (Coralville, Iowa, USA). All DNA sequencing was performed by Genewiz (Plainfield, NJ, USA).

S.3. DNA Mutagenesis Protocols

QuikChange type Site-Directed Mutagenesis

Non-library site directed mutagenesis for routine cloning and mutation of tRNA anticodons and GFP fluorophore codons was accomplished using a method adapted from the Stratagene QuikChange Multi-Site Directed Mutagenesis protocol. For each 25 µl reaction, 125 ng of mutagenic oligonucleotide was combined with 100 ng of plasmid DNA isolated from NEB 5-alpha cells in 1x Q5 High Fidelity Polymerase with 1x High GC Buffer, 200 µM dNTPs, and 1 unit of Q5 High-Fidelity Polymerase. Reactions were cycled in a PeqSTAR thermocycler (peqLab) for 18 cycles: 98 °C for 20 seconds, and 72 °C for 15 seconds/kb of template plasmid. A final extension at 72 °C was carried out for 5 minutes. 20 units of DpnI enzyme was added directly to the PCR reaction and incubated at 37 °C for 2 hours. Reactions were transformed into electrocompetent NEB 5-alpha without cleanup.

Isolation of single stranded deoxy-uridine containing DNA

DNA libraries were produced with Kunkel mutagenesis using a method adapted from Sidhu and Weiss (Sidhu, S., and Weiss, G. (2004) *Phage display: a practical approach* (Clackson, T. and Lowman, H.B., Ed.). Oxford University Press, USA). Briefly, cultures of CJ236 cells harboring the phagemid to be mutated were grown to an OD_{600} of 0.5 and infected with M13K07 helper phage at a multiplicity of infection of 10:1. The infected culture was transferred into 50-100 mL of LB media with 5 µg/mL chloramphenicol, appropriate antibiotic to maintain the phagemid, and 0.25 µg/mL uridine. Cultures were grown overnight at 30 °C or 37 °C. Cells were pelleted at 17,000 x*g* at 4 °C for 12 minutes in a Sorvall RC 6+ with a Thermo FIBERLite F13-14x50cy rotor. Phage particles were isolated by decanting the supernatant from the pelleted cells into 1/5th volume of 20% 8,000 molecular weight polyethylene glycol and 2.5M NaCl in water. Solutions were incubated on ice or at 4 °C for 2 hours or overnight (respectively). Phage particles were isolated by pelleting at 17,000 x*g* at 4 °C for 20 minutes in a Sorvall RC 6+. The supernatant was decanted and the phage pellet was spun for an additional minute to collect

remaining supernatant, which was then removed. The phage pellet was resuspended in approximately 1 mL of phosphate or Tris buffered saline, pH=7.4. Insoluble material was pelleted out of the phage solution at 17,000 x*g* for 5 minutes. Single-stranded, uridine-enriched DNA (ss dU DNA) was isolated from phage particles using a Qiagen M13 spin kit.

Kunkel Site-Directed Mutagenesis

Mutagenic primers were phosphorylated for 1 hour at 37 °C using T4 polynucleotide kinase. Phosphorylated primers were annealed to 1 μg of single stranded template DNA at a 10:1 molar ratio by incubating at 90 °C for 2 minutes, 50 °C for 3 minutes, and then 25 °C for 5 minutes. The annealed mixture was extended using T7 DNA polymerase in the presence of T4 DNA Ligase and 670 μM ATP and 330 μM dNTPs (each) at room temperature overnight. Reactions were cleaned up using a PCR spin kit and transformed into electrocompetent NEB-5alpha cells. This procedure typically yields 10^5 transformants with 40-80% mutation efficiency.

S.4. GFP Fluorescence-Based Sense Codon Reassignment Efficiency Assays

Superfolder green fluorescent protein (GFP) reporter plasmids (pGFP66xxx, where xxx specifies the codon at position 66) were co-transformed with vectors expressing the modified orthogonal translational components (pWB_Ultra-Tyr-yyy, where yyy indicates the anticodon on the tRNA) into SB3930 *E. coli*. After overnight growth, colonies were picked into 200 μL LB media in a 96 well plate. Cells were grown to saturation (usually 12 hours) with shaking at 37 °C. Cells were diluted 10-fold into LB media with antibiotics to maintain the pWB_Ultra and pGFP plasmids and 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for induction of aaRS and GFP reporter. Assays were performed in a Fluorotrac 200 clear bottom 96 well plate (Greiner 655096) and monitored in a BioTek Synergy H1 plate reader at 37 °C with continuous double orbital shaking. The optical density (OD600) and fluorescence of each well was measured every 20 minutes; optical density was measured at 600 nm, and fluorescence was measured with an excitation at 485 nm and detection at 515 nm with an 8 nm band pass. The relative fluorescence of each 200 μL culture was calculated by dividing the LB media-blanked fluorescence by the OD600. The 100% relative fluorescence unit (reported as fluorescence units per unit optical density at

600 nm, RFU) value for sense codon reassignment efficiency was defined by taking an average of three cultures expressing wild type sfGFP (pGFP66tat) in cells harboring a negative control translational machinery plasmid (pWB_Ultra-Tyr-CUA) to maintain a similar metabolic burden on the cell. Sense codon reassignment efficiency for each tRNA variant was calculated by dividing the individual RFU values from each of three colonies by the average 100% reference RFU value and then averaging the results.

S.5. Phenol:Chloroform Extraction of RNA

NEB 5-alpha cells harboring a pWB_Ultra vector were grown overnight with shaking at 37 °C in LB/spectinomycin (50 μg/mL) media. Cell cultures were diluted 10-fold in fresh media to an approximate OD 600 nm = 0.1) and grown with shaking at 37 °C. When cellular cultures reached an OD600 of 0.5, 4 mL of cell culture were harvested by centrifugation in a microcentrifuge at 17000 x*g* for 1 minute. The supernatant was decanted, and the cell pellets were frozen at -20 °C overnight. The following day, cell pellets were thawed and mixed with 200 μL of B-PER lysis reagent (Thermo Scientific) at room temperature for 30 minutes. 200 μL of 0.3 M NaOAc, 10 mM EDTA, pH 4.5 were added to the lysis reactions to ensure the pH was close to 4.5. 450 μL of water-buffered phenol:chloroform (5:1 ratio) was then added to the lysate and vortexed three times for 60 seconds, 60 seconds, and 30 seconds, with lysates held on ice for 60 seconds between each step. The layers were then separated by centrifugation at 15000 x*g* at 4 °C for 15 minutes. The aqueous layer containing RNA was transferred 2.5 volumes of 100% EtOH for precipitation of nucleic acids. The mixture was vortexed briefly and stored overnight at -20 °C. The nucleic acids were recovered by centrifugation at 15000 x*g* at 4 °C for 30 minutes. The supernatant was aspirated, and the pellet was allowed to air dry prior to resuspension in 100 μL of 1x DNAseI Buffer (Thermo Scientific). One unit of DNaseI was added, and the solution was incubated at 37 °C for 20 minutes. DNaseI was inactivated by phenol:chloroform extraction and ethanol precipitation as described above. RNA was resuspended in 20 μL of 0.3 M NaOAc, 10 mM EDTA, pH 4.5 and quantified using a BioTek Synergy H1 plate reader. Each RNA extraction yielded 12-20 μg of RNA.

Table S1: pWB_Ultra aaRS/tRNA Expression Vectors

aaRS/tRNA Plasmid	Anticodon	Codon Read via Watson-Crick Interactions	Use
pWB_Ultra-Tyr-AUG	AUG	CAU	Reassignment of His CAU codons to Tyr[a]
pWB_Ultra-Tyr-(XhoI)	N/A	N/A	Starting template for tRNA library
pWB_Ultra-Tyr-CUA	CUA	UAG	Vector for sense codon reassignment controls[b]. Used in combination with a GFP reporter with a sense codon at position 66.

a) Biddle, W., Schmitt, M. A., and Fisk, J. D. (2015) Evaluating sense codon reassignment with a simple fluorescence screen, *Biochemistry*, **54**, 7355-7364.

b) Chatterjee, A., Sun, S. B., Furman, J. L., Xiao, H., & Schultz, P. G. (2013). A versatile platform for single-and multiple-unnatural amino acid mutagenesis in *Escherichia coli*. *Biochemistry*, **52**, 1828-1837.

Table S2: Reporter Protein Plasmid Designations and Usage

Reporter Plasmid	Codon at Position 66	Canonically Encoded Amino Acid	Use
pGFP66tat	UAU	Tyr	Wild type GFP vector for 100% fluorescence reference
pGFP66cac	CAC	His	Vector for evaluating CAC codon discrimination
pGFP66cau	CAU	His	Vector for evaluating reassignment of CAU sense codon
pGFP66caa	CAA	Gln	Vector for evaluating CAA codon discrimination

Vector design and construction reported in: Biddle, W., Schmitt, M. A., and Fisk, J. D. (2015) Evaluating sense codon reassignment with a simple fluorescence screen, *Biochemistry* 54, 7355-7364.

Table S3: Oligonucleotide Primers

Primer for focused anticodon loop library of the *M. jannaschii* tRNA$^{Opt}_{AUG}$

Primer code	Primer sequence	Primer name
JQ	CGG CGG ANT atg NNT CCG CAT GGC AGG GGT TCA AAT CC	MjtRNA ATG His Lib JQ

Primers for reverse transcription, amplification, and sequencing of *E. coli* tRNA^{Arg2} and *M. jannaschii* tRNA$^{Opt}_{AUG}$ variants

Primer code	Primer sequence	Primer name
LW	GAC ACG GTA CCA CAC AAC TGG GCA ACG CAA CCT AGC TAA TGG TGC ATC CGG GAG GAT TCG	RT-PCR tRNA-Arg2 LW
LX	GAC ACG GTA CCA CAC AAC TGG	Wolf[a] Arg2 Primer LX
LY	GCA TCC GTA GCT CAG CTG G	Wolf[a] Arg3 Primer LY
LZ	GAC ACG GTA CCA CAC AAC TGG GCA ACG CAA CCT AGC TAA TGG TCC GGC GGA GGG GAT TTG	RT-PCR tRNA_Opt LZ
CC	GGC GTA TCT GCG CAG TAA GAT GCG CCC CGC ATT CCG GCG GTA GTT CAG CAG GGC AGA ACG	tRNA TBIO CC

a) Wolf, J., Gerber, A. P., and Keller, W. (2002) TadA, an essential tRNA-specific adenosine deaminase from Escherichia coli, *Embo Journal* **21**, 3841-3851.

Figure S1

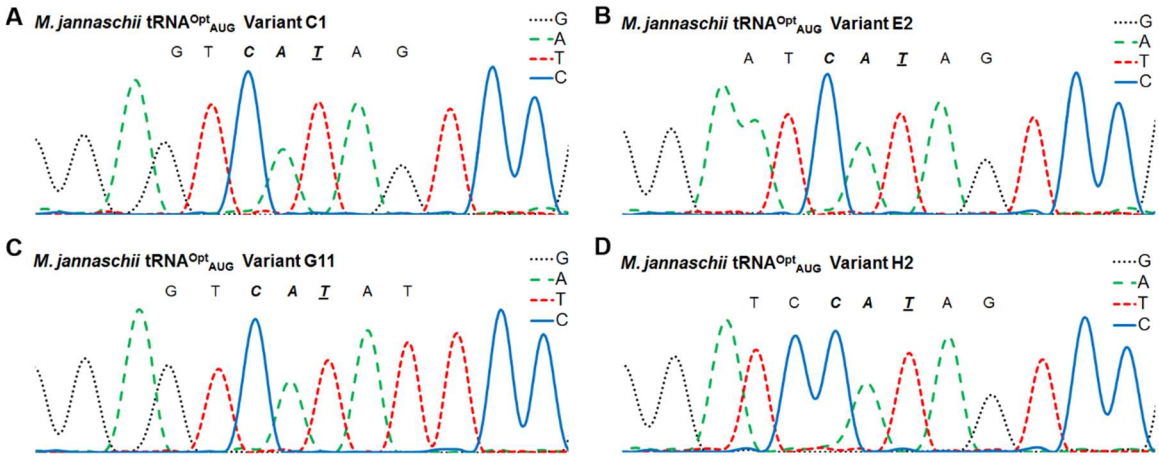


Figure S1. Sequence traces of the additional reverse transcribed *M. jannaschii* $tRNA^{Opt}_{AUG}$ variants isolated from the tRNA anticodon loop library. The portion of the chromatogram that corresponds to the tRNA anticodon sequence is italicized and bolded. In all cases the DNA encoded A34 is not modified to inosine.

Figure S2

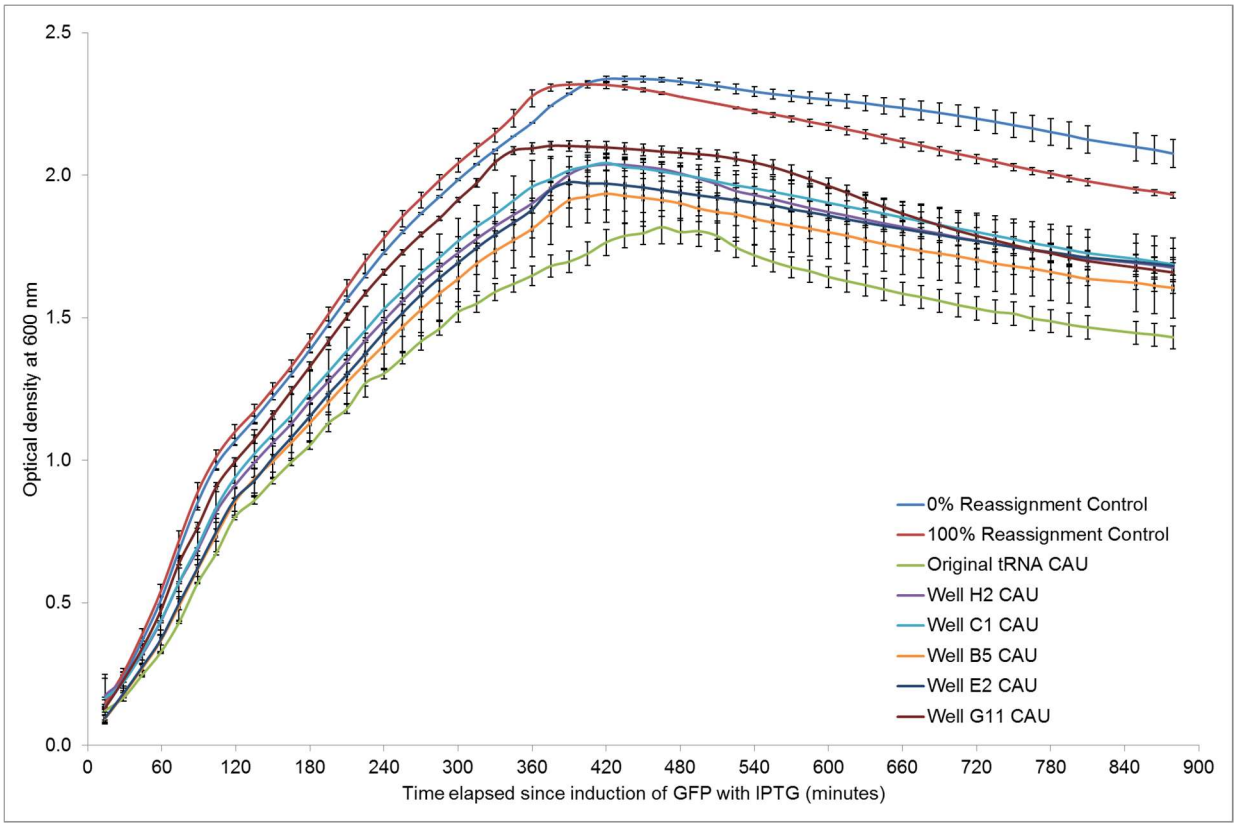


Figure S2: Optical density at 600 nm (OD_{600}) versus time profile for the *M. jannaschii* $tRNA^{Opt}_{AUG}$ variants evaluated in this manuscript. The original tRNA variant in which only the anticodon was changed exhibits the least carrying capacity. Each of the variants that result in improved discrimination between the CAU and CAC codons exhibit improved growth profiles more similar to those of the control systems.