



MACQUARIE
University
SYDNEY · AUSTRALIA

Reprogramming the genetic code: defining a new start codon in *Escherichia coli*

Russel Miranda Vincent

**A thesis submitted in partial fulfilment of the
degree of Master of Research**

Supervisor: Dr. Paul R. Jaschke

Department of Chemistry and Biomolecular Sciences

Faculty of Science

Macquarie University

9th October, 2017

Statement of Originality

This thesis entitled “**Identifying and defining a new start codon in Escherichia coli**” is representative of the research study conducted between January 2017 and October 2017 for the completion of Master of Research degree in Chemistry and Biomolecular Sciences at Macquarie University, New South Wales, Australia. The work presented in this thesis is certified to be original by the author, unless otherwise referenced in the literature and/or acknowledged of personal advice and suggestions.

This thesis is formatted according to Master of Research guidelines prescribed by the Faculty of Science & Engineering and Department of Chemistry & Biomolecular Sciences and has not been submitted for qualification or assessment to any other institution.

Sincerely,

Russel Miranda Vincent

Student ID: 43971199

Table of Contents

Acknowledgements.....	I
Abbreviations	II
Abstract.....	III
Chapter 1: Introduction	1
1.1 Translation is a fundamental cellular process	1
1.1.1 Major components of translation in prokaryotes and eukaryotes.....	2
1.1.1.1 Ribosome	2
1.1.1.2 Messenger RNA.....	3
1.1.1.3 Transfer RNA	4
1.2 The standard genetic code.....	4
1.2.1 Non-overlapping codons.....	5
1.2.2 Code degeneracy.....	5
1.2.3 Variation from the 'standard' genetic code.....	5
1.2.3.1 Amber codons	5
1.2.3.2 Natural variation in the genetic code	6
1.3 Translation initiation in prokaryotes.....	6
1.3.1 Initiation modes	7
1.3.1.1 30S binding mode	7
1.3.1.2 70S-scanning mode	7
1.3.1.3 Leaderless mRNA mode	7
1.3.2 Accessory features of the initiation complex	8
1.3.2.1 The initiation codon	8
1.3.2.2 Role of initiation factors.....	9
1.3.3 Unique features of initiator tRNA	10

1.3.3.1 Conserved anticodon stem enables P-site discrimination between initiator and elongator tRNA.....	10
1.3.3.2 Conserved acceptor stem of initiator tRNA enables formylation and IF2 recognition	10
1.3.3.3 Codon-anticodon pair specificity	11
1.4 Re-engineering Translation	12
1.4.1 Engineering sense codons.....	12
1.4.2 A new strain for amber codon engineering	13
1.4.3 Engineering start codons	13
1.4.4 An orthogonal translation initiation system using amber initiator tRNA ^{fmet2} _{CUA}	14
1.5 Project aims and objectives	14
Chapter 2: Methods	15
2.1 Bacterial strains, cultures and glycerol stock.....	15
2.1.1 Bacterial strains.....	15
2.2.1 Bacterial culture preparation.....	15
2.2.2 Glycerol stock preparation and storage.....	15
2.2 Plasmid construction.....	15
2.2.1 Synthesized oligonucleotides and dsDNA.....	15
2.2.2 Plasmids	15
2.2.3 Plasmid assembly	16
2.2.4 Transformation	17
2.2.5 Colony PCR	17
2.2.6 Plasmid extraction.....	17
2.2.7 Sequencing	18
2.3 Sequential double transformation	18
2.3.1 Transformation of amber initiation plasmids	18
2.3.2 Transformation of reporter plasmids	18

2.4 Fluorescence measurements	18
2.4.1 Bulk fluorescence measurements.....	18
2.4.2 Dynamic fluorescence measurements.....	19
2.4.3 Data analysis	19
2.4.4 Flow cytometry	19
2.5 Proteomic analysis	20
2.5.1 Cell growth and harvesting	20
2.5.2 Sample preparation for LC-MS.....	20
2.5.3 LC-MS	21
2.5.4 LC-MS data analysis.....	21
2.6 Fitness analysis.....	22
2.6.1 Growth measurements	22
2.6.2 Data analysis	22
Chapter 3: Results	23
3.1 Defining an amber initiation tRNA and modular fluorescent reporter system in E. coli	23
3.1.1 Designing plasmids to express an amber translation initiation system	23
3.1.2 Designing plasmids to express fluorescent reporters compatible with amber initiation plasmids	24
3.1.3 The amber initiation system translates a sfGFP reporter with an amber start codon.....	25
3.1.4 Flow cytometry reveals $\text{tRNA}_{\text{CUA}}^{\text{fmet2}}$ initiates translation across a uniform population	28
3.1.5 Dynamic induction of $\text{tRNA}_{\text{CUA}}^{\text{fmet2}}$ potentially reveals effect of tRNA maturation on start codon interaction fidelity.....	29
3.2 Measuring whole-cell proteomic effects from $\text{tRNA}_{\text{CUA}}^{\text{fmet2}}$ expression	30
3.2.1 Production of $\text{tRNA}_{\text{CUA}}^{\text{fmet2}}$ causes few changes to chromosomal protein expression in E. coli C321. Δ A.exp	30
3.2.2 Production of $\text{tRNA}_{\text{CUA}}^{\text{fmet2}}$ results in no detectable off-target peptides produced from UAG open reading frames in E. coli C321. Δ A.exp.....	32

3.3 Amber initiator efficiency in Generally Recognized as Safe (GRAS) E. coli strains	34
3.3.1 Expression of tRNA _{CUA} ^{fmet2} results in efficient translation initiation from sfGFP(UAG) reporter in five common laboratory E. coli strains	34
3.3.2 Strain-specific amber start codon initiation capacity	35
3.3.3 Production of tRNA _{CUA} ^{fmet2} causes minimal growth defects	36
Chapter 4: Discussion	38
4.1 Creating plasmids to measure new aspects of the amber initiator	38
4.2 Effect of shifting the stoichiometric ratio of tRNA _{CUA} ^{fmet2} to wild type tRNA ^{fmet}	39
4.3 Amber initiator expression dynamics potentially reveals tRNA _{CUA} ^{fmet2} maturation process	40
4.4 Proteomic analysis	41
4.5 Amber initiator expression across five common laboratory E. coli strains	43
4.5.1 The amber initiator tRNA is functional in all five E. coli GRAS strains	43
4.5.2 C321.ΔA.exp is the only strain with growth defect due to tRNA _{CUA} ^{fmet2} expression	45
4.6 Off-target effects of amber initiator tRNA _{CUA} ^{fmet2} expression	46
4.7 Conclusion and future directions	47
References	i
Supplementary Material	viii

Acknowledgements

My Masters of Research has been an inspirational journey of self-discovery and a major highlight of my life as a student. None of it would have been remotely possible without the constant guidance and support of my supervisor **Dr. Paul R. Jaschke**. I was a mere graduate student inquisitive and naïve, but interactions with him over the course of this journey, has been the reason why I can proudly call myself a research graduate now. He has been a wonderful person and mentor throughout with invaluable academic advice and expertise that has enabled this work. A source of inspiration for me and many others in the way he confidently juggles a million tasks still finding the time to be unperturbed with all the burden. I would also like to acknowledge his sincere efforts in helping me compile this thesis and being available all the time.

The remarkable team of faculty members in the Master of Research committee including **Dr. Louise Brown, A/Prof Bridget Mabbutt, Prof Paul A. Haynes, A/Prof Mark Molloy, Prof Helena Nevalainen** and **Prof Ian Paulsen** is the reason that I could enter the current research degree. Their constant encouragement and belief in my abilities has been a great motivator and I appreciate their continuous assistance and support.

During my research year, I have made many friends who have helped lower the stress levels along with excellent inputs to help me along. I would like to thank the **SynBio** group post-docs, **Dr. Heinrich Kroukamp, Dr. Hugh Goold, Dr. Thomas Williams** and **Dr. Niel van Wyk** for their excellent inputs and advice along with the rest of the group (**Elizabeth Daniels, Monica Espinoza Gomes, Alex Carpenter, Elizabeth Wightman, Xin Xu, and Kai Peng**) for being so welcoming and helpful every bit of the way. I would also like to thank **Dominic Logel** and **Bradley Wright** for their help in laboratory work and **Dana Pascovici** (APAF Bioinformatician) for making down-stream analysis so simple.

I would like to especially thank **Ms. Sayantani Chatterjee** for tolerating and supporting me throughout this life-changing journey. It has been a great pleasure experiencing and sharing the highs and lows of life with her.

Lastly, and most importantly, the two women who have nurtured and inculcated the precious virtues to build my personality, my mom **Ms. Monica Vincent** and sister **Mrs. Belinda Vincent-D’Rozario**. I would love to thank them for supporting every decision that I’ve made and sharing every moment separated miles across the ocean.

Abbreviations

A	Adenine	ng	Nano-gram
Ala	Alanine	nL	Nanolitre
bp	Base pair	NSAF	Normalized spectral abundance factor
C	Cytosine	PBS	Phosphate buffered saline
cDNA	Complementary deoxyribonucleic acid	Post-PT QC	Post-peptidyltransferase quality control
CDS	Coding sequence	RF1	Release factor 1
Da	Dalton	RF2	Release factor 2
DNA	Deoxyribonucleic acid	RF3	Release factor 3
DTT	Dithiothreitol	RBS	Ribosome binding site
FITC	Fluorescein isothiocyanate	RNA	Ribonucleic acid
fmol	Femtomole	Rpm	Rotation per minute
G	Guanine	SDS	Sodium dodecyl sulfate
Glu	Glutamic Acid	sfGFP	Superfolder green fluorescent protein
Gly	Glycine	Ser	Serine
IAA	Iodoacetamide	T	Thymine
IF1	Initiation factor 1	TBE	Tris-borate ethylene-diamine(tetra)-acetic acid buffer
IF2	Initiation factor 2	TBS	Tris-buffered saline
IF3	Initiation factor 2	Thr	Threonine
IPTG	Isopropyl β -D-1-thiogalactopyranoside	Tris-HCl	Tris-(hydroxymethyl)-aminomethane hydrochloric acid
kDa	Kilo-Dalton	tRNA	Transfer ribonucleic acid
LC	Liquid chromatography	tRNA ^{fmet2} _{CUA}	Amber initiator tRNA
m ⁷ G	7-methylguanosine	Tyr	Tyrosine
mRNA	Messenger ribonucleic acid		
mL	Mililitre		
MS	Mass spectrometry		
MS/MS	Tandem mass spectrometry		
ms ² i ⁶	2-methylthio-N ⁶ -isopentenyladenosine		

Abstract

Translation maps triplet RNA codons onto single amino acids within a protein. Bacterial translation initiation occurs at a conserved AUG start codon recognised by the initiator tRNA anticodon, with studies demonstrating that codon-anticodon complementarity is sufficient for initiation. Thus far, engineering initiation codons have been neglected aside from one identified mutant amber initiator tRNA ($\text{tRNA}_{\text{CUA}}^{\text{fmet}2}$) that can start protein synthesis from the UAG stop codon. In this study, I create a new inducible system to express $\text{tRNA}_{\text{CUA}}^{\text{fmet}2}$ and a new set of fluorescent reporter plasmids to enable population-level studies for the first time. Fluorescence measurements confirmed that $\text{tRNA}_{\text{CUA}}^{\text{fmet}2}$ initiates translation from UAG start codon with 200-fold increase on inducing $\text{tRNA}_{\text{CUA}}^{\text{fmet}2}$ compared to the repressed condition with 30-fold increase in reporter expression from an AUG start codon in similar conditions. Time-course measurements indicate different initiation effects as $\text{tRNA}_{\text{CUA}}^{\text{fmet}2}$ matures. Proteomic analysis of cells expressing $\text{tRNA}_{\text{CUA}}^{\text{fmet}2}$ reveal a minimal effect on the host proteome with no evidence of peptides initiating from genomic UAG codons. I demonstrate that my $\text{tRNA}_{\text{CUA}}^{\text{fmet}2}$ -expressing plasmid is transportable in all five *E. coli* strains with no adverse fitness defects. In this thesis, I define $\text{tRNA}_{\text{CUA}}^{\text{fmet}}$ characteristics with future scope for refinement for this orthogonal translation initiation system.

Chapter 1: Introduction

1.1 Translation is a fundamental cellular process

Translation is a fundamental molecular process occurring in every living cell that is responsible for the formation of proteins, which comprise the structural and catalytic machinery of life (Figure 1A). The translation process is divided in three phases known as initiation, elongation, and termination (Figure 1B) (Cooper 2000).

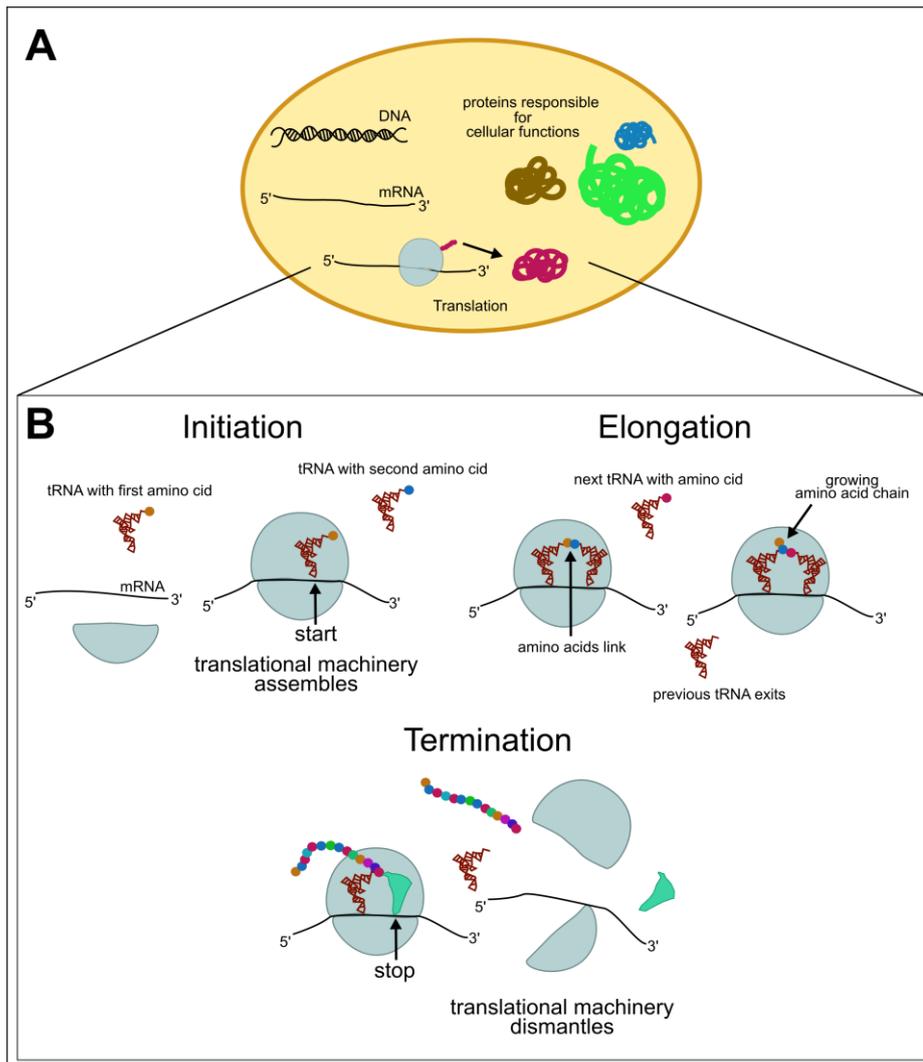


Figure 1: Translation is a fundamental process in the cell.

(A) Translation is a crucial process that synthesizes proteins responsible for various cellular functions. (B) Translation occurs in three phases namely initiation, elongation, and termination. The different coloured circles represent different amino acids.

Translation initiation brings all the parts of the translation machinery together to link the first two amino acids and start synthesis of a protein from the beginning of a gene, defined in the genetic code by a start codon. The elongation phase continues to extend the amino acid chain until the end of the gene at a stop codon. The process of translation ends at the termination phase when the growing chain of amino acids stops and the translation machinery dismantles (Hirokawa et al. 2005). The ribosome plays a central role in all these phases, reading the genetic message encoded in messenger RNAs (mRNAs) to produce a polymerised chain of amino acids using transfer RNAs (tRNAs) as adaptor molecules (Figure 1B).

1.1.1 Major components of translation in prokaryotes and eukaryotes

1.1.1.1 Ribosome

The ribosome is made of multiple proteins and RNAs forming a large complex and functioning as the site of protein production. Its catalytic activity links amino acids via peptide bonds to form polypeptide chains, which mature to form a protein. During translation initiation, the ribosome attaches to the mRNA and exposes the genetic message to the tRNA adaptor molecule. The general structure and function of the ribosome is conserved in all domains of life with minor variations (Ramakrishnan 2002).

The large ribosomal subunit has catalytic activity to form peptide bonds between the C-terminus of the growing polypeptide chain (chain of amino acids) and N-terminus of the next amino acid to be attached. The prokaryotic large ribosomal subunit is called the 50S subunit while the eukaryotic version is referred to as the 60S subunit due to different sedimentation rates (Cooper 2000; Hirokawa et al. 2005). The prokaryotic 50S large subunit consists of 5S and 23S rRNAs and 21 proteins, whereas the eukaryotic 60S large subunit consists of 5S, 5.8S, and 28S rRNAs and 50 proteins (Figure 2). The smaller ribosomal subunit in both prokaryotes and eukaryotes is responsible for recognizing and binding the mRNA ribosomal binding site via conserved ribosomal RNA (rRNA) sequences. In prokaryotes, the small 30S ribosomal subunit consists of the 16S rRNA which recognises and binds to a ribosomal binding site (RBS), commonly called the Shine-Dalgarno sequence, on prokaryotic mRNA and 21 proteins (Shine and Dalgarno 1975). In eukaryotes, the small 40S subunit consists of the 18S rRNA (Figure 2) which recognizes and binds a RBS called the Kozak sequence, on the eukaryotic mRNA (Kozak 1987) and 33 proteins. The small subunit after binding to the mRNA at the RBS such that the start codon is aligned to interact with the incoming

initiator tRNA and start protein synthesis. Subsequently, the large ribosomal subunit binds to complete the ribosomal machinery and releases the complex to begin the elongation phase.

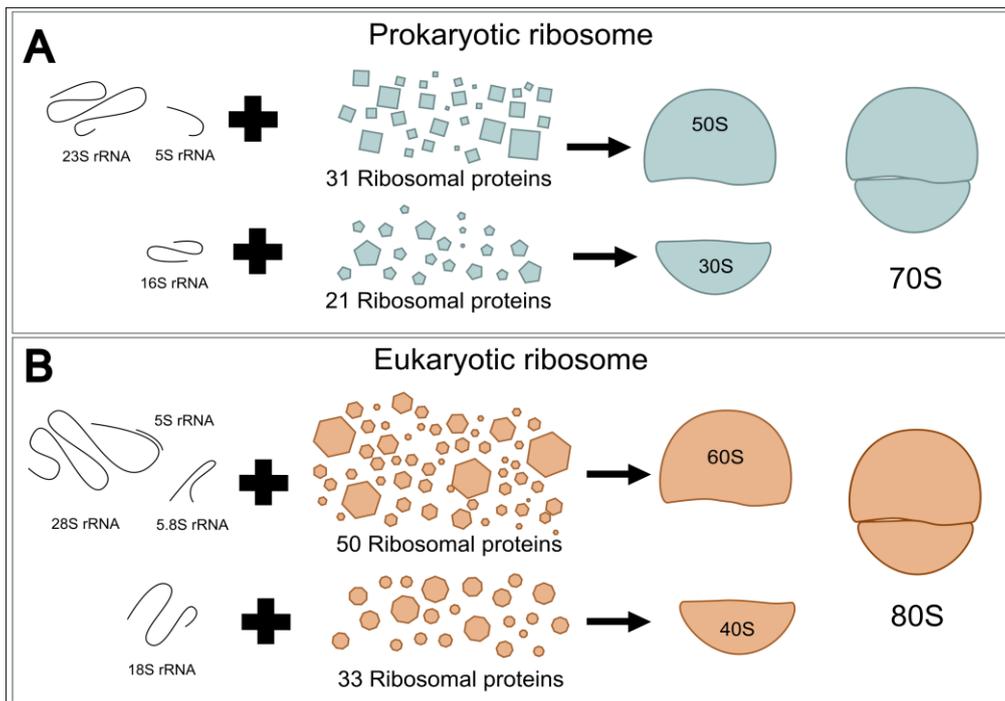


Figure 2: Components of the prokaryotic and eukaryotic ribosome.

(A) Prokaryotic 70S ribosome is made up of two subunits: the 50S and 30S. The large 50S subunit is made up of two rRNAs (23S rRNA and 5S rRNA) and 31 ribosomal proteins. The small 30S is made up of a 16S rRNA and 21 ribosomal proteins.

(B) Eukaryotic 80S ribosome is made up of two subunits: the 60S and 40S. The large 60S subunit is made up of three rRNAs (28S rRNA, 5.8S, and 5S rRNA) and 50 ribosomal proteins. The small 40S is made up of a 18S rRNA and 33 ribosomal proteins.

1.1.1.2 Messenger RNA

The mRNA is a string of ribonucleotides with genetic information (i.e. gene) transferred from the DNA via the process of transcription. The process of translation converts the mRNA message to have a functional meaning as a chain of amino acids (protein) serving as a temporary signal for genetic expression due to its short half-life (Kushner 2002). The beginning of each gene encoded in the mRNA starts from a group of the three consecutive nucleotides (triplet) called a start codon. Succeeding RNA base triplets are called codons (Tsugita et al. 1960; Crick et al. 1961). The region (~3-15 nt) upstream of the start codon consists of the RBS with the Shine-Dalgarno sequence in prokaryotes (Shine and Dalgarno 1975; Hyatt et al. 2010). Eukaryotic mRNAs are more complex with special secondary structures at the 5'-terminus (5'-7-methylguanosine capping) and 3'-

terminus (3' polyA tail) along with the upstream Kozak sequence in the RBS (Kozak 1987), which increases half-life of the mRNA and stimulates translation initiation, respectively (Mitchell et al. 2010).

1.1.1.3 Transfer RNA

The tRNA is an adaptor molecule that connects nucleic acids with amino acids in the ribosome. The tertiary structure of tRNAs is described as a compact L-shape but is frequently modelled as a cloverleaf structure. The structure of all tRNA molecules is formed through complementary pairing between nucleotide bases in the stem regions. The cloverleaf structure depicts three loops: D-loop (Dihydro-uridine loop), T Ψ C loop (Pseudo-uridine loop) and the anticodon loop with an additional variable loop (Figure 3). The anticodon loop contains the three nucleotides that specifically bind to the codon on the mRNA, forming the physical link within the tRNA between the codon and amino acid bound to the 3'-end.

The specific amino acid attached to each tRNA is also based on direct interaction with the anticodon loop. An amino-acyl synthetase charges the tRNA with an amino acid by recognising the same anticodon (Normanly and Abelson 1989). Together, these two recognition events are major determinants of overall translation fidelity (Anantharamaiah et al. 2003).

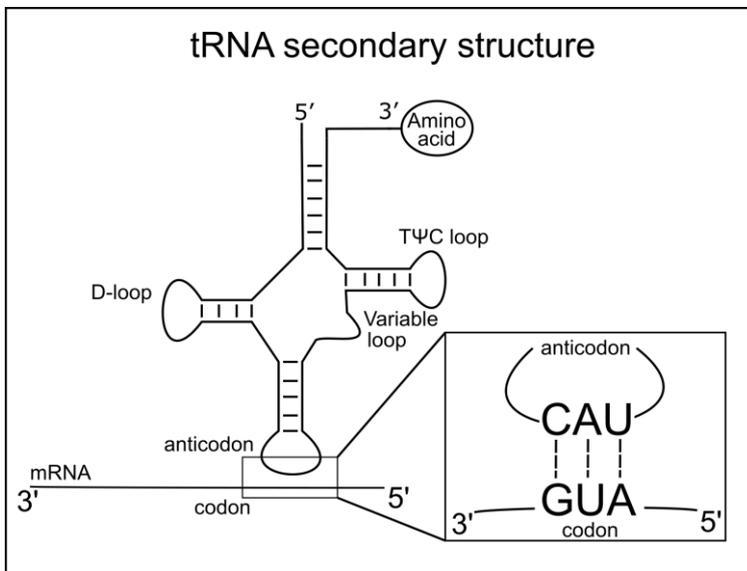


Figure 3: Cloverleaf structure of tRNA bound to an mRNA via codon-anticodon pairing.

The secondary cloverleaf structure of tRNA depicts the D-loop, T Ψ C loop, anticodon loop, and variable loop. The anticodon of the tRNA binds to the codon on the mRNA via specific complementary base pairing.

1.2 The standard genetic code

The standard genetic code is conserved across all domains of life and acts as the link that enables information transfer from DNA to RNA to protein. In the early 1960s, the general characteristics of the standard genetic code that helped information transfer between DNA, RNA, and protein were

described in the central dogma of molecular biology (Crick et al. 1961). The general nature of the genetic code solved previously disputed features regarding codon arrangement in terms of overlapping or non-overlapping nature and code degeneracy.

1.2.1 Non-overlapping codons

The non-overlapping feature of the genetic code was discovered while studying mutant Tobacco Mosaic Viruses. Scientists observed that proteins created from mutant viruses exhibited a single amino acid change (Tsugita et al. 1960). If the genetic code was overlapping, a mutation in one nucleotide base would give rise to changes in three consecutive amino acids whereas a non-overlapping genetic code would only have one (Crick et al. 1961).

1.2.2 Code degeneracy

In the genetic code there are 64 possible permutations of three nucleotide codons, but there are only 20 standard amino acid, indicating either that some codons do not code for anything or that each amino acid can be coded by multiple codons (Crick et al. 1961). By the late 1960s, scientists found out all possible combinations of nucleotide triplets and their corresponding amino acid through a series of ingenious experiments (Nirenberg et al. 1962) giving the world its first iteration of the standard genetic code (Woese et al. 1966). This redundancy of the genetic code is now referred to as codon degeneracy.

1.2.3 Variation from the 'standard' genetic code

Although the genetic code was initially hypothesised to be conserved across all domains because of its fundamental use, this idea was disproved fairly early on as scientists found discrepancies in the genetic code.

1.2.3.1 Amber codons

Early experimental evidence for variation in the genetic code came from the use of amber mutants. An amber mutant was defined as bacteriophage T4 which acquired a conditional lethal nonsense mutation (Epstein et al. 2012). One of the first uses of amber mutants was through the creation of bacteriophage T4 mutants that could not lyse wild-type *Escherichia coli* strain, but these mutants could be grown on a mutant *E. coli* strain called an amber suppressor (because it suppressed the negative amber mutation). It was discovered later that the suppressor strain possesses a mutant tRNA which recognises the amber (UAG) codon and allows translation of this codon which is normally read as 'stop' in other strains. The amber suppressor strain effectively reassigns the function of UAG from a stop codon to a sense codon (Brenner et al. 1965).

1.2.3.2 Natural variation in the genetic code

The first instance of a distinct natural genetic code was in human mitochondria where AUA, instead of encoding Leucine as in the standard code, encoded methionine (Barrell et al. 1979). Successive discoveries observed stop codon (UAG and UGA) reassignments as well. For example, the UGA stop codon is reassigned to encode tryptophan in both animal mitochondria and *mycoplasma* species (Barrell et al. 1979; Panicker et al. 2015). In contrast, UAG is assigned to encode leucine, alanine and glutamine in the mitochondria of several plant and fungal species but in particular, the mitochondrial genetic code seems to be especially variable (Knight et al. 2001).

There have been various attempts to explain genetic code variation as an evolutionary phenomenon with two theories gaining the most prominence. The codon reassignment theory explains that a mutational pressure can lead to the disappearance of a codon from the genetic code of an organism. The mutational pressure, such as a change in the GC content of the organism's genome, could over time mutate and replace all instances of a given codon, freeing up the corresponding tRNA (Osawa and Jukes 1989).

Conversely, the ambiguous intermediate theory proposes that disappearance of a codon from all gene sequences in an organism is not a necessary pre-condition for codon reassignment. In this theory, reassignment is possible due to a mutation in the functional centre of a tRNA gene copy. This mutation leads to an ambiguous tRNA molecule that can add to or change the meaning of a codon, causing a change in the genetic code. The intermediate variation becomes permanent if the mutation confers a selective advantage (Schultz and Yarus 1996). The CUG codon in *Candida* species that can encode both serine and leucine supports this ambiguous intermediate model (Knight et al. 1999, 2001).

1.3 Translation initiation in prokaryotes

Protein synthesis in prokaryotes occurs after the entire ribosomal complex is formed in the initial phase of translation. The ribosomal complex formation serves as a check point because it is the rate-limiting step of translation (Hauryliuk and Ehrenberg 2006). The process of translation initiation in prokaryotes has been widely studied over decades and the bulk of evidence to date supports three distinct modes that can operate simultaneously in the cell at different rates.

1.3.1 Initiation modes

1.3.1.1 30S binding mode

This is the widely-accepted dominant mode of translation initiation in prokaryotes where the 30S ribosomal subunit initiates the process (Cooper 2000; Laursen et al. 2005). In this binding mode, the 30S subunit recognises the initiation site on the mRNA, which contains the Shine-Dalgarno sequence upstream from the start codon (Figure 4A). The initiator tRNA together with initiation factors, accessory proteins that help in translation initiation, interact with the 30S subunit to form the 30S initiation complex (30S IC) allowing the anticodon of the initiator tRNA to interact with the initiation codon on the mRNA. The attachment of the 50S ribosomal subunit and subsequent release of the initiation factors complete the initiation complex maturation and allows translation to enter the elongation phase (Cooper 2000; Clancy and Brown 2008).

1.3.1.2 70S-scanning mode

Contrasting with the 30S-binding model in which ribosomal subunits are recycled after every translation event, the 70S-scanning mode model proposes that after successfully completing translation of a coding sequence and encountering a stop codon, the 70S ribosomal complex does not dissociate and instead continues to move downstream along the mRNA (Figure 4B). With the help of the initiator tRNA^{fmet} and initiation factors bound, the 70S -scanning complex encounters the next Shine-Dalgarno sequence and start codon downstream ejecting the initiation factors entering translation elongation as in the 30S-binding model.

The 70S scanning mode is mainly observed in a bicistronic mRNA where the first gene is translated via the 30S-mode while translation of the second gene downstream occurs via the 70S-scanning mode since it is energy efficient compared to the 30S mode (Yamamoto et al. 2016). Translational-coupling and bicistronic constructs (Mutalik et al. 2013) have been documented to enhance heterologous protein expression lending support for the existence of a 70S scanning mode for prokaryotic ribosomes (Makoff and Smallwood 1990; Mendez-Perez et al. 2012). Recent experimental evidence also suggests that the initiator tRNA can trigger scanning mode in 70S ribosomes and that neither the presence of competing mRNA nor ribosomal release factors can prevent this (Yamamoto et al. 2016).

1.3.1.3 Leaderless mRNA mode

A rare subset of mRNAs in prokaryotes lack the Shine-Dalgarno sequence and instead have a start codon within 5 nucleotides of the 5' end. These mRNAs are called 'leaderless' and they can

surprisingly still initiate translation through an alternative mode where a 30S—IF2-initiator-tRNA complex (Figure 4C) is formed (Moll et al. 2002). Efficient leaderless mRNA initiation is possible even in the absence of initiation factors (Udagawa et al. 2004) and the 5′-terminal start codon was described to provide a major feature for ribosome recognition (Brock et al. 2008). However, this mode of initiation may be specific for the λ-phage leaderless mRNA employed in *E. coli*.

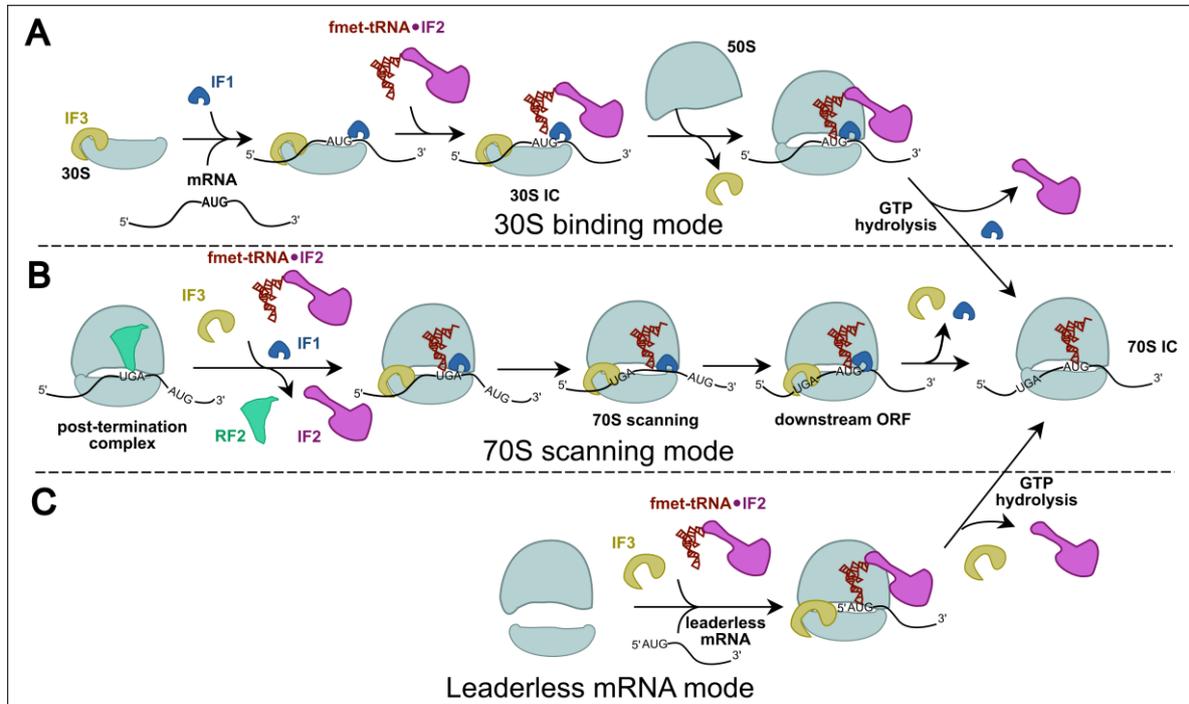


Figure 4: Three proposed translation initiation modes in prokaryotes.

(A) 30S binding mode requires IF1 and IF3 to aid initiator fmet-tRNA binding, recruited by IF2.

(B) 70S scanning mode is influenced by initiator fmet-tRNA and IF1 to scan downstream for the next coding sequence to initiate translation in a bicistronic mRNA.

(C) Leaderless mRNA mode requires a special feature to initiate translation without a translation initiation region (TIR) heavily dependent on initiator fmet-tRNA and IF3 (adapted from Yamamoto et al. 2016).

1.3.2 Accessory features of the initiation complex

1.3.2.1 The initiation codon

The start codon of a gene is the first nucleotide triplet (codon) in the mRNA which is translated by the ribosome and sets the reading frame (Crick et al. 1961). Bioinformatic analysis of model bacterial genome annotations suggest that AUG (81.8%) is the most common initiation codon, followed by GUG (13.8%), and UUG (4.3%) (Hecht et al. 2017).

A distinct feature of the genetic code is the dual function of AUG as a both the major start codon as well as a sense codon internal to genes. When positioned as start codon, AUG initiates

translation with a formyl-methionine, but when AUG is in-frame downstream of the start codon it behaves as a normal sense codon directing methionine incorporation into the protein sequence (Lobanov et al. 2010).

The AUG codon duality creates issues in gene prediction and genetic design. For example, in the bacteriophage ϕ X174 gene A* was one of the last genes discovered in this virus because it is an N-terminally truncated version of gene A that uses an internal AUG start codon (Colasanti and Denhardt 1985). This problem is also seen in recombinant protein design where in-frame AUG sense codons can behave as start codons if the DNA sequence upstream resembles a Shine-Dalgarno site (Whitaker et al. 2015).

1.3.2.2 Role of initiation factors

The initiation factors are a group of proteins that aid in 30S IC formation. Initiation Factor 1 (IF1) and Initiation Factor 3 (IF3) control the fidelity of the initiation process while Initiation Factor 2 (IF2) recruits the initiator tRNA (Figure 4)(Julián et al. 2011).

IF1 is the smallest initiation factor with a single β -domain (Sette et al. 1997) that binds to the ribosomal A-site (Moazed et al. 1995; Carter et al. 2001) and is encoded by the *infA* gene. IF1 is essential for cell viability (Cummings and Hershey 1994) because of its critical role preventing unwanted tRNA binding during the initiation process. IF1 also modulates IF2 and IF3 interaction with the ribosome by providing a key anchoring point (Gualerzi and Pon 1990; Hussain et al. 2016).

IF2 is a multi-domain GTPase encoded by the *infB* gene and is thought to recruit and anchor the initiator tRNA on the 30S subunit (Simonetti et al. 2008) with a highly conserved C-terminal domain that contains GTP-binding and initiator tRNA binding sites (Guenneugues et al. 2000; Caserta et al. 2006; Julián et al. 2011). IF2 plays a crucial role in stimulating the leaderless mRNA initiation mode of initiation (Grill et al. 2000).

IF3 ensures the fidelity of the codon-anticodon interaction between the start codon and initiator tRNA anticodon (Hartz et al. 1990; Ayyub et al. 2017). Additionally, IF3 affects the rate of tRNA association with, and disassociation from, the P-site. Lastly, IF3 is also thought to be involved in the process of enabling the ribosome to detect translation initiation regions on mRNA (Milon et al. 2008).

1.3.3 Unique features of initiator tRNA

1.3.3.1 Conserved anticodon stem enables P-site discrimination between initiator and elongator tRNA

It has long been an open question how the P-site of the ribosome is able to discriminate between the similar methionyl-initiator tRNA and methionyl-elongator tRNA (Figure 5). The initiator tRNA has three uniquely conserved G:C base pairs in the anticodon stem which is absent in the elongator (Figure 5) (Marck and Grosjean 2002). The significance of this conserved anticodon stem sequence in translation initiation was established when an elongator methionyl-tRNA initiated translation after the anticodon stem was mutated to harbour the three conserved G:C base pairs (Varshney et al. 1993). The conserved base pairs are essential for P-site binding with G:C base pairs 30-40 and 29-41 forming type I and type II A-minor interactions with 16S rRNA nucleotides G1338 and A1339 (Lancaster and Noller 2005) conserved in the prokaryotic domain. Recent structural studies of the 70S-complex of *Thermus thermophilus* showed the conserved interaction points in the ribosomal P-site (Selmer et al. 2006; Barraud et al. 2008). These studies also discovered that the G:C base pairs along with an unusual Cm32:A38 base pair causes a distinct anticodon stem-loop conformation of the initiator tRNA (Figure 6) projecting the anticodon loop outward (Barraud et al. 2008).

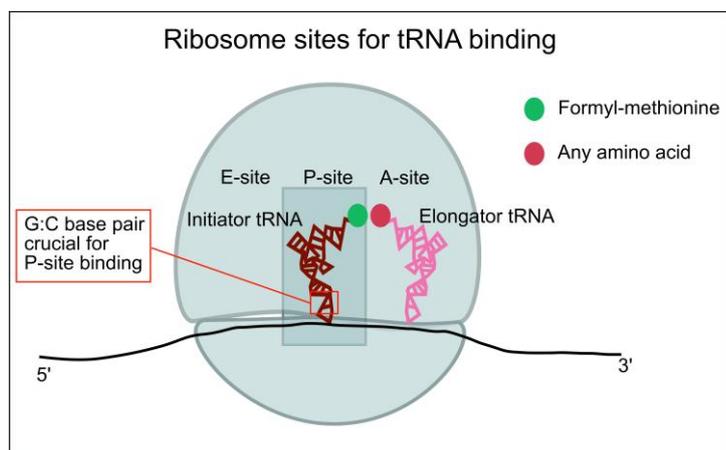


Figure 5: The initiator tRNA binds to the specifically to the P-site in the ribosome.

The G:C base pair in the initiator tRNA is specifically recognised by elements in the ribosome P-site and the elongator tRNA can only enter the ribosome from the A-site.

1.3.3.2 Conserved acceptor stem of initiator tRNA enables formylation and IF2 recognition

The methionyl-initiator tRNA and methionyl-elongator tRNA are both charged with a methionine by the methionyl-tRNA synthetase (MetRS) but the methionyl-tRNA N-formyltransferase enzyme only formylates the methionine attached to the *initiator* tRNA. Formylation has been suggested to be crucial for initiator tRNA functioning since variants with an acceptor stem mutation with reduced formylation have lowered initiation activity *in vivo* (Varshney and RajBhandary 1992; Guillon et al. 1993). Subsequently, IF2 was discovered to exhibit a higher binding affinity towards

formylated aminoacyl-tRNAs and play a major role in initiator tRNA recruitment for translation initiation (Mayer et al. 2003) suggesting that the formyl group may be essential for IF2 mediated initiator tRNA recruitment.

However, recent structural analysis of the 30S IC shows IF2 interacting with the conserved nucleotide sequence (CAACCA) in the tRNA acceptor stem (Figure 6) that programs formylation by methionyl-tRNA N-formyltransferase, rather than the formyl group itself. This result suggests that the conserved acceptor stem sequence (CAACCA) of initiator tRNA differentiates it elongator tRNAs rather than the formyl group on the amino acid attached (Guenneugues et al. 2000; Mayer and RajBhandary 2002; Simonetti et al. 2008).

1.3.3.3 Codon-anticodon pair specificity

Translation initiation most frequently starts at AUG because the cytoplasmic initiator tRNA in all organisms has a conserved anticodon (CUA) that recognises AUG start codons. Therefore, initiation fidelity may not depend on the start codon sequence (AUG) itself but rather depends on a stable codon-anticodon interaction. The codon-anticodon pair should be considered as a unit, when discriminated by IF3 at the initiation checkpoint, rather than individual elements (Hartz et al. 1990). Further evidence based on Cryo-EM structures of the translation initiation complexes showed that IF3 does not recognise the start codon or the anticodon as individual elements (Allen et al. 2005; Julián et al. 2011). Instead, the C-terminal domain of IF3 causes a conformational change in the 30S IC, putting a larger constraint on the codon-anticodon pair, which can only be sustained by energetically favourable cognate codon-anticodon interactions (Hussain et al. 2016). In this manner the IF3 destabilizes any tRNA with non-cognate and pseudo-cognate codon-anticodon pairs (Petrelli et al. 2001).

Methanococcus jannaschii (Wang et al. 2000), which has been used to incorporate 71 different non-standard amino acids into proteins (Liu and Schultz 2010).

An additional advantage of *in vivo* incorporation of non-standard amino acids is the development of a novel biocontainment strategy. With recent large-scale deployment of genetically modified organisms in agriculture, therapeutics, bioremediation, and bioenergy (Moe-Behrens et al. 2013), biocontainment strategies aim to prevent unwanted proliferation of genetically modified organisms or related compounds in the environment. One strategy, using UAG sense codons, introduces non-standard amino acids into key sites on essential proteins to establish synthetic auxotrophy, preventing these organisms from growing without intentional supplementation with the non-standard amino acid (Mandell et al. 2015; Schmidt and Pei 2015).

1.4.2 A new strain for amber codon engineering

With large-scale advances in recoding and non-standard amino acid incorporation, scientists have looked to create a platform for improving the purity and yield of proteins containing non-standard amino acids. The *E. coli* MG1655 derivative-strain C321.ΔA.exp was modified to remove all 321 instances of the UAG stop codon. Additionally, release factor 1 (RF1) was deleted since it is the only release factor that recognises the UAG codon. This results in two advantages if a site-specific UAG amber sense codon is introduced: firstly there is no RF1 to recognise and terminate translation, and secondly the orthogonal tRNA does not face any competition for UAG codon binding. Thus, this strain serves as an ideal host with an altered genetic code possessing a devoted UAG sense codon for robust *in vivo* site-specific incorporation of non-standard amino acids (Lajoie et al. 2013). Alternatively, this strain also serves as an excellent biocontainment host with experimental evidence showing that it obstructs horizontally transferred genetic elements such as viruses since the altered genetic code with RF1 deletion prevents efficient termination of the viral genes (Ma and Isaacs 2016).

1.4.3 Engineering start codons

The first instance of reprogramming the initiation codon came in the early 1990s when protein synthesis from a UAG start codon was successfully initiated in cell extracts (Varshney and RajBhandary 1990). This result was achieved by supplementing the extract with an initiator tRNA with modified anticodon sequence CUA that was complementary to the amber codon UAG. This so-called amber initiator tRNA ($\text{tRNA}_{\text{CUA}}^{\text{fmet}}$) was later shown to work *in vivo* as well, with efficiency approximately 50-60% that of canonical AUG initiation measured by western blot analysis and

chloramphenicol acetyl-transferase (CAT) assay (Varshney and RajBhandary 1990). However, unlike the engineered amber sense codon systems, the amber initiator has not found widespread use in applications, instead being used as a tool to determine features essential for translation initiation (Rajbhandary et al. 1967; Varshney and RajBhandary 1992; Varshney et al. 1993; Mangroo et al. 1995; Wu and Rajbhandary 1997).

1.4.4 An orthogonal translation initiation system using amber initiator $tRNA_{CUA}^{fmet}$

With recent scientific advances in recoding and codon reassignment technology it is a perfect time to develop a well-defined and robust orthogonal translation initiation system in *E. coli* using the amber initiator $tRNA_{CUA}^{fmet}$. Unfortunately, many questions regarding how the $tRNA_{CUA}^{fmet}$ operates within cells have not been addressed to this date. In this thesis I chose to develop new tRNA expression and reporter tools to study the function of $tRNA_{CUA}^{fmet}$ at the population level, as well as to use proteomics and phenotypic growth measurements to identify deleterious off-target effects that may limit the amber initiator's use in future applications.

1.5 Project aims and objectives

The overall objective of this thesis is to define and measure the amber initiator tRNA ($tRNA_{CUA}^{fmet2}$) activity in *E. coli* as a first step to developing an orthogonal translation initiation system capable of overexpressing proteins independently of the host translation system and providing genetic isolation and biocontainment through the use of an alternative genetic code.

Specifically, I aim to:

1. Measure $tRNA_{CUA}^{fmet2}$ amber (UAG) start codon translation initiation efficiency in *E. coli* strain C321.ΔA.exp through the creation of improved tRNA expression and reporter plasmids.
2. Identify host proteome changes from $tRNA_{CUA}^{fmet2}$ expression in *E. coli* strain C321.ΔA.exp to determine off-target effects.
3. Determine the portability of the amber initiator $tRNA_{CUA}^{fmet2}$ genetic system to four other commonly used *E. coli* strains.
4. Measure fitness defects caused by $tRNA_{CUA}^{fmet2}$ expression in five *E. coli* strains to determine the tRNA's suitability in an efficient and robust orthogonal translation initiation system.

Chapter 2: Methods

2.1 Bacterial strains, cultures and glycerol stock

2.1.1 Bacterial strains

I used five strains of *Escherichia coli* in this study: C321.ΔA.exp (K-12 strain) obtained from Addgene (strain #49018) (Lajoie et al. 2013), BL21 (B strain) obtained from ATCC (strain # BAA-1025), NCTC122 (C strain) obtained from Public Health England (strain #NCTC122), Crooks obtained from ATCC (strain #8739), and W obtained from ATCC (strain #9637).

2.2.1 Bacterial culture preparation

All *E. coli* C321.ΔA.exp strains were grown in lysogeny broth Lennox (LB^L) or on LB^L agar, supplemented with 100µg/ml zeocin (Thermo Fisher Scientific #R25001) and 100 µg/ml carbenicillin (Sigma-Aldrich #C9231) and/or 100µg/ml spectinomycin (Sigma-Aldrich #S4014-5G) as appropriate for plasmid content. All other *E. coli* strains were grown in lysogeny broth Miller (LB^M) or LB^M agar, supplemented with 100 µg/ml carbenicillin and/or 100µg/ml spectinomycin for selection.

2.2.2 Glycerol stock preparation and storage

To make glycerol stocks of all *E. coli* strains, 900µL of an overnight culture was mixed with 900µL of glycerol (60%), mixed, and stored at -80° C. Recovery of strains from glycerol stocks involved streaked out onto LB^L and LB^M agar plates with the appropriate antibiotics, overnight incubation at 37° C, followed by storage at 4° C for later use.

2.2 Plasmid construction

2.2.1 Synthesized oligonucleotides and dsDNA

All oligonucleotides used in this study were synthesised by Integrated DNA Technologies (IDT) (Supplementary Table 1). I designed and synthesized a 1971bp IDT gblock (supplementary information) containing the *metY* gene for initiator tRNA from MG1655 (GenBank: U00096.3) with an anticodon mutation (CAU>CUA). This DNA fragment was used to amplify the *metY*(CUA) gene fragments for plasmid construction.

2.2.2 Plasmids

I designed reporter plasmids containing superfolder GFP (*sfGFP*) with different start codons and they were commercially constructed (Genscript). I designed and constructed the amber initiation plasmids from 2 parts using in vitro homologous assembly (Gibson et al. 2009).

Plasmids created and purchased for this study are listed in Table 1 and represented in Figure 7.

Table 1: Strains and plasmids used in this study

Strain/ Plasmids	Description	Source
<i>E. coli</i> strains		
<i>E. coli</i> C321.ΔA.exp	Recoded <i>E. coli</i> genome with all instances of the UAG codon and ΔRF1 (UAG termination function removed)	Addgene
<i>E. coli</i> BL21	<i>E. coli</i> B strain with all instances of UAG codon and intact termination function.	ATCC
<i>E. coli</i> C122	<i>E. coli</i> C strain with all instances of UAG codon and intact termination function	Public Health England
<i>E. coli</i> W (ATCC 9637)	<i>E. coli</i> W strain with all instances of UAG codon and intact termination function	ATCC
<i>E. coli</i> Crooks (ATCC 8736)	<i>E. coli</i> Crooks strain with all instances of UAG codon and intact termination function	ATCC
Plasmids		
pQE-60 (AUG)	Reporter plasmid; pQE-60 based vector, Amp ^R , <i>sfGFP</i> with AUG start codon driven by <i>T5</i> promoter.	This study
pQE-60 (UAG)	Reporter plasmid; pQE-60 based vector, Amp ^R , <i>sfGFP</i> with UAG start codon driven by <i>T5</i> promoter.	This study
pQE-60 (GCC)	Reporter plasmid; pQE-60 based vector, Amp ^R , <i>sfGFP</i> with GCC start codon driven by <i>T5</i> promoter.	This study
pULTRA-<i>metYp1p2-metYCUA</i>	Amber initiator plasmid; pULTRA based vector, Spec ^R , tRNA _{CUA} ^{fmet2} driven by <i>metYp1p2</i> promoter	This study
pULTRA-tac-<i>metYCUA</i>	Amber initiator plasmid; pULTRA based vector, Spec ^R , tRNA _{CUA} ^{fmet2} driven by <i>tac</i> promoter	This study

[^] Native *E. coli metY* gene promoter

[†] Strong inducible promoter

2.2.3 Plasmid assembly

The pULTRA-CNF vector (Addgene plasmid #48215) (Schultz et al. 2006) was used as the backbone for the amber initiation plasmid construction (Table 1). Polymerase chain reaction (PCR) with

Phusion High-Fidelity PCR Master Mix (Life Technologies #F-530) was used to produce the vector backbone and oligonucleotide fragments with common 20 bp overlaps. All PCR products were subjected to *DpnI* (NEB #R0176S) digestion at 37° C for 60 minutes followed by heat inactivation of *DpnI* at 80° C for 20 minutes. PCR products were cleaned up using the GenElute™ PCR Clean-Up kit (Sigma-Aldrich #NA1020-1KT). Two-fragment one-step isothermal *in vitro* DNA assembly was used for all constructed plasmids (Gibson et al. 2009). To a total reaction volume of 20µL containing 10µL NEBuilder® HiFi DNA Assembly Master Mix (NEB #E2621L), 30 fmol of PCR amplified vector backbone with 30-60 fmol of one PCR amplified DNA fragment was added. The entire reaction mixture was incubated at 50° C for 30 minutes. All PCR and assembly reactions were performed on a Mastercycler® proS (Eppendorf International).

2.2.4 Transformation

In vitro assembled plasmids were transformed into chemically competent NEB® Turbo Competent *E. coli* (NEB #C2984H) where 2.5µL of the assembly mixture was mixed with 25µL thawed competent *E. coli* NEBTurbo and chilled on ice for 30 minutes. Cells were heat-shocked at 42° C for 30 seconds and further chilled on ice for 5 minutes. A total of 975µL of SOC media preheated at 37° C was added to the cells and shaken at 250 rpm for 60 minutes. Transformants were plated onto pre-warmed LB^L and LB^M agar plates with 100µg/ml spectinomycin.

2.2.5 Colony PCR

Each transformant colony was picked and solubilized/homogenized in 50µL of sterile. The screening primers (Table 1) were used to amplify segments of the both amber initiation plasmids via PCR with Kappa2G Robust Hotstart Ready Mix (Sigma-Aldrich #KK5701). The PCR products were run on a 1% agarose gel in Tris/Borate/EDTA (TBE) buffer (Sigma-Aldrich # T4415) to determine and differentiate successfully assembled amber initiation plasmids as the PCR products were of different sizes (Supplementary Table S1).

2.2.6 Plasmid extraction

Plasmids were isolated from the transformants using the QIAprep Spin Miniprep Kit (Qiagen #27106) according to the manufacturer's instructions and stored in 30 µL of 10 mM Tris-HCl buffer (pH 8.0) at 4° C for frequent use and - 20° C for prolonged storage.

2.2.7 Sequencing

Purified plasmids (100ng/μL) along with sequencing primers (Supplementary Table S1) were sent to Macrogen Inc. for standardized Sanger sequencing. Sequencing data were aligned against expected sequence using software tool Benchling (<https://benchling.com/>).

2.3 Sequential double transformation

To obtain double transformants of each strain with both the amber initiation plasmid and reporter plasmid, sequential double transformation was performed since the conventional method of simultaneous double transformation (Weston et al. 1979) did not result in required strains. The sequential double transformation method involved two steps of transformation preceded by preparing competent cells.

2.3.1 Transformation of amber initiation plasmids

Competent *E. coli* cells of each strain (C321.ΔA.exp, BL21, NCTC122, Crooks, and W) used were prepared using the Mix & Go *E. coli* Transformation Buffer Set (Zymo Research #T3002) following manufacturer instructions. Each strain was individually transformed by the constructed amber initiation plasmids and plated on LB^L and LB^M agar plates with 100μg/mL spectinomycin to isolate transformants.

2.3.2 Transformation of reporter plasmids

Competent *E. coli* cells containing the amber initiation plasmids were prepared using the Mix & Go *E. coli* Transformation Buffer Set (Zymo Research #T3002) following manufacturer instructions. Each strain was individually transformed with individual reporter plasmids and plated on LB^L and LB^M agar plates with 100μg/mL spectinomycin and 100μg/mL carbenicillin to isolate double transformants.

2.4 Fluorescence measurements

2.4.1 Bulk fluorescence measurements

Measurements of fluorescence intensity from the amber initiation plasmid system was adapted from a previous method (Hecht et al. 2017). Three individual colonies for each bacterial culture were used to inoculate 2mL of LB^L and LB^M containing appropriate antibiotics: zeocin only for *E. coli* C321.ΔA.exp strains and spectinomycin (amber initiation plasmid) and/or carbenicillin (reporter plasmids) in 15 mL tubes (Labtek #FB50). Tubes were sealed and grown overnight at 37° C shaking at 250 rpm. After overnight growth, each culture was diluted 1:100 into into 400 μL of

fresh media in a 96-well deep well plate and grown for 2 hours at 37° C (250 rpm). Following regrowth, either 4µL of freshly prepared 100mM isopropyl-beta-D-thiogalactoside (IPTG) (Sigma-Aldrich #I6758) was added to cultures to induce *metY*(CUA) expression, or 40µL 20% glucose was added to cultures to repress *metY*(CUA) expression. Following inducer/repressor addition, cultures were grown for an additional 5 hours at 37°C (250 rpm).

A total of 250 µL of each cell culture was centrifuged at 2,240g for 12 minutes (swinging bucket rotor) and the supernatant was aspirated. The pelleted cells were resuspended in 200 µL of 1X phosphate buffered saline (PBS) (Sigma-Aldrich #P4417) and transferred to Greiner CELLSTAR® 96 well plate (Sigma-Aldrich #M0312). Absorbance was measured at 600 nm (OD₆₀₀) to estimate culture density, followed by fluorescence (excitation= 485 nm, emission = 520 nm, bandwidth = 9 nm) measured at a single gain setting (high gain sensitivity=111), on a PHERAstar FSX (BMG Labtech) plate reader.

2.4.2 Dynamic fluorescence measurements

After overnight growth as mentioned previously, each culture was diluted 1:100 into 30 mL of fresh media in a 250 ml Erlenmeyer flask and induced as previously described above. At 30-minute intervals post-induction, 250 µL of each culture was transferred to a Greiner CELLSTAR® 96 well plate to measure absorbance and fluorescence.

2.4.3 Data analysis

Raw fluorescence (arbitrary units) data and optical density (OD₆₀₀ arbitrary units) measurements for each culture were imported to Microsoft Office 2016 Excel. Wells filled with 1X PBS or culture media were used to subtract background optical density and fluorescence. Normalized fluorescence was calculated by dividing the fluorescence by the optical density. Normalized fluorescence for each combination of amber initiator and reporter plasmid in each strain was calculated by averaging the normalized fluorescence measurements from three biological replicates. Fold change due to dynamic induction by IPTG was calculated by dividing the normalized fluorescence at each time point by the normalized fluorescence at the start of induction (time point 0).

2.4.4 Flow cytometry

Bacterial cultures were transferred to a 96-well plate and measured on Cytoflex S (Beckman Coulter) using FITC fluorescence channel (488 nm excitation laser with a 525/40 nm emission band-pass filter). Measured events were triggered on a side scattering threshold and 10,000

events were measured from *E. coli* strain C321.ΔA.exp cultures containing the inducible amber initiation plasmid system (pULTRA-tac-*metY*CUA). Flow cytometry data was processed using FlowJo v10 (FlowJo, LLC).

2.5 Proteomic analysis

2.5.1 Cell growth and harvesting

Three individual colonies of each *E. coli*C321.ΔA.exp strain was used to inoculate 2mL of LB^L broth containing appropriate antibiotics: zeocin, spectinomycin (amber initiation plasmid) and carbenicillin (reporter plasmids) in 15 mL tubes. Tubes were sealed and grown overnight at 37° C (250 rpm). After overnight growth, 300 μL of each culture was inoculated in 30mL of fresh media with appropriate antibiotics in 250 mL flasks and grown at 37° C (250 rpm). After 1 hour of growth, cultures were induced for *metY*(CUA) expression by addition of 300μL of 100mM IPTG and grown for further 5 hours at 37° C (250 rpm). Cells were harvested at late-log growth phase by pelleting the culture at 3,500g at 4° C for 10 minutes via centrifugation. The supernatant was discarded and the pellets were stored at -20° C overnight to aid in cell lysis.

Each bacterial cell culture pellet was lysed using the CellLytic™ B Cell Lysis kit (Sigma-Aldrich #B7310) according to manufacturer specification. A total of 1 mL of CellLytic B 2X solution was added to resuspend each pellet followed by addition of lysozyme (3.3μL or 0.2mg/mL), protease inhibitor (50μL), and benzonase (0.5 μL or 50 units/mL) to increase efficiency of lysis and decrease viscosity of solution. The extraction mixture was incubated at room temperature for 15 minutes followed by centrifugation at 16,000g for 10 minutes and the supernatant containing the soluble protein fraction from each culture was collected.

2.5.2 Sample preparation for LC-MS

LC-MS sample preparation was adapted from previous methods (Hecht et al. 2017). A total of 250 μL of each sample was reduced and alkylated by incubation with 10mM of dithiothreitol (DTT) at 60° C for 30 minutes, followed by incubation with 30mM of iodoacetamide (IAA) at room temperature, in the dark for 20 minutes. Excess IAA was quenched by incubation with 30mM DTT at room temperature for 15 minutes. Proteins were precipitated out by addition of 1mL ice-cold acetone and incubated at - 80° C for 2 hours followed by centrifugation at 15,000g at 4° C for 10 minutes. The supernatant was discarded and the protein pellet was dried under the chemical hood. The pellet was resuspended in tris-HCl buffer containing 8M urea. Protein concentration of each sample was determined using the bicinchoninic acid assay (BCA) protein assay (Pierce™, cat#

23225). Bovine serum albumin protein standards (0-2mg/mL) were used for the generation of standard curves to calculate sample protein concentrations. Protein standards, blanks, and samples were assayed in duplicates. Protein samples were enzymatically cleaved by addition of 60ng of trypsin (Sigma-Aldrich #T2600000) to each 30µg of sample (1:50 ratio of enzyme to protein) and incubated overnight at 37° C. After overnight incubation, formic acid (100%) was added to the samples at a final concentration of 1% to acidify the peptides and stop digestion. The acidified digests were C₁₈ stage tip purified using micro-spin columns and dried in a speed vac to be finally stored as tryptic peptide samples at – 20° C.

2.5.3 LC-MS

The EASY-nLC was coupled on-line to a QExactive with a nanoelectrospray ion source (Proxeon). Tryptic peptide samples were reconstituted to 0.1 mg/mL in solvent A (2% acetonitrile, 0.1% formic acid) and 3µg (10 µL) of each sample was injected onto a C₁₈ reversed phase analytical column (10cm long, 75 µm inner diameter) at a constant flow rate of 300 nL/minute using an auto-sampler. The peptides would bind to the reversed phase column and elution was achieved with increasing concentration of solution B (95% acetonitrile, 0.1% formic acid) over a 120-minute runtime along the following gradient: from 1% solvent B to 95% solvent B within 120 min, from 1-50%% solution B within 110 minutes, 50-85% solution B for the next 2 minutes and 85-95% solution B for remaining 8 minutes. Electrospray ionization at a spray voltage of 2,400V was used to produce positively charged ions for MS/MS analysis and detection of peptides.

2.5.4 LC-MS data analysis

MS/MS data was extracted using MaxQuant v1.6.0.16 and searched against the complete *E. coli* K-12 MG1655 proteome from Uniprot (version from July 2017). The MS mass tolerance was set to 20 ppm, MS/MS fragment ion deviation was set to 0.5 Da and FDR < 0.1 (Cox et al. 2008). For label-free quantitative shotgun proteomics peptide hits were filtered to proteins with peptides in all three replicates. Normalized spectral abundance factor (NSAF) was calculated by the Scrappy program to quantify differential expression of filtered proteins (Neilson et al. 2013). For “off-target peptide” detection MaxQuant v1.6.0.16 was used to search the MS/MS data against a combined off-target and K-12 protein database.

2.6 Fitness analysis

2.6.1 Growth measurements

Strains were grown overnight as previously described and 1.5µL of each overnight culture (1:100 dilution) was transferred to a flat-bottom 96-well plate containing 150µL LB^L or LB^M with appropriate antibiotics at 37 °C (300 rpm). Cultures were induced after 1 hour by addition of 1.5 µL of 100mM IPTG and repressed by addition of 1.5 µL of 20% glucose to the growth medium. Culture light scattering (OD₆₀₀) over time was measured on a SPECTROstar NANO (BMG Labtech) plate reader at 5-minute intervals.

2.6.2 Data analysis

Analysis method was adapted from previous work (Lajoie et al. 2013) and used to compare ratios of doubling time and maximal cell density for individual strains. Doubling time was calculated by $t_{\text{double}} = \frac{\ln(2) \times t}{\ln(N/N_0)}$, where N= OD₆₀₀ at 160 minutes, N₀= OD₆₀₀ at 80 minutes, and t= difference between time-points of OD₆₀₀ measurements. Ratios of doubling time and maximal cell densities were determined for strains with the amber initiation plasmid versus strains without the amber initiation plasmid.

Chapter 3: Results

3.1 Defining an amber initiation tRNA and modular fluorescent reporter system in *E. coli*

I aim to build from the previous discovery of an amber initiation tRNA (Varshney and RajBhandary 1990) by creating a modular plasmid-based amber initiator tRNA expression and reporter system that can measure functional parameters of the system that have been previously inaccessible. The reason for creating a new expression system for the amber initiator and fluorescent reporter was two-fold: (1) to enable dynamic measurements of amber tRNA $\text{tRNA}_{\text{CUA}}^{\text{fmet2}}$ expression, and (2) to enable measurements of amber initiator $\text{tRNA}_{\text{CUA}}^{\text{fmet2}}$ population-level and real-time *in vivo* expression.

3.1.1 Designing plasmids to express an amber translation initiation system

To initiate translation at a UAG codon instead of AUG, I designed a plasmid that could express a tRNA that binds to an amber start codon and was compatible with a second reporter plasmid. The two initiator tRNA genes from *E. coli* K-12 MG1655 (*metZ* and *metY*) only differ in sequence by one nucleotide at position 46 and encode the tRNA molecules $\text{tRNA}^{\text{fmet1}}$ and $\text{tRNA}^{\text{fmet2}}$, respectively. I have chosen to focus on $\text{tRNA}^{\text{fmet2}}$ (*metY*) because previous work showed that this initiator is not essential in *E. coli* and was capable of being recoded to initiate translation at UAG start codons (Varshney and RajBhandary 1990). My amber initiator plasmid design features a recoded *metY* gene (*metY(CUA)*) that expresses $\text{tRNA}_{\text{CUA}}^{\text{fmet2}}$ with an altered anticodon loop that is the reverse complement of the desired start codon, UAG (Figure 7).

To identify the ideal system for production of $\text{tRNA}_{\text{CUA}}^{\text{fmet2}}$ in *E. coli*, I designed two variants of the amber initiation plasmids. One plasmid drives $\text{tRNA}_{\text{CUA}}^{\text{fmet2}}$ expression from the native K-12 *metY* locus constitutive *metYp1p2* promoter (Figure 7A). A second plasmid expresses the $\text{tRNA}_{\text{CUA}}^{\text{fmet2}}$ from an IPTG-inducible *tac* promoter (Figure 7B). Both amber initiation plasmids harbour the cloDF13 (medium copy) origin of replication (Stuitje et al. 1979) and spectinomycin resistance gene (Uhlin and Nordstrom 1975).

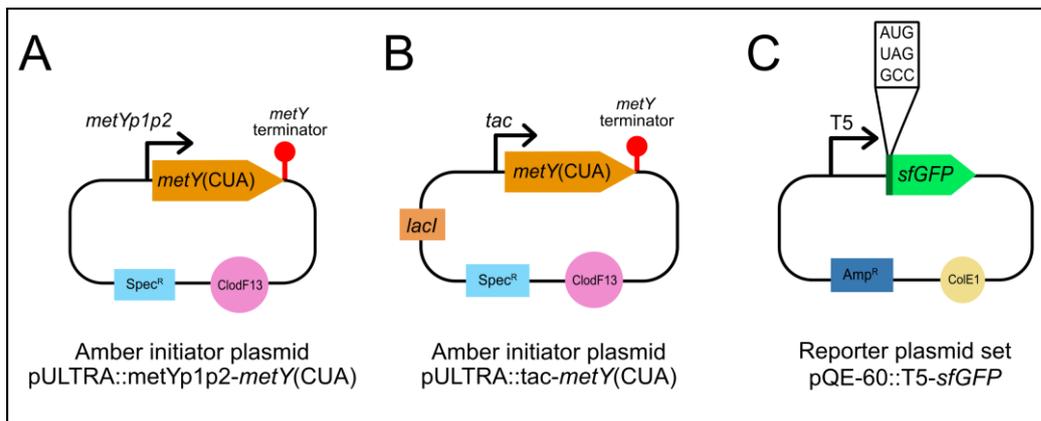


Figure 7: Amber initiator tRNA and reporter plasmid designs.

(A) Amber initiator plasmid pULTRA::metYp1p2-*metY(CUA)*. The plasmid encodes the medium-copy ClodF13 origin, spectinomycin resistance (*Spec^R*) marker, and native *E. coli* K-12 promoters *metYp1p2*. (B) Amber initiator plasmid pULTRA::*tac-metY(CUA)*. The plasmid encodes the medium-copy ClodF13 origin, spectinomycin resistance (*Spec^R*) marker, and inducible *tac* promoter. (C) pQE-60 reporter plasmids. The reporter plasmid set has a medium-copy ColE1 origin, *T5* promoter and superfolder green fluorescent protein gene (*sfGFP*) with three different start codons (AUG, UAG, and GCC). All *metY(CUA)* and *sfGFP* gene variants have a Rho-independent terminator after individual gene sequences.

3.1.2 Designing plasmids to express fluorescent reporters compatible with amber initiation plasmids

I designed a set of three reporter plasmids to measure translation initiation events caused by expression of $\text{tRNA}_{\text{CUA}}^{\text{fmet2}}$ from my amber initiation plasmids (Section 3.1.1). The reporter plasmids featured the super-folder green fluorescent protein gene (*sfGFP*) with three start codon variants. I created a reporter with UAG start codon (*sfGFP(UAG)*) to measure translation initiation efficiency from amber (UAG) start codons. As a negative control, I created a *sfGFP(GCC)* variant to measure translation initiation events from a codon that is normally initiated from the wild-type $\text{tRNA}^{\text{fmet2}}$ at a rate below background (Hecht et al. 2017) and was expected to perform similarly for the modified $\text{tRNA}_{\text{CUA}}^{\text{fmet2}}$. Lastly, I created a reporter plasmid with the canonical start codon, *sfGFP(AUG)* as a positive control for reporter expression. The *T5* constitutive promoter controls the expression of all reporter gene variants by continuously transcribing mRNA molecules containing the *sfGFP* variants.

I designed all the reporter plasmid variants to have the ColE1 origin of replication (Chansb et al. 1985) and ampicillin antibiotic resistance marker so they were compatible (Hashimoto-Gotoht and Inselburg 1979) with the amber initiation plasmid features.

3.1.3 The amber initiation system translates a sfGFP reporter with an amber start codon

To measure the functional parameters of the amber initiation system, I built the two plasmid variants and three reporter plasmids (Figure 7) using synthetic DNA and *in vitro* assembly, followed by transformation into *E. coli*. I chose a host strain of *E. coli* for this work, called C321.ΔA.exp, that was engineered to lack all instances of UAG stop codons in the genome (Lajoie et al. 2013). I chose this strain of *E. coli* for these experiments because I reasoned that if the modified amber initiator tRNA_{CUA}^{fmet2} was to have any detrimental effects on the host it would be reduced in C321.ΔA.exp.

I measured translation initiation of the sfGFP reporters under three conditions: (1) when tRNA_{CUA}^{fmet2} was absent from the system, (2) when tRNA_{CUA}^{fmet2} was expressed by a native constitutive promoter, and (3) when tRNA_{CUA}^{fmet2} was expressed by an inducible promoter. To control for the effects of the chemicals used for induction and repression of the *tac* promoter, I used the same repressed (2% glucose) and induced (1mM IPTG) culture conditions for all plasmid combinations, regardless of whether they contained a plasmid with a *tac* promoter or not.

To measure the effect of tRNA_{CUA}^{fmet2} expression, I grew cells up containing both the amber initiation plasmid with the native *metYp1p2* promoter as well as the reporter plasmids. Fluorescence from the sfGFP(UAG) reporter increased 7-fold compared to sfGFP(GCC) when tRNA_{CUA}^{fmet2} was expressed from the native promoter (Figure 8, middle). Surprisingly, the fluorescence from the sfGFP(AUG) reporter also increased slightly (1-fold) under the same conditions. These data seem to indicate that the tRNA_{CUA}^{fmet2} enables initiation from UAG start codons as expected, but that it can also cause initiation from AUG codons.

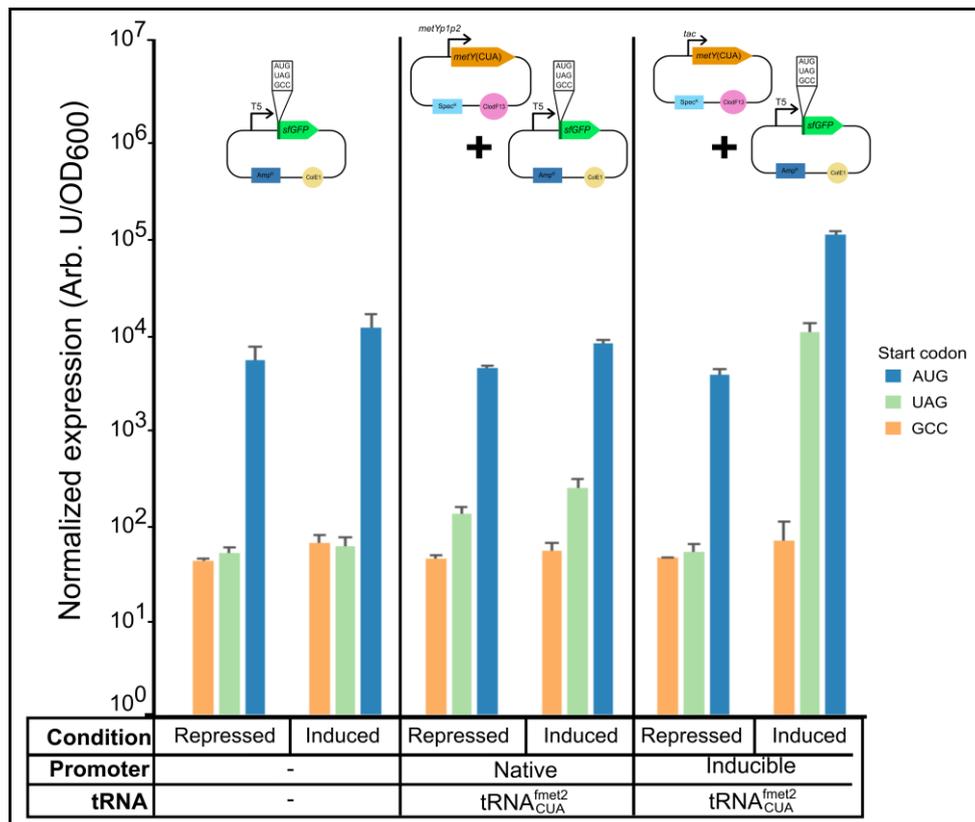


Figure 8: Expression of tRNA^{fmet2}_{CUA} from amber initiator plasmids causes increase in reporter fluorescence from both amber UAG and canonical AUG start codons. Normalized expression levels from sfGFP reporters under different tRNA^{fmet2}_{CUA} conditions. Condition: measurement performed on cells that had been treated with 2% glucose (repressed) or 1 mM IPTG (induced). tRNA: presence or absence of tRNA^{fmet2}_{CUA} expressing amber initiator plasmid. Promoter: identity of promoter controlling expression of tRNA^{fmet2}_{CUA}. Each column is the average of three biological replicates with the error bars showing one standard deviation.

To measure the effect of tRNA^{fmet2}_{CUA} expression from the inducible amber initiator plasmid (pULTRA::*tac-metY(CUA)*), I grew cells up under both repressing (2% glucose) and inducing (1mM IPTG) conditions. Under repressing conditions, the sfGFP(UAG) reporter fluorescence was indistinguishable from the negative control sfGFP(GCC) fluorescence (Figure 8, right). Under inducing conditions, the sfGFP(UAG) fluorescence increased 3-fold over sfGFP(AUG) fluorescence in the repressing conditions, indicating that tRNA^{fmet2}_{CUA} was capable of providing as much initiator tRNA to the reporter plasmid transcripts as normal tRNA^{fmet} is providing to sfGFP(AUG) transcripts in the absence of initiator tRNA. The tRNA^{fmet2}_{CUA} expression again caused an increase in the sfGFP(AUG) fluorescence, reinforcing the previous result and showing that the amber initiator tRNA can initiate from AUG start codons in addition to UAG. I also found that normalized

expression level from the sfGFP(GCC) reporter with a non-canonical start codon was equivalent to fluorescence levels reported previously for this initiation codon (Hecht et al. 2017). Fluorescence levels from cells harbouring the sfGFP(GCC) reporter did not increase when $tRNA_{CUA}^{fmet2}$ was expressed from either the native promoter or the inducible promoter (Figure 8). This result indicates that $tRNA_{CUA}^{fmet2}$ initiates translation at GCC codons at the same low frequency as wild-type $tRNA^{fmet2}$.

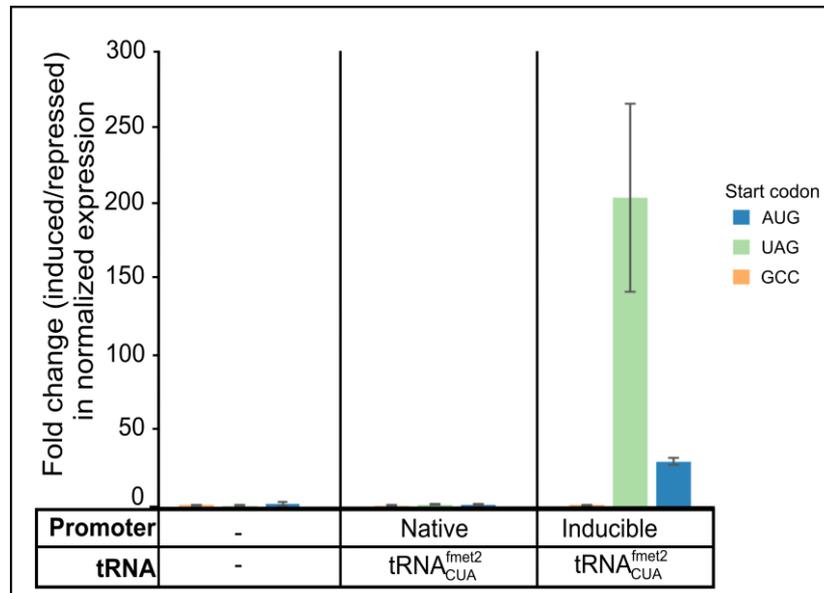


Figure 9: Expression of $tRNA_{CUA}^{fmet2}$ from inducible pULTRA::tac-metY(CUA) amber initiator plasmid causes significantly greater increase in reporter fluorescence from amber UAG start codon versus canonical AUG start codons. Fold change in normalized expression of sfGFP reporter due to IPTG induction. Promoter: identity of promoter controlling expression. tRNA: presence or absence of $tRNA_{CUA}^{fmet2}$ expressing amber initiator plasmid. Each column represents average of three biological replicates with the error bars showing one standard deviation.

I compared fold change between the induced and repressed condition (Figure 9). I found that due to expression of $tRNA_{CUA}^{fmet2}$, the increase in translation from an amber start codon exceeds that of a canonical start codon in *E. coli strain C321.ΔA.exp* by 70-fold. This gave me confidence in using the inducible system for further analysis. With the knowledge that the $tRNA_{CUA}^{fmet2}$ expressed from my plasmids initiates translation at amber start codons I next wanted to determine if the observed fluorescence was from a small population of highly fluorescence cells or whether the cell population was more uniform.

3.1.4 Flow cytometry reveals $tRNA_{CUA}^{fmet2}$ initiates translation across a uniform population

Using flow cytometry, I measured the distribution of fluorescence (FITC-H) within the population of *E. coli* strain C321.ΔA.exp cells harbouring the inducible amber initiation plasmid pULTRA::*tac-metY*(CUA) and sfGFP reporters. I found that all measured cell populations had unimodally distributed fluorescence (Figure 10A). Furthermore, analysis of average fluorescence from the populations of repressed and induced cultures (Figure 10B and 10C) matched closely to that from previous experiments on bulk culture fluorescence (Figure 8) with similar trends in fluorescence fold change due to IPTG induction.

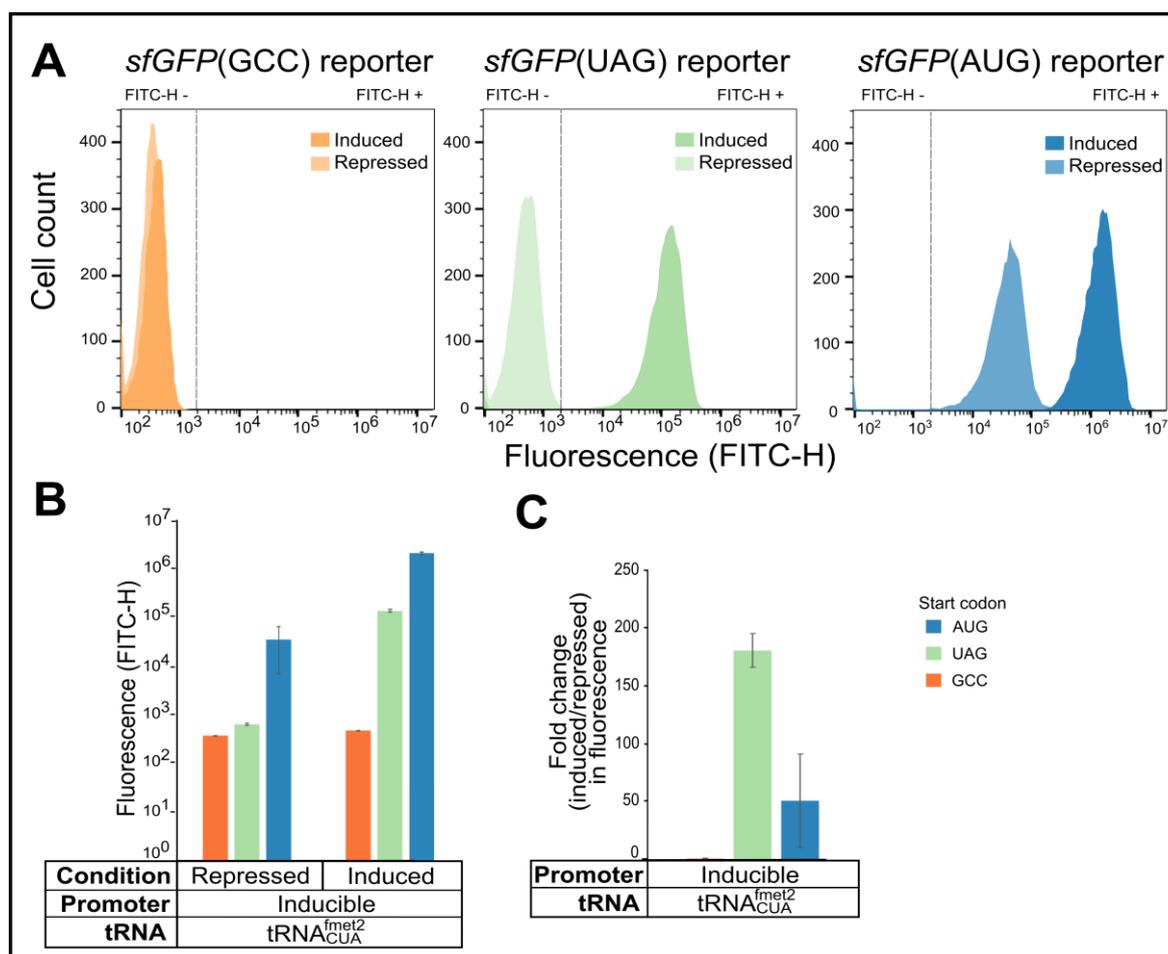


Figure 10: Flow cytometry analysis shows unimodal fluorescent *E. coli* C321.ΔA.exp population expressing $tRNA_{CUA}^{fmet2}$. (A) Representative fluorescence (FITC-H) histogram of gated populations. Induced cell population (darker shades) and repressed cell population (lighter shades) shown for each reporter. Vertical dashed lines demarcate fluorescent and non-fluorescent populations based on the 99th percentile of the control population (fluorescence of *E. coli* C321.ΔA.exp population expressing $tRNA_{CUA}^{fmet2}$ with sfGFP(GCC) reporter). (B) Average fluorescence (FITC-H) for three biological replicates. Error bars represent one standard deviation. (C) Fold change in fluorescence intensities between repressed and induced cultures.

3.1.5 Dynamic induction of $tRNA_{CUA}^{fmet2}$ potentially reveals effect of tRNA maturation on start codon interaction fidelity

To determine the dynamic behaviour of the $tRNA_{CUA}^{fmet2}$ amber initiator, I performed an induction time-course using the *tac*-inducible pULTRA::*tac-metY*(CUA) plasmid. The results of this experiment showed that the $tRNA_{CUA}^{fmet2}$ seemed to have a different effect on the translation initiation from the sfGFP(UAG) compared to the sfGFP(AUG) reporter over time.

Taking the point of induction as a baseline, I observed a rapid 3-fold expression increase from the canonical start codon sfGFP(AUG) reporter by 1.5 hours post-induction (Figure 11). This 3-fold increase stayed relatively constant throughout the next 5.5 hours of the time-course. In contrast, expression from the amber start codon sfGFP(UAG) reporter stayed approximately unchanged until 3-hours post-induction when it began steadily increasing until the end of the 7-hour mark when it achieved 7-fold increased expression (Figure 11).

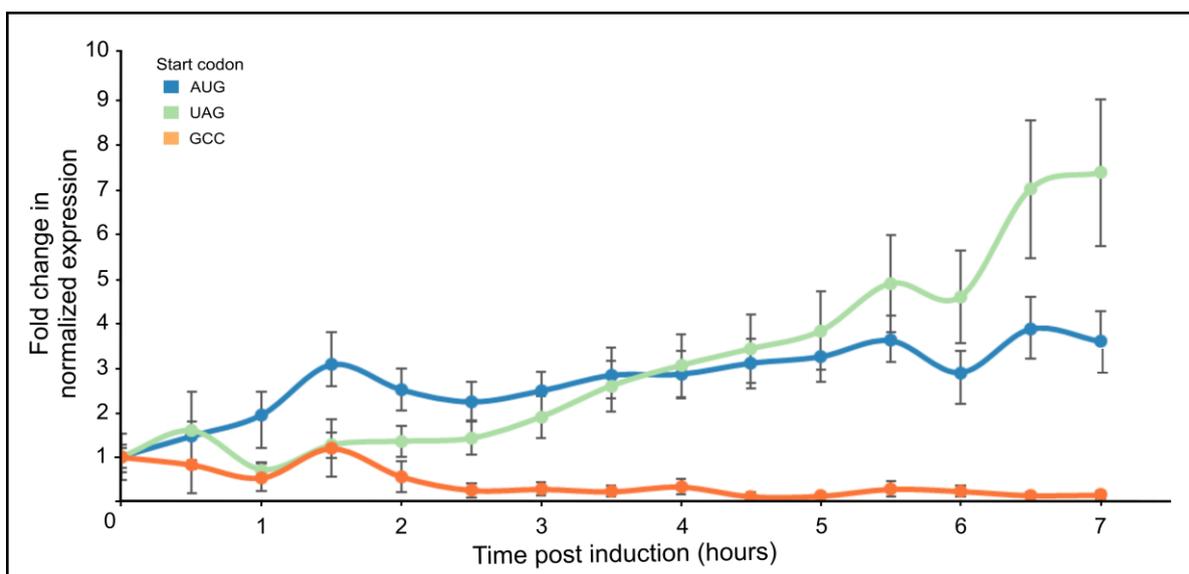


Figure 11: Fold change in normalized reporter expression shows differential effects on UAG versus AUG start codons over time. Each line represents trend line over the mean of fold-changes to normalized expression for three biological replicates at each time-point. Error bars represent one standard deviation.

These data seem to suggest there is some process that takes approximately ~2-3-hours to occur that is necessary for the $tRNA_{CUA}^{fmet2}$ to initiate translation from the sfGFP(UAG) reporter. In contrast, the $tRNA_{CUA}^{fmet2}$ is able to initiate translation from the canonical sfGFP(AUG) start codon reporter very soon after induction, but then this process ceases around the 2-3-hour mark.

3.2 Measuring whole-cell proteomic effects from $tRNA_{CUA}^{fmet2}$ expression

To define the full range of cellular effects from expression of the amber initiator $tRNA_{CUA}^{fmet2}$, I used mass spectrometry based whole cell shotgun proteomics to measure protein production from chromosomally encoded and amber initiator plasmid-borne sequences in *E. coli* cells.

3.2.1 Production of $tRNA_{CUA}^{fmet2}$ causes few changes to chromosomal protein expression in *E. coli* C321. Δ A.exp

Because I consistently observed $tRNA_{CUA}^{fmet2}$ expression caused an increase in fluorescence from the canonical start codon sfGFP(AUG) reporter (Section 3.1), I reasoned that this effect may also extend to the chromosomally-encoded *E. coli* genes with AUG start codons. To measure global cellular protein production, I conducted shotgun proteomic measurements on *E. coli* C321. Δ A.exp cells containing the sfGFP(UAG) reporter plasmid and the $tRNA_{CUA}^{fmet2}$ inducible plasmid. Comparing uninduced to induced cells I detected a total of 759 proteins present in all three biological replicates with at least 2 unique peptides. Of the 759 proteins, I observed only 23 proteins (3.0%) differentially expressed (p-value ≤ 0.05) in cells expressing the $tRNA_{CUA}^{fmet2}$ compared to the cells without the $tRNA_{CUA}^{fmet2}$ (Figure 12). Of the 23 differentially regulated proteins, only 12 proteins (1.6%) were up-regulated, and of those only 10 proteins (1.3%) were encoded in the *E. coli* genome. The most highly up-regulated protein was the LacI (lactose operon repressor) expressed from the amber initiator plasmid. The second most highly up-regulated protein was sfGFP with UAG as the start codon from the reporter plasmid. Together, these data show that $tRNA_{CUA}^{fmet2}$ does not cause wide-spread global up-regulation of *E. coli* proteins produced from genes with the canonical start codons (AUG, GUG, and UUG).

I analysed the remaining up- and down-regulated proteins (Table 2) and interestingly observed up-regulation of peptide chain release factor 3 (RF3) which is known to stimulate translation termination by interacting with class I release factors RF1 and RF2 (O'Connor 2015). I investigated the identified protein using the Gene Ontology annotation database PANTHER – Gene List Analysis (GO annotation) to determine evidence of functional or pathway clustering, but could find only weak pathway clustering with no apparent correlation in functions.

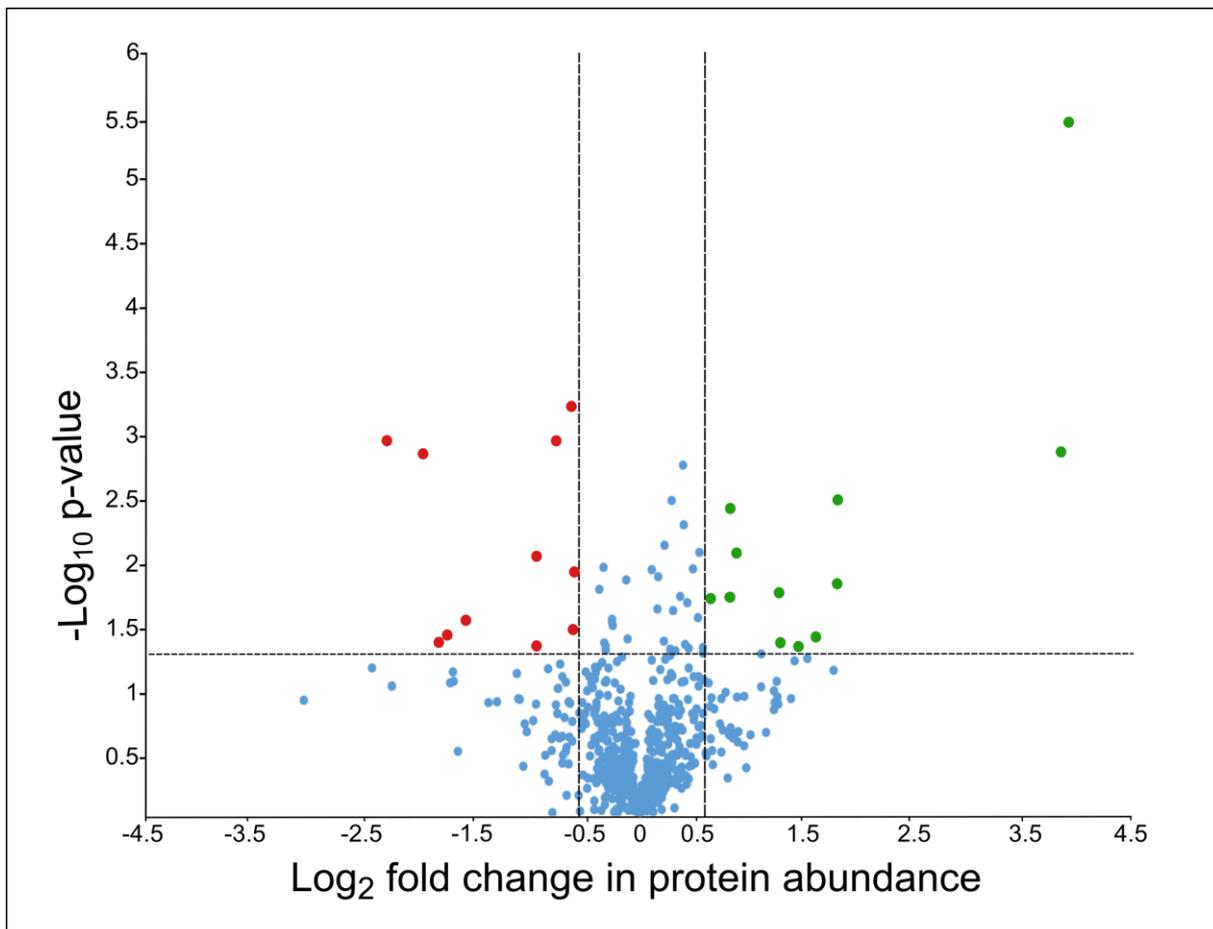


Figure 12: The production of tRNA_{CUA}^{fmet2} in *E. coli* C321.ΔA.exp(pQE60::sfGFP(UAG)) cells results in 22 differentially expressed proteins. Volcano plot of 759 quantified proteins from *E. coli* C321.ΔA.exp cells harbouring the sfGFP(UAG) reporter plasmid and expressing tRNA_{CUA}^{fmet2} versus cells lacking tRNA_{CUA}^{fmet2}. Green dots represent up-regulated proteins and red dots represent down-regulated proteins with both statistically significant differential expression (\log_2 fold change ≤ -0.58 or ≥ 0.58) and with p-value ≤ 0.05 .

Table 2: Differentially expressed proteins in *E. coli* C321.ΔA.exp(pQE60::sfGFP(UAG)) cells expressing tRNA_{CUA}^{fmet2}. Proteins with both a fold change <0.66 (down-regulated) or >1.5 (up-regulated) and p-value <0.05.

Gene	Protein name	Fold change	p-value	Unique Peptides
Up-regulated proteins				
<i>lacI</i>	Lactose operon repressor*	15.32	0.000003	8
<i>sfGFP</i>	Superfolder green fluorescent protein*	14.59	0.001385	11
<i>prfC</i>	Peptide chain release factor RF3	3.50	0.003333	25
<i>htpG</i>	Chaperone protein	1.76	0.003835	38
<i>yfeX</i>	Probable deferrochelataase/peroxidase	1.84	0.008759	10
<i>tolQ</i>	Protein TolQ	3.48	0.015211	3
<i>bioD1</i>	ATP-dependent dethiobiotin synthetase	2.41	0.018149	8
<i>sdaC</i>	Serine transporter	1.71	0.019466	4
<i>glpK</i>	Glycerol kinase	1.56	0.020220	37
<i>lhgO</i>	L-2-hydroxyglutarate oxidase	3.06	0.039808	3
<i>pal</i>	Peptidoglycan-associated lipoprotein	2.44	0.045382	5
<i>cysQ</i>	3'(2'),5'-bisphosphate nucleotidase	2.73	0.047743	5
Down-regulated proteins				
<i>degQ</i>	Periplasmic pH-dependent serine endoprotease	0.52	0.048357	11
<i>cheA</i>	Chemotaxis protein	0.28	0.044040	8
<i>rdgC</i>	Recombination-associated protein	0.29	0.038790	5
<i>moeA</i>	Molybdopterin molybdenumtransferase	0.65	0.035180	10
<i>cheW</i>	Chemotaxis protein	0.33	0.029896	3
<i>fliC</i>	Flagellin	0.66	0.012223	26
<i>lolD</i>	Lipoprotein-releasing system ATP-binding protein	0.52	0.009247	4
<i>tar</i>	Methyl-accepting chemotaxis protein II	0.25	0.001423	14
<i>hemN</i>	Oxygen-independent coproporphyrinogen III oxidase	0.59	0.001113	3
<i>nema</i>	N-ethylmaleimide reductase	0.20	0.001109	4
<i>gltX</i>	Glutamate--tRNA ligase	0.65	0.000592	23

* Genes encoded in the amber initiation and reporter plasmids

3.2.2 Production of tRNA_{CUA}^{fmet2} results in no detectable off-target peptides produced from UAG open reading frames in *E. coli* C321.ΔA.exp

I next wanted to determine whether production of tRNA_{CUA}^{fmet2} was resulting in translation initiation from chromosomally-encoded amber UAG codons. Although the *E. coli* C321.ΔA.exp genome has had all 321 instances of genes with UAG stop codons changed to UAA (Lajoie et al. 2013), there still remains over 10,000 UAG codons within the genome. Expression of the tRNA_{CUA}^{fmet2} may cause

a translation initiation event if a ribosome-binding site (Shine-Dalgarno sequence) were located on an mRNA at a suitable distance upstream of one of these UAG codons.

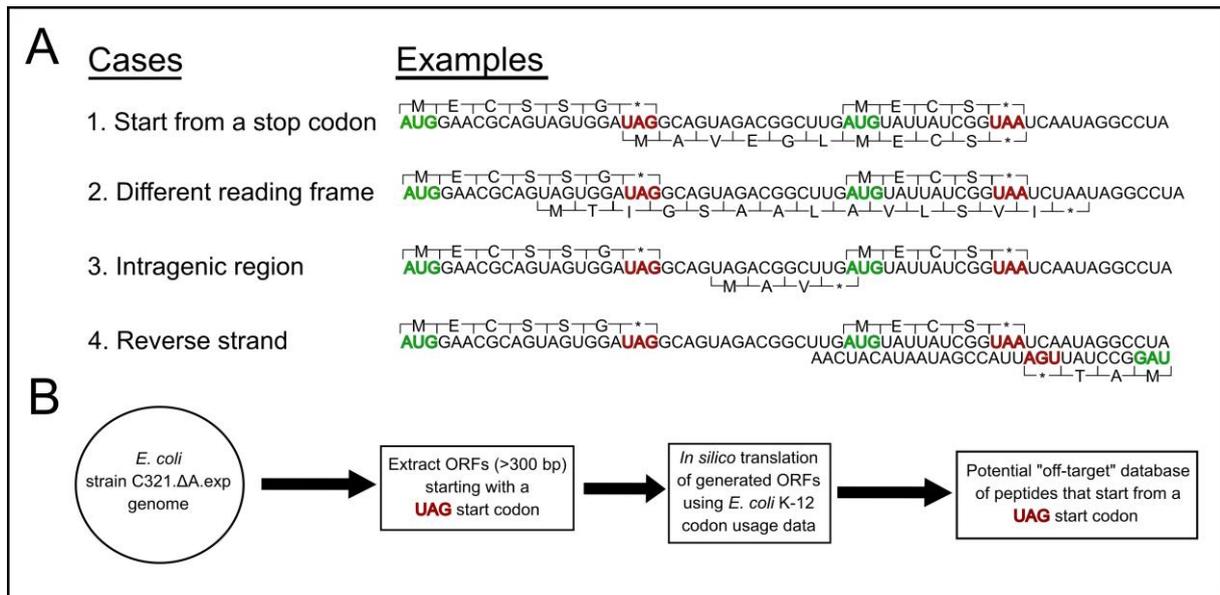


Figure 13: Method used to create UAG open reading frame off-target peptide database. (A) The various possibilities that may result in ORFs with UAG start codon. The top peptides represent annotated sequences translated from AUG start codon and bottom peptides represent hypothetical off-target peptide sequences translated from a UAG start codon. **(B)** Algorithmic workflow to create an ‘off-target’ peptide database from *E. coli* C321.ΔA.exp genome.

I modelled the four different scenarios that would result in an open reading frame beginning with a UAG start codon being present in the C321.ΔA.exp genome (Figure 13A). Case #1 is not possible in C321.ΔA.exp because all 321 UAG stop codons were previously removed (Lajoie et al. 2013). I used a script to extract all possible UAG open reading frames ≥ 300 bp from the C321.ΔA.exp genome sequence (Figure 13B). The resulting ‘off-target’ protein database was merged with the *E. coli* K-12 proteome database to form a resultant ‘combined’ K-12 protein search space to match observed peptides (I call this the ‘combined K-12 database’).

I performed mass spectrometry on *E. coli* C321.ΔA.exp cells both with and without $\text{tRNA}_{\text{CUA}}^{\text{fmet2}}$ expression and matched peptides to the combined K-12 database. I detected 257 peptides that mapped to 196 proteins in the combined K-12 database. The false discovery rate of my measurements was within acceptable limits (less than 5% of the detected peptides were from the reverse (false) database). To determine if translation initiation was occurring at UAG start codons, I filtered out all peptides matching ORFs with AUG start codons, but not ORFs with UAG start codons. I further filtered out peptides identified in only one replicate, which left three proteins

produced in *E. coli* C321.ΔA.exp from ORFs initiating from UAG start codons. The remaining 3 peptides (Table 3) had a peptide level score > 50 and PEP value below 0.01 providing strong evidence that these peptides arose from translated proteins. However, all these peptides were produced in both the presence and absence of $tRNA_{CUA}^{fmet2}$ expression, suggesting that the amber initiator did not cause the translation initiation events producing these peptides.

Table 3: Proteins and peptides identified exclusive to the ‘off-target’ database in *E. coli* C321.ΔA.exp.

Protein ID	Protein Score	Unique Peptide Sequences	Peptide level			Peptide count	
			Length	Score	PEP	$tRNA_{CUA}^{fmet2-}$	$tRNA_{CUA}^{fmet2+}$
CP006698.1.2209887-2210546	2.3817	LTSSICLAGR [#]	10	58.917	0.006141	3	2
CP006698.1.36227-36982	3.5928	AIVIVADLR [#]	9	70.912	0.001556	2	3
CP006698.1.987948-987106	2.8696	DIFTAQAAR [#]	9	68.422	0.000535	3	3

Peptides identified in both presence and absence of $tRNA_{CUA}^{fmet2}$.

Protein level Score = Andromeda based score on MS/MS spectrum.

Peptide level Score = Andromeda score for the best associated MS/MS spectrum.

PEP = Posterior Error Probability

3.3 Amber initiator efficiency in Generally Recognized as Safe (GRAS) *E. coli* strains

3.3.1 Expression of $tRNA_{CUA}^{fmet2}$ results in efficient translation initiation from sfGFP(UAG) reporter in five common laboratory *E. coli* strains

After showing that production of the amber initiator $tRNA_{CUA}^{fmet2}$ in *E. coli*C321.ΔA.exp was able to cause translation initiation from the sfGFP(UAG) reporter, I next wanted to know whether this system could be used, without modification, in other common *E. coli* lineages. I tested the amber initiation system in four other Generally Recognized as Safe (GRAS) (Archer et al. 2011) *E. coli* strains namely *E. coli* BL21 (B-strain), *E. coli* C122 (C-strain), ATCC 9637 (W-strain) and ATCC 8739 (Crooks-strain). All five *E. coli* strains (including *E. coli* C321.ΔA.exp from the K-12 lineage) are designated risk group 1 (Sims and Kim 2011) but arise from distinct lineages (Bauer et al. 2007).

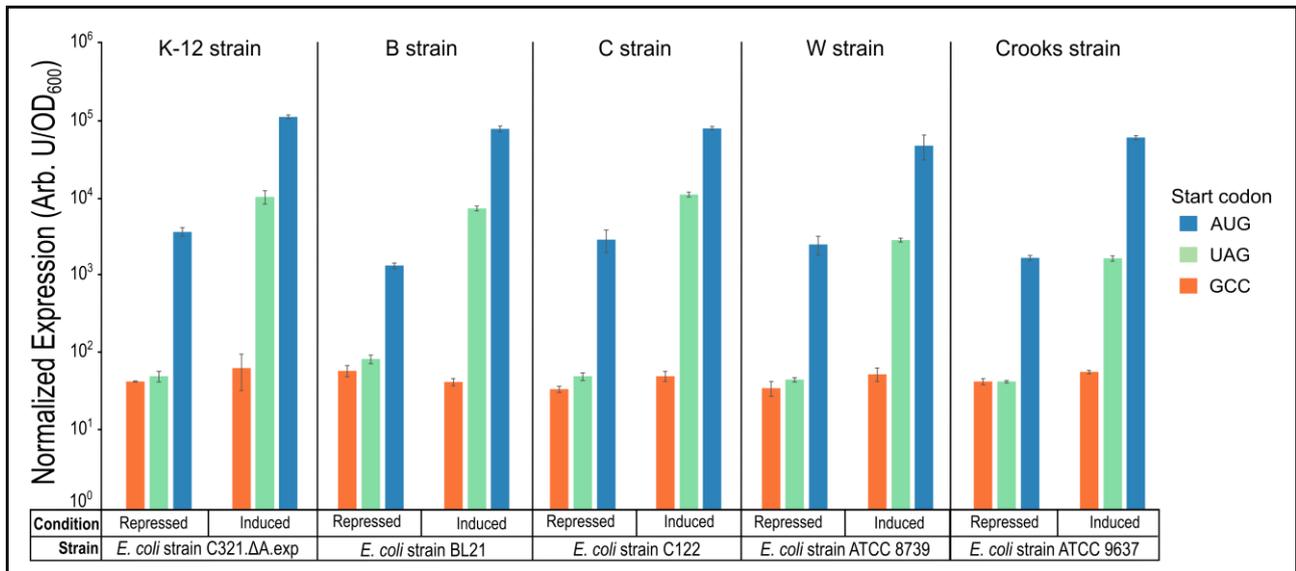


Figure 14: Amber initiator tRNA^{fmet2}_{CUA} can initiate translation from UAG start codons in all five GRAS *E. coli* lineages. The normalized expression levels from sfGFP with either one of the three start codons (AUG, UAG, or GCC) is shown when tRNA^{fmet2}_{CUA} is expressed by an inducible promoter in *E. coli* strain C321.ΔA.exp (K-12 strain), *E. coli* strain BL21 (B strain), *E. coli* strain C122 (C strain), *E. coli* strain ATCC 8739 (W strain) and *E. coli* strain ATCC 9637 (Crooks strain). Each bar displays the average of three replicates and the error bars represent one standard deviation.

After transforming both inducible amber initiator plasmid (pULTRA::*tac-metY*(CUA)) and reporter plasmids into all strains, I measured bulk fluorescence under repressing and inducing conditions. I observed a similar trend in all tested *E. coli* strains: tRNA^{fmet2}_{CUA} expression caused strong sfGFP(UAG) expression and fluorescence (Figure 14). As before, I also observed increased non-specific initiation from sfGFP(AUG) reporter when tRNA^{fmet2}_{CUA} was expressed. I found that the K-12, B, and C strains expressed relatively higher amounts of sfGFP(UAG) compared to the W and Crooks strains. In contrast, expression from sfGFP(AUG) was quite consistent across all the five strains.

3.3.2 Strain-specific amber start codon initiation capacity

To determine if the relative effect of tRNA^{fmet2}_{CUA} expression on UAG versus AUG start codons was similar across the five strains, I compared the expression of sfGFP with amber start codon (sfGFP(UAG)) to a canonical start codon (sfGFP(AUG)) in each *E. coli* species.

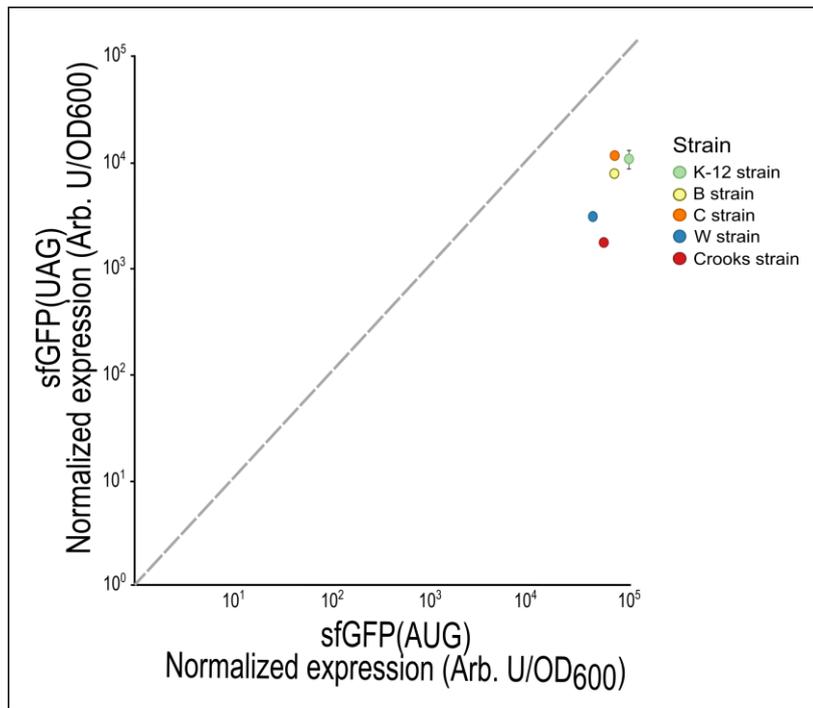


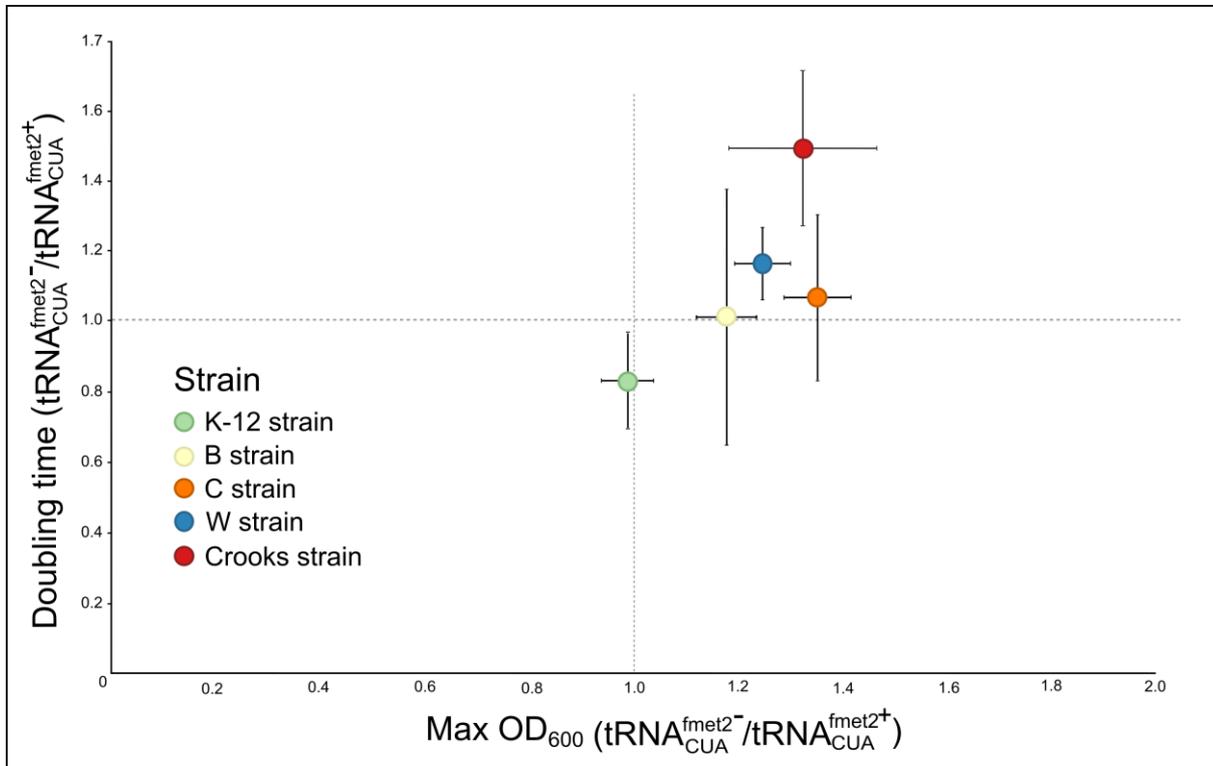
Figure 15: Expression of $tRNA_{CUA}^{fmet2}$ results in higher UAG/AUG start codon affinity in *E. coli* K-12, C and B strains versus W and Crooks strains. The data points represent expression level of sfGFP(UAG) versus sfGFP(AUG) for each strain from $tRNA_{CUA}^{fmet2}$ induction. Error bars represent one standard deviation from the mean (some error bars are too small to visualize).

These data show that $tRNA_{CUA}^{fmet2}$ has the highest affinity for translation initiation from amber UAG start codons versus canonical AUG start codons in the *E. coli* C strain (Figure 15). *E. coli* K-12 and B strains are the next best host system for expressing the amber initiator with strains W and Crooks initiating low levels of reporter expression from both UAG and AUG start codons (Figure 15). After determining the amber plasmid system's propensity for UAG versus AUG start codons in common laboratory *E. coli* GRAS strains, I next wanted to identify any growth defects caused by expression of the $tRNA_{CUA}^{fmet2}$ in these strains.

3.3.3 Production of $tRNA_{CUA}^{fmet2}$ causes minimal growth defects

To determine the fitness effects on *E. coli* strains from amber initiator expression, I measured doubling time and maximal optical density for each strain expressing the $tRNA_{CUA}^{fmet2}$ grown in LB^L/LB^M with antibiotic (spectinomycin) compared to respective strains without pULTRA::tac-*metY*(CUA) plasmid (plasmid-free strains devoid of $tRNA_{CUA}^{fmet2}$) grown in LB^L/LB^M excluding antibiotics. I found that only the *E. coli* strain K-12 (C321.ΔA.exp) showed 18% impaired

fitness specifically slower growth upon $\text{tRNA}_{\text{CUA}}^{\text{fmet2}}$ expression (Figure 16). The C321. Δ A.exp fitness defect only affected growth rate as this strain reached the same maximal cell density regardless of $\text{tRNA}_{\text{CUA}}^{\text{fmet2}}$ expression. Surprisingly, I observed no fitness impairment in any of the other *E. coli* strains (C, B, W, and Crooks) and instead observed that Crooks and W strain exhibited faster



growth due to $\text{tRNA}_{\text{CUA}}^{\text{fmet2}}$ expression. *E. coli* B and C strain growth was unaffected by $\text{tRNA}_{\text{CUA}}^{\text{fmet2}}$ expression.

Figure 16: Effect of $\text{tRNA}_{\text{CUA}}^{\text{fmet2}}$ expression on GRAS *E. coli* strain fitness.

Ratios of maximal cell densities (X-axis) and doubling time (Y-axis) for strains without $\text{tRNA}_{\text{CUA}}^{\text{fmet2}}$ versus strains expressing $\text{tRNA}_{\text{CUA}}^{\text{fmet2}}$. Strains exhibiting slower growth expressing $\text{tRNA}_{\text{CUA}}^{\text{fmet2}}$ are below the horizontal grey line and strains exhibiting higher maximal cell density expressing $\text{tRNA}_{\text{CUA}}^{\text{fmet2}}$ are to the right of the vertical grey line. Each data point is the average of three biological replicates with the error bars showing one standard deviation.

Chapter 4: Discussion

Reassignment of UAG as a sense codon has enabled a wide variety of application like non-standard amino acid incorporation (Liu and Schultz 2010) and novel strategies in biocontainment (Mandell et al. 2015). A genetically recoded organism with an altered genetic code serves as a vehicle specifically for UAG codon reassignment (Lajoie et al. 2013). Previous studies have identified an amber initiation tRNA that can initiate protein synthesis from a UAG start codon (Varshney and RajBhandary 1990), but have not studied the effects of the tRNA at the population and organismal level.

In this study, I measure the efficiency of an amber initiator plasmid expressing tRNA^{fmet2}_{CUA} to reassign the UAG as a start codon in five common laboratory *E. coli* strains from different lineages. In doing so, I examine several features of the amber initiation plasmid such as effect of tRNA^{fmet2}_{CUA} on the fitness of each strain and duration of time the tRNA requires for complete maturation. Additionally, using proteomic analysis, I confirm that tRNA^{fmet2}_{CUA} expression does not cause adverse effects at the host proteome level.

4.1 Creating plasmids to measure new aspects of the amber initiator

In this work I created a new plasmid system based on the pULTRA-backbone to inducibly express the amber initiator tRNA^{fmet2}_{CUA} to measure several aspects of this tRNA that had not been studied before. I built two amber initiation plasmids, one plasmid expresses tRNA^{fmet2}_{CUA} constitutively similar to their plasmid system under the control *metYp1p2* promoter and the other expresses tRNA^{fmet}_{CUA} under the control of a *tac* promoter. Additionally, I designed a fluorescence reporter system using sfGFP to enable a wider dynamic range and real-time measurements at the population level (Martin et al. 1996; Lewis et al. 1998) that were not possible with the previously used CAT reporter (Varshney and RajBhandary 1990).

Bulk fluorescence intensity measurements from cells containing pULTRA::*metYp1p2-metY(CUA)* showed very low levels of sfGFP(UAG) reporter expression when tRNA^{fmet2}_{CUA} was constitutively produced from the *metYp1p2* promoter with 7-fold higher expression compared to sfGFP(GCC) (negative control). In contrast, sfGFP(AUG) expression was about 129-fold higher than sfGFP(GCC) background (Figure 8). This indicates that tRNA^{fmet2}_{CUA} expressed from the native *metYp1p2* promoter can initiate translation at UAG start codons about 5% as frequently as compared to AUG

initiation. This value is also surprisingly low compared to previous estimates of 50-60% expressing amber initiator plasmid from the *metYp1* promoter.

My results with the native *metYp1p2* promoter (proximal and distal promoters) likely differed from previous results because of a *CRP* binding site which represses $\text{tRNA}_{\text{CUA}}^{\text{fmet2}}$ production (Krin et al. 2003). The *CRP* binding site is absent in the *metYp1* promoter (Varshney and RajBhandary 1990) but is present in the distal portion of *metYp1p2* promoter in $\text{pULTRA}::\text{metYp1p2-metY(CUA)}$. Therefore, my results indicated low levels of initiation from sfGFP(UAG) (Figure 8) with surprisingly low efficiency.

Expressing $\text{tRNA}_{\text{CUA}}^{\text{fmet2}}$ from the inducible tac promoter enormously improved sfGFP(UAG) expression, generating 240-fold higher fluorescence compared to sfGFP(GCC). Using the inducible plasmid enabled UAG translation initiation equivalent to sfGFP(AUG) expression levels without $\text{tRNA}_{\text{CUA}}^{\text{fmet2}}$ expression (Figure 8 and 9).

Using the fluorescent sfGFP reporter, for the first time, I could determine the population distribution of *E. coli* cells expressing $\text{tRNA}_{\text{CUA}}^{\text{fmet2}}$ (Figure 10A). I found normally distributed cell populations expressing $\text{tRNA}_{\text{CUA}}^{\text{fmet2}}$, similar in structure to wild-type cells with $\text{tRNA}_{\text{CUA}}^{\text{fmet2}}$. Furthermore, averages of cell fluorescence from flow cytometry measurements closely matched average values determined using bulk culture fluorescence plate reader measurements (Figure 10B and 10C), lending further support to the validity of those results.

4.2 Effect of shifting the stoichiometric ratio of $\text{tRNA}_{\text{CUA}}^{\text{fmet2}}$ to wild type $\text{tRNA}^{\text{fmet}}$

I observed an enormous increase (up to 196-fold) in expression from sfGFP(UAG) in the induced state compared to the repressed state by overexpressing $\text{tRNA}_{\text{CUA}}^{\text{fmet2}}$ via a strong inducible tac promoter (Figure 9). Previous studies highlighted the importance of intracellular competition between mutant $\text{tRNA}_{\text{CUA}}^{\text{fmet2}}$ and the wild type initiator $\text{tRNA}^{\text{fmet}}$ molecules to the amber initiator tRNA access to the P-site of the ribosome (Samhita et al. 2012). To overcome the problem of low $\text{tRNA}_{\text{CUA}}^{\text{fmet2}}$ expression, studies have overexpressed accessory proteins such as aminoacyl-synthetases or IF2 to increase initiation efficiency from UAG codons (Varshney and RajBhandary 1990; Thanedar et al. 2000). In contrast, I observed overexpressing $\text{tRNA}_{\text{CUA}}^{\text{fmet2}}$ can shift the stoichiometric ratio of initiator tRNAs in competition for P-site binding to be more favourable for $\text{tRNA}_{\text{CUA}}^{\text{fmet2}}$. This shift resulted in an enormous increase in expression from sfGFP(UAG) but with a

minor side-effect of sfGFP(AUG) expression increase of 30-fold (Figure 9) similar to what was observed in previous studies (Varshney and RajBhandary 1990).

4.3 Amber initiator expression dynamics potentially reveals $tRNA_{CUA}^{fmet2}$ maturation process

By creating an inducible $tRNA_{CUA}^{fmet}$ expression system I was able to visualize for the first time the dynamics of the amber initiator induction. I observed that $tRNA_{CUA}^{fmet2}$ expression resulted in increased sfGFP(UAG) initiation efficiency only after 2-3 hours. Furthermore, the sfGFP(AUG) reporter showed an initial spike in expression (up to 3-fold) but then remained constant (2-3 fold) over time. However, sfGFP(UAG) expression increased rapidly after the initial 2-3 hour latency, overtaking the fold change in sfGFP(AUG) expression by the end of the 7 hours (Figure 11).

It is known that many tRNAs acquire an isopentenylated adenine modification at base 37 during their maturation that is thought to stabilize weak interactions between anticodon-codon A:U base pair in the first position (Esberg et al. 1999; Schweizer et al. 2017). In *E. coli*, the MiaA/B enzymes are responsible for the modification of an adenine at position 37 to 2-methylthio- N^6 -(Δ^2 -isopentenyl) adenosine (ms^2i^6A) (Esberg et al. 1999). The tRNA helical stem-loop containing the A36-A37-A38 motif are determinants for MiaA recognition (Zhou et al. 2015). The $tRNA_{CUA}^{fmet2}$ has been shown to acquire the ms^2i^6A modification in a MiaA-dependent fashion and lack of this modification disrupted the ability of the tRNA to initiate translation from UAG start codons (Mangroo et al. 1995).

It is plausible that I observed this MiaA/B-dependent maturation process in the fluorescence time-course (Figure 11) and that the delay in sfGFP(UAG) reporter fluorescence is due to the time required for the $tRNA_{CUA}^{fmet2}$ to mature and be able to productively interact with the ribosome to initiate translation. It would be interesting to allow $tRNA_{CUA}^{fmet2}$ to mature before introducing mRNA with a UAG start codon or conversely modulate the expression of $tRNA_{CUA}^{fmet2}$ so that most of the tRNAs can be efficiently modified by MiaA/B.

The ms^2i^6A37 base modification (Figure 17) has an added implication towards an orthogonal $tRNA_{CUA}^{fmet2}$ since previously it was proposed that the unmodified A37 is necessary for $tRNA^{fmet}$ to recognise the canonical start codons (Dube et al. 1968). Concomitantly, recent structural evidence shows that $tRNA^{fmet}$ forms a unique base-triple interaction between the unmodified A37 and the G29-C41 pair which stabilises the conformation of the anticodon loop (Barraud et al. 2008),

suggesting that the unmodified base directly enables AUG start codon recognition. However, the ms^2i^6 modification is likely to cause A37 in $tRNA_{CUA}^{fmet2}$ to base pair with U33 (Mangroo et al. 1995) instead of the base-triple seen in normal initiator $tRNA^{fmet}$. In this scenario, when the $tRNA_{CUA}^{fmet2}$ acquires the ms^2i^6A37 base modification after 2-3 hours of latency it loses non-specific activity towards the AUG start codon, preventing further off-target expression from the sfGFP(AUG) reporter (Figure 11). At the same time, the ms^2i^6A37 modified $tRNA_{CUA}^{fmet2}$ now has increased affinity towards the sfGFP(UAG) reporter and fluorescence increases at the 2-3 hour mark (Figure 11). Control of the ms^2i^6A37 base modification may be a crucial element to engineer a truly orthogonal amber initiator in *E. coli* since seems to be important to increase specificity towards UAG while reducing affinity towards AUG.

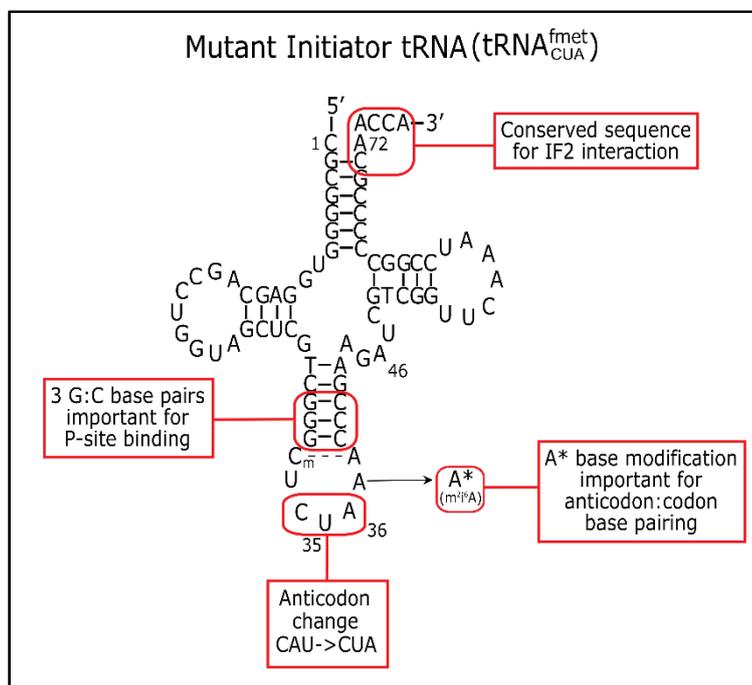


Figure 17: Cloverleaf structure of the amber initiator tRNA ($tRNA_{CUA}^{fmet2}$) denoting functional elements. The acceptor stem for IF2 recognition and 3 G:C base pair essential for P-site binding is conserved in $tRNA_{CUA}^{fmet2}$. Nucleotide A37 is shown with possible modification to ms^2i^6A37 (adapted from Mangroo et al. 1995).

4.4 Proteomic analysis

My proteomic analysis of C321. Δ A.exp revealed less than 6% of identified proteins were differentially expressed due to $tRNA_{CUA}^{fmet}$ expression. It remains to be seen whether this small effect is limited to the C321. Δ A.exp strain due to the advantageous lack of UAG stop codons and RF1 deletion, but this minimal alteration to the *E. coli* proteome is an encouraging result for future creation of an orthogonal translation system.

The lack of RF1 in the C321. Δ A.exp strain is beneficial for the amber initiation system because RF1 is a class I release factor that recognises the UAG codon when it is presented in the ribosomal A-site. I would expect that if RF1 were present it would compete with $tRNA_{CUA}^{fmet2}$ for UAG start codon

binding. This effect would be especially pronounced if the ribosome were using the 70S scanning mode because RF1 binding to the UAG start codon would have a dual negative effect on tRNA_{CUA}^{fmet} efficiency: firstly by outcompeting tRNA_{CUA}^{fmet2} and secondly by dismantling the 70S ribosome lower 70S-scanning mode efficiency (Figure 5). However, I would not expect any difference in reporter expression due to deficiency in the 70S-scanning mode since I used a monocistronic reporter system (single sfGFP reporter in each plasmid) which is not affected by the 70S-scanning mode (Yamamoto et al. 2016).

The release factor 3 (RF3) protein showed a significant increase in protein abundance (3.5 fold-change and p-value < 0.005) from tRNA_{CUA}^{fmet2} expression (Table 2). RF3 is a class II release factor with GTPase activity which primarily interacts with class I termination factors RF1 and RF2 accelerating their dissociation after peptide release in prokaryotes (Freistroffer et al. 1997). Overexpression studies of RF3 implicates this protein in reducing frameshifting and increasing termination efficiency within cells (Freistroffer et al. 1997). Subsequent work described how RF3 plays an indirect role since overexpression of the class I release factors would exhibit the same effect (Crawford et al. 1999). RF3 is now regarded as a recycling factor which maintains the concentration of free class I termination factors in the cell. It is important to note that *E. coli* C321.ΔA.exp has a less effective RF2 variant due to a Threonine at position 246 which has been shown to be less efficient in translation termination (Dinçbas-Renqvist et al. 2000; O'Connor 2015). RF2 defects can be compensated for by RF1 (O'Connor 2015) but the C321.ΔA.exp strain relies solely on the RF2 variant as it has had RF1 deleted from the genome (Lajoie et al. 2013).

In my proteomics experiments I observed unchanged RF2 protein abundance and an increased RF3 abundance, indicating that RF3 up-regulation may be a cellular response to translation termination deficiencies. Firstly, considering the RF3 recycling factor model, it may be plausible that RF3 upregulation compensates for the lack of RF1 in C321.ΔA.exp by attempting to replenish the pool of free RF2 in the cell more rapidly. Secondly, RF3 up-regulation may be caused by the presence of large amounts of tRNA_{CUA}^{fmet} which may not be adequately aminoacylated, resulting in a pool of un-aminoacylated tRNA_{CUA}^{fmet} and increasing 30S ribosomal initiation complex formation. Subsequent accumulation of non-functional 70S complexes (including un-aminoacylated tRNA) would require a higher abundance of free RF2 molecules, which the cell would compensate for by increasing the abundance of the recycling factor RF3.

A secondary, and more controversial role, of RF3 is in post-peptidyltransferase quality control (post-PT QC) where it is thought to be responsible for selectively terminating polypeptides with misincorporated amino acids to maintain translational fidelity (Zaher and Green 2011).

Contradicting *in vivo* studies have shown this model is restricted to *E. coli* K-12 strains with a defective RF2 (O'Connor 2015) implying that RF2 instead of RF3 plays a direct role in post-PT QC. Nonetheless, all these studies indicate an indirect role of RF3 in post-PT QC because RF3 deletion causes an increase in genome-wide recoding and read-through events (Baggett et al. 2017).

Previous work has reported that tRNA_{CUA}^{fmet2} is aminoacylated with glutamine, although the precise fraction with methionine versus glutamine is not known (Varshney and RajBhandary 1990; Varshney et al. 1993). If large numbers of sfGFP(UAG) proteins with N-terminal glutamine are produced in cells expressing the reporter and tRNA_{CUA}^{fmet2} then the increased RF3 expression may be from increasing post-PT QC events due to the large number of sfGFP(UAG) proteins with glutamine as its first amino acid. However, if RF3 up-regulation was a result of post-PT QC, I would have expected to also see other post-PT QC related proteins such as ribosomal release factor or elongation factor G increase in abundance, which I did not (Heurgué-Hamard et al. 1998).

4.5 Amber initiator expression across five common laboratory E. coli strains

4.5.1 The amber initiator tRNA is functional in all five E. coli GRAS strains

Using the pULTRA::*tac-metY*(CUA) plasmid I designed and built, I found that the tRNA_{CUA}^{fmet2} was capable of initiating translation across all five *E. coli* strains tested (Figure 14). Translation initiation from UAG start codons was equally efficient in C321.ΔA.exp (K-12 strain) as in the *E. coli* C and B strains (Figure 14 and 15). In contrast, UAG start codon initiation efficiency was 10-fold lower in *E. coli* W and Crooks strains (Figure 14).

The difference in amber start codon translation initiation efficiency may be due to the different tRNAs encoded by each strain. The genome of C321.ΔA.exp contains four copies of the initiator tRNA located in two loci: The *metZWV*-loci contains three identical copies of tRNA^{fmet1} (Kenri et al. 1994), while the *metY*-loci contains one copy of tRNA^{fmet2} which differs by a single base in the variable region (Komine et al. 1990) from the *metZWV* genes. The major essential tRNA^{fmet1} has methylated guanosine (m⁷G) at position 46 whereas the non-essential minor tRNA^{fmet2} species has an A at this position (Kenri et al. 1991). In contrast, *E. coli* BL21 contains only the major tRNA^{fmet1} species in the *metZWV* locus and an extra copy of the major tRNA^{fmet1} species in the *metY* locus (Figure 18) (Mandal and RajBhandary 1992).

To determine the initiator tRNA composition of *E. coli* W, C, and Crooks I compared sequences of currently annotated *metZWV* and *metY* genes using BLAST and multiple sequence alignments. I found that only the *E. coli* K-12 (C321.ΔA.exp) strain has the minor tRNA^{fmet2} species while all the other strains used in this study have 4 copies of the tRNA^{fmet1} species (Figure 18) similar to *E. coli* BL21.

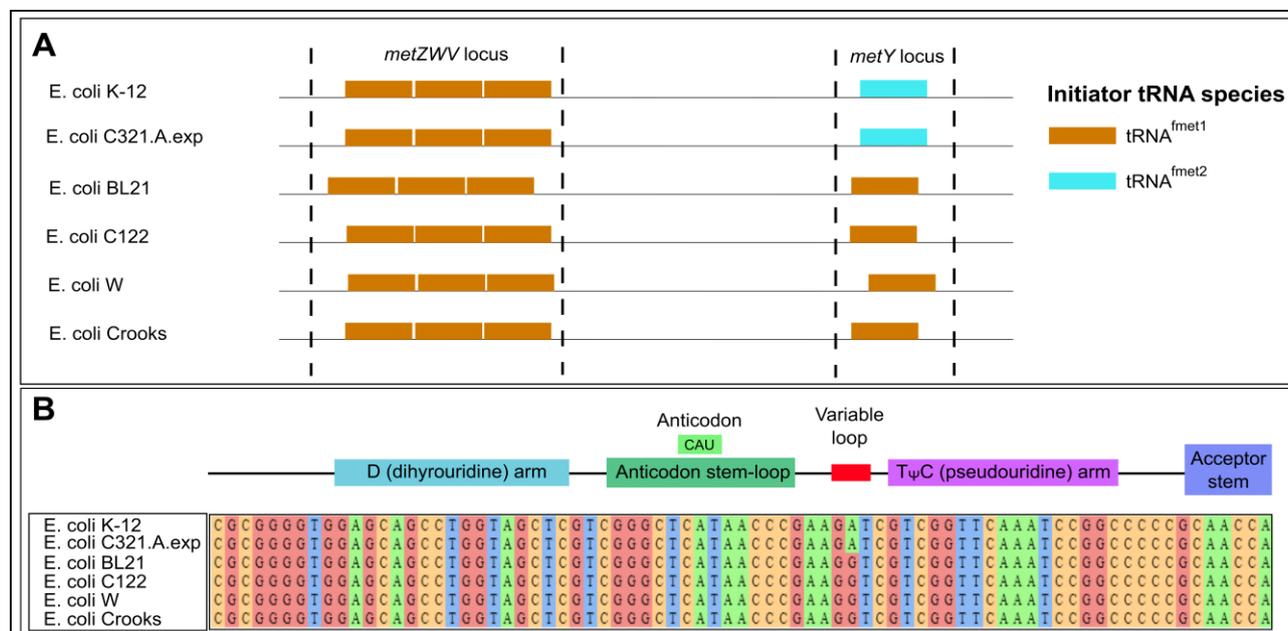


Figure 18: Composition of initiator tRNA species in *E. coli* strains.

(A) Occurrence of initiator tRNA genes in the *metZWV* and *metY* loci (Ishii et al. 1984).

(B) Multiple sequence analysis of tRNA gene sequences in the *metY* loci of each *E. coli* strain. The schematic gene sequence of a tRNA is shown above. Each nucleotide base is represented by a different colour.

Unexpectedly, the sfGFP expression levels (Figure 8) did not correlate with the tRNA^{fmet} composition (Figure 14) since even though *E. coli* strain C and B do not have the minor tRNA^{fmet2} species, they have similar expression from UAG start codon as the *E. coli* C321.ΔA.exp strain. This result implies that the base variation at position 46 between tRNA^{fmet1} and tRNA^{fmet2} is insignificant for translation initiation. Similarly, the low sfGFP(UAG) and sfGFP(AUG) reporter expression levels from in *E. coli* strains W and Crooks may be due to other strain-dependent variables because I observed a lower expression for sfGFP(AUG) as well (Figure 14 and 15). In summary, I showed that the amber initiator tRNA plasmid expression system I designed can be used without modification within five common laboratory strains of *E. coli* and may prove useful for future applications that leverage the unique properties of each strain.

4.5.2 C321.ΔA.exp is the only strain with growth defect due to $tRNA_{CUA}^{fmet2}$ expression

Growth experiments with all five *E. coli* strains expressing the $tRNA_{CUA}^{fmet2}$ showed that *E. coli* W, C, and Crooks had enhanced growth rate and maximum cell density, while C321.ΔA.exp showed a small (18%) growth defect (Figure 16). Previous studies have shown growth impairment due to $tRNA^{fmet2}$ deletion (Kenri et al. 1991) but to my knowledge there have been no growth experiments on strains expressing plasmid-borne copies of an initiator tRNA or any of its mutants.

This growth defect in C321.ΔA.exp was surprising considering that this strain has all UAG stop codons removed. I had reasoned that this strain would have been less affected by $tRNA_{CUA}^{fmet2}$ expression due to less opportunity for ribosome recycling at UAG start codons from the removal of RF1. It may be that ribosome stalling is still occurring in C321.ΔA.exp due to incorporation of un-aminoacylated (uncharged) or misaminoacylated (wrongly charged) $tRNA_{CUA}^{fmet2}$ instead causing perturbed growth. On the contrary, $tRNA_{CUA}^{fmet2}$ interfering $tRNA^{fmet2}$ activity (specifically found in C321.ΔA.exp) could explain minor growth impairment.

Concomitantly, all the experiments were performed in a rich growth media (LB^L and LB^M) with a good carbon, nitrogen and energy source previously described to encourage rapid growth cells by switching off biosynthetic pathways to focus on macromolecular synthesis, prominently protein synthesis. Conversely, cells grown in minimal media starve for building blocks like amino acids and this burden is overcome by switching on the biosynthetic pathways and elevating expression of regulators of cellular processes (Tao et al. 1999). It is probable that the rich LB^L and LB^M growth media negates some of the fitness effects caused by expression of $tRNA_{CUA}^{fmet2}$, exhibiting minimal growth defect across all five *E. coli* strains. Analysing fitness of the strains expressing $tRNA_{CUA}^{fmet2}$ grown in minimal media could demonstrate more variability among the five *E. coli* strains.

However, as proteomic analysis suggests, a compensatory mechanism may control this effect to minimize the reduction in growth rate. Another reason for reduced growth rate in *E. coli* C321.ΔA.exp compared to the rest of the strains may be implicated by a K-12 specific Thr246 RF2 variant whereas all the remaining strains have the fully functional Ala246 RF2 protein (O'Connor 2015; Baggett et al. 2017). The K-12 specific RF2 variant has shown to reduce termination efficiency and efficiency was restored when the Thr246 was restored to Ala246 (Dinçbas-Renqvist et al. 2000). Together, these disadvantages appear to make it harder for *E. coli* C321.ΔA.exp to cope with $tRNA_{CUA}^{fmet2}$ expression and the concomitant burden to the translation machinery. It will be interesting in the future to measure proteome level changes in one of the four strains with a

more efficient translation termination machinery expressing $tRNA_{CUA}^{fmet2}$ to see if RF3 is still overexpressed due to $tRNA_{CUA}^{fmet2}$, implying direct interactions between $tRNA_{CUA}^{fmet2}$ and RF3.

4.6 Off-target effects of amber initiator $tRNA_{CUA}^{fmet2}$ expression

Although my experiments showed that $tRNA_{CUA}^{fmet2}$ expression is capable of directing translation initiation from UAG start codons, I also observed an increase in expression from AUG start codons (Figure 8). It is known that due to wobble base pairing in the 1st position, the regular $tRNA^{fmet}$ can initiate translation from GUG and UUG start codons in some cases. In fact, of annotated genes in model bacterial genomes, 13.8% have GUG start codons and 4.3% have UUG start codons (Hecht et al. 2017). As a result of the known promiscuity of the regular $tRNA^{fmet}$ (Hecht et al. 2017), I did not expect that the amber initiator $tRNA_{CUA}^{fmet2}$ would only initiate translation at UAG start codons. Furthermore, previous results with the amber initiator suggested initiation from a AUG start codon was possible, but were inconclusive due to the narrow dynamic range of the CAT assay (Varshney and RajBhandary 1990). Despite the increased sfGFP(AUG) expression due to $tRNA_{CUA}^{fmet2}$ expression my proteomic results showed only 42 proteins (5.6% of total) changed abundance, suggest this effect is not occurring to any great extent within cells. Furthermore, all of the detected proteins with differential regulation were from genes beginning with AUG, suggesting that the amber initiator $tRNA_{CUA}^{fmet2}$ may not have the same affinity with GUG or UGG codons as the normal $tRNA^{fmet}$ does. I plan to further investigate the precise interactions of the $tRNA_{CUA}^{fmet2}$ with all other codons in the future.

The minor $tRNA_{CUA}^{fmet2}$ interaction with AUG start codons in my assays may to be governed by two factors. First, the A37 base modification next to the anticodon (Figure 17) is known to occur in tRNAs with NNA anticodons and enables interaction with UNN which enhances specificity of $tRNA_{CUA}^{fmet2}$ to recognise UAG instead of AUG. The modified ms^2i^6A37 base stabilizes codon recognition by stacking interactions on the first Watson-Crick base pair (Schweizer et al. 2017). Secondly, the underlying AUG initiation activity may be enhanced when the reporter transcript is at a high copy number due to the medium-copy (20-40) plasmid with a strong T5 promoter (Morelli et al. 2011). Although a previous study using the regular $tRNA^{fmet}$ showed little difference in rank order initiation strength from 16 different codons when reporter copy number was varied (Hecht et al. 2017), we do not know if this same result would hold true for $tRNA_{CUA}^{fmet2}$. It has already been shown that competition between the amber initiator and the regular $tRNA^{fmet}$ can affect accessibility to the ribosome P-site. Shifts in the stoichiometric ratio between amber and

native tRNAs has been shown to increase initiation efficiency of the amber tRNA_{CUA}^{fmet2} (Samhita et al. 2012). Similarly, I have shown that overexpressing tRNA_{CUA}^{fmet2} introduces a shift in the stoichiometric ratio increasing tRNA_{CUA}^{fmet2} efficiency above that seen in other experiments using the amber initiator tRNA (Varshney and RajBhandary 1990). In my experiments there was no detectable translation initiation from UAG codons in the C321.ΔA.exp genome, but I measured the proteome in the presence of the sfGFP(UAG) reporter plasmid that provided abundant transcripts to 'soak up' the available tRNA_{CUA}^{fmet2}. It is possible that in the absence of a large number of transcripts with a strong RBS and *sfGFP* gene with UAG start codon I may have observed peptides from genome-encoded UAG open reading frames. In the future, I could assess these predictions by using reporter plasmids with lower copy numbers, plasmids with inducible promoters that allow the number of transcripts to be more precisely controlled, and comparative proteomics in the absence of reporter plasmids.

4.7 Conclusion and future directions

In this study I defined several properties of the amber initiator tRNA_{CUA}^{fmet2} that will enable further development of an efficient orthogonal translation initiation system in *E. coli*. I have demonstrated the first experimental evidence that overexpressing tRNA_{CUA}^{fmet2} does not cause any major physiological or proteome level defects in *E. coli*. Furthermore, for the first time, I showed that cells expressing the amber initiator tRNA form a uniform population distribution. Lastly, the modular two-plasmid system I have designed and constructed in this work will form a platform from which to engineer an improved mutant tRNA_{CUA}^{fmet2} which can effectively initiate protein synthesis from the UAG start codon while reducing non-specific interaction with the canonical (AUG) start codon.

References

- Allen GS, Zavialov A, Gursky R, et al (2005) The cryo-EM structure of a translation initiation complex from *Escherichia coli*. *Cell* 121:703–712. doi: 10.1016/j.cell.2005.03.023
- Anantharamaiah GM, Garber DW, Segrest JP (2003) Structure of the Plasma Lipoproteins. *Atlas Atheroscler* 51–67. doi: 10.1007/978-1-4615-6484-3_4
- Archer CT, Kim JF, Jeong H, et al (2011) The genome sequence of *E. coli* W (ATCC 9637): comparative genome analysis and an improved genome-scale reconstruction of *E. coli*. *BMC Genomics* 12:9. doi: 10.1186/1471-2164-12-9
- Ayyub SA, Dobriyal D, Varshney U (2017) Contributions of the N- and C-Terminal Domains of Initiation Factor 3 to Its Functions in the Fidelity of Initiation and Antiassociation of the Ribosomal Subunits. *J Bacteriol* 199:e00051-17. doi: 10.1128/JB.00051-17
- Baggett NE, Zhang Y, Gross CA, Ibba M (2017) Global analysis of translation termination in *E. coli*. *PLOS Genet*. doi: 10.1371/journal.pgen.1006676
- Barraud P, Schmitt E, Mechulam Y, et al (2008) A unique conformation of the anticodon stem-loop is associated with the capacity of tRNA^{fMet} to initiate protein synthesis. *Nucleic Acids Res* 36:4894–4901. doi: 10.1093/nar/gkn462
- Barrell BG, Bankier a T, Drouin J (1979) A different genetic code in human mitochondria. *Nature* 282:189–194. doi: 10.1038/282189a0
- Bauer AP, Dieckmann SM, Ludwig W, Schleifer K-H (2007) Rapid identification of *Escherichia coli* safety and laboratory strain lineages based on Multiplex-PCR. *FEMS Microbiol Lett* 269:36–40. doi: 10.1111/j.1574-6968.2006.00594.x
- Brenner S, Stresson AOW, Kaplan S (1965) Genetic Code: The “Nonsense” Triplets for Chain Termination and Their Suppression. *Nature* 206:994–998. doi: 10.1038/206994a0
- Brock JE, Pourshahian S, Giliberti J, et al (2008) Ribosomes bind leaderless mRNA in *Escherichia coli* through recognition of their 5'-terminal AUG. *RNA* 14:2159–69. doi: 10.1261/rna.1089208
- Carter AP, Clemons WM, Brodersen DE, et al (2001) Crystal structure of an initiation factor bound to the 30S ribosomal subunit. *Science* 291:498–501. doi: 10.1126/SCIENCE.1057766
- Caserta E, Tomšić J, Spurio R, et al (2006) Translation Initiation Factor IF2 Interacts with the 30 S Ribosomal Subunit via Two Separate Binding Sites. *J Mol Biol* 362:787–799. doi: 10.1016/j.jmb.2006.07.043
- Chansb PT, Ohmorill11 H, Tomizawall J-I, Lebowitzs J (1985) Nucleotide Sequence and Gene Organization of ColE1 DNA*. *J Biol Chem* 260:8925–8935.
- Clancy S, Brown W (2008) Translation: DNA to mRNA to Protein. *Nat Educ* 1:101.
- Colasanti J, Denhardt DT (1985) Expression of the cloned bacteriophage phi X174 A* gene in *Escherichia coli* inhibits DNA replication and cell division. *J Virol* 53:807–13.
- Cooper G (2000) Translation of mRNA. In: *The Cell: A Molecular Approach*. 2nd edition. Sinauer Associates,
- Cox J, Hubner NC, Mann M (2008) How Much Peptide Sequence Information Is Contained in Ion Trap Tandem Mass Spectra? *J Am Soc Mass Spectrom* 19:1813–1820. doi:

10.1016/j.jasms.2008.07.024

- Crawford D-JG, Ito K, Nakamura Y, Tate WP (1999) Indirect regulation of translational termination efficiency at highly expressed genes and recoding sites by the factor recycling function of *Escherichia coli* release factor RF3. *EMBO J* 18:727–732. doi: 10.1093/emboj/18.3.727
- Crick FH, Barnett L, Brenner S, Watts-Tobin RJ (1961) General nature of the genetic code for proteins. *Nature* 192:1227–1232. doi: 10.1038/1921227a0
- Cummings HS, Hershey JW (1994) Translation initiation factor IF1 is essential for cell viability in *Escherichia coli*. *J Bacteriol* 176:198–205.
- Dinçbas-Renqvist V, Engström Å, Mora L, et al (2000) A post-translational modification in the GGQ motif of RF2 from *Escherichia coli* stimulates termination of translation. *EMBO J* 19:6900–6907. doi: 10.1093/emboj/19.24.6900
- Dube SK, Marcker K a, Clark BF, Cory S (1968) Nucleotide sequence of N-formyl-methionyl-transfer RNA. *Nature* 218:232–233. doi: 10.1038/218232a0
- Epstein RH, Bolle A, Steinberg CM (2012) Amber mutants of bacteriophage T4D: their isolation and genetic characterization. *Genetics* 190:833–40. doi: 10.1534/genetics.112.138438
- Esberg B, Leung H-CE, Tsui H-CT, et al (1999) Identification of the miaB Gene, Involved in Methylthiolation of Isopentenylated A37 Derivatives in the tRNA of *Salmonella typhimurium* and *Escherichia coli*. *J Bacteriol* 181:7256.
- Freistroffer D V., Pavlov MY, MacDougall J, et al (1997) Release factor RF3 in *E. coli* accelerates the dissociation of release factors RF1 and RF2 from the ribosome in a GTP-dependent manner. *EMBO J* 16:4126–4133. doi: 10.1093/emboj/16.13.4126
- Gibson DG, Young L, Chuang R-Y, et al (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6:343–5. doi: 10.1038/nmeth.1318
- Grill S, Gualerzi CO, Londei P, Bläsi U (2000) Selective stimulation of translation of leaderless mRNA by initiation factor 2: evolutionary implications for translation. *EMBO J* 19:4101–10. doi: 10.1093/emboj/19.15.4101
- Gualerzi CO, Pon CL (1990) Initiation of mRNA translation in prokaryotes. *Biochemistry* 29:5881–5889. doi: 10.1021/bi00477a001
- Guenneugues M, Caserta E, Brandi L, et al (2000) Mapping the fMet-tRNA^{fMet} binding site of initiation factor IF2. *EMBO J* 19:5233–5240. doi: 10.1093/emboj/19.19.5233
- Guillon JM, Mechulam Y, Blanquet S, Fayat G (1993) Importance of formylability and anticodon stem sequence to give a tRNA(Met) an initiator identity in *Escherichia coli*. *J Bacteriol* 175:4507–4514.
- Hartz D, Binkley J, Hollingsworth T, Gold L (1990) Domains of initiator tRNA and initiation codon crucial for initiator tRNA selection by *Escherichia coli* IF3. *Genes Dev* 4:1790–1800. doi: 10.1101/gad.4.10.1790
- Hashimoto-Gotoht T, Inselburg J (1979) ColE1 Plasmid Incompatibility: Localization and Analysis of Mutations Affecting Incompatibility. *J Bacteriol* 139:608–619.
- Haurlyuk V, Ehrenberg M (2006) Two-Step Selection of mRNAs in Initiation of Protein Synthesis.

Mol. Cell 22:155–156.

- Hecht A, Glasgow J, Jaschke PR, et al (2017) Measurements of translation initiation from all 64 codons in *E. coli*. *Nucleic Acids Res.* doi: 10.1093/nar/gkx070
- Heurgué-Hamard V, Karimi R, Mora L, et al (1998) Ribosome release factor RF4 and termination factor RF3 are involved in dissociation of peptidyl-tRNA from the ribosome. *EMBO J* 17:808–16. doi: 10.1093/emboj/17.3.808
- Hirokawa G, Nijman RM, Raj VS, et al (2005) The role of ribosome recycling factor in dissociation of 70S ribosomes into subunits. *RNA* 11:1317–28. doi: 10.1261/rna.2520405
- Hussain T, Licer JL, Wimberly BT, et al (2016) Large-Scale Movements of IF3 and tRNA during Bacterial Translation Initiation. *Cell* 167:133–144.e13. doi: 10.1016/j.cell.2016.08.074
- Hyatt D, Chen G-L, Locascio PF, et al (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119
- Ishii S, Kuroki K, Imamoto F (1984) tRNAMet^{f2} gene in the leader region of the nusA operon in *Escherichia coli*. *Proc Natl Acad Sci U S A* 81:409–413.
- Julián P, Milon P, Agirrezabala X, et al (2011) The cryo-EM structure of a complete 30S translation initiation complex from *Escherichia coli*. *PLoS Biol* 9:e1001095. doi: 10.1371/journal.pbio.1001095
- Kenri T, Imamoto F, Kano Y (1994) Three tandemly repeated structural genes encoding tRNA(f1Met) in the metZ operon of *Escherichia coli* K-12. *Gene* 138:261–262. doi: 10.1016/0378-1119(94)90821-4
- Kenri T, Kohno K, Goshima N, et al (1991) Construction and characterization of an *Escherichia coli* mutant with a deletion of the metZ gene encoding tRNA (f1Met). *Gene* 103:31–36. doi: 10.1016/0378-1119(91)90387-Q
- Knight RD, Freeland SJ, Landweber LF (2001) Rewiring the keyboard: evolvability of the genetic code. *Nat Rev Genet* 2:49–58. doi: 10.1038/35047500
- Knight RD, Freeland SJ, Landweber LF (1999) Selection, history and chemistry: The three faces of the genetic code. *Trends Biochem Sci* 24:241–247. doi: 10.1016/S0968-0004(99)01392-4
- Komine Y, Adachi T, Inokuchi H, Ozeki H (1990) Genomic organization and physical mapping of the transfer RNA genes in *Escherichia coli* K12. *J Mol Biol* 212:579–598. doi: 10.1016/0022-2836(90)90224-A
- Kozak M (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res* 15:8125–48.
- Krin E, Laurent-Winter C, Bertin PN, et al (2003) Transcription regulation coupling of the divergent argG and metY promoters in *Escherichia coli* K-12. *J Bacteriol* 185:3139–46. doi: 10.1128/JB.185.10.3139-3146.2003
- Kushner SR (2002) mRNA decay in *Escherichia coli* comes of age. *J Bacteriol* 184:4658–65; discussion 4657. doi: 10.1128/JB.184.17.4658-4665.2002
- Lajoie MJ, Rovner AJ, Goodman DB, et al (2013) Genomically Recoded Organisms Expand Biological Functions. *Science* (80-) 342:357–360. doi: 10.1126/science.1241459

- Lancaster L, Noller HF (2005) Involvement of 16S rRNA Nucleotides G1338 and A1339 in Discrimination of Initiator tRNA. *Mol Cell* 20:623–632. doi: 10.1016/j.molcel.2005.10.006
- Laursen BS, Sorensen HP, Mortensen KK, Sperling-Petersen HU (2005) Initiation of Protein Synthesis in Bacteria. *Microbiol Mol Biol Rev* 69:101–123. doi: 10.1128/MMBR.69.1.101-123.2005
- Lewis JC, Feltus A, Ensor CM, et al (1998) Peer Reviewed: Applications of Reporter Genes. *Anal Chem* 70:579A–585A. doi: 10.1021/ac9819638
- Liu CC, Schultz PG (2010) Adding New Chemistries to the Genetic Code. *Annu Rev Biochem* 79:413–444. doi: 10.1146/annurev.biochem.052308.105824
- Liu DR, Magliery TJ, Schultz PG (1997) Characterization of an “orthogonal” suppressor tRNA derived from *E. coli* tRNA^{2 Gln}.
- Lobanov A V, Turanov AA, Hatfield DL, Gladyshev VN (2010) Dual functions of codons in the genetic code. *Crit Rev Biochem Mol Biol* 45:257–65. doi: 10.3109/10409231003786094
- Ma NJ, Isaacs FJ (2016) Genomic Recoding Broadly Obstructs the Propagation of Horizontally Transferred Genetic Elements. *Cell Syst* 3:199–207. doi: 10.1016/j.cels.2016.06.009
- Makoff AJ, Smallwood AE (1990) The use of two-cistron constructions in improving the expression of a heterologous gene in *E. coli*. *Nucleic Acids Res* 18:1711–8.
- Mandal N, RajBhandary UL (1992) *Escherichia coli* B lacks one of the two initiator tRNA species present in *E. coli* K-12. *J. Bacteriol.* 174:7827–7830.
- Mandell DJ, Lajoie MJ, Mee MT, et al (2015) Biocontainment of genetically modified organisms by synthetic protein design. *Nature* 518:55–60. doi: 10.1038/nature14121
- Mangroo D, Limbach PA, McCloskey JA, RajBhandary UL (1995) An anticodon sequence mutant of *Escherichia coli* initiator tRNA: possible importance of a newly acquired base modification next to the anticodon on its activity in initiation. *J Bacteriol* 177:2858–62. doi: 10.1128/JB.177.10.2858-2862.1995
- Marck C, Grosjean H (2002) tRNomics: analysis of tRNA genes from 50 genomes of Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific features. *RNA* 8:1189–232.
- Martin CS, Wight PA, Dobretsova A, Bronstein I (1996) Dual luminescence-based reporter gene assay for luciferase and β -galactosidase. *Biotechniques* 21:520–524.
- Mayer C, Köhrer C, Kenny E, et al (2003) Anticodon sequence mutants of *Escherichia coli* initiator tRNA: Effects of overproduction of aminoacyl-tRNA synthetases, methionyl-tRNA formyltransferase, and initiation factor 2 on activity in initiation. *Biochemistry* 42:4787–4799. doi: 10.1021/bi034011r
- Mayer C, RajBhandary UL (2002) Conformational change of *Escherichia coli* initiator methionyl-tRNA(fMet) upon binding to methionyl-tRNA formyl transferase. *Nucleic Acids Res* 30:2844–50.
- Mendez-Perez D, Gunasekaran S, Orler VJ, Pflieger BF (2012) A translation-coupling DNA cassette for monitoring protein translation in *Escherichia coli*. *Metab Eng* 14:298–305. doi: 10.1016/j.ymben.2012.04.005

- Milon P, Konevega AL, Gualerzi CO, Rodnina M V. (2008) Kinetic Checkpoint at a Late Step in Translation Initiation. *Mol Cell* 30:712–720. doi: 10.1016/j.molcel.2008.04.014
- Mitchell SF, Walker SE, Algire MA, et al (2010) The 5'-7-methylguanosine cap on eukaryotic mRNAs serves both to stimulate canonical translation initiation and to block an alternative pathway. *Mol Cell* 39:950–962. doi: 10.1016/j.molcel.2010.08.021
- Moazed D, Samaha RR, Gualerzi C, Noller HF (1995) Specific protection of 16 S rRNA by translational initiation factors. *J Mol Biol* 248:207–210. doi: 10.1016/S0022-2836(95)80042-5
- Moe-Behrens GHG, Davis R, Haynes KA (2013) Preparing synthetic biology for the world. *Front Microbiol* 4:5. doi: 10.3389/fmicb.2013.00005
- Moll I, Grill S, Gualerzi CO, Bläsi U (2002) Leaderless mRNAs in bacteria: Surprises in ribosomal recruitment and translational control. *Mol Microbiol* 43:239–246. doi: 10.1046/j.1365-2958.2002.02739.x
- Morelli MJ, Allen RJ, Rein ten Wolde P (2011) Effects of Macromolecular Crowding on Genetic Networks. *Biophys J* 101:2882–2891. doi: 10.1016/J.BPJ.2011.10.053
- Mutalik VK, Guimaraes JC, Cambray G, et al (2013) Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat Methods* 10:354–360. doi: 10.1038/nmeth.2404
- Neilson KA, Keighley T, Pascovici D, et al (2013) Label-Free Quantitative Shotgun Proteomics Using Normalized Spectral Abundance Factors. Humana Press, Totowa, NJ, pp 205–222
- Nirenberg MW, Matthae JH, Jones OW (1962) an Intermediate in the Biosynthesis of Polyphenylalanine Directed By Synthetic Template Rna. *Proc Natl Acad Sci* 48:104–109. doi: 10.1073/pnas.48.1.104
- Normanly J, Abelson J (1989) tRNA Identity. *Annu Rev Biochem* 58:1029–1049. doi: 10.1146/annurev.bi.58.070189.005121
- O'Connor M (2015) Interactions of release factor RF3 with the translation machinery. *Mol Genet Genomics* 290:1335–1344. doi: 10.1007/s00438-015-0994-x
- Osawa S, Jukes TH (1989) Codon reassignment (codon capture) in evolution. *J Mol Evol* 28:271–8. doi: 10.1007/BF02103422
- Panicker IS, Browning GF, Markham PF (2015) The effect of an alternate start codon on heterologous expression of a PhoA fusion protein in mycoplasma gallisepticum. *PLoS One* 10:1–10. doi: 10.1371/journal.pone.0127911
- Petrelli D, LaTeana A, Garofalo C, et al (2001) Translation initiation factor IF3: two domains, five functions, one mechanism? *EMBO J* 20:4560–9. doi: 10.1093/emboj/20.16.4560
- Rajbhandary UL, Chang SH, Stuart A, et al (1967) Studies on polynucleotides, lxviii the primary structure of yeast phenylalanine transfer RNA. *Proc Natl Acad Sci U S A* 57:751–8.
- Ramakrishnan V (2002) Ribosome Structure and the Mechanism of Translation. *Cell* 108:557–572. doi: 10.1016/S0092-8674(02)00619-0
- Samhita L, Shetty S, Varshney U (2012) Unconventional initiator tRNAs sustain Escherichia coli. *Proc Natl Acad Sci U S A* 109:13058–63. doi: 10.1073/pnas.1207868109

- Schmidt M, Pei L (2015) Improving Biocontainment with Synthetic Biology: Beyond Physical Containment. Springer, Berlin, Heidelberg, pp 185–199
- Schultz DW, Yarus M (1996) On malleability in the genetic code. *J Mol Evol* 42:597–601. doi: 10.1007/bf02352290
- Schultz KC, Supekova L, Ryu Y, et al (2006) A Genetically Encoded Infrared Probe. *J Am Chem Soc* 128:13984–13985. doi: 10.1021/ja0636690
- Schweizer U, Bohleber S, Fradejas-Villar N (2017) The modified base isopentenyladenosine and its derivatives in tRNA. *RNA Biol.* 1–12.
- Selmer M, Dunham CM, Murphy F V, et al (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* 313:1935–42. doi: 10.1126/science.1131127
- Sette M, van Tilborg P, Spurio R, et al (1997) The structure of the translational initiation factor IF1 from *E.coli* contains an oligomer-binding motif. *EMBO J* 16:1436–43. doi: 10.1093/emboj/16.6.1436
- Shine J, Dalgarno L (1975) Determinant of cistron specificity in bacterial ribosomes. *Nature* 254:34–38. doi: 10.1038/254034a0
- Simonetti A, Marzi S, Myasnikov AG, et al (2008) Structure of the 30S translation initiation complex. *Nature* 455:416–420. doi: 10.1038/nature07192
- Sims GE, Kim S-H (2011) Whole-genome phylogeny of *Escherichia coli*/*Shigella* group by feature frequency profiles (FFPs). *Proc Natl Acad Sci* 108:8329–8334. doi: 10.1073/pnas.1105168108
- Stuitje AR, Veltkamp E, Weijers PJ, Nijkamp HJJ (1979) Nucleic Acids Research Origin and direction of replication of the bacteriocinogenic plasmid Clo DF13.
- Tao H, Bausch C, Richmond C, et al (1999) Functional genomics: Expression analysis of *Escherichia coli* growing on minimal and rich media. *J Bacteriol* 181:6425–6440.
- Thanedar S, Vinay Kumar N, Varshney U (2000) The Fate of the Initiator tRNAs Is Sensitive to the Critical Balance between Interacting Proteins*. doi: 10.1074/jbc.M001238200
- Tsugita A, Gish T, Young J, et al (1960) The Complete Amino Acid Sequence of the Protein of Tobacco Mosaic Virus. *Proc Natl Acad Sci* 46:1–5. doi: 10.1073/pnas.46.11.1463
- Udagawa T, Shimizu Y, Ueda T (2004) Evidence for the Translation Initiation of Leaderless mRNAs by the Intact 70 S Ribosome without Its Dissociation into Subunits in Eubacteria. *J Biol Chem* 279:8539–8546. doi: 10.1074/jbc.M308784200
- Uhlin BE, Nordstrom K (1975) Plasmid Incompatibility and Control of Replication: Copy Mutants of the R-Factor Ri in *Escherichia coli* K-12. *J Bacteriol* 124:641–649.
- Varshney U, Lee CP, RajBhandary UL (1993) From elongator tRNA to initiator tRNA. *Proc Natl Acad Sci U S A* 90:2305–9. doi: 10.1073/pnas.90.6.2305
- Varshney U, RajBhandary UL (1992) Role of methionine and formylation of initiator tRNA in initiation of protein synthesis in *Escherichia coli*. *J Bacteriol* 174:7819–7826.
- Varshney U, RajBhandary UL (1990) Initiation of protein synthesis from a termination codon. *Proc Natl Acad Sci* 87:1586–1590. doi: 10.1073/pnas.87.4.1586

- Wang L, Magliery TJ, Liu DR, Schultz PG (2000) A new functional suppressor tRNA/aminoacyl-tRNA synthetase pair for the in vivo incorporation of unnatural amino acids into proteins [16]. *J Am Chem Soc* 122:5010–5011. doi: 10.1021/ja000595y
- Weston A, Humphreys GO, Brown MGM, Saunders JR (1979) Simultaneous transformation of *Escherichia coli* by pairs of compatible and incompatible plasmid DNA molecules. *MGG Mol Gen Genet* 172:113–118. doi: 10.1007/BF00276222
- Whitaker WR, Lee H, Arkin AP, Dueber JE (2015) Avoidance of Truncated Proteins from Unintended Ribosome Binding Sites within Heterologous Protein Coding Sequences. *ACS Synth Biol* 4:249–257. doi: 10.1021/sb500003x
- Woese CR, Dugre DH, Saxinger WC, Dugre SA (1966) The molecular basis for the genetic code. *Proc Natl Acad Sci U S A* 55:966–74. doi: 10.1073/pnas.55.4.966
- Wu XQ, Rajbhandary UL (1997) Effect of the amino acid attached to *Escherichia coli* initiator tRNA on its affinity for the initiation factor IF2 and on the IF2 dependence of its binding to the ribosome. *J Biol Chem* 272:1891–1895. doi: 10.1074/jbc.272.3.1891
- Yamamoto H, Wittek D, Gupta R, et al (2016) 70S-scanning initiation is a novel and frequent initiation mode of ribosomal translation in bacteria. *Proc Natl Acad Sci U S A* 113:E1180-9. doi: 10.1073/pnas.1524554113
- Zaher HS, Green R (2011) A primary role for release factor 3 in quality control during translation elongation in *Escherichia coli*. *Cell* 147:396–408. doi: 10.1016/j.cell.2011.08.045
- Zhou M, Long T, Fang Z-P, et al (2015) Identification of determinants for tRNA substrate recognition by *Escherichia coli* C/U34 2'-O-methyltransferase. *RNA Biol* 12:900–11. doi: 10.1080/15476286.2015.1050576

Supplementary Material

Table S1: Synthesized oligonucleotides

Synthesized DNA	Length (bp)	Template	Description/Sequence
A80-tRNA ^{fmet} CUA	1971	-	Synthesized gblock containing <i>metY</i> (CUA) gene and <i>metYp1p2</i> promoter
544-pULTRA_linear-FOR	40	pULTRA-CNF	Forward primer for to amplify pULTRA backbone to assemble pULTRA:: <i>metYp1p2-metY</i> (CUA)
545-pULTRA_linear-REV	40	pULTRA-CNF	Reverse primer to amplify pULTRA backbone to assemble pULTRA:: <i>metYp1p2-metY</i> (CUA)
546- <i>metY</i> (CUA)-FOR	40	A80	Forward primer for to amplify <i>metY</i> (CUA) gene to assemble pULTRA:: <i>metYp1p2-metY</i> (CUA)
547- <i>metY</i> (CUA)-REV	40	A80	Reverse primer to amplify <i>metY</i> (CUA) gene to assemble pULTRA:: <i>metYp1p2-metY</i> (CUA)
548-pULTRA_back-FOR	48	pULTRA-CNF	Forward primer for to amplify pULTRA backbone to assemble pULTRA:: <i>tac-metY</i> (CUA)
549-pULTRA_back-REV	41	pULTRA-CNF	Reverse primer to amplify pULTRA backbone to assemble pULTRA:: <i>tac-metY</i> (CUA)
550- <i>metY</i> (CUA)_2-FOR	42	A80	Forward primer for to amplify <i>metY</i> (CUA) gene to assemble pULTRA:: <i>tac-metY</i> (CUA)
551- <i>metY</i> (CUA)_2-REV	47	A80	Reverse primer to amplify <i>metY</i> (CUA) gene to assemble pULTRA:: <i>tac-metY</i> (CUA)
552-Amber_screen-FOR [^]	22	Amber initiator plasmids	Forward primers for screening assembled amber initiator plasmids [#]
553-Amber_screen-REV [^]	18	Amber initiator plasmids	Forward primers for screening and sequencing of assembled amber initiator plasmids [#]

[#]PCR with these primers gave a 1.2 kb product for pULTRA::*metYp1p2-metY*(CUA) and 0.7 kb product for pULTRA::*tac-metY*(CUA).

Table S2: Normalized fluorescence expression in *E. coli* C321.ΔA.exp

Condition	Amber initiator plasmid	Reporter Plasmid	<i>E. coli</i> strain	Normalized fluorescence (Arb.U./OD ₆₀₀)	Standard deviation
Repressed	-	<i>sfGFP</i> (GCC)	<i>E. coli</i> C321.ΔA.exp	50.3	15.2
Induced	-	<i>sfGFP</i> (GCC)	<i>E. coli</i> C321.ΔA.exp	62.7	13.9
Repressed	-	<i>sfGFP</i> (UAG)	<i>E. coli</i> C321.ΔA.exp	42.9	6.5
Induced	-	<i>sfGFP</i> (UAG)	<i>E. coli</i> C321.ΔA.exp	58.2	13.7
Repressed	-	<i>sfGFP</i> (AUG)	<i>E. coli</i> C321.ΔA.exp	7143.0	817.3
Induced	-	<i>sfGFP</i> (AUG)	<i>E. coli</i> C321.ΔA.exp	111793.6	4329.3
Repressed	pULTRA: <i>metYp1p2-metY</i> (CUA)	<i>sfGFP</i> (GCC)	<i>E. coli</i> C321.ΔA.exp	62.5	17.2
Induced	pULTRA: <i>metYp1p2-metY</i> (CUA)	<i>sfGFP</i> (GCC)	<i>E. coli</i> C321.ΔA.exp	52.2	11.9
Repressed	pULTRA: <i>metYp1p2-metY</i> (CUA)	<i>sfGFP</i> (UAG)	<i>E. coli</i> C321.ΔA.exp	429.3	243.3
Induced	pULTRA: <i>metYp1p2-metY</i> (CUA)	<i>sfGFP</i> (UAG)	<i>E. coli</i> C321.ΔA.exp	240.4	59.6
Repressed	pULTRA: <i>metYp1p2-metY</i> (CUA)	<i>sfGFP</i> (AUG)	<i>E. coli</i> C321.ΔA.exp	7854.3	247.8
Induced	pULTRA: <i>metYp1p2-metY</i> (CUA)	<i>sfGFP</i> (AUG)	<i>E. coli</i> C321.ΔA.exp	8083.3	528.8
Repressed	pULTRA: <i>tac-metY</i> (CUA)	<i>sfGFP</i> (GCC)	<i>E. coli</i> C321.ΔA.exp	48.7	17.5
Induced	pULTRA: <i>tac-metY</i> (CUA)	<i>sfGFP</i> (GCC)	<i>E. coli</i> C321.ΔA.exp	43.7	10.3
Repressed	pULTRA: <i>tac-metY</i> (CUA)	<i>sfGFP</i> (UAG)	<i>E. coli</i> C321.ΔA.exp	128.9	22.8
Induced	pULTRA: <i>tac-metY</i> (CUA)	<i>sfGFP</i> (UAG)	<i>E. coli</i> C321.ΔA.exp	10503.4	2518.1
Repressed	pULTRA: <i>tac-metY</i> (CUA)	<i>sfGFP</i> (AUG)	<i>E. coli</i> C321.ΔA.exp	45202.1	3924.9
Induced	pULTRA: <i>tac-metY</i> (CUA)	<i>sfGFP</i> (AUG)	<i>E. coli</i> C321.ΔA.exp	112895.6	7786.1

Table S2: Normalized fluorescence expression in *E. coli* BL21, *E. coli* C122, *E. coli* W and *E. coli* Crooks.

Condition	Amber initiator plasmid	Reporter Plasmid	<i>E. coli</i> strain	Average Normalized fluorescence (Arb.U./OD ₆₀₀)	Standard deviation
Repressed	pULTRA: <i>tac-metY</i> (CUA)	<i>sfGFP</i> (GCC)	<i>E. coli</i> BL21	1379.3	100.4
Induced	pULTRA: <i>tac-metY</i> (CUA)	<i>sfGFP</i> (GCC)	<i>E. coli</i> BL21	79265.4	1138.7
Repressed	pULTRA: <i>tac-metY</i> (CUA)	<i>sfGFP</i> (UAG)	<i>E. coli</i> BL21	86.9	13.4
Induced	pULTRA: <i>tac-metY</i> (CUA)	<i>sfGFP</i> (UAG)	<i>E. coli</i> BL21	7544.6	716.5
Repressed	pULTRA: <i>tac-metY</i> (CUA)	<i>sfGFP</i> (AUG)	<i>E. coli</i> BL21	61.2	12.0
Induced	pULTRA: <i>tac-metY</i> (CUA)	<i>sfGFP</i> (AUG)	<i>E. coli</i> BL21	44.0	5.5
Repressed	pULTRA: <i>tac-metY</i> (CUA)	<i>sfGFP</i> (GCC)	<i>E. coli</i> C122	3047.0	273.5
Induced	pULTRA: <i>tac-metY</i> (CUA)	<i>sfGFP</i> (GCC)	<i>E. coli</i> C122	80276.3	1339.4
Repressed	pULTRA: <i>tac-metY</i> (CUA)	<i>sfGFP</i> (UAG)	<i>E. coli</i> C122	52.0	7.6
Induced	pULTRA: <i>tac-metY</i> (CUA)	<i>sfGFP</i> (UAG)	<i>E. coli</i> C122	11322.7	812.6
Repressed	pULTRA: <i>tac-metY</i> (CUA)	<i>sfGFP</i> (AUG)	<i>E. coli</i> C122	35.2	3.4
Induced	pULTRA: <i>tac-metY</i> (CUA)	<i>sfGFP</i> (AUG)	<i>E. coli</i> C122	51.9	4.5
Repressed	pULTRA: <i>tac-metY</i> (CUA)	<i>sfGFP</i> (GCC)	<i>E. coli</i> W	2582.9	638.2
Induced	pULTRA: <i>tac-metY</i> (CUA)	<i>sfGFP</i> (GCC)	<i>E. coli</i> W	49017.6	3640.7
Repressed	pULTRA: <i>tac-metY</i> (CUA)	<i>sfGFP</i> (UAG)	<i>E. coli</i> W	46.8	1.5
Induced	pULTRA: <i>tac-metY</i> (CUA)	<i>sfGFP</i> (UAG)	<i>E. coli</i> W	2928.7	166.6
Repressed	pULTRA: <i>tac-metY</i> (CUA)	<i>sfGFP</i> (AUG)	<i>E. coli</i> W	36.5	1.5
Induced	pULTRA: <i>tac-metY</i> (CUA)	<i>sfGFP</i> (AUG)	<i>E. coli</i> W	55.0	10.3
Repressed	pULTRA: <i>tac-metY</i> (CUA)	<i>sfGFP</i> (GCC)	<i>E. coli</i> Crooks	1728.7	58.1
Induced	pULTRA: <i>tac-metY</i> (CUA)	<i>sfGFP</i> (GCC)	<i>E. coli</i> Crooks	60601.7	379.8
Repressed	pULTRA: <i>tac-metY</i> (CUA)	<i>sfGFP</i> (UAG)	<i>E. coli</i> Crooks	44.1	1.7
Induced	pULTRA: <i>tac-metY</i> (CUA)	<i>sfGFP</i> (UAG)	<i>E. coli</i> Crooks	1700.7	2.7
Repressed	pULTRA: <i>tac-metY</i> (CUA)	<i>sfGFP</i> (AUG)	<i>E. coli</i> Crooks	44.7	4.2
Induced	pULTRA: <i>tac-metY</i> (CUA)	<i>sfGFP</i> (AUG)	<i>E. coli</i> Crooks	58.6	3.8

Table S3: Average t_{double} and maximal OD₆₀₀ for fitness analysis scatter plot.

Strain	Amber initiator plasmid	Average t_{double} (mins)	Standard deviation	Average max. OD ₆₀₀ (A.U.)	Standard deviation
<i>E. coli</i> C321.ΔA.exp	-	51.5	2	0.58	0.02
<i>E. coli</i> BL21	-	42.4	9	0.46	0.02
<i>E. coli</i> C122	-	67.4	17	0.54	0.03
<i>E. coli</i> W	-	146.4	11	0.58	0.02
<i>E. coli</i> Crooks	-	159.1	13	0.76	0.07
<i>E. coli</i> C321.ΔA.exp	pULTRA:: <i>tac-metY</i> (CUA)	60.4	9	0.57	0.02
<i>E. coli</i> BL21	pULTRA:: <i>tac-metY</i> (CUA)	47.7	10	0.55	0.02
<i>E. coli</i> C122	pULTRA:: <i>tac-metY</i> (CUA)	64.1	2	0.73	0.01
<i>E. coli</i> W	pULTRA:: <i>tac-metY</i> (CUA)	119.9	3	0.73	0.0
<i>E. coli</i> Crooks	pULTRA:: <i>tac-metY</i> (CUA)	103.1	10	1.01	0.06

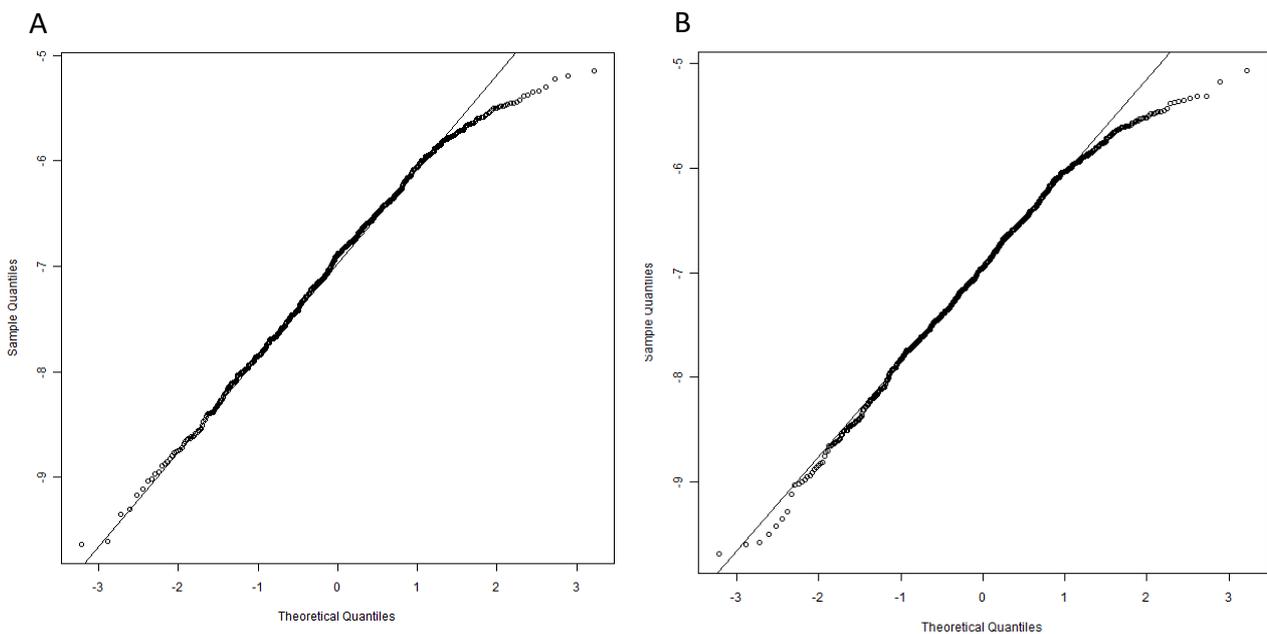


Figure S1: Normalized Q-Q plots of proteins with differential abundances in three replicates of *E. coli* C321.ΔA.exp (A) control without tRNA_{CUA}^{fmet2} (B) expressing tRNA_{CUA}^{fmet2} from pULTRA::*tac-metY*(CUA) plasmids.