Characterising N-terminal amino acid incorporation and protein stability using a mutant i-tRNA^{AAC}

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

Dominic Scopelliti

Supervisor: **Dr. Paul R. Jaschke**

Department of Molecular Sciences

Faculty of Science and Engineering

Macquarie University

24th October 2019

Statement of Originality

This thesis entitled **"Characterising N-terminal amino acid incorporation and protein stability using a mutant i-tRNA^{AAC}"** is representative of the research study conducted between January 2019 and October 2019 for the completion of Master of Research degree in the Molecular Science department 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 Molecular Sciences and has not been submitted for qualification or assessment to any other institution.

Sincerely,

Dominic Scopelliti

Acknowledgements

This final year in the Master of Research program has been a life changing experience that I will hold close to my heart and truly cherish. The program challenged me and forced me out of my comfort zone and really helped me grow into a better researcher and into a better person. None of this journey would have been possible for me if it wasn't for the guidance, support, and mentorship of my supervisor **Dr. Paul R. Jaschke** for which I am forever grateful. His work ethic and unwavering dedication to science has been a massive inspiration to me and it has sincerely been a pleasure learning and working alongside with him.

I would also like to thank the Master of Research advisor **A/Prof Louise Brown**, the members of the Master of Research committee including **A/Prof Bridget Mabbutt**, **Dr. Yuling Wang**, **Dr. Alf Garcia-Bennett**, **Prof. Paul Haynes**, and the thesis convener **Morten Thaysen-Andersen**. The structure and constant support that this remarkable team have provided throughout the year is the reason I was able to conduct this work.

I would like to especially thank the members of the Jaschke lab, **Andras Hutvagner, Bradley Wright, Dominic Logel** and **Russel Vincent**. They were all so incredibly welcoming to me at the beginning of the year and have also continuously supported me. They were always more than willing to sit down with me and answer all of my questions and help me get through all of my troubles. Without them, none of this journey would have been possible. I would also like to thank the members of the Yeast 2.0 group with whom I shared a lab space with and the members of APAF for constantly providing me with guidance within the lab.

Finally, I would like to thank the loved ones in my life. My mother **Gloria Scopelliti**, my father **Orazio Scopelliti**, and my brother **Salvatore Scopelliti** have all supported and encouraged me throughout the entire year, making me feel like I really can achieve my dreams. And lastly, I would like to thank my girlfriend **Carolina Farias** for always believing in me and lending an ear to me in the hardest of times. Without any of you I wouldn't be where I am today.

TABLE OF CONTENTS

Abstract

Synthetic biology is a rapidly evolving field that harnesses principles derived from computational and biomolecular science to create novel biological systems. However, current hurdles such as reduced circuit control and efficiency prevent its full potential being realised for applications in industry, medicine, and research. One prospective solution to circumvent these issues is the idea of developing an orthogonal central dogma of biology. Orthogonal translation initiation from a non-canonical start codon is possible by simultaneously introducing a non-canonical start codon and a mutant initiator tRNA with a complimentary anticodon into cells, but little is known about which amino acids are incorporated into translated proteins. In this thesis, I explore the initiation fidelity of a mutant initiator tRNA with an AAC anticodon which specifically initiates from a non-canonical GUU start codon. Proteomic analysis shows that the mutant initiator tRNA substitutes valine for methionine as the first amino acid in reporter proteins, resulting in improved protein stability. This study serves as a roadmap for the measurement of other non-canonical initiator tRNA activities and the development of a system allowing for improved control of protein levels *in vivo*, bringing synthetic biology closer to achieving an orthogonal central dogma.

Chapter 1: Introduction

1.1 Applications and limitations of synthetic biology

Synthetic biology is a discipline of molecular science that has emerged as an amalgamation of molecular science, genetic engineering and computational science with the overarching goal of constructing novel biological systems or of modifying existing ones. Synthetic biology achieves this goal through the development and implementation of underlying principles that have been adapted from computational science and engineering. These principles include: masking the complex nature of the genetic code through devised interchangeable parts or circuits with defined functions, the ability to insert and remove these circuits without any adverse reactions to the host organism, and to implement these systems in such a fashion whereby the inserted circuit operates independently from the native system (Endy, 2005).

Despite being a relatively new field, synthetic biology is progressing rapidly with numerous techniques and systems that are already being used in bioremediation efforts, development of pharmaceutical agents, and the production of biofuels and biomaterials (Chang and Keasling, 2006, Schmidt, 2012, Paddon and Keasling, 2014, Goody et al., 2002)Thus, synthetic biology may present the necessary tools to reinvent entire industries, improve therapeutics, and the way we conduct research.

However, there are a number of hurdles that need to be overcome in order for synthetic biology to reach its true potential and become further applicable to the aforementioned fields. In its current state, products of synthetic biology often lack the efficacy and yield that is required to make them a feasible alternative to current methods used in the large-scale production of goods. For example, recent work has refactored a nitrogen fixation gene cluster from *Klebsiella oxytoca* for conversion of nitrogen to ammonia, which may hold potential for future use in agriculture, but it currently lacks efficiency for large-scale application (Temme et al., 2012). Furthermore, introducing synthetic circuitry places a burden on host cells, as they use the host cell's machinery and resources for expression, ultimately reducing cellular host viability and increases mutation rates (Andrianantoandro et al., 2006, Wu et al., 2016, Borkowski et al., 2016).

A major issue being tackled at the moment by synthetic biologists is reducing the cross talk between the native DNA within the host and exogenous genetic circuits as this holds the potential to address all of the aforementioned issues. The concept of developing an orthogonal central dogma is currently at the forefront of reducing this cross talk and holds the potential to prevent introduced synthetic circuitry from impeding the native processes of the host cell and vice versa (Liu et al., 2018).

1.2 The Central Dogma

In molecular biology, the central dogma is described as the flow of information from DNA to RNA and then through to protein expression (Crick, 1970). This information is most stably encoded in an organism's DNA and is present across all domains of life. The genetic code is the language of the central dogma and is composed of the four nucleotide bases, adenine, thymine, cytosine, and guanine when in its DNA form. When transcribed to RNA, thymine is substituted to uracil. These bases exist as triplicates known as codons and can form 64 different combinations that code for the 20 amino acids. There is however redundancy in the genetic code as 64 codons encode for only 20 amino acids, implying that some codons encode for the same amino acid. This occurrence became known as codon degeneracy and is thought to mitigate the effects of point mutations, as changing one base pair in the codon will likely not change the amino acid incorporated into a protein (Lagerkvist, 1978). These 20 amino acids can be assembled in an incredibly vast number of combinations which produce proteins, the functional units of life. The conservation and flow of this information through the central dogma is achieved through the processes of DNA replication, transcription, and translation (Crick, 1970)

DNA replication is the biological process in which one DNA molecule is replicated into a complimentary strand, a process that is essential for the propagation of genetic material through to progeny and thus, is highly conserved. DNA exists as a helical structure and is first unwound at an origin of replication by DNA helicase, forming a replication fork where both strands are used as templates for replication. The separated strands are then stabilised by single stranded binding proteins to prevent the complimentary strands from reannealing. Once separated and stabilised, DNA polymerase binds with one of the template strands called the leading strand and reads it in a 5' to 3' direction towards the growing replication fork and begins synthesising the new

complimentary strand of DNA. The other strand is the lagging strand and is synthesised away from the replication fork in short fragments called Okazaki fragments which are ultimately joined together by DNA ligase (Ogawa and Okazaki, 1980)

Transcription is defined as the process of producing an RNA copy from a DNA template (Browning and Busby, 2004). During transcription, the enzyme RNA polymerase reads the DNA template and produces a primary RNA transcript that can be present in many different forms, including messenger RNA (mRNA), transfer RNA (tRNA), transfer-messenger RNA hybrids (tmRNA), ribosomal RNA (rRNA), and small nuclear RNA (snRNA) (Sharp, 2009). Each of these RNA copies perform different and specific functions that are crucial to the efficient functioning of cells.

Following transcription, the final process in the central dogma of biology, translation, commences. During translation, an mRNA transcript is read and decoded by the ribosome in order to produce a protein using individual amino acid building blocks. The primary role of tRNA molecules is to read in-frame codons present on the mRNA template and bridge the RNA and protein codes by incorporating amino acids to a growing polypeptide chain. This process continues until a stop signal in the form of a stop codon is reached. Here, the entire complex is dissociated, releasing the completed protein (Gualerzi and Pon, 1990).

1.3 An Orthogonal Central Dogma

1.3.1 The development of orthogonal systems

The proposal of developing an orthogonal central dogma came about to resolve one of the most fundamental issues currently facing synthetic biology, that is, how to decouple a cell's housekeeping functions from expressing synthetic circuits. In this context, orthogonality refers to systems that exist within the same space and share the same 'hardware' whilst they operate in a completely segregated fashion. This concept has been described by using the analogy of having a virtual machine installed onto a computer (Liu et al., 2018). Both the virtual machine and original operating system function as intended using the same hardware, however there is limited interaction between the two systems themselves. Numerous research efforts have already devised synthetic designs that aim to incorporate orthogonality into cellular systems (Figure 1). These efforts have been directed towards all major aspects of the central dogma including, DNA replication, transcription, and translation (Rackham and Chin, 2005, Ravikumar et al., 2014, Tabor and

Richardson, 1985). In addition, there has also been extensive work conducted on altering the way in which the genetic code is read, the incorporation of non-standard amino acids, as well as expanding the genetic code itself using synthetic nucleotide bases to further improve orthogonality (Wang et al., 2014, Hoshika et al., 2019).

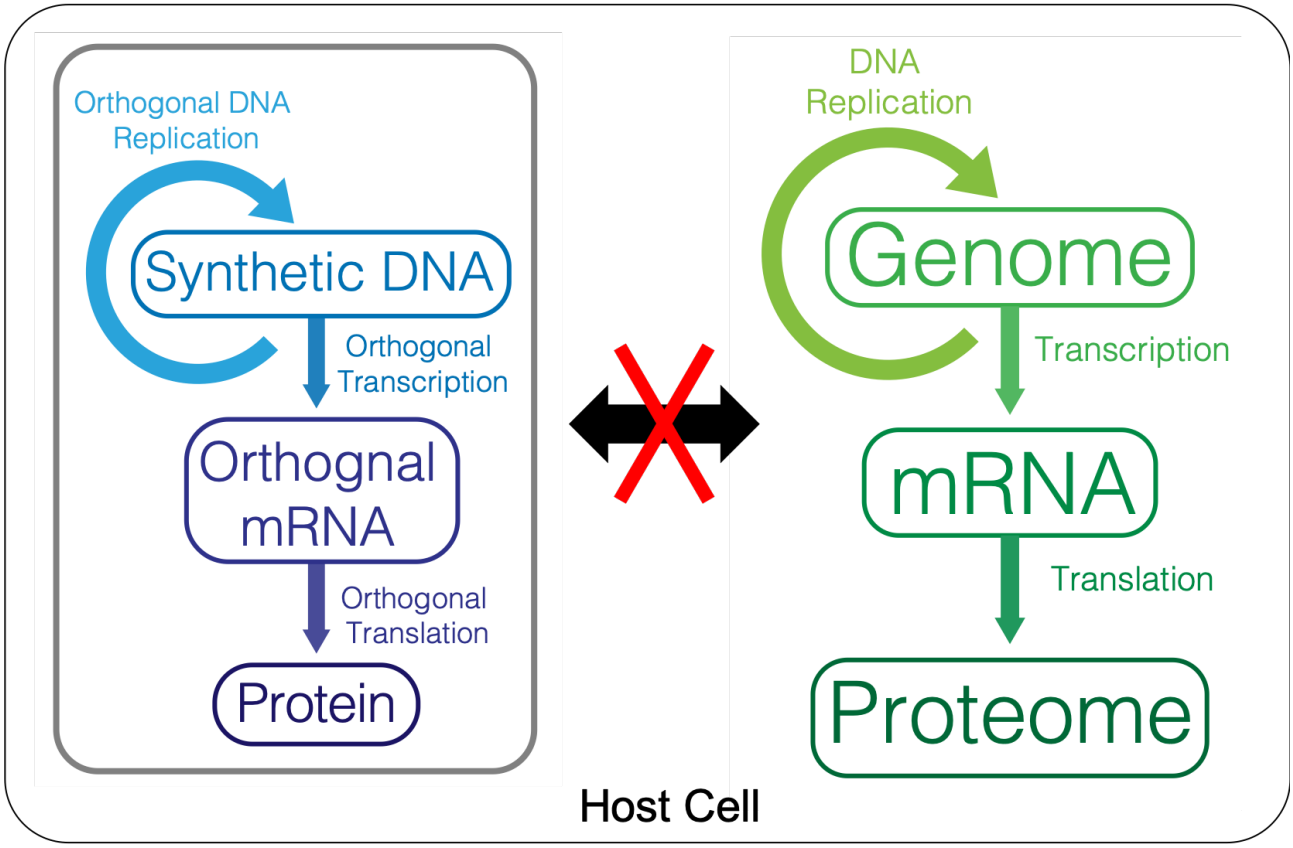


Figure 1. An illustration depicting the concept behind an orthogonal central dogma co-existing within the same host cell as the native central dogma. The phases of DNA replication, transcription and translation are included. Image represents that there is no cross talk between the two systems.

1.3.2 Orthogonal genetic code

Inspiration behind the concept of introducing orthogonality into the central dogma arose from natural occurrences observed in nature. It was initially believed that the genetic code was universal and that any changes in this code would be of great detriment, if not lethal to an organism. However, a study suggested that this was not always the case as changes to the genetic code were identified in Human mitochondrial DNA (Barrell et al., 1979). Ordinarily the AUG codon encoded for the amino acid methionine, however in this

study it was identified that the AUA codon encoded for methionine and not its canonical amino acid isoleucine. They also suggested that Opal stop codons (UGA) encoded for tryptophan and did not signal for termination. Another instance in which this was the case was observed in echinoderm mitochondria where the codon, AAA, that normally encodes for lysine, has been found to be translated as asparagine (Tomita et al., 1999). It has been proposed that these deviations from the standard genetic code or codon reassignments are instances that have risen due to environmental pressures on these specific codons or are due to tRNA mutations (Osawa and Jukes, 1989, Schultz and Yarus, 1996). These deviations from the standard genetic code has led to attempts in achieving non-natural codon reassignment. The degeneracy of the genetic code has been in the focus of synthetic biologists to implement non-natural codon reassignment as it may be exploited to free up codons for other purposes, removing the inherent restrictions of the natural genetic code.

A previous study have managed to design and test an orthogonal tRNA that incorporates the synthetic amino acid *O*-methyl-L-tyrosine using this codon degeneracy (Wang et al., 2001). This feat was achieved by designing a tRNA which recognised the amber stop codon (UAG) and was loaded with *O*-methyl-L-tyrosine using a repurposed synthetic tyrosine-tRNA synthetase. Although this foundational work provided novel insights into codon reassignment, it lacked efficiency due to the tRNA competing with a release factor specific for UAG stop codons. The inherent issue of competition for these amber stop codons was later circumvented by developing an *E. coli* strain that lacked the competing release factor but was still viable through other genetic modifications (Mukai et al., 2010). This work was crucial as it supported the notion that the genetic code was more flexible than previously thought. Codon reassignment later progressed by creating a mutant *E. coli* strain that had a recoded genome such that it could only read 57 of the 64 codons (Ostrov et al., 2016). They designed and assembled an *E. coli* genome that replaced approximately 62 000 instances of degenerate codons with codons that encoded for the same amino acid, making seven codons available for orthogonal uses such as the incorporation of non-standard amino acids.

Concerns regarding biocontainment of engineered organisms may also be resolved through the use of genomes that use an altered genetic code. A strain of *E. coli* has been developed in which it has had all of its UAG stop codons mutated to the synonymous UAA stop codon and deleted release factor 1 which normally aids in terminating translation (Lajoie et al., 2013). This reassignment impaired horizontal gene transfer by hindering the

propagation of this genetic material as the receiving organisms lacked the ability to efficiently express recoded genes (Ma and Isaacs, 2016). Preventing horizontal gene transfer in engineered organisms may provide a suitable means of containing synthetic circuits within intended hosts, protecting native populations. Another approach that has been explored for achieving biocontainment is the introduction of an altered genetic code that confers non-standard amino acid dependence to organisms through redesigning of essential enzymes (Mandell et al., 2015). Producing organisms that are metabolically dependent on non-standard amino acids prevents the use of compounds present in the environment as well as the genetic escape through horizontal gene transfer or random mutations.

1.3.3 Orthogonality in DNA replication

In recent years, many advances have strived towards introducing orthogonality into the central dogma of molecular biology. Orthogonality was introduced into DNA replication through the development of an orthogonal DNA polymerase that replicated specific sections of DNA in *Saccharomyces cerevisiae* (Ravikumar et al., 2014). This developed orthogonal DNA plasmid-DNA polymerase system employs an altered DNA polymerase designed to exclusively identify and replicate the associated plasmid. The DNA polymerase in this system was designed such that it was error prone, allowing the plasmid partner in these hosts to be subjected to mutations and thus, evolution.

Expanding the genetic code through the engineering of unnatural nucleotides has emerged as another avenue of introducing orthogonality. This field however has faced many limitations such as maintaining cellular viability, availability of such unnatural nucleotides within the host, and the ability of host machinery to recognise and incorporate these unnatural nucleotides into a stable DNA sequence. Recently these limitations have been addressed and the first organism that is able to replicate and propagate an expanded genetic code has been developed (Malyshev et al., 2014).

1.3.4 Orthogonality in Transcription

Possessing a system of orthogonal transcription has proven to be beneficial to synthetic biologists as there is a reduced level of competition for resources and machinery, thus benefiting the host. Orthogonal design in transcription is a concept that has been explored for a relatively long time. In the mid-1980s, a system utilising a T7 RNA

polymerase was developed and was one of the first instances of orthogonal transcription (Tabor and Richardson, 1985). This was possible due to the specific nature of the T7 RNA polymerase to exclusively recognise T7 promoters. When the system was developed and introduced into an *E. coli* host, the T7 polymerase could only recognise its associated T7 promoter regions and not the native *E. coli* promoters, resulting in the exclusive transcription of genes under the control of a T7 promoter. This system however, is not totally orthogonal as it is still reliant on host nucleotides and energy sources.

Recent technological advances have opened doors and allowed for new strategies to be implemented towards developing orthogonal systems. One such technological advancement was the development of CRISPR-Cas9 which enabled a new method of genome engineering (Jinek et al., 2012, Cong et al., 2013). Using this technology, a transactivator was created that is directed to specific DNA sequences using guide RNA molecules, allowing for targeted expression of endogenous genes in human cells (Perez-Pinera et al., 2013).

More recently, the genetic code was expanded further in the form of Hachimoji DNA. Hachimoji DNA is composed of two new purine analogues (P and B) and two new pyrimidine analogues (Z and S) that would bind with one another as P:Z and B:S pairs (Hoshika *et al.*, 2019). These new nucleotide bases were structurally designed such that they did not disrupt the DNA helix, were sufficiently and predictably thermostable, and supported the natural course of evolution. It was also demonstrated in this study that the Hachimoji DNA could be employed to produce a fully functional fluorescence aptamer, indicating that Hachimoji DNA was successfully transcribed into its RNA form.

1.3.5 Orthogonal Translation

Translation of RNA to proteins is another front that has been explored to introduce orthogonal design. The development of an orthogonal ribosome and mRNA pair in *E. coli* is one example of this work (Rackham and Chin, 2005). This orthogonality was achieved by designing custom sequences in the 16S rRNA portion of the ribosome and the Shine-Dalgarno sequence upstream of the promoter in mRNA. The orthogonal binding of the modified 16S rRNA to the modified Shine-Dalgarno sequence resulted in the translation of mRNA in parallel to native translation. Orthogonal ribosomes were subjected to further work where a rationally designed RNA staple was developed, linking specific and orthogonal ribosomal subunits and preventing them from associating with native ribosomal

subunits present in the host (Fried et al., 2015). A novel orthogonal translation mechanism that utilises the elongator amber tRNA and evolved tRNAs that possess the ability to decode quadruplet codons (codons composed of four nucleotide bases) has recently emerged. These tRNA molecules were then used to successfully incorporate a multitude of differing non-standard amino acids such as tetrazines, azides, alkenes, and alkynes, which could be used as reactive handles, into a growing polypeptide chains with superior efficiency to previous techniques (Wang *et al.*, 2014).

1.3.6 Orthogonal tRNA

tRNA molecules have also been the subject of genetic engineering efforts in order to introduce orthogonality into translation and to incorporate non-natural amino acids into proteins (Wang et al., 2007, Chatterjee et al., 2012, Wang et al., 2014, Liu et al., 2018). Orthogonal tRNA molecules working in conjunction with mutated ribosomes and aminoacyl-RNA synthetases (aaRS) are essential if a truly orthogonal central dogma and expanded genetic code is to be developed. Once transcribed, tRNA molecules undergo numerous modifications, one of which is the charging of an amino acid in a process known as aminoacylation. This process is carried out by the aaRS enzyme and has been in the sight of synthetic biologists in recent years. tRNA-aaRS pairs have been designed that are able to co-exist with native machinery and are compatible with existing orthogonal systems (Neumann et al., 2010). In addition to its orthogonality, using this type of technology provides another means for incorporating non-natural amino acids into proteins, further enhancing the reach of synthetic biology.

The initiation of translation can potentially be made orthogonal through the modification of initiator tRNAs (i-tRNA). Pioneering studies in this field successfully modified the anticodon region of an i-tRNA from a CAU to CUA (Varshney and RajBhandary, 1990). The newly designed i-tRNA with a CUA anticodon could recognise the amber stop codon, UAG. Furthermore, they demonstrated that these modified i-tRNAs were able to initiate translation from an amber stop codon at relatively high efficiencies (Varshney and RajBhandary, 1990). This study may have laid the foundation required to devise a new method of introducing orthogonality into translation through the use of non-canonical start codons (Hecht et al., 2017, Vincent, 2017) However, for this method to be fully realised, more work needs to be done. For example, the identity of which amino acid is loaded onto the amber initiator tRNA was shown to be only methionine when the

reporter was expressed from a low-copy vector in *E. coli* (Vincent et al., 2019) while another showed it to be both methionine and glutamine when the reporter was expressed in *mycobacteria* (Govindan et al., 2018).

1.4 Translation Initiation

1.4.1 The Initiator tRNA

The i-tRNA molecule is chiefly involved in bringing the first amino acid to the peptidyl-site (P-site) of a ribosome at the commencement of translation. In *E. coli*, the first amino acid is normally methionine, however methionine is also used within protein sequences by elongator tRNA. The i-tRNA is distinctive from related elongator tRNAs (e-tRNA) that are involved with the elongation phase of translation. i-tRNAs possesses a unique non-Watson-Crick base pairing between the C1 and A72 nucleotides (Figure 2A) and three sequential G–C base pairs (Figure 2B) in the anticodon stem (Selmer et al., 2006). The three G-C base pairs present in the anticodon loop bind with the 16S rRNA region on the 30S complex and is thought to be a critical for binding of the i-tRNA with the P-site of the ribosome. Additionally, it has been shown that these G-C base pairs orientate the anticodon loop outwards enhancing the binding and interaction between the codon and anticodon (Barraud et al., 2008). Due to these highly conserved regions being composed of G-C pairing and not A-U, it is likely that i-tRNAs require more stabilising factors in comparison to elongator tRNAs.

Unlike methionine bearing e-tRNA, the methionine on i-tRNAs are formylated through the activity of methionyl-tRNA formyltransferase (MTF). The proposed purpose of methionine formylation was that it contributes to initiation factor 2 (IF2) binding to the translation initiation complex. Early studies suggested that mutated acceptor stem regions on i-tRNA resulted in reduced formylation and in turn IF2 binding affinity (Varshney and RajBhandary, 1992). However, later studies had determined that the acceptor stem, more specifically the CAACCA region (Figure 2C), exhibited more than one role. The first role being that it influenced the affinity of IF2 binding alongside the formylated methionine. However, the CAACCA sequence was also found to interact with MTF, ultimately leading to the formylation of methionine on i-tRNAs. This suggests that the CAACCA sequence on the acceptor stem of i-tRNAs is responsible for both the formylation of methionine as well

as increasing the affinity of IF2 binding to the translation initiation complex (Mayer and RajBhandary, 2002).

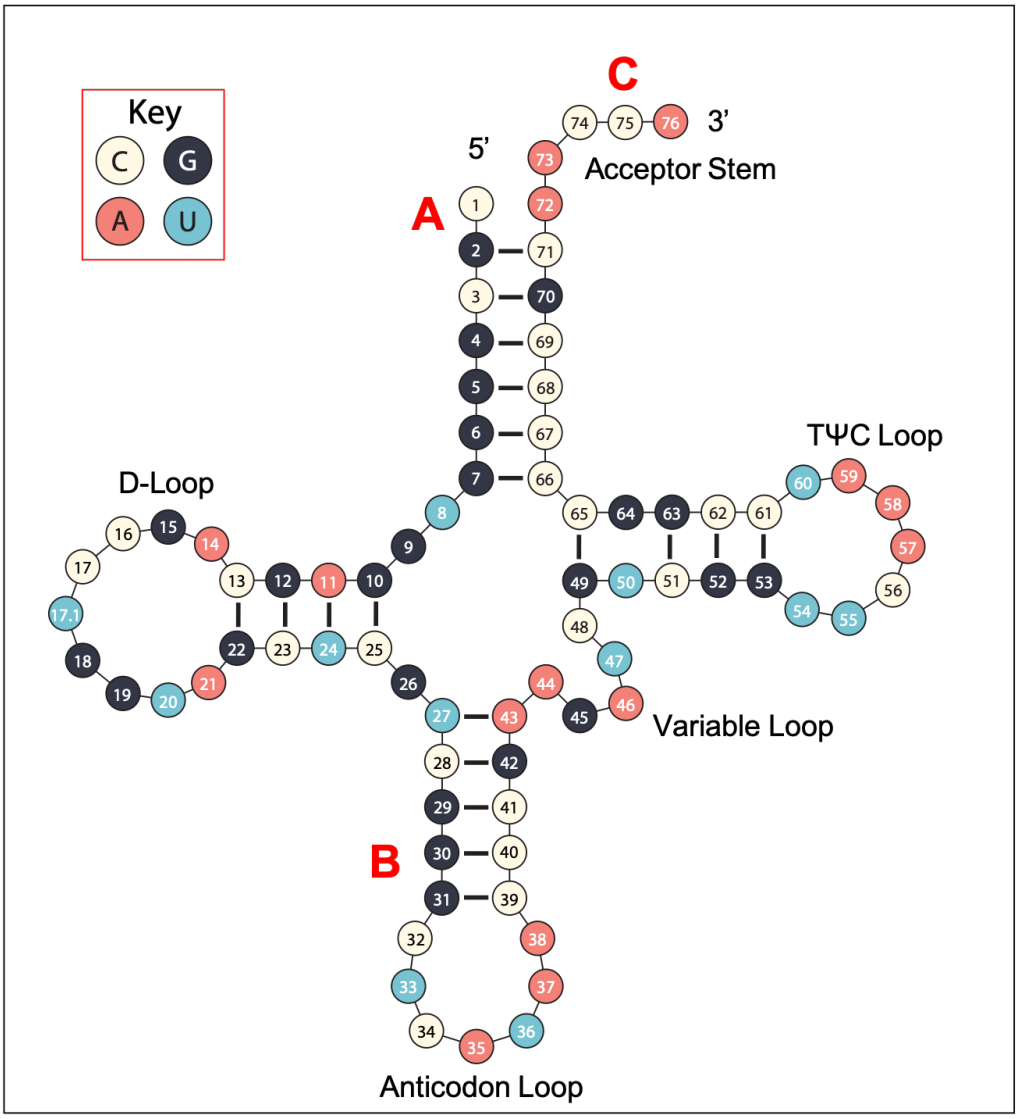


Figure 2. The *E. coli* initiator tRNA. (A) The C1-A72 base pair binding. (B) three highly conserved G-C pairs. (C) The AACCA sequence on the acceptor stem. All these regions are unique to the i-tRNA.

1.4.2 Mechanisms of translation initiation in prokaryotic cells

Translation initiation through the formation of a ribosomal complex is a highly regulated process and is the primary rate limiting step in protein translation (Hauryliuk and Ehrenberg, 2006). In prokaryotes, there are three different translation initiation mechanisms that have been observed to occur in cells concurrently, all of which initiate translation at different rates.

The 30S binding mode is the first and most dominant mechanism for translation initiation where it initiates protein translation through the 30S ribosomal subunit (Figure 3A) (Laursen et al., 2005). In this mechanism, the 30S ribosomal subunit recognises a specific region known as the Shine-Dalgarno sequence on the template mRNA that is normally situated upstream of a start codon. The i-tRNA along with initiation factor proteins interact with the 30S ribosomal subunit to form the 30S initiation complex. Following this the 50S ribosomal subunit attaches to the 30S initiation complex forming the 70S ribosomal complex. Initiation factors are then released, allowing for the elongation phase of translation to commence. Upon reaching a stop codon in the mRNA sequence and subsequent termination of translation, the entire complex dissociates, and the ribosomal subunits are recycled to translate different genes (Laursen et al., 2005).

The second mechanism of translation initiation is the 70S scanning mode (Figure 3B). Once translation has reached a termination codon using this mechanism, the 70S ribosomal complex does not undergo dissociation and continues scanning down the mRNA template until another Shine-Dalgarno sequence and start codon is encountered. Bound i-tRNA and initiation factors aid in the identification of this start codon, once identified these factors are ejected and translation enters the elongation phase (Yamamoto et al., 2016).

Leaderless mRNA initiation is the third and final mechanism of translation initiation and is a rare occurrence relative to the two aforementioned mechanisms (Figure 3C). This mechanism is employed exclusively to a rare set of mRNA templates that lack a Shine-Dalgarno sequence but possess a start codon 5 bases upstream from the beginning of a gene. Translation initiation using this mechanism relies on the formation of a complex consisting of a 30S ribosomal subunit, an i-tRNA, and IF2 (Moll et al., 2002, Udagawa et al., 2004).

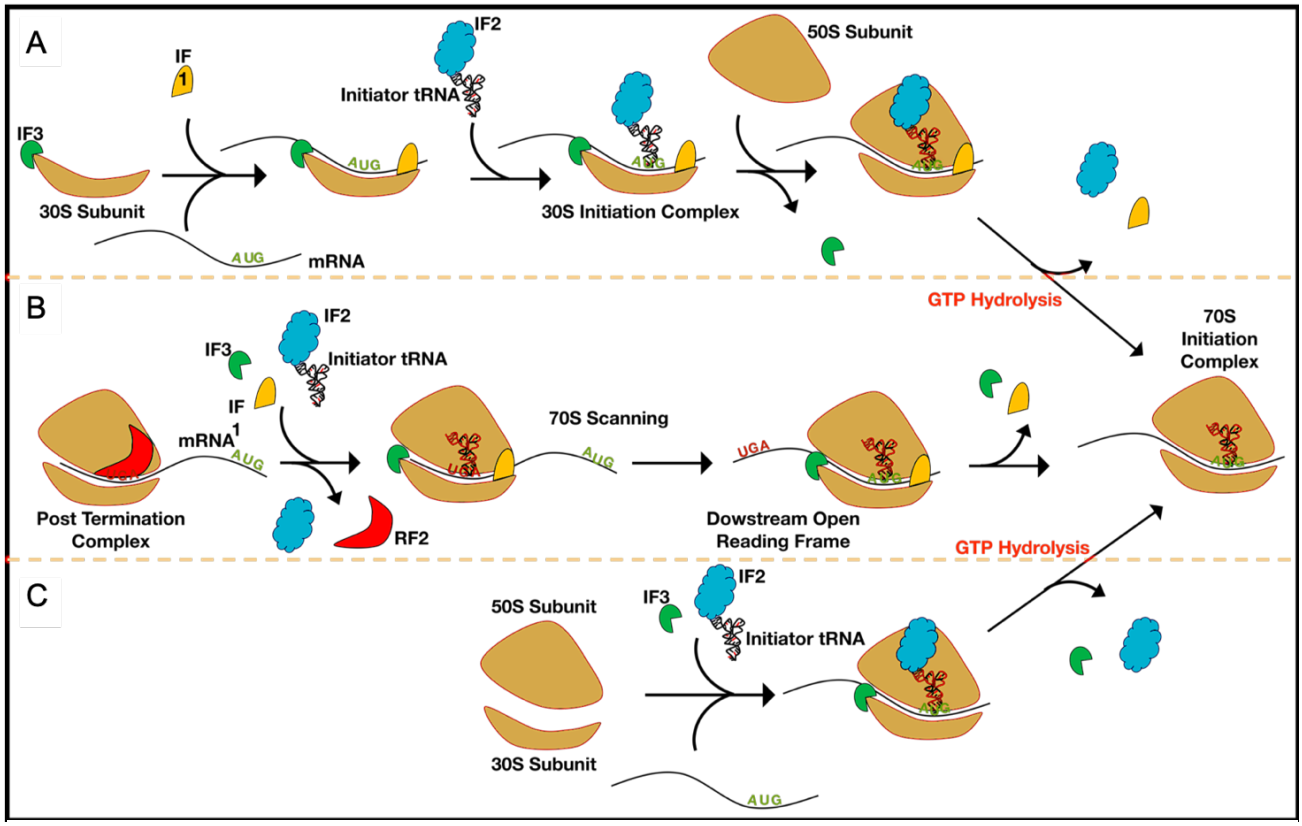


Figure 3. Diagram of the three different mechanisms of translation initiation in prokaryotes. (A) 30S binding mode, IF1, IF2, and IF3 form a 30s initiation complex. (B) 70S scanning mode, where the 70S ribosomal complex continuously reads an mRNA transcript until the next site of initiation is reached. (C) Leaderless mRNA initiation. IF2 and IF3 participate in initiating translation of a gene lacking a Shine-Dalgarno sequence. Image adapted from Hutvagner (2018).

1.4.3 Initiation factors in translation initiation

Initiation factors are accessory proteins that are essential in translation initiation as they assist in the stringent formation of the 30S initiation complex in the correct location. The three initiation factors that participate in translation initiation are initiation factor 1 (IF1), initiation factor 2 (IF2), and initiation factor 3 (IF3). IF1 and IF3 are responsible for maintaining fidelity in the initiation complex whilst the i-tRNA is recruited to the complex by IF2 (Julián et al., 2011).

The smallest initiation factor, IF1 is encoded by the *infA* gene and binds with the A site of a ribosome. IF1 is indispensable as it aids in the formation 30S initiation complex, facilitates 70s complex dissociation, and acts as a fail-safe mechanism, ensuring the correct binding of i-tRNA (Dottavio-Martin et al., 1979). In addition, IF1 acts as an anchorage point for IF2 and IF3, in turn regulating the entire initiation process (Gualerzi and Pon, 1990, Hussain et al., 2016).

IF2 is encoded by the *infB* gene and is selective for i-tRNAs correctly charged with formyl-methionine, ensuring the correct i-tRNA is attached to the 30S ribosomal subunit (Simonetti et al., 2008). This attachment to the ribosome is facilitated through GTPase activity located on the highly conserved C-terminal domain (Caserta et al., 2006).

The primary role of IF3 in translation initiation has been recognised as ensuring the fidelity of codon/anticodon pairing between the i-tRNA and the start codon of the mRNA template (Ayyub et al., 2017). IF3 has also been shown to aid the ribosome in detecting initiation signalling features on mRNA transcripts and through modulating the dissociation of 70S complexes upon the termination of translation, allowing for recycling of 30S ribosomal subunits (Milon et al., 2008).

1.4.4 The importance of the start codon in translation initiation

Start codons are an imperative component of mRNA and are responsible for setting the reading frame and for translation initiation. The canonical start codon is composed of an AUG triplet sequence in most organisms and initiates translation most strongly, although UUG and GUG are also frequently used (Hecht et al., 2017). All three of the strongest start codons share U and G nucleotides in the second and third positions, suggesting a degree of flexibility in terms of translation initiation.

The AUG codon encodes for the amino acid methionine, an amino acid residue which possesses dual functions as both a start codon and an in-sequence sense codon. This function is dependent on the position of the AUG codon relative to the Shine-Dalgarno sequence that is situated seven to nine nucleotides upstream of start codons on mRNA transcripts (Malys, 2012). When the Shine-Dalgarno sequence is situated just upstream from the AUG codon, a formyl methionine is incorporated. In the absence of the Shine-Dalgarno sequence, an un-formylated methionine residue is incorporated into the polypeptide sequence (Lobanov et al., 2010). At times, this dual function of the AUG codon has been shown to be problematic in recombinant protein expression. It has been shown that these issues arise when sequences resembling Shine-Dalgarno sequences are located just upstream of AUG codons located within a gene, resulting in unintended ribosomal binding and subsequent translation initiation (Whitaker et al., 2014).

1.4.5 The codon/anticodon interaction in translation initiation

The interaction between the anticodon of i-tRNAs and start codons of mRNA transcripts is stabilised through Watson-Crick base pairing. The pairing between these two elements is crucial for translation initiation, consequently the process is tightly regulated through the activity of IF3 (Hartz et al., 1990). Through this activity of IF3, the anticodon/codon interaction acts as a check point for stable binding, ergo the interaction should be recognised as a whole and not by its two separate components. Structural studies conducted using Cryo-EM techniques determined that IF3 does not interact with the start codon or anticodon individually but rather causes a conformation change of the 30S initiation complex (Allen et al., 2005). The change in conformation strains the interaction between the anticodon/codon pair, thus this interaction can only be sustained through the energetically favourable binding of cognate start codons and anticodons. Due to these conformational changes, IF3 can be considered to destabilise the pairing of non-matching anticodon/codon pairs and maintain the stringency of translation initiation.

The GUG and UUG near cognate codons that can act as suitable start codons in a minority of cases may possess the stability required to overcome the conformational strain issued by IF3 during translation initiation as the U nucleotide in the second position contributes a large degree of stability to the binding event (Sussman et al., 1996). A pioneering study had successfully quantified the translation rates of a green fluorescent protein (GFP) reporters using all 64 possible start codon combinations with the native i-tRNA (Hecht et al., 2017). Of the 64 combinations, 47 start codons were identified emitting fluorescence above background levels. This suggests that the GFP reporter was being expressed using 47 different codons with varying efficiencies, likely due to near cognate or wobble-base pairing between the anticodon/codon pair. Other binding events between anticodon/codon pairs have been observed in nature and are most likely due to the occurrence of wobble base pairing occurring in the third position of the codon (Figure 4). Although significantly weaker, this type of pairing may still provide a sufficient degree of stability and thus allows translation initiation to occur (Murphy IV and Ramakrishnan, 2004, Cochella and Green, 2004).

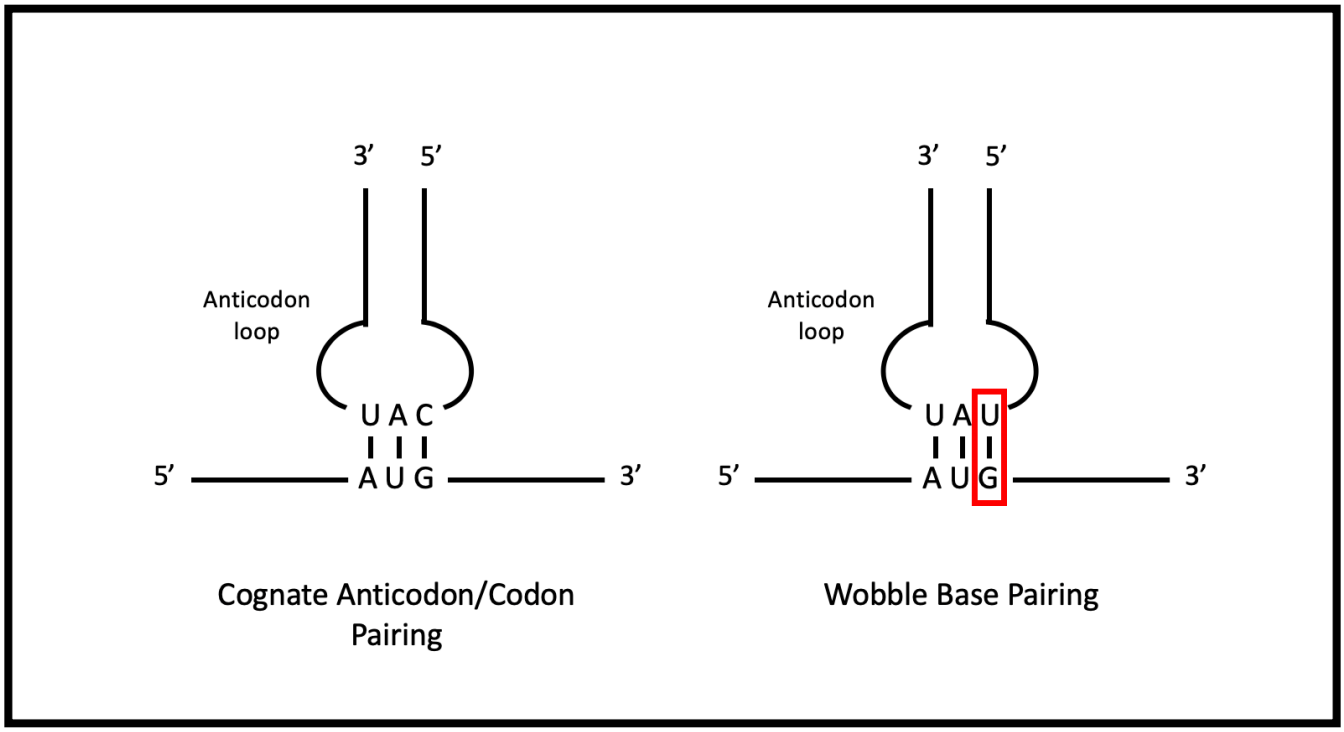


Figure 4. Cognate anticodon/codon pairing between the i-tRNA and the mRNA transcript. Interaction provides stability to resist conformation strains applied by IF3. Wobble base pairing between i-tRNA and mRNA transcript. Nucleotide bases in the third position (boxed in red) undergo wobble base pairing.

1.4.6 Introducing orthogonality to translation initiation in *E. coli*

The notion of initiating translation the use of non-canonical start codons (codons other than AUG, GUG, and UUG) has been long explored (Varshney and RajBhandary, 1990, Mayer et al., 2003, Vincent, 2017, Hutvagner, 2018). The compelling findings prompted the exploration of orthogonal translation initiation through modifications to the anticodon loop of i-tRNAs and cognate start codons (Figure 5). The work conducted by Vincent (2017) and Hutvagner (2018) investigated the feasibility of introducing orthogonality into translation initiation through the mutation of the anticodon present on i-tRNAs. Their findings indicated successful expression of GFP through a mutant i-tRNA/start codon system, suggesting that translation initiation through the use of mutant i-tRNAs with cognate start codons was indeed possible.

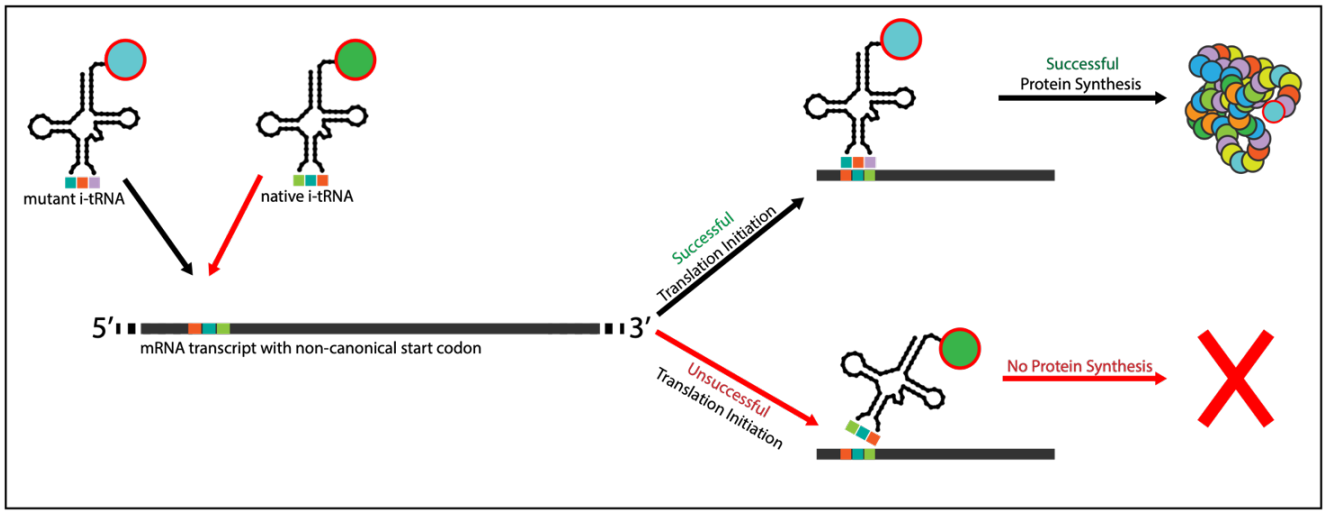


Figure 5. Concept of orthogonal translation initiation. Conceptual illustration of developing a system in which a protein begins with a non-canonical start codon and is co-expressed with a mutant i-tRNA with a cognate anticodon. This system would allow for translation initiation to commence from the mutant tRNA and not the native tRNA.

1.5 Protein Degradation

1.5.1 Possible off-target effects from changing the i-tRNA anticodon

Designing mutant i-tRNA molecules that have altered anticodons may present a plausible avenue to achieving orthogonal translation initiation. However, mutating this region may potentially lead to unintended off-target effects in translation initiation. One such example is changes in the aminoacylation pattern of these i-tRNAs such that formyl methionine is no longer charged onto the mutant i-tRNA. This phenomenon was demonstrated by swapping the canonical CAU anticodon for the CUA anticodon (Schulman and Pelka, 1985). This mutation resulted in the increased charging of glutamine instead of methionine *in vitro*. Later studies further supported this work, demonstrating that altering the anticodon region of i-tRNAs to GAC or GAA, resulted in aminoacylation of the i-tRNA with valine or phenylalanine, respectively *in vivo* (Chattapadhyay et al., 1990). Mis-aminoacylation of these mutant i-tRNA$^{GAC\ or\ GAA}$ is due to the anticodon sequence being recognised by the corresponding aaRS enzymes that charge amino acids on to i-tRNAs (Mayer et al., 2003). Thus, mutating the anticodon sequence of i-tRNAs results in different aminoacylation patterns of i-tRNA. Currently, it is still unclear which amino acid residues are being charged onto mutated i-tRNAs across the possible 64 combinations of anticodons. Therefore, to introduce orthogonality in

translation initiation, these effects must be further explored to ensure that we can predict and control the function of mutated i-tRNA molecules.

1.5.2 The N-end rule degradation pathway

The N-end rule pathway is a highly conserved degradation mechanism that has been observed in both prokaryotic and eukaryotic organisms (Dougan et al., 2010). According to the N-end rule degradation pathway, having certain amino acids, known as N-degrons, on the N terminus of a protein may drastically alter overall protein stability. The N-end rule dictates that the presence of an N-degron on the N-terminus of a protein can recruit protease machinery and subsequently lead to the degradation of that protein (Humbard et al., 2013). The protease responsible for the degradation of N-end rule substrates is the ClpAP protease and is one of five ATP dependent proteases naturally produced in *E. coli* (Dougan et al., 2010). ClpAP is a complex of two proteins, ClpA and ClpP. ClpA is described as a regulatory particle that binds with specific substrates and unfolds proteins tagged for degradation. The unfolded protein is then translocated through an ATPase ring into ClpP where degradation occurs. However, N-end rule substrates are not directly targeted by ClpA, but by the adaptor protein ClpS. ClpS identifies N-degrons on tagged proteins and delivers them to ClpA, which later delivers the substrate to ClpP (Figure 6) (De Donatis et al., 2010, Humbard et al., 2013).

In prokaryotes, these N-degrons can be classified into either primary or secondary destabilising residues. Primary residues include leucine, phenylalanine, tyrosine, and tryptophan and can be directly recognised by ClpS. Secondary residues, lysine and arginine, require further modifications before undergoing degradation. Secondary residues are recognised by amino acyltransferase, an enzyme that facilitates the addition of primary residues leucine or phenylalanine onto the N-terminal end of proteins, allowing for binding with ClpS (Humbard et al., 2013).

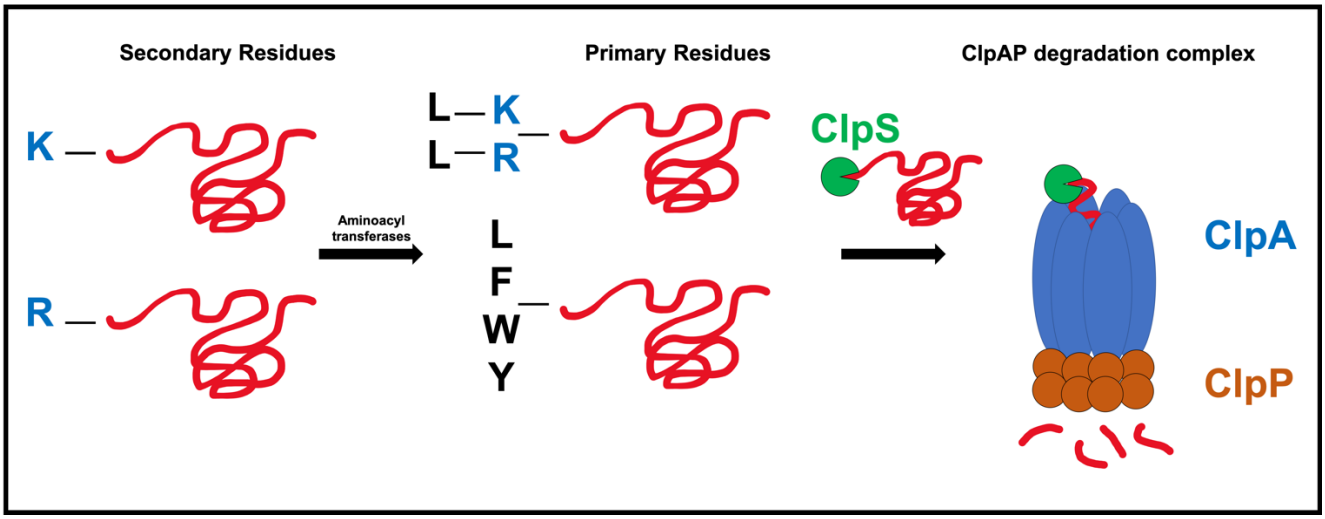


Figure 6. N-end rule degradation pathway. Proteins possessing secondary destabilising residues lysine and arginine undergo the addition of a primary destabilising residue to the N-terminus. The reaction is catalysed by the enzyme aminoacyl transferase. Once proteins containing primary destabilising residues are formed/recognised, degradation of the protein occurs through the activity of adaptor molecule ClpS and protease complex ClpAP.

1.5.3 The N-end rule and orthogonal translation initiation

Several studies have shown that modified anticodon regions of mutant i-tRNAs leads to the incorporation of amino acids other than methionine, such as phenylalanine or glutamine (Schulman and Pelka, 1985, Chattapadhyay et al., 1990, Mayer et al., 2003). This may prove to be problematic in the development of orthogonal translation initiation systems, as amino acids such as phenylalanine are subjected to N-end rule degradation pathways. Expression of proteins through the use of mutant i-tRNAs in orthogonal systems may potentially result in the expression of proteins that are not comparable with native systems in terms of protein stability. On the contrary, increased degradation of proteins may be a beneficial tool to synthetic biologists as it provides a means of controlling intracellular levels of synthetic circuits introduced into hosts. Nonetheless, patterns of amino acid incorporation through the use of mutant i-tRNAs needs to be further investigated in order for the development of a fully controllable and customisable orthogonal translation initiation system.

1.6 Project aims

The overall goal of this thesis project was to identify the amino acid incorporated onto the N-terminus of a reporter protein when using the orthogonal i-tRNA mutant with a mutated anticodon in an *E. coli* host. In addition, I examined the interaction of proteins expressed using the mutated i-tRNA with N-end degradation pathways to identify the effect that using this orthogonal translation initiation system has on protein stability *in vivo*.

Chapter 2: Methods

2.1 Bacterial strain handling, glycerol stocking, making competent cells

2.1.1 Bacterial strains used in this study

In this study, two strains of *Escherichia coli* were used. NEB® Turbo cells were obtained from New England BioLabs (NEB) (NEB, # C29841) and were used as a cloning strain for plasmid assembly and confirmation due to its desirable DNA expression levels and fast growth rate. The second strain used, BL21 (DE3) pLysS, was sourced from NEB (NEB, Catologue #C2527I) and was used as an expression strain. The BL21 (DE3) pLysS strain carries the gene for T7 RNA polymerase that is under the control of the IPTG inducible lacUV5 promoter, allowing for expression of genes under T7 promoter control. Furthermore, this strain is deficient of the Ion protease and OmpT protease to reduce protein degradation via these pathways. It also possesses the pLysS plasmid harbouring chloramphenicol resistance and the T7 lysozyme to reduce basal level expression of gene of interest.

2.1.2 Growth conditions

All the aforementioned *E. coli* strains were grown in lysogeny broth Miller (LBM) broth or on LBM agar plates. Growth media was supplemented with either 100 µg/mL spectinomycin (Sigma-Aldrich, #S4014), 100 µg/mL carbenicillin (Sigma-Aldrich, #C9231), or 25 µg/mL chloramphenicol (Sigma-Aldrich, #C0378). Antibiotics were supplemented to media as required for selection of desired strains.

2.1.3 Glycerol stocking

All *E. coli* strains were glycerol stocked for long-term storage by mixing 900 µL of overnight cultures with 900µl of glycerol (60%) and stored at -80^{o} C. Glycerol stocked strains were recovered by streaking out the stock onto LBM plates with necessary antibiotics and incubated overnight at 37^{o} C.

2.1.4 Preparing competent cells

The *E. coli* strains used in this study were made competent using the *Mix and Go* Transformation and buffer kit (Zymo, #T3001) from Zymo Research. Cells were grown at

37° until an OD of 0.4 - 0.6 was reached and were then made competent using the kit as per the manufacturer's instructions. Competent cells were then stored at -80°C until required for transformation.

2.2 DEseq2 Analysis

2.2.1 Data Acquisition and analysis using DEseq2

Mutant initiator tRNA translation initiation efficiency raw data of 9 different codon/anticodon combinations was supplied by Hutvagner (2018). Analysis and visualisation of this raw data was performed using the DEseq2 (Version 1.24.0) and ggplot2 (Version 3.2.1) packages in Rstudio (Version 1.1.456)

2.3 Plasmid construction

2.3.1 Construction of pET20B:T7-NanoLuciferase/6xhis-tagged reporter plasmid

2.3.1.1 Linearisation of pET20B:T7-sfGFP plasmid

Synthesis of all oligonucleotides and gBlocks used within this study was outsourced to Integrated DNA Technologies (IDT). The pET20B:T7-Nanoluciferase/6xhis-tagged plasmids were designed and constructed using the pET20B:T7-sfGFP expression plasmid obtained from a previous study as a template (Hutvagner, 2018). Polymerase chain reaction (PCR) with Q5® High-Fidelity 2x Master Mix (New England BioLabs, #M0492S) was used to linearise the plasmid and to remove the *sfGFP* gene. Plasmid linearisation was achieved through the use of oligonucleotide primers that were designed using the Primer wizard tool powered by Primer3 in the online tool Benchling (https://benchling.com/). Primers were designed to possess a 20bp overlap with regions flanking the *sfGFP* gene.

2.3.1.2 Dpn1 Digestion and PCR clean up

To ensure removal of the original pET20B:T7-sfGFP plasmid, PCR products were subjected to DPN1 (NEB, #R0176S) digestion at 37°C for 60 minutes and then heat inactivated at 80°C for 20 minutes. PCR products were subjected to PCR clean up using

the GenElute™ PCR Clean-Up kit (Sigma-Aldrich, #NA1020-1KT) following manufacturers protocol. Concentration of PCR cleaned products were measured using the NanoDrop™ 2000 (ThermoFisher Scientific, #ND2000).

2.3.1.3 Gibson assembly of Nanoluciferase variants with linear pET20B backbone

Synthesised gBlocks possessing the gene encoding for Nanoluciferase with either an AUG or GUU start codon and a C-terminal 6xhis-tag were assembled with the linearised pET20B:T7 backbone. On the flanking regions of this gBlock were 30bp overlaps that complement the linear pET20B plasmid to allow for efficient Gibson assembly. A reaction was set up such that there was a 5:1 ratio of gBlock inserts to PCR amplified vector backbone. The reaction consisted of 10 µL of NEBuilder® HiFi DNA Assembly Master Mix (NEB, #E2621L), 30 fmol of the linearised pET20B:T7 vector and 150 fmol of Nanoluciferase gBlock variants. Total volume was brought up to 20 µL using sterile water. The reaction mixture was then incubated for 30 minutes at 50°C.

Following plasmid construction, 5 µL of the plasmid assembly was added to 200 µL of competent NEB turbo *E. coli* cells and was then incubated on ice for 15 minutes. 1 mL of S.O.C media (2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM glucose) was added and cells were incubated at 37°C for 1 hour. Cells were spread plated out onto pre-heated LBM agar plates with Carbenicillin and incubated at 37°C overnight.

2.3.1.4 Colony PCR

Colony PCR was performed on five randomly selected transformants to ensure successful assembly of the pET20B: Nanoluciferase plasmids. Colonies were resuspended in 100 µL of sterile water to be used as a DNA template in PCR and once confirmed, to form a starter culture. Using screening primers and KAPA2G Master mix (Sigma Aldrich, #KK5512) transformant colonies were screened. PCR products were run on a 2% agarose gel in Tris Acetate-EDTA buffer to identify products of the correct size (Approx. 1kb). Successful transformants were used to inoculate 5 mL of LBM with Carbenicillin and grown overnight at 37°C with shaking (250rpm). A proportion of the overnight cultures were glycerol stocked for long term storage while the remainder of the cultures were subjected to plasmid extraction.

2.3.1.5 Plasmid extraction and sequencing

Plasmid extraction was conducted using QIAprep Spin Miniprep kit (Qiagen #27106) according to manufactures instructions. The concentration of extracted plasmids was measured using the NanoDrop™ 2000 (ThermoFisher Scientific, #ND2000) to ensure concentration was above the 100 ng/µL concentration required for Sanger sequencing. Plasmids along with the relevant sequencing primers were sent to Macrogen Inc. Sequencing data was analysed by performing an alignment (MAFFT algorithm) against a hypothetical expected sequence on the online tool Benchling.

2.3.2 Construction of pET20B:T7-sfGFP-ssrA tagged reporter plasmids

To assist with protein stability assays, sfGFP reporter proteins with *ssrA* degradation tags were designed and constructed. The pET20B-T7: sfGFP plasmid was again used as a template for the construction of the pET20B-T7: sfGFP-ssrA plasmid. This plasmid was constructed using the same protocols as described above in the construction of the pET20B:T7-Nanoluciferase/6xhis-tagged plasmids however with some minor changes. These changes include using oligonucleotides that linearised the template plasmid whilst maintaining the *sfGFP* gene and using a gBlock encoding for a ssrA degradation tag, with the flanking regions of the gBlock complementing the new linear plasmid. In addition to the sfGFP variants with AUG and GUU start codons, a third sfGFP variant was included with a GCC start codon to be used as a control as it was shown in the study by Hecht *et al.* (2017) that it produces levels of fluorescence consistent with background fluorescence.

2.4 Strain formation

2.4.1 Sequential double transformation

Strains utilising the orthogonal translation initiation system require to be transformed with two plasmids, one plasmid harbouring the *metY* gene encoding for the initiator tRNA with a mutated anticodon region and another plasmid harbouring the *NanoLuciferase* gene with an altered start codon that is complimentary to that of the mutated initiator tRNA. The Zymo transformation protocol loses efficiency for simultaneous double transformations, and thus a sequential double transformation protocol was employed.

2.4.2 Transformation of pULTRA mutant i-tRNA plasmids

Two forms of the pULTRA plasmid were used in this study. One form of the plasmid (pULTRA: empty) did not possess the *metY* gene and was used as a control. The second form (pULTRA: metY (AAC)) possessed the *metY* gene with an AAC anticodon rather than the native CAU anticodon. Plasmids were obtained from previous studies (Vincent *et al.*, 2019; Hutvagner, 2018). The two pULTRA plasmids were transformed into competent BL21(DE3) pLysS *E. coli* cells and plated out onto LBM plates with chloramphenicol and spectinomycin to select for transformants and grown overnight at 37°C. Glycerol stocks of these cells were made for long-term storage.

2.4.3 Transformation of pET20B: Nanoluciferase plasmids

E. coli cells that were successfully transformed with the pULTRA plasmids were made competent. The two forms of the pET20B: Nanoluciferase plasmids with either a native AUG or mutated GUU start codon previously assembled was transformed into the competent pULTRA cells. Cells were plated out onto LBM plates with chloramphenicol, spectinomycin, and carbenicillin to select for transformants and grown overnight at 37°C.

2.4.4 Transformation of pET20B-T7: sfGFP-ssrA tagged plasmids

The two *E. coli* cell strains possessing the pULTRA (empty) and pULTRA: metY (AAC) plasmids were recovered from glycerol stocks and made competent. These two strains were transformed with one of the three sfGFP-ssrA tagged plasmids previously constructed or untagged sfGFP plasmids. Cells were plated out onto LBM plates with chloramphenicol, spectinomycin, and carbenicillin to select for transformants and grown overnight at 37°C.

2.5 Proteomic analysis

2.5.1 Cell Growth and harvesting

Biological triplicates of BL21(DE3) pLysS strains possessing both the pET20B: Nanoluciferase and the pULTRA plasmids were used to inoculate 5 mL of LBM broth with spectinomycin, chloramphenicol, and carbenicillin and incubated overnight at 37°C with

shaking (250rpm). Following the incubation, 300 µL of these cultures were used to inoculate 30 mL of fresh LBM with the necessary antibiotics in 250 mL Erlenmeyer flasks and incubated for 1 hour at 37°C with shaking (250rpm). The *metY (AAC)* gene on the pULTRA plasmid were then induced using 1 mM IPTG to produce mutant i-tRNAs. Cultures were then returned to the incubator and grown for a further 6 hours to reach the late log growth phase. Cells were harvested by centrifugation at 3500g for 10 minutes at 4°C and supernatant was discarded. Cell pellets were stored overnight at -20°C.

2.5.2 Cell lysis

To lyse cell pellets a master mix was made up (2 mL per pellet) containing: 400 µL CelLytic ™ B Cell Lysis Reagent (Sigma Aldrich, #B7310), 120µl of lysozyme, 80µl of protease inhibitor, and 1 mL of sterile water. 2 mL of Cell lysis master mix was added to each of the pellets and 100 units/mL of benzonase was also added to degrade DNA. Cell pellets were resuspended and left to incubate at room temperature for 20 minutes. The mixture was then centrifuged for 10 minutes at 16 000g and the supernatant containing all the soluble proteins was collected.

2.5.3 Purification of Nanoluciferase

Protein fraction containing the 6xhis-tagged Nanoluciferase proteins were then purified using immobilised metal affinity chromatography. Column was made using 800 µL of HisPur™ Ni-NTA Resin (ThermoFisher scientific, #88221). Equilibration, wash, and elution buffers were made using 300 mM of PBS with 20 mM, 60 mM, and 500 mM concentration of imidazole respectively. Purification was carried according to manufacturer's protocol.

2.5.4 LC-MS sample preparation

In preparation for LC-MS analysis, purified protein samples were first reduced through the addition of 10 mM Dithiothreitol (DTT) and incubated at 60°C for 30 minutes. Reduced proteins were then alkylated with 30 mM iodoacetamide (IAA) and incubated at room temperature for 1 hour in the dark. To quench the excess IAA, 30 mM of DTT was added and the sample was incubated at room temperature for 15 minutes in the dark. Proteins were then precipitated out of solution through the addition of 800 µL of ice cold 100% acetone and incubated overnight at -20°C. Proteins were pelleted through

centrifugation at 16 000g for 10 minutes at 4°C and then washed using a 4:1 wash solution (Acetone: water). Proteins were re-pelleted, and supernatant was discarded. The protein pellet was dissolved in 8 M Urea 50mM Tris-HCl buffer (pH 8.0) and then diluted 5-fold through the addition of Urea 50 mM Tris-HCl buffer (pH 8.0). The protein concentration was determined using the Pierce™ bicinchoninic acid assay (BCA) Protein assay kit (ThermoFisher Scientific, #23225). Bovine serum albumin was used as protein standards in concentrations ranging from 0-2 mg/mL and all samples were prepared as technical duplicates. 30 µg of purified Nanoluciferase was then subjected to a 1:50 enzymatic digestion using 60ng of trypsin and incubated overnight at 37°C. Peptides were acidified and digestion was then stopped through the addition of formic acid to reach a final concentration of 1%. Peptide samples were purified using c18 stage tips that were assembled using Empore™ Octadecyl C18 extraction disks (Empore, #2215) and 100 µL pipette tips according to manufacturer's protocol. Peptides were dried in a speed vac and resuspended to reach 1 µg/µL in 2% acetonitrile (ACN) / 0.1% formic acid solution. Peptides were diluted 5-fold through the addition of 2% ACN / 0.1% formic acid and 20 µL of sample was transferred into glass vials (Waters, #186000385c) for mass spectrometry analysis.

2.5.5 Parallel reaction monitoring

To identify the amino acid incorporated onto the N-terminus of proteins expressed using the orthogonal translation initiation system, the targeted mass spectrometry method parallel reaction monitoring was employed. The unscheduled inclusion list was generated using Skyline (Version 19.1.0.193) and a FASTA file of the Nanoluciferase protein sequence with all 20 possible amino acids present at the N-terminus. Purified Nanoluciferase proteins were analysed using an Easy-nLC1000 liquid chromatography system coupled with a high-resolution Q-Exactive mass spectrometer (ThermoFisher Scientific). Peptide samples were injected onto a C^{18} reverse phase column that was 10 cm in length and had an inner diameter of 75 µm to separate peptides. Peptides were gradually eluted over a 120-minute linear gradient by increasing concentration of solution B (95%, ACN, 0.1% formic acid). The full MS (precursor) scan was performed using 70 000 resolution to identify desirable precursor ions. Automatic gain control (AGC) target was set to 3×10^6 with a maximum injection time (IT) of 100 ms. Scan range was set to 320 to 1800 *m/z*. Identified target precursors were subjected to high collision dissociation

fragmentation using a normalised collision energy of 27. Following precursor fragmentation, an MS/MS scan was used to measure product ions. MS/MS scan parameters were set at 17 500 resolution, AGC target of 3 x $10^{6,}$ maximum IT of 60ms, and an isolation window of 2.0 *m/z*. Generated spectra were processed using the program MaxQuant (Version 1.6.7.0). MaxQuant output was imported into Skyline (Version 19.1.0.193) allowing for analysis of precursor and product ion spectra.

2.6 Protein stability assays

2.6.1 Flow cytometry

2.6.1.1 Protein degradation experiment

BL21(DE3) pLysS cell strains transformed with different combinations of the two pULTRA plasmids and the three pET20B-T7: sfGFP-ssrA tagged plasmids or pET20B-T7: sfGFP plasmids were used to inoculate 5 mL of LBM broth with chloramphenicol, spectinomycin, and carbenicillin. Overnight cultures were used to inoculate fresh 10 mL LBM with chloramphenicol, spectinomycin, and carbenicillin and grown for 1 hour at 37^{o}C with shaking (250rpm) then induced with 1 mM IPTG. Cultures were diluted 100-fold in PBS solution in a 96-well plate and measured on a Cytoflex S (Beckman coulter) using a FITC fluorescence channel. Cytoflex S acquisition settings were set to FSC: 264, SSC: 2000, FITC: 299, and ECD: 150. Flow rate was 10 µL/min and an even threshold of 10 000 events was set. Data was processed and analysed using CytExpert (Version 2.3.0.84). The fluorescence data was analysed by calculating the mean fluorescence across 10 000 individual events measured and used to construct a plot illustrating trends in the data across time.

Chapter 3: Results

3.1 Identifying a suitable candidate anticodon/codon pair for further analysis

3.1.1 Identification of the mutant i-tRNA with largest effect on translation initiation

The overall goal of this entire thesis is to determine the amino acid incorporated onto the N-terminus of proteins when using an i-tRNA with a mutated anticodon and to see its effects on protein stability. I initiated this part study to determine codon/anticodon pairs that significantly improved translation initiation and exhibited complimentary base pairing to identify a potential candidate for further analysis. Previous work created 10 new i-tRNAs with anticodons AUG, AAC, AUA, UAC, GAU, AAU, UAU, CAG, CAA, CAC and measured their translation initiation efficiency against reporters containing all 64 start codons (Figure 6A) (Vincent, 2017, Hutvagner, 2018). Although the heatmap presentation of the data was useful to evaluate trends across all start codons, it was not sufficient to systematically determine the mutant i-tRNA/codon pairs that exhibited improved translation the most. To identify the most efficient pair to be used in this study, I re-analysed the raw count bulk fluorescence data from the study by Hutvagner (2018) using the software package DEseq2 (Love et al., 2014a). My new analysis showed that a total of 11 anticodon/codon pairs exhibited significantly different translation initiation efficiency when compared to native i-tRNA^{CAU}. Nine of the 10 pairs showed an increase in translation initiation efficiency while two showed a significant decrease (Figure 6B). Of these anticodon/codon pairs, UAC/GUA, AAC/GUU, and CUA/UUG were the three most and highly upregulated and statistically significant combinations compared to the native CAU/AUG pair. The CUA anticodon / UUG start codon pair showed a highly significant increase in translation initiation but, there is unspecific binding between the two uracil nucleotides in the second position and so this pair was not selected for this study as I wanted to focus on increased translation initiation between cognate pairs. The two remaining pairs showed a 4-fold increase in translation initiation with significant p-values, cognate pairing, and both encoded for the amino acid valine. I decided to proceed with the AAC/GUU anticodon/codon pair because the BL21 (DE3) pLysS *E. coli* strains used in this study normally lack a tRNA with an AAC anticodon, thus reducing the potential for unintended off target effects. Furthermore, I reasoned that using a mutant i-tRNA that

doesn't clash with any of the native translation machinery allows for the use of a mutant i-tRNA that is consistent with the original goal of introducing orthogonality into translation initiation. Having decided to proceed with the i-tRNA^{AAC} for further study, I next wanted to identify the amino acid that this i-tRNA was incorporating onto the N-terminus of proteins.

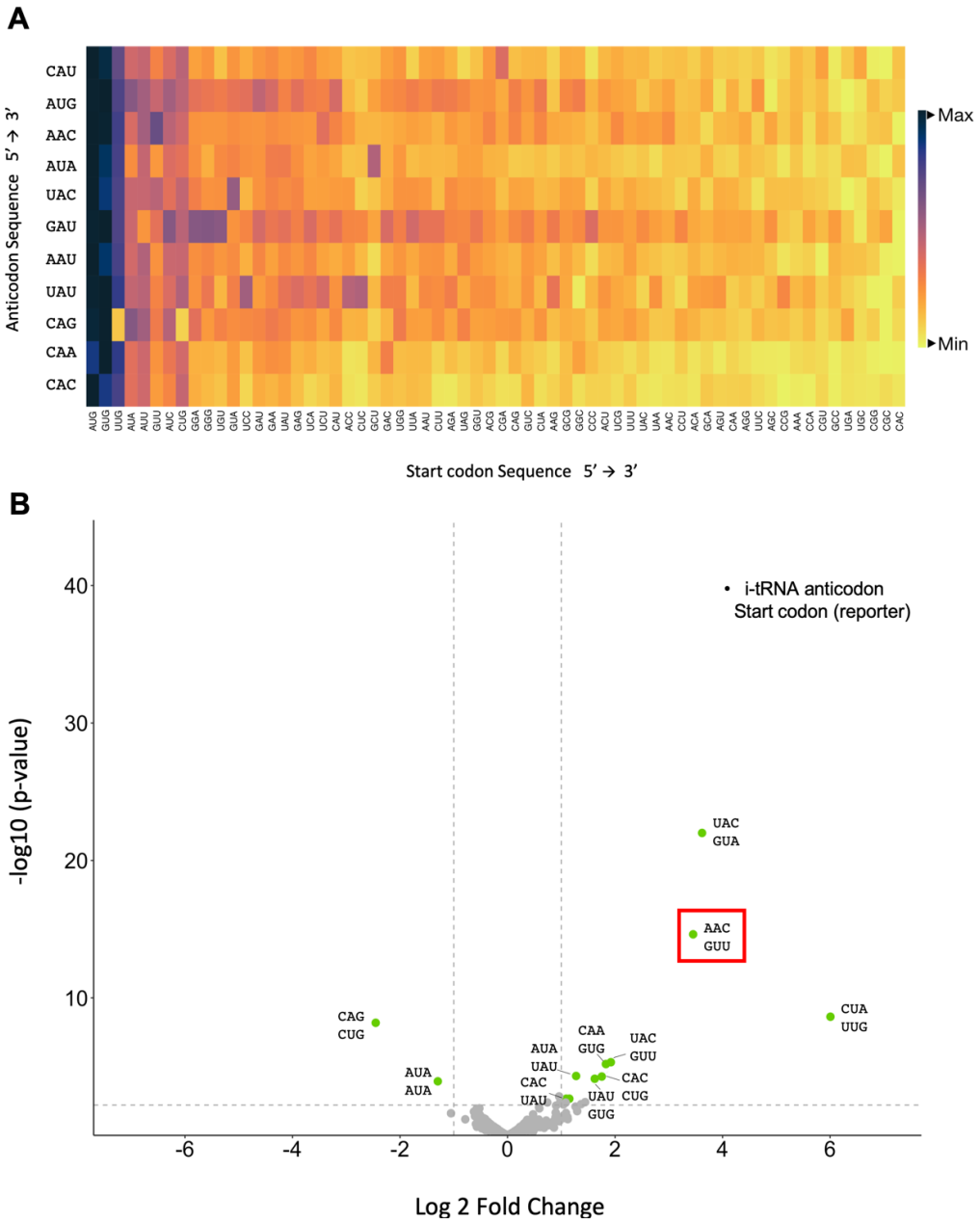


Figure 7. Anticodon mutant i-tRNA initiation efficiencies. (A) Heat map of normalised bulk fluorescence data of interaction between mutant i-tRNAs and sfGFP reporters against start codons. Image from Hutvagner (2018). **(B)** Volcano plot of bulk fluorescence data re-analysed using DEseq2. Fold-change threshold set to -1 and +1 and p-value threshold was set to ≤ 0.05. A total of nine anticodon/codon pairs were identified to be upregulated whilst two were down regulated. Data point outlined in the red square was selected for this study. Top line of each label represents the anticodon of that pair and the bottom line shows the start codon interacting with the anticodon. Labels in 5' to 3' orientation.

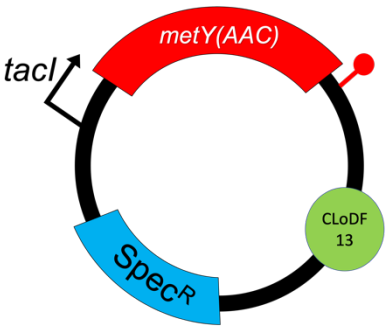
3.2.1 Plasmids used in N-terminal amino acid identification

To identify the amino acid incorporated onto the N-terminus of a protein expressed using the i-tRNA^{AAC}, I created a new purpose-built reporter. I needed to build this new reporter to be compatible with both proteomic analysis and translation initiation using the i-tRNA^{AAC}. This new reporter was the pET20B:T7: Nanoluc-6xhis plasmid (Figure 7). It was designed to harbour a *NanoLuc^{TM}* gene that possesses a C-terminal 6xhis tag and is under the control of a T7 promoter. NanoLuc^{TM} was selected as a reporter protein due to its favorable tryptic cleavage patterns, producing an N-terminal peptide 13 amino acids in length with a mass of 1614 Da, which is ideal for proteomic analysis. Two variants of the NanoLuc reporter were designed, one with a canonical AUG start codon and the second having a GUU start codon that is complimentary to mutant i-tRNA^{AAC}. The C-terminal 6xhis tag allows for downstream protein purification of the Nanoluciferase reporter, reducing sample complexity for proteomic analysis. The pET20B carries a carbenicillin resistance gene to allow for selection and a pBR322 origin of replication that is compatible with the pULTRA plasmid used in this study to expresses the i-tRNAs and the pLysS plasmid.

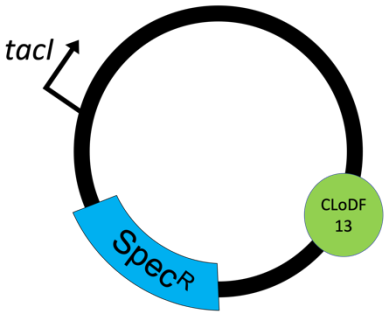
The pULTRA:*tac*:*metY*(AAC) or pULTRA:*tac*:empty and pET20B:*T7*:*NanoLuc* (AUG/GUU)-6xhis plasmids were transformed into the BL21 (DE3) pLysS *E. coli* strain creating i-tRNA^{AAC}/NanoLuc^{GUU} and i-tRNA^{CAU}/NanoLuc^{AUG} strains for protein expression and subsequent extraction (Table 1). The BL21 (DE3) pLysS strain possesses a pLysS plasmid harbouring the gene for T7 lysozyme, reducing the background expression of the Nanoluciferase reporter under the control of a T7 promoter.

Table 1. Strains created and used in this study. Each strain was transformed with two plasmids. A pULTRA plasmid possessing *metY(*AAC) or left empty to use native i-tRNA^{CAU}. Reporter plasmids pET20B contain reporter with start codon complimentary to anticodon of i-tRNA.

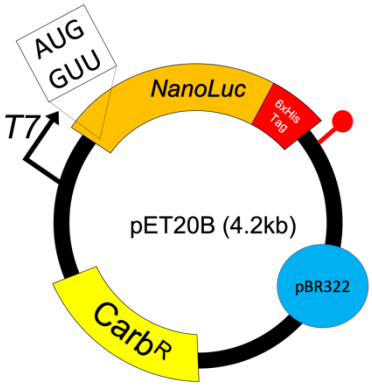
N-terminal amino acid identification		
Strain	*i-tRNA plasmid*	*Reporter plasmid*
BL21 (DE3) pLysS	pULTRA:*tac:metY* (AAC)	pET20B:*T7*:*NanoLuc*(GUU)-6xhis
BL21 (DE3) pLysS	pULTRA:*tac:empty* (CAU)	pET20B:*T7*:*NanoLuc*(AUG)-6xhis
Protein stability assay		
BL21 (DE3) pLysS	pULTRA:*tac:empty* (CAU)	pET20B:*T7*:*sfGFP*(AUG)-ssrA
BL21 (DE3) pLysS	pULTRA:*tac:empty* (CAU)	pET20B:*T7*:*sfGFP*(AUG)
BL21 (DE3) pLysS	pULTRA:*tac:metY* (AAC)	pET20B:*T7*:*sfGFP*(GUU)-ssrA
BL21 (DE3) pLysS	pULTRA:*tac:metY* (AAC)	pET20B:*T7*:*sfGFP*(GUU
BL21 (DE3) pLysS	pULTRA:*tac:empty* (CAU)	pET20B:*T7*:*sfGFP*(GCC)



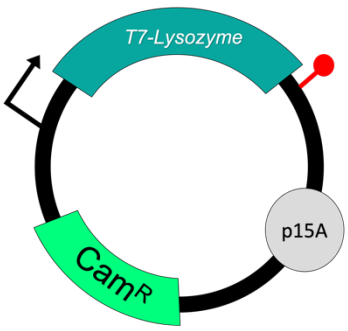
pULTRA:*tacI:metY(AAC)*



pULTRA:*tacI:Empty*



pET20B:*T7:*NanoLuc(AUG/GUU)-6xhis



pLysS

Figure 8. The plasmids used to identify N-terminal amino acids carried by i-tRNA^{AAC}. pULTRA:tac:*metY* plasmid possesses the *metY gene* variant encoding for i-tRNAs with AAC anticodon region, spectinomycin resistance, and CLoDF13 origin of replication. pULTRA:*tacI*:*Empty* used as control and relies on native i-tRNA^{CAU}. pET20B:T7:*NanoLuc* - 6xhis plasmid possesses gene variants for 6xhis tagged Nanoluciferase reporter protein with either AUG or GUU start codons, carbenicillin resistance, and pBR322 origin of replication. pLysS plasmid possesses gene for T7 -Lysozyme, chloramphenicol resistance, and a p15A origin of replication.

3.2.2 Using parallel reaction monitoring to sequence N-terminal peptides

Using mass spectrometry, I wanted to identify if an amino acid other than methionine is being incorporated onto the N-terminus of NanoLuc when using the mutant i-tRNA^{AAC} to initiate translation. Parallel reaction monitoring (PRM) is a form of tandem mass spectrometry which allows for the targeted identification of peptides using a predefined list of desired targets. PRM performs two scans, the first being the survey scan (MS1) that serves to detect the ionised peptides (precursor ions) that match with the predefined list. The targeted precursor ions are then fragmented and the resultant ions (product ions) are identified in a second scan (MS2). Identifying the mass of the precursor ion specifically associated with the N-terminal peptide of the reporter protein will allow for the elucidation of the peptide's composition while the MS2 scan will provide information on the specific order of these amino acids. This analysis was conducted on two protein samples, one expressed using the native i-tRNA^{CAU} and the other using i-tRNA^{AAC} co-expressed with NanoLuc (AUG or GUU) (Table 1).

The MS1 scan of the native AUG start codon system, identified three distinct precursor ions between 807 and 809 m/z and within retention times of 66 to 73 minutes. The precursor with 807.8849 m/z (M) possessed the most intense peak and correlated with the expected mass (807.8847m/z) of the doubly charged N-terminal peptide of NanoLuc composed of methionine, 2 x valine, 2 x phenylalanine, threonine, leucine, glutamic acid, 2 x aspartic acid, glycine, tryptophan, and arginine amino acid residues (Figure 9A). The other two precursor ions identified with a m/z of 808.3851 (M + 1) and 808.8855 (M + 2) were both approximately 0.5 m/z heavier than the M-ion and are isotopes of the N-terminal peptide, synthesised using the heavy carbon isotope C^{13} which is known to occur (Senko et al., 1995). Although the composition of the N-terminal peptide can be deduced using the precursor mass, the amino acid sequence can only be determined through the fragmentation of these ions and a deeper analysis of the resulting product ions. Once fragmentation of the precursor ion is achieved, two types of product ions known as b- and y-ions are formed. The b-ions extend from the N-terminus, making them essential for the identification of the N-terminal amino acid, while y-ions extend from the C terminus. For example, in a 10 amino acid peptide, when a *b6* ion is formed, so is a *y4* ion. By identifying the m/z of these product ions in the MS2 scan, the sequence of these product ions can be confidently identified and pieced together to sequence the entire peptide. Formation and analysis of the b- and y-ions provides this confidence by eliminating the chance of the m/z of precursor contaminants matching that of a desired

precursor. The 807.8849 m/z precursor from the NanoLuc (AUG) sample was fragmented and both b- and y-ions were identified (Figure 9B). The *b4* ion was measured at 479.2323 m/z with a signal intensity of $2x10^6$, consistent with a peptide sequence of methionine, valine, phenylalanine, and threonine residues as was expected using native translation initiation. The y4 to y7 ions were also identified, confirming the sequence of the precursor ion to be MVFTLEDFVGDWR.

In contrast, the MS1 scan from PRM analysis of the NanoLuc (GUU) / i-$tRNA^{AAC}$ sample identified 4 peaks within this retention time (Figure 9C). The precursor ion with 791.8990 m/z (M) correlated with the expected mass of the doubly charged N-terminal peptide of NanoLuc composed of 3 x valine, 2 x phenylalanine, threonine, leucine, glutamic acid, 2 x aspartic acid, glycine, tryptophan, and arginine amino acids (791.8986 m/z). Similar to the NanoLuc (AUG) / i-$tRNA^{CAU}$ sample, two C^{13} isotopes of this precursor were also detected. The precursor ion at 791.8990 m/z from the NanoLuc (GUU) / i-$tRNA^{AAC}$ sample was fragmented and a *b4* ion at 447.2602 m/z was detected, indicating the N-terminal amino acid sequence is valine, valine, phenylalanine, and threonine. Further analysis of the *y*-series ions confirmed the sequence of the entire precursor ion to be VVFTLEDFVGDWR.

Although substantially lower in its relative peak intensity, the fourth precursor identified at 807.8921 m/z in the NanoLuc (GUU) / i-$tRNA^{AAC}$ sample is similar to the mass identified in the NanoLuc (AUG) / i-$tRNA^{CAU}$ sample that correlated with a methionine bearing N-terminal peptide (807.8849 m/z). When fragmented, this precursor yielded *b4* ions at 479.2323 m/z, confirming methionine bearing N-terminal peptides. This would support the idea that a heterogenous population of NanoLuc is present in NanoLuc (GUU) / i-$tRNA^{AAC}$ samples, with both valine and methionine amino acids at the N-terminal position.

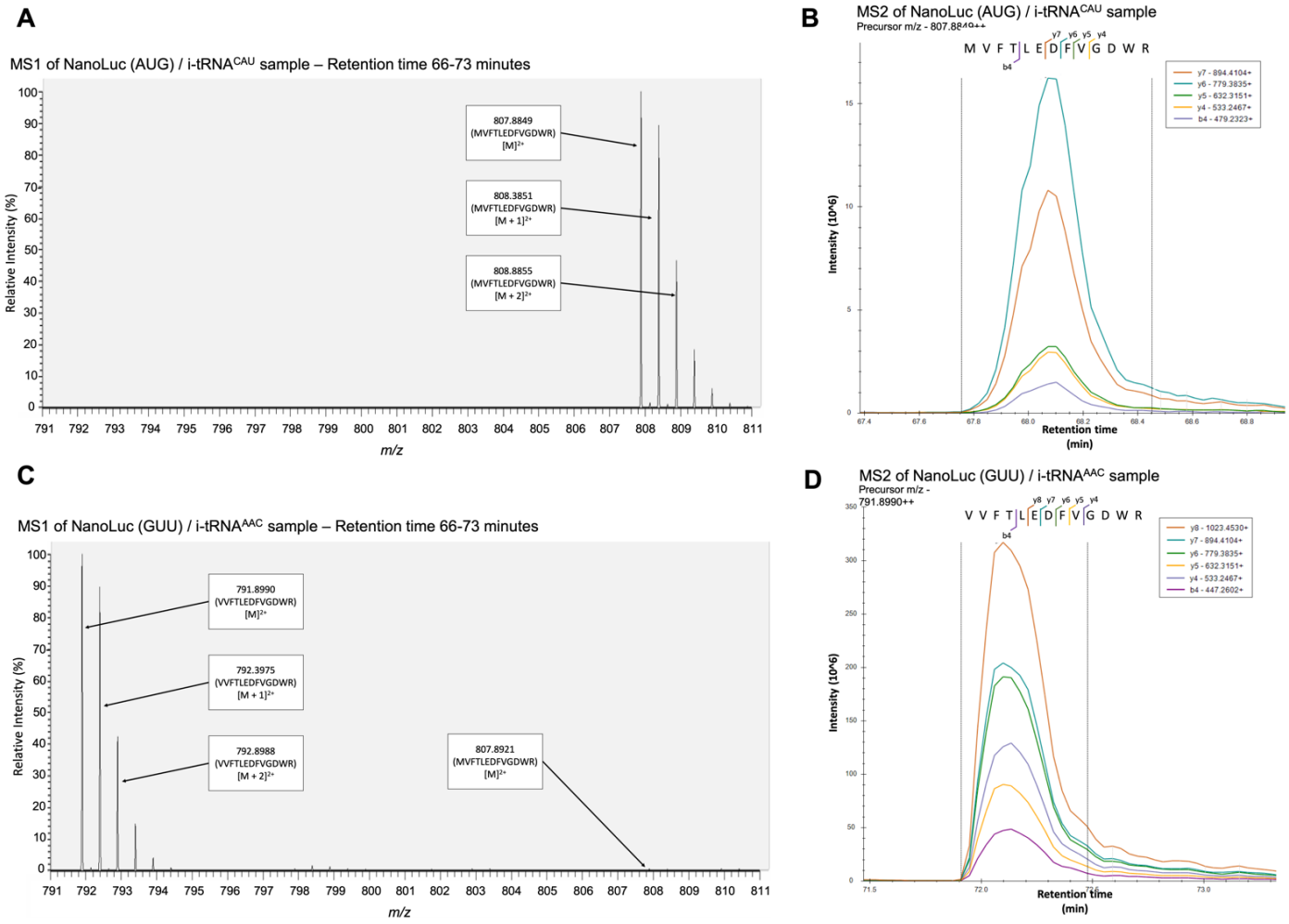


Figure 9. Identification of N-terminal peptides from NanoLuc (AUG/GUU) using the native i-tRNA or the mutant i-tRNA^{AAC}. (A) MS1 scan of NanoLuc (AUG) / i-tRNA^{AAC}. Precursor for methionine bearing N-terminal peptide and its carbon isotopes are annotated. (B) MS2 of NanoLuc (AUG) / i-tRNA^{CAU}. Peak intensity of product ions following fragmentation is shown. b4 ion indicates amino acid sequence to be methionine, valine, phenylalanine, and threonine. (C) MS1 scan of NanoLuc (GUU)/ i-tRNA^{AAC} . Precursor for valine bearing N-terminal peptide and its carbon isotopes are annotated. (D) MS2 of NanoLuc (GUU) / i-tRNA^{AAC}. b4 ion indicates amino acid sequence to be valine, valine, phenylalanine, and threonine. Precursor and product ion analysis conducted using data acquired between retention times 66-73 minutes.

3.2.3 Comparing methionine bearing peptides across both NanoLuc (AUG and GUU)

To determine the relative quantities of methionine bearing N-terminal peptides in both NanoLuc (AUG) / i-tRNA^{CAU} and NanoLuc (GUU) / i-tRNA^{AAC} samples, the peak area under the curve of each of these peptides was normalised against an internal peptide (sequence VVYPVDDH) of NanoLuc detected across all samples. This internal peptide should not be affected by using different i-tRNAs and would be present regardless of the identity of the N-terminal amino acid. Comparing only methionine bearing peptides

between the NanoLuc (AUG) and NanoLuc (GUU) samples indicates there is approximately a 200-fold decrease in NanoLuc production with an N-terminal methionine using the mutant i-tRNA^{AAC} (Figure 10A)

Comparing the proportion of methionine bearing peptides against valine bearing peptides within the heterogenous NanoLuc (GUU) samples indicates that valine is incorporated in 96-99% of all N-terminal peptides, with either methionine or oxidized methionine making up between 1-4% (Figure 10B). Oxidised methionine was included in this analysis to prevent skewing of the data as this is a common modification of N-terminal methionine (Kim et al., 2014). Methionine and cysteine residues are the most sensitive and likely residues to be oxidised, thus oxidised valine was not included, nor found in this analysis (Dunlop et al., 2009).

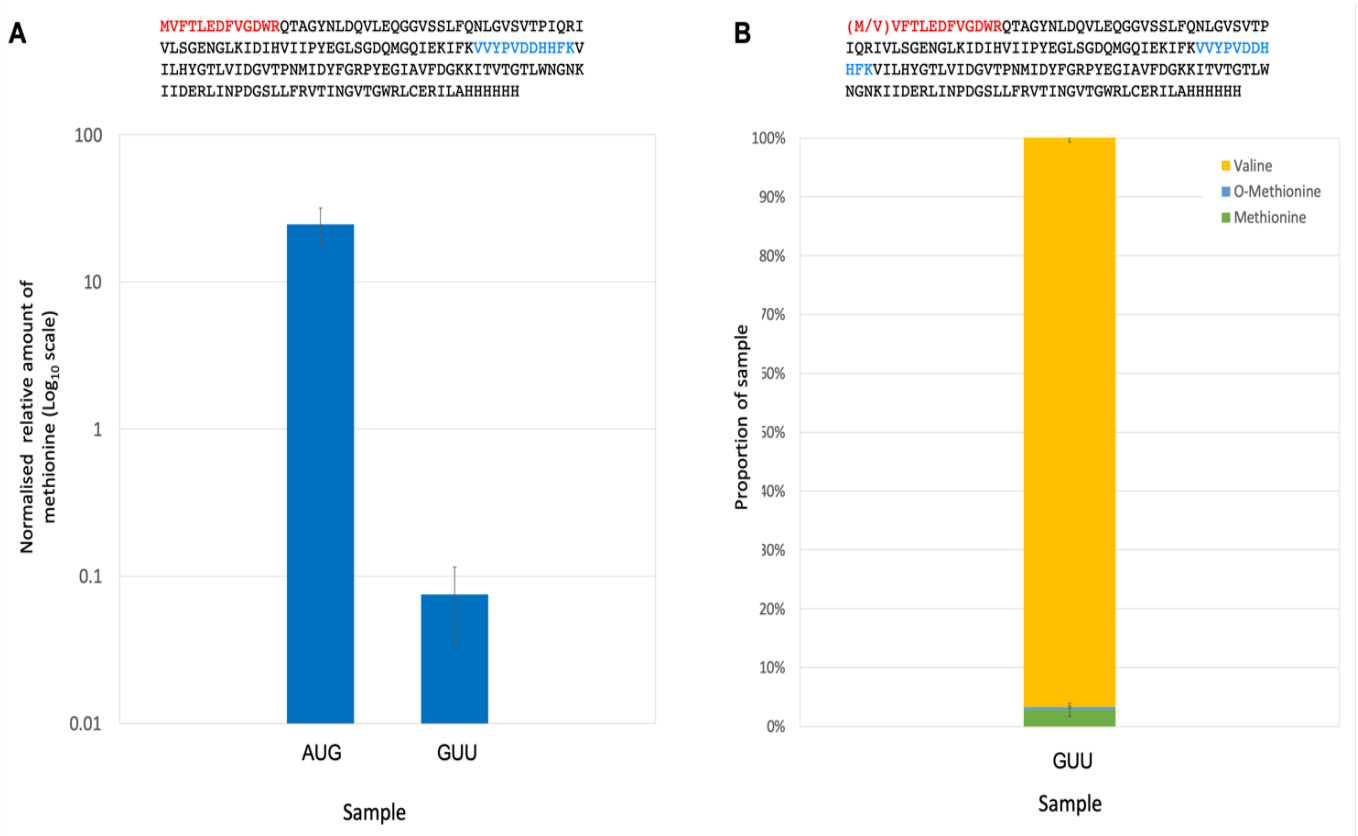


Figure 10. Relative quantitation of N-terminal peptides. (A) Relative quantitation of methionine bearing N-terminal amino acids across NanoLuc (AUG and GUU) samples implementing native i-tRNA^{CAU} or mutant i-tRNA^{AAC}. (B) Relative quantitation comparing proportions of N-terminal peptides bearing a methionine or valine N-terminal residue in NanoLuc (GUU) / i-tRNA^{AAC} samples. Amino acid sequences are displayed above graphs. Amino acids in red represent the N-terminal amino acid sequence and amino acids in blue represent the internal peptide used for normalisation of the data.

3.3 Determining stability of proteins expressed using mutant i-tRNA^{AAC}

3.3.1 Plasmids constructed and used in protein stability assays

With valine being incorporated onto the N-terminus of proteins expressed using this orthogonal translation initiation system, the overall protein stability may be affected by changing the proteins interaction with proteases involved with N-end degradation. The pET20B:T7:*sfGFP* plasmid variants harboring the gene for the fluorescent reporter sfGFP with AUG, GUU, and GCC start codons were sourced from a previous study (Hecht et al., 2017). This plasmid was used as a template for the design of the pET20B:T7:*sfGFP-ssrA* plasmid which is essentially the pET20B:T7:*sfGFP* plasmid however with an added ssrA degradation tag on the C-terminus of the sfGFP gene to potentially hasten degradation as sfGFP is highly stable (Figure 11). Different combinations of the pULTRA plasmid variants carrying the *metY* gene and pET20B plasmids harbouring the sfGFP genes were transformed into the BL21 (DE3) pLysS strains (Table 1).

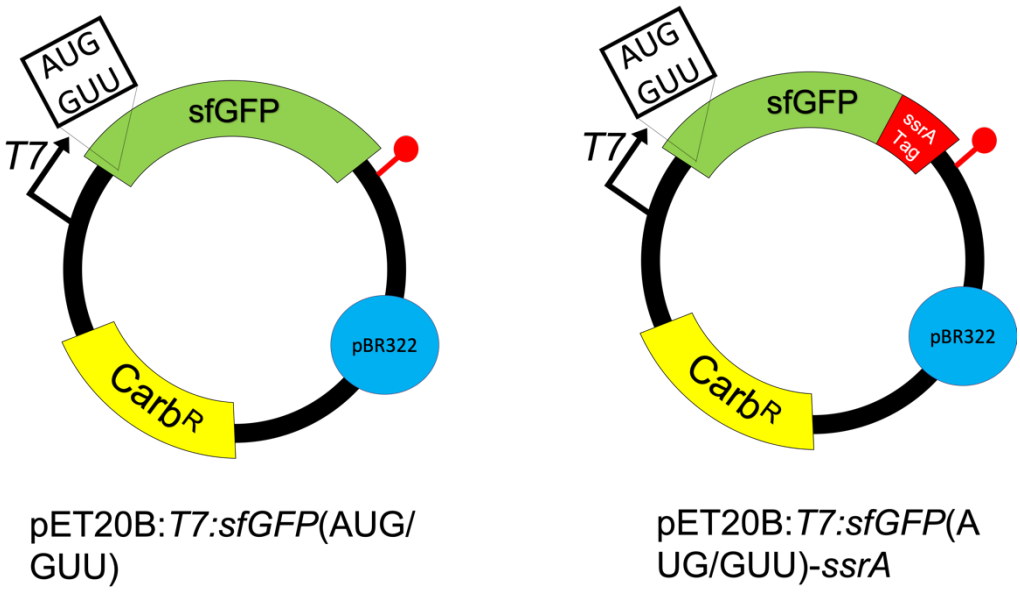


Figure 11. Two reporter plasmids used in protein stability assays. The pET20B:*T7:sfGFP*(AUG/GUU) plasmid contains the *sfGFP* gene under the control of a T7 promoter, carbenicillin resistance gene and a pBR322 origin of replication. The pET20B:T7:*sfGFP-ssrA* plasmid is essentially the same plasmid however, with a ssrA degradation tag on the C-terminus of the sfGFP gene.

3.3.2 Using the ssrA degradation tag to hasten sfGFP degradation

Next I wanted to determine if incorporating valine onto the N-terminus instead of methionine had an impact on the proteins overall stability as it may change its interaction with the N-end degradation pathway. Using fluorescence of sfGFP as a proxy for protein concentration, I monitored protein levels over time using flow cytometry to assess protein degradation. The sfGFP reporter protein is inherently a very stable protein and thus a ssrA degradation tag was added onto the C-terminus of sfGFP proteins to hasten the degradation process (Pédelacq et al., 2006). In sfGFP (AUG) / i-tRNA^{CAU} samples, degradation commenced at approximately seven to eight hours post-induction (Figure 12A). The ssrA tagged sfGFP (AUG) reporter showed degradation to commence at a similar time point but exhibited a decrease in overall fluorescence, failing to accelerate degradation.

Monitoring the sfGFP (GUU) / i-tRNA^{AAC} samples with an N-terminal valine, indicates that degradation does not seem to commence within the 9-hour time frame of this study and is on a general upward trend. The ssrA tagged version of this protein showed minimal fluorescence throughout the course of this study suggesting degradation from the ssrA tag was approximately equal to the expression rate of sfGFP (GUU) (Figure 12B).

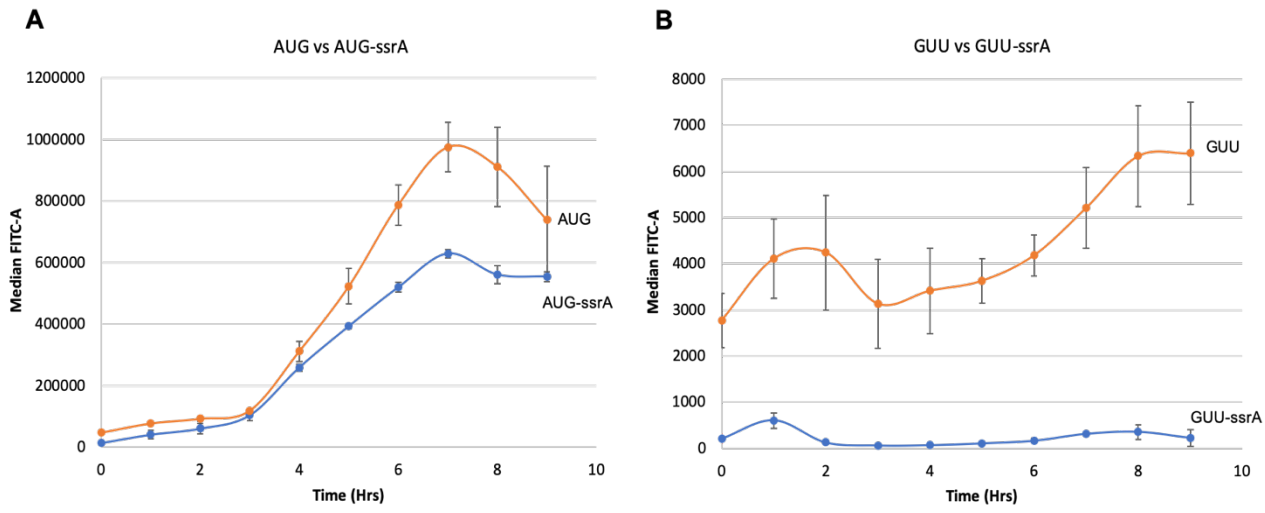


Figure 12. Identifying the applicability of the ssrA degradation tags in protein stability assays. (A) Comparison of sfGFP (AUG) reporter proteins synthesised using native translation initiation with and without the ssrA degradation tag. (B) Comparison tagged and untagged sfGFP (GUU) synthesised using the orthogonal i-tRNA^{AAC} mutant.

3.4.2 Measuring stability of sfGFP with N-terminal valine

Since ssrA tag degradation masked sfGFP (GUU) expression, the untagged sfGFP (AUG and GUU) were used to monitor degradation over a 33-hour period. To better determine the trend of sfGFP degradation, the fold-change of fluorescence in comparison to the point of induction was calculated. Expression of sfGFP occurs between 0 and 9 hours across all strains, except for the negative control, sfGFP (GCC) / i-tRNA^{CAU} (pULTRA:*tac*:empty) strain which has been shown to emit background levels of fluorescence (Hecht et al., 2017) (Figure 13).

Fluorescence of sfGFP (AUG) / i-tRNA^{CAU} samples peaks at 9 hours and steadily declines for 24 hours at which point fluorescence is only 70% of initial fluorescence at the time of induction. There is no observable point where degradation occurred in the sfGFP (GUU) / i-tRNA^{AAC} samples over the 33-hour time course, at which point the sample is 10-fold more fluorescent than at the time of its induction. Suggesting that proteins with an N-terminal valine that are synthesised using the AAC mutant i-tRNA are more stable than in native translation initiation with a methionine N-terminus.

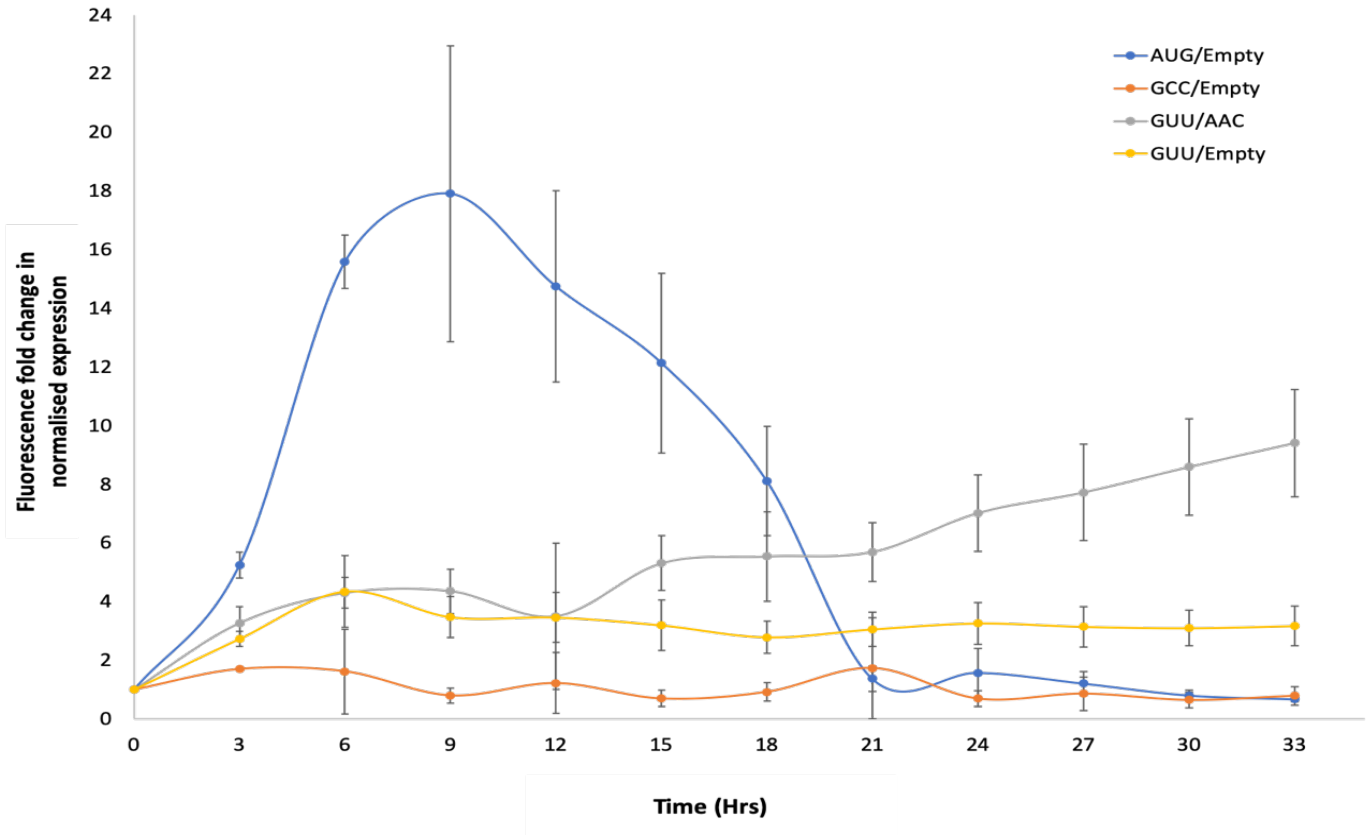


Figure 13. Protein stability assay of proteins expressed using native and orthogonal translation initiation systems. Fold change (y-axis) in fluorescence of sfGFP reporters from time of induction up until 33 hours post induction (x-axis). All cell strains were induced with 1mM IPTG. Fluorescence of AUG/Empty strain begins to decline at 9-hour time point and steady declines across the duration of this study until fluorescence is 70% of fluorescence at time of induction. GUU/AAC sample shows a steady incline in fluorescence over the 33-hour period indicating it is more stable in comparison to the AUG/Empty sample. GUU/Empty strain included as a control for expression of sfGFP with a GUU start codon but lacking AAC mutant i-tRNA. GCC/Empty sample used as negative control as it is known to be at background fluorescence levels.

Chapter 4: Discussion

Developing an entirely orthogonal central dogma is one of the grand goals of synthetic biology that would allow us to circumvent the lack of efficiency and control in circuits inserted into synthetic organisms. All aspects of the central dogma including DNA replication, transcription and translation are currently being investigated to identify a means of introducing orthogonality (Malyshev et al., 2014, Hoshika et al., 2019, Rackham and Chin, 2005).

In this thesis project, I used targeted mass spectrometry to identify the amino acid incorporated onto the N-terminus of NanoLuc initiated from i-tRNA^{AAC}. To do this I designed a NanoLuc reporter protein with a GUU start codon to be expressed with a mutant i-tRNA^{AAC}. NanoLuc was selected and redesigned such that it was compatible with proteomic sample preparations and analysis. Using the optimized reporter platform, I identified valine, rather than the usual methionine, was incorporated onto the N-terminus of NanoLuc (GUU) initiated from mutant i-tRNA^{AAC}. Finally, I showed that when valine occupies the N-terminus of sfGFP, the protein has great stability over a 33 hour time-course.

4.1 Deseq2 analysis of bulk fluorescence data to determine a suitable starting point

With the increasing use of technologies allowing for high-throughput sequencing (HTS), the identification of quantitative differences between transcriptomes has become a commonly asked question. For this reason, the R package DEseq2 was developed to statistically identify differentially expressed genes between samples (Love et al., 2014b). The statistical models in DEseq2 take into account the common assumptions needed to assess count data (Love et al., 2014a). Some of these include non-normality of data distribution, variance estimation, and in the case of HTS the low number of replicates per sample condition (Love et al., 2014b). DEseq2 is now commonly employed for the differential analysis of numerous HTS approaches such a chromatin immunoprecipitation sequencing or comparing taxa in metagenomics (Love et al., 2014b, Chabbert et al., 2015). The only instance of using DEseq2 for reasons other than HTS, is its use in cluster

analysis of tandem MS data to identify differential expression using count data (Moon et al., 2018), implying that DEseq2 can only be applied in a limited number of circumstances.

Similar to HTS data, the raw bulk fluorescence data used to determine translation initiation efficiency of previously designed mutant i-tRNAs against all 64 of the possible start codons is in the form of count data as integers. By setting the canonical i-tRNA with a CAU anticodon as a sample condition and the mutant i-tRNA with a varying anticodon as another condition, comparison between conditions was now possible. Furthermore, by having each of the conditions in triplicates aligns well with the low number of replicates expected in HTS experiments, allowing the statistical models set in place for DEseq2 analysis to hold true. This allows for the comparison of translation initiation efficiency of codon/anticodon pairs that are either up or down regulated in comparison to the native i-tRNA. This new way of analysing the bulk fluorescence data of the 10 anticodons identified a difference in translation initiation in 11 pairs, nine being up regulated and two being down regulated (Figure 7). Of these samples, the AAC/GUU anticodon/codon pair was selected for further analysis.

To my knowledge, this is the first instance of the DEseq2 software package being used to identify differential translation initiation efficiency using bulk fluorescence data. Reformatting the bulk fluorescence data into a matrix of integer values provides an unconventional use of the DEseq2 package, potentially expanding its applications beyond high-throughput sequencing analysis and quantitative proteomics.

4.2 Identifying the amino acid incorporated on the protein N-terminus by the mutant i-tRNA^{AAC}

Although the efficiency of translation initiation has previously been described using sfGFP (Hutvagner, 2018), the system itself lacked characterisation in terms of aminoacylation patterns and stability of protein products. To define the aminoacylation patterns of mutant i-tRNA^{AAC}, I designed a new NanoLuc reporter compatible with common MS preparation and that can be expressed by the mutant i-tRNA^{AAC} (Hecht et al., 2017, Vincent, 2017). I employed a targeted mass spectrometry approach called PRM to sequence the amino acid composition of the N-terminal peptide produced through tryptic digestion of NanoLuc (GUU). This work showed that valine was loaded onto the mutant i-tRNA^{AAC} and incorporated onto the N-terminus of NanoLuc.

This result follows a similar trend with earlier studies that found mutating the anticodon region of native i-tRNA and the start codons of proteins to compliment these mutants, changed aminoacylation from methionine to other amino acids. In an *in vitro* study, changing the anticodon from CAU to CUA caused glutamine to be charged onto the mutated i-tRNA rather than the canonical methionine (Schulman and Pelka, 1985). It was postulated in this study, but not shown, that this would also be the case *in vivo*. A previous study found that the addition of glutaminyl-tRNA synthase to *E. coli* strains possessing mutated i-tRNA^{CUA} greatly increased the synthesis of chloramphenicol acetyltransferase modified to have a UAG start codon (Varshney and RajBhandary, 1990). This provided indirect evidence that glutamine was being inserted onto the protein N-terminus. The idea that changing the anticodon region of i-tRNAs would lead to its aminoacylation with the amino acid normally encoded by that anticodon in e-tRNAs was confirmed in a later study. Changing the canonical anticodon of i-tRNA^{CAU} to GAU, GAC, and GAA led to the aminoacylation of the i-tRNA with isoleucine, valine and phenylalanine respectively, by the corresponding aaRS (Pallanck and Schulman, 1991). This was confirmed by using these mutated i-tRNAs to express dihydrofolate reductase using start codons complimentary to the mutant i-tRNAs. The dihydrofolate reductases were then purified and the first three N-terminal amino acids were sequenced using Edman degradation. Later it was concluded that not only are these mutant i-tRNAs loaded with non-methionine amino acids, but they also appear to be formylated (Varshney et al., 1991). MTF overexpression in strains with the same mutant i-tRNAs increased the production of a chloramphenicol acetyltransferase reporter with complimentary start codons (Mayer et al., 2003). When it comes to the aminoacylation of tRNAs, the anticodon region is clearly a key identity element used by the cognate aaRS to ensure stringent aminoacylation (Cusack, 1997). Thus, it is likely that mutating the anticodon region of the i-tRNA from CAU to AAC reduces its recognition by methionyl-tRNA synthetases (MetRS) and increases recognition by valyl-tRNA synthetase (ValRS), in turn causing the mutant i-tRNA^{AAC} to incorporate valine onto the N-terminus of proteins (Figure 10).

In *E. coli* strains transformed with the mutated i-tRNA^{AAC} and NanoLuc (GUU), a heterogenous population of N-terminal peptides bearing either methionine or valine at the extreme N-terminus were identified. Relative quantitation of these peptides revealed that between 96 and 99 percent of the proteins present within the sample bore an N-terminal valine, whilst methionine and oxidized methionine made up the rest of the population (Figure 10). One possible cause for the mixed N-terminal amino acid population is that

along with the mutated i-tRNA^{AAC}, native i-tRNA^{CAU} is still present within this strain and may still be initiating translation from the NanoLuc (GUU) reporter through non-specific binding. The phenomenon of the native i-tRNA^{CAU} initiating translation from all 64 potential start codons has previously been characterised and initiation from a GUU start codon was identified as the 10th most efficient at initiating translation (Hecht et al., 2017).

My novel analysis using DEseq2 showed that flexible binding between i-tRNAs and non-cognate codons is not exclusive to the native i-tRNA^{CAU}. For example, translation of sfGFP with a UUG start codon using the mutant i-tRNA^{CUA} showed an increase in translation initiation in comparison to the native i-tRNA^{CAU} (Figure7B). In this interaction between the mutant i-tRNA^{CUA} and sfGFP (UUG), there is non-specific binding in the middle position of the interaction between the two uracil residues. The same may be occurring in my i-tRNA^{AAC} / NanoLuc (GUU) sample with the native i-tRNA^{CAU} normally expressed by the cell, weakly interacting with the GUU start codon. The flexible binding of native i-tRNA^{CAU} with the GUU start codon in NanoLuc (GUU) may cause methionine incorporation and is one possible reason for the observed heterogenous amino acid population on the N-terminus of NanoLuc (GUU).

Another possible explanation for the heterogenous protein population observed in NanoLuc (GUU) / i-tRNA^{AAC} samples would be the misacylation of the i-tRNA^{AAC} with methionine by the MetRS enzyme. Although it seems to have the largest influence on aminoacylation, the anticodon is not the sole identity element used by aaRS (Schulman and Pelka, 1988). In addition to the anticodon, the G2:C71 and C3:G70 nucleotide interactions and the C32, U33, A37 nucleotides in native i-tRNA^{CAU} (Figure 2) are also identity elements used by MetRS (Giegé et al., 1998). The mutant i-tRNA^{AAC} was designed using the native i-tRNA with all nucleotides the same, including the identity elements, apart from the anticodon. These other identity elements may be recognised by the MetRS and may be loading a small proportion of the mutant i-tRNA^{AAC} with methionine despite having an AAC anticodon. This small amount of misacylation could account for the one percent of methionine bearing proteins. Misacylation rates of e-tRNAs have been estimated to be between 10^{-4} (Parker, 1989) and 10^{-6} (Kramer and Farabaugh, 2007) in *E. coli*, which is significant lower than what I observed. However, these figures are for e-tRNAs. The identity elements unique to the i-tRNA may be causing a higher rate of misacylation by MetRS and may account for the increased rate of misacylation of the mutant i-tRNA^{AAC}.

To identify whether non-specific recognition of the GUU start codon by the native i-tRNA^{CAU} or misacylation of the mutant i-tRNA^{AAC} is responsible for expressing methionine bearing reporter proteins, further proteomic analysis could be conducted. Performing PRM on a new strain possessing both the pULTRA-empty plasmid and pET20B:T7: Nanoluc (GUU)-6xhis plasmid would help identify the source of the two amino acids being incorporated. If only methionine was being incorporated onto the N-terminus, we could infer that translation initiation is occurring through the non-specific binding interactions between the native i-tRNA^{CAU} and the GUU codon. However, if valine was present, we could speculate that the GUU start codon may be misread by an e-tRNA^{Val}. If a mixture of methionine and valine was found, it may be possible that it is a combination of non-specific binding of native i-tRNA^{CAU} and misreading by an e-tRNA^{Val}.

4.3 Determining the feasibility of using a ssrA degradation tag to explore N-end degradation

The amino acid on the extreme N-terminus of a protein influences its recognition by the N-end degradation pathway with some amino acids degrading more rapidly than others. Amino acids are categorised as either stabilising or destabilising residues, with destabilising residues having a half-life of less than three minutes and stabilising residues greater than 10 hours (Tobias et al., 1991). However, there has been little work discriminating degradation times between amino acids of the same category. Both methionine and valine are classified as stabilising residues, with half-lives greater than 10 hours in *E. coli* (Tobias et al., 1991). In this study, I identified that valine was predominantly incorporated onto the N-terminus of proteins using the i-tRNA^{AAC} mutant at a 99:1 ratio with methionine. This result enabled me to evaluate the difference in stability between methionine and valine N-end substrates.

Unfortunately, using luminescence readings from the NanoLuc protein as a proxy for protein concentration levels requires lysis of the cells and is not ideal for time-course assays. Thus, I used sfGFP synthesised using the native i-tRNA^{CAU} and the mutant i-tRNA^{AAC} and analysed fluorescence on a flow cytometer. One issue with using sfGFP was that it is inherently very stable due to its favorable folding kinetics, making measuring it's degradation a challenge (Pédelacq et al., 2006). To circumvent this inherent stability, a version of sfGFP tagged with a ssrA degradation on the C-terminus was constructed. ssrA

tagged proteins are recognised and degraded by the ClpXP protease and the N-degrons are identified by a ClpAP protease (Farrell et al., 2005, Varshavsky, 2011). Degradation of proteins that possess both signals has not been well characterised and so, I attempted to identify whether the degradation of these proteins through these two pathways were cumulative and predictable. Comparison of sfGFP and sfGFP-ssrA expressed using AUG start codons revealed that the ssrA tag only caused an overall decrease in fluorescence and failed to reduce the time taken to see degradation (Figure 12A). The tagged and untagged sfGFP samples expressed using the GUU start codons were harder to evaluate as the ssrA tagged sample failed to display any fluorescence above background levels (Figure 12B). This suggests that degradation from the tag commences promptly after expression and that the degradation rate was greater than or equal to the expression rate.

Using reporters that were not ssrA-tagged proved to be sufficient to evaluate the relative stability of methionine and valine N-terminal sfGFP reporters. I showed that in strains containing untagged sfGFP with AUG, GUU, GCC start codons and different i-tRNAs, the protein stability varied. Peaking at nine hours, the fluorescence of sfGFP (AUG) expressed using the native i-tRNA^{CAU} was approximately 80-fold higher than sfGFP (GUU) strains with a mutant i-tRNA^{AAC} (Supplementary figure 1). Indicating that orthogonal translation initiation using a mutant i-tRNA^{AAC} is not as efficient as the native system. However, at 33 hours post-induction, fluorescence between the two samples was almost equal, suggesting sfGFP (GUU) expressed with mutant i-tRNA^{AAC} are more stable than sfGFP expressed using native translation initiation *in vivo*. Further supporting this notion that sfGFP (GUU) is more stable, fluorescence from sfGFP (GUU) was 10-fold higher than at the time of its induction whilst fluorescence from sfGFP^{AUG} was only 70% of its initial fluorescence.

Rather than using the ssrA tag, a system overexpressing the mediator molecule ClpS potentially presents a superior alternative to measuring N-end degradation in future studies of the i-tRNA^{AAC} system. ClpS is a critical component of recognising N-degrons, as ClpS deficient *E. coli* strains were unable to recognise otherwise strong N-degrons (Erbse et al., 2006). Furthermore, ClpS inhibits the degradation of other ClpAP substrates such as ssrA tagged proteins or ClpA itself, dedicating the ClpAP machinery to the degradation of N-degrons (Dougan et al., 2002, Kunjapur et al., 2018). Thus, ClpS overexpression will increase available ClpS to deliver protein substrates to the ClpAP protease complex and inhibit degradation of other substrates by the ClpAP protease, making them more available to our targeted N-degrons. Accelerating degradation using

this system would improve the study as fluorescence may not need to be measured well into the stationary phase of *E. coli* growth where translation rates and machinery vary.

One further consideration for this study, is that the i-tRNA^{AAC} was overexpressed using a plasmid in these strains whilst the strains using i-tRNA^{CAU} were reliant solely on native levels of i-tRNA. This could have potentially resulted in a higher concentration of i-tRNA^{AAC} in cells as they progressed past the log-phase in their growth cycle, causing translation initiation rates to be increased in these strains. A chloramphenicol repression assay would prevent all translation from occurring as it inhibits peptidyl transferase from elongating the protein chain, even in the presence of excessive i-tRNAs and may be a useful technique to remove this variable in future studies. The interplay of i-tRNA levels present within the cell and their formylation are critical to their binding of the ribosomal P-site and to the fidelity of protein synthesis (Kapoor et al., 2010, Varshney and RajBhandary, 1992). Mutating the promoter sequence of the *metZWV* operon in *E. coli* drastically lowered the production and levels of native i-tRNA^{CAU} and reduced initiation fidelity. Doing so allowed for initiation to occur from a mutated i-tRNA lacking the highly conserved three G-C pairs in the anticodon stem and an e-tRNA that normally encodes for glutamine (Kapoor et al., 2010). This indicates that high levels of mutant i-tRNA is imperative for its proper binding with the P-site of the ribosome.

Overexpression of the mutant i-tRNA^{AAC} may also be causing unanticipated changes in the proteome of *E. coli* (Frumkin et al., 2018). Several proteomic changes were identified in *E. coli* expressing the mutant amber i-tRNA^{CUA} (Vincent et al., 2019). Proteins directly involved with ribosomal biogenesis and assembly were among the upregulated proteins, including rRNA modification enzymes and both IF2 and IF3 (Vincent et al., 2019). Overexpression of the mutant i-tRNA^{AAC} may have similar effects to overexpression of the amber i-tRNA^{CUA}, increasing the concentration of mature ribosomes and initiation factors. Thus, the overexpression of the mutant i-tRNA^{AAC} may be having a direct impact on the level of expressed sfGFP and may further exaggerate protein production through indirectly causing the upregulation of ribosomal machinery.

Formylated methionine on the N-terminus of proteins is thought to serve as a degradation signal that is required to control protein quality (Piatkov et al., 2015). Thus, proteins expressed using the mutant i-tRNA^{AAC} lack formylated methionine on the N-terminus and may result in the misfunctioning of quality control systems in place for *E. coli* and ultimately have deleterious effects. All these things need to be taken into account when considering the concentration of proteins produced using mutant i-tRNAs.

The engineering of mutant i-tRNAs to control the amino acid incorporated onto the N-terminus of proteins has potential for novel applications and provides a new tool for synthetic biology. If the incorporation of an amino acid and its effect on protein stability is predictable, such a system could be used to tune the concentrations of individual components within an engineered synthetic biological pathway or load predetermined amino acids onto the protein N-terminus. For example, specific loading or amino acids onto the N-terminus of proteins may have benefits in the development of therapeutics. Peptides such as fMLP produced in bacteria are chemo attractants that elicit a proinflammatory response in humans, recruiting leukocytes to the site of infection to eliminate the invading bacteria (Pan et al., 2000). Specifically, the formylation of the N-terminal methionine of these peptides is a feature unique to bacteria that is recognised by the human immune system to identify invading pathogens (Le et al., 2002). Further refinement of this orthogonal translation initiation system to incorporate other amino acids than formylated methionine may allow for the development of therapeutics in *E. coli*, without the concern of a formyl-methionine being recognised, mounting an immune response in drug recipients.

4.4 Enhancing the efficiency of orthogonal tRNA translation initiation

Although having valine present on the N-terminus of a protein seems to improve its stability in comparison to the canonical methionine, protein expression through orthogonal translation initiation using a mutant i-tRNA^{AAC} is still unable to achieve initiation rates comparable to native i-tRNA^{CAU} (Supplementary figure 1).

Aminoacylation of the mutant i-tRNA^{AAC} may be limiting translation initiation due to a lack of availability of cognate valyl-synthetases (Chiapello et al., 1999, Sharp and Li, 1987). It has been previously shown that the over expression of valyl-tRNA synthetase enhances the production of chloramphenicol acetyltransferase (GUC) using mutant i-tRNA^{GAC} (Mayer et al., 2003), suggesting that it is indeed possible to enhance efficiency by aaRS overexpression. Both the second A and third C nucleotides in the anticodon of tRNAs are the main identity elements for valyl-tRNA synthase recognition (Pallanck and Schulman, 1991). The mutant i-tRNA^{AAC} used in this study possesses both those nucleotides in the correct positions, thus it would be reasonable to assume that over expression of standard valyl-synthetases will enhance translation initiation efficiency.

One factor identified as limiting translation initiation from mutant i-tRNAs, is formylation (Mayer et al., 2003). MTF is the enzyme chiefly involved with the formylation of the methionine on charged i-tRNAs. It has been shown that MTF is not as specific to methionine as once thought and may allow binding of other amino acids such as valine, phenylalanine, lysine, and isoleucine (Mayer et al., 2003). Phenylalanine is particularly interesting as not only does it get formylated (Mayer et al., 2003), but according to the N-end degradation pathway it should be a destabilising residue and rapidly degraded within a couple minutes (Humbard et al., 2013). Evidence has shown that this was not the case when phenylalanine is incorporated onto the protein N-terminus by mutant i-tRNAs (Mayer et al., 2003). The formylation of this otherwise strong N-degron may be preventing it's recognition by ClpS. It would be interesting to investigate whether deformylase enzymes that normally recognise formylated methionine would also recognise a formylated phenylalanine, as this may potentially be developed into a switch for rapid protein degradation (Meinnel and Blanquet, 1995). Overexpression of MTF in orthogonal translation initiation may be a potential avenue to enhance the formylation of mutant i-tRNAs, in turn improving the efficiency of orthogonal translation initiation. Due to PRM being limited to an inclusion list that is constrained by the performance and speed of the mass spectrometer, formylated valine residues were not searched for in this study. However, future proteomic experiments which also include a search for formylated valine at the N-terminus would provide valuable information regarding formylation capabilities of MTF for i-tRNA^{AAC}

It is thought that IF2, which carries i-tRNA to the ribosome P-site, recognises both the highly conserved region on the acceptor stem of i-tRNAs, and the formylated amino acid (Simonetti et al., 2008, Guenneugues et al., 2000). The IF2 affinity for three formylated amino-acid bearing i-tRNAs was tested and it was found that formylated valine bound to the P-site of a ribosome just as well as formylated methionine (Wu and RajBhandary, 1997). Thus, as with valyl-synthetases and MTF, overexpression of IF2 may also improve the efficiency of translation initiation the mutant i-tRNA^{AAC} used in this study.

4.5 Future considerations

The results from this study identified that the amino acid valine is being incorporated onto the N-terminus of proteins when using orthogonal translation initiation with a mutant i-tRNA^{AAC}. Furthermore, I explored the degradation of this protein bearing an N-terminal valine through the N-end degradation pathway and have identified it to be more stable than the canonical methionine. Relative quantitation was used in this study to identify the proportion of methionine and valine bearing N-terminal peptides and found that the majority of the population possessed valine at the extreme N-terminus.

To improve our understanding and provide a more holistic view of orthogonal translation initiation, the methods and techniques used in this study, in addition to those suggested in the discussion should be performed on the remaining 62 possible mutant i-tRNAs. Doing so may allow for the development of a suite of mutant i-tRNAs allowing for greater control of protein levels *in vivo*. Successfully designing and characterising this full suite of mutant i-tRNAs will provide comprehensive insights into how each anticodon influences translation initiation efficiency and allow us to assess each of the possible 64 anticodons against all the possible 64 start codons.

The mutant i-tRNAs used in this study and all others so far are only semi-orthogonal as they are still dependent on cellular machinery. For example, both mutant and native tRNAs are sharing a pool of aaRS for aminoacylation. In order to reduce the cellular dependence of mutant i-tRNA, two goals must be achieved. The first goal is to develop a method of preventing the recognition of the mutant i-tRNAs by the endogenous cognate aaRS and prevent amino acid misincorporation. The second goal is to develop an orthogonal aaRS that is specific to the mutant i-tRNA and does not recognise any endogenous tRNAs. Further characterising all the possible anticodon/codon pairs and developing a truly orthogonal i-tRNA and aaRS will bring the field of synthetic biology one step closer to achieve the grand vision of an entirely orthogonal central dogma.

Chapter 5: Conclusion

In this thesis, I characterised the amino acid incorporation using an orthogonal initiator tRNA with an AAC anticodon and try to define the effects that using orthogonal translation initiation has on the N-terminus of proteins. The results within this thesis indicated that the amino acid valine is being loaded onto the protein N-terminus when using the mutant i-tRNA^{AAC} and that this improves the stability of proteins in *E. coli*. This work presents a potentially new avenue for protein engineering using orthogonal translation initiation. Further work to characterise the full suite of anticodon/codon pairs may allow for the development of a repertoire of mutant i-tRNAs, all incorporating different amino acids and with different stabilities *in vivo*, providing greater control over protein levels in synthetic circuits or pathways.

References

ALLEN, G. S., ZAVIALOV, A., GURSKY, R., EHRENBERG, M. & FRANK, J. 2005. The cryo-EM structure of a translation initiation complex from Escherichia coli. *Cell,* 121**,** 703-712.

ANDRIANANTOANDRO, E., BASU, S., KARIG, D. K. & WEISS, R. 2006. Synthetic biology: new engineering rules for an emerging discipline. *Molecular systems biology,* 2**,** 2006.0028.

AYYUB, S. A., 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. *Journal of bacteriology,* 199**,** e00051-17.

BARRAUD, P., SCHMITT, E., MECHULAM, Y., DARDEL, F. & TISNE, C. 2008. A unique conformation of the anticodon stem-loop is associated with the capacity of tRNAfMet to initiate protein synthesis. *Nucleic acids research,* 36**,** 4894-4901.

BARRELL, B., BANKIER, A. & DROUIN, J. 1979. A different genetic code in human mitochondria. *Nature,* 282**,** 189.

BORKOWSKI, O., CERONI, F., STAN, G.-B. & ELLIS, T. 2016. Overloaded and stressed: whole-cell considerations for bacterial synthetic biology. *Current opinion in microbiology,* 33**,** 123-130.

BROWNING, D. F. & BUSBY, S. J. 2004. The regulation of bacterial transcription initiation. *Nature Reviews Microbiology,* 2**,** 57.

CASERTA, E., TOMŠIC, J., SPURIO, R., LA TEANA, A., PON, C. L. & GUALERZI, C. O. 2006. Translation initiation factor IF2 interacts with the 30 S ribosomal subunit via two separate binding sites. *Journal of molecular biology,* 362**,** 787-799.

CHABBERT, C. D., ADJALLEY, S. H., KLAUS, B., FRITSCH, E. S., GUPTA, I., PELECHANO, V. & STEINMETZ, L. M. 2015. A high-throughput ChIP-Seq for large-scale chromatin studies. *Molecular systems biology,* 11.

CHANG, M. C. & KEASLING, J. D. 2006. Production of isoprenoid pharmaceuticals by engineered microbes. *Nature chemical biology,* 2**,** 674.

CHATTAPADHYAY, R., PELKA, H. & SCHULMAN, L. H. 1990. Initiation of in vivo protein synthesis with non-methionine amino acids. *Biochemistry,* 29**,** 4263-4268.

CHATTERJEE, A., XIAO, H. & SCHULTZ, P. G. 2012. Evolution of multiple, mutually orthogonal prolyl-tRNA synthetase/tRNA pairs for unnatural amino acid mutagenesis in Escherichia coli. *Proceedings of the National Academy of Sciences,* 109**,** 14841-14846.

CHIAPELLO, H., OLLIVIER, E., LANDÈS-DEVAUCHELLE, C., NITSCHKÉ, P. & RISLER, J.-L. 1999. Codon usage as a tool to predict the cellular location of eukaryotic ribosomal proteins and aminoacyl-tRNA synthetases. *Nucleic acids research,* 27**,** 2848-2851.

COCHELLA, L. & GREEN, R. 2004. Wobble during decoding: more than third-position promiscuity. *Nature structural & molecular biology,* 11**,** 1160.

CONG, L., RAN, F. A., COX, D., LIN, S., BARRETTO, R., HABIB, N., HSU, P. D., WU, X., JIANG, W. & MARRAFFINI, L. A. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science,* 339**,** 819-823.

CRICK, F. 1970. Central dogma of molecular biology. *Nature,* 227**,** 561.

CUSACK, S. 1997. Aminoacyl-tRNA synthetases. *Current opinion in structural biology,* 7**,** 881-889.

DE DONATIS, G. M., SINGH, S. K., VISWANATHAN, S. & MAURIZI, M. R. 2010. A single ClpS monomer is sufficient to direct the activity of the ClpA hexamer. *Journal of Biological Chemistry,* 285**,** 8771-8781.

DOTTAVIO-MARTIN, D., SUTTLE, D. P. & RAVEL, J. M. 1979. The effects of initiation factors IF-1 and IF-3 on the dissociation of Escherichia coli 70 S ribosomes. *FEBS letters,* 97**,** 105-110.

DOUGAN, D., TRUSCOTT, K. & ZETH, K. 2010. The bacterial N-end rule pathway: expect the unexpected. *Molecular microbiology,* 76**,** 545-558.

DOUGAN, D. A., REID, B. G., HORWICH, A. L. & BUKAU, B. 2002. ClpS, a substrate modulator of the ClpAP machine. *Molecular cell,* 9**,** 673-683.
DUNLOP, R. A., BRUNK, U. T. & RODGERS, K. J. 2009. Oxidized proteins: mechanisms of removal and consequences of accumulation. *IUBMB life,* 61**,** 522-527.
ENDY, D. 2005. Foundations for engineering biology. *Nature,* 438**,** 449.
ERBSE, A., SCHMIDT, R., BORNEMANN, T., SCHNEIDER-MERGENER, J., MOGK, A., ZAHN, R., DOUGAN, D. & BUKAU, B. 2006. ClpS is an essential component of the N-end rule pathway in Escherichia coli. *Nature,* 439**,** 753.
FARRELL, C. M., GROSSMAN, A. D. & SAUER, R. T. 2005. Cytoplasmic degradation of ssrA-tagged proteins. *Molecular microbiology,* 57**,** 1750-1761.
FRIED, S. D., SCHMIED, W. H., UTTAMAPINANT, C. & CHIN, J. W. 2015. Ribosome subunit stapling for orthogonal translation in E. coli. *Angewandte Chemie International Edition,* 54**,** 12791-12794.
FRUMKIN, I., LAJOIE, M. J., GREGG, C. J., HORNUNG, G., CHURCH, G. M. & PILPEL, Y. 2018. Codon usage of highly expressed genes affects proteome-wide translation efficiency. *Proceedings of the National Academy of Sciences,* 115**,** E4940-E4949.
GIEGÉ, R., SISSLER, M. & FLORENTZ, C. 1998. Universal rules and idiosyncratic features in tRNA identity. *Nucleic acids research,* 26**,** 5017-5035.
GOODY, R. S., ALEXANDROV, K. & ENGELHARD, M. 2002. Combining chemical and biological techniques to produce modified proteins. *ChemBioChem,* 3**,** 399-403.
GOVINDAN, A., MIRYALA, S., MONDAL, S. & VARSHNEY, U. 2018. Development of assay systems for amber codon decoding at the steps of initiation and elongation in mycobacteria. *Journal of bacteriology,* 200**,** e00372-18.
GUALERZI, C. O. & PON, C. L. 1990. Initiation of mRNA translation in prokaryotes. *Biochemistry,* 29**,** 5881-5889.
GUENNEUGUES, M., CASERTA, E., BRANDI, L., SPURIO, R., MEUNIER, S., PON, C. L., BOELENS, R. & GUALERZI, C. O. 2000. Mapping the fMet-tRNAfMet binding site of initiation factor IF2. *The EMBO journal,* 19**,** 5233-5240.
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 & development,* 4**,** 1790-1800.
HAURYLIUK, V. & EHRENBERG, M. 2006. Two-step selection of mRNAs in initiation of protein synthesis. *Molecular cell,* 22**,** 155-156.
HECHT, A., GLASGOW, J., JASCHKE, P. R., BAWAZER, L. A., MUNSON, M. S., COCHRAN, J. R., ENDY, D. & SALIT, M. 2017. Measurements of translation initiation from all 64 codons in E. coli. *Nucleic acids research,* 45**,** 3615-3626.
HOSHIKA, S., LEAL, N. A., KIM, M.-J., KIM, M.-S., KARALKAR, N. B., KIM, H.-J., BATES, A. M., WATKINS, N. E., SANTALUCIA, H. A. & MEYER, A. J. 2019. Hachimoji DNA and RNA: A genetic system with eight building blocks. *Science,* 363**,** 884-887.
HUMBARD, M. A., SURKOV, S., DE DONATIS, G. M., JENKINS, L. M. & MAURIZI, M. R. 2013. The N-degradome of Escherichia coli LIMITED PROTEOLYSIS IN VIVO GENERATES A LARGE POOL OF PROTEINS BEARING N-DEGRONS. *Journal of Biological Chemistry,* 288**,** 28913-28924.
HUSSAIN, T., LLÁCER, J. L., WIMBERLY, B. T., KIEFT, J. S. & RAMAKRISHNAN, V. 2016. Large-scale movements of IF3 and tRNA during bacterial translation initiation. *Cell,* 167**,** 133-144. e13.
HUTVAGNER, A. J. 2018. *Engineered semi-orthogonal initiator tRNAs initiate translation at non-canonical start codons.*, Master of Research, Macqurie University.
JINEK, M., CHYLINSKI, K., FONFARA, I., HAUER, M., DOUDNA, J. A. & CHARPENTIER, E. 2012. A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *science,* 337**,** 816-821.

JULIÁN, P., MILON, P., AGIRREZABALA, X., LASSO, G., GIL, D., RODNINA, M. V. & VALLE, M. 2011. The Cryo-EM structure of a complete 30S translation initiation complex from Escherichia coli. *PLoS biology,* 9**,** e1001095.

KAPOOR, S., DAS, G. & VARSHNEY, U. 2010. Crucial contribution of the multiple copies of the initiator tRNA genes in the fidelity of tRNA fMet selection on the ribosomal P-site in Escherichia coli. *Nucleic acids research,* 39**,** 202-212.

KIM, G., WEISS, S. J. & LEVINE, R. L. 2014. Methionine oxidation and reduction in proteins. *Biochimica et Biophysica Acta (BBA)-General Subjects,* 1840**,** 901-905.

KRAMER, E. B. & FARABAUGH, P. J. 2007. The frequency of translational misreading errors in E. coli is largely determined by tRNA competition. *Rna,* 13**,** 87-96.

KUNJAPUR, A. M., STORK, D. A., KURU, E., VARGAS-RODRIGUEZ, O., LANDON, M., SÖLL, D. & CHURCH, G. M. 2018. Engineering posttranslational proofreading to discriminate nonstandard amino acids. *Proceedings of the National Academy of Sciences,* 115**,** 619-624.

LAGERKVIST, U. 1978. " Two out of three": an alternative method for codon reading. *Proceedings of the National Academy of Sciences,* 75**,** 1759-1762.

LAJOIE, M. J., ROVNER, A. J., GOODMAN, D. B., AERNI, H.-R., HAIMOVICH, A. D., KUZNETSOV, G., MERCER, J. A., WANG, H. H., CARR, P. A. & MOSBERG, J. A. 2013. Genomically recoded organisms expand biological functions. *science,* 342**,** 357-360.

LAURSEN, B. S., SØRENSEN, H. P., MORTENSEN, K. K. & SPERLING-PETERSEN, H. U. 2005. Initiation of protein synthesis in bacteria. *Microbiol. Mol. Biol. Rev.,* 69**,** 101-123.

LE, Y., MURPHY, P. M. & WANG, J. M. 2002. Formyl-peptide receptors revisited. *Trends in immunology,* 23**,** 541-548.

LIU, C. C., JEWETT, M. C., CHIN, J. W. & VOIGT, C. A. 2018. Toward an orthogonal central dogma. *Nature chemical biology,* 14**,** 103.

LOBANOV, A. V., TURANOV, A. A., HATFIELD, D. L. & GLADYSHEV, V. N. 2010. Dual functions of codons in the genetic code. *Critical reviews in biochemistry and molecular biology,* 45**,** 257-265.

LOVE, M., ANDERS, S. & HUBER, W. 2014a. Differential analysis of count data–the DESeq2 package. *Genome Biol,* 15**,** 10.1186.

LOVE, M. I., HUBER, W. & ANDERS, S. 2014b. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology,* 15**,** 550.

MA, N. J. & ISAACS, F. J. 2016. Genomic recoding broadly obstructs the propagation of horizontally transferred genetic elements. *Cell systems,* 3**,** 199-207.

MALYS, N. 2012. Shine-Dalgarno sequence of bacteriophage T4: GAGG prevails in early genes. *Molecular biology reports,* 39**,** 33-39.

MALYSHEV, D. A., DHAMI, K., LAVERGNE, T., CHEN, T., DAI, N., FOSTER, J. M., CORRÊA, I. R. & ROMESBERG, F. E. 2014. A semi-synthetic organism with an expanded genetic alphabet. *Nature,* 509**,** 385.

MANDELL, D. J., LAJOIE, M. J., MEE, M. T., TAKEUCHI, R., KUZNETSOV, G., NORVILLE, J. E., GREGG, C. J., STODDARD, B. L. & CHURCH, G. M. 2015. Biocontainment of genetically modified organisms by synthetic protein design. *Nature,* 518**,** 55.

MAYER, C., KÖHRER, C., KENNY, E., PRUSKO, C. & RAJBHANDARY, U. L. 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.

MAYER, C. & RAJBHANDARY, U. L. 2002. Conformational change of Escherichia coli initiator methionyl-tRNA fMet upon binding to methionyl-tRNA formyl transferase. *Nucleic acids research,* 30**,** 2844-2850.

MEINNEL, T. & BLANQUET, S. 1995. Enzymatic properties of Escherichia coli peptide deformylase. *Journal of bacteriology,* 177**,** 1883-1887.
MILON, P., KONEVEGA, A. L., GUALERZI, C. O. & RODNINA, M. V. 2008. Kinetic checkpoint at a late step in translation initiation. *Molecular cell,* 30**,** 712-720.
MOLL, I., GRILL, S., GUALERZI, C. O. & BLÄSI, U. 2002. Leaderless mRNAs in bacteria: surprises in ribosomal recruitment and translational control. *Molecular microbiology,* 43**,** 239-246.
MOON, C., STUPP, G. S., SU, A. I. & WOLAN, D. W. 2018. Metaproteomics of colonic microbiota unveils discrete protein functions among colitic mice and control groups. *Proteomics,* 18**,** 1700391.
MUKAI, T., HAYASHI, A., IRAHA, F., SATO, A., OHTAKE, K., YOKOYAMA, S. & SAKAMOTO, K. 2010. Codon reassignment in the Escherichia coli genetic code. *Nucleic acids research,* 38**,** 8188-8195.
MURPHY IV, F. V. & RAMAKRISHNAN, V. 2004. Structure of a purine-purine wobble base pair in the decoding center of the ribosome. *Nature structural & molecular biology,* 11**,** 1251.
NEUMANN, H., SLUSARCZYK, A. L. & CHIN, J. W. 2010. De novo generation of mutually orthogonal aminoacyl-tRNA synthetase/tRNA pairs. *Journal of the American Chemical Society,* 132**,** 2142-2144.
OGAWA, T. & OKAZAKI, T. 1980. Discontinuous DNA replication. *Annual review of biochemistry,* 49**,** 421-457.
OSAWA, S. & JUKES, T. H. 1989. Codon reassignment (codon capture) in evolution. *Journal of molecular evolution,* 28**,** 271-278.
OSTROV, N., LANDON, M., GUELL, M., KUZNETSOV, G., TERAMOTO, J., CERVANTES, N., ZHOU, M., SINGH, K., NAPOLITANO, M. G. & MOOSBURNER, M. 2016. Design, synthesis, and testing toward a 57-codon genome. *Science,* 353**,** 819-822.
PADDON, C. J. & KEASLING, J. D. 2014. Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. *Nature Reviews Microbiology,* 12**,** 355.
PALLANCK, L. & SCHULMAN, L. H. 1991. Anticodon-dependent aminoacylation of a noncognate tRNA with isoleucine, valine, and phenylalanine in vivo. *Proceedings of the National Academy of Sciences,* 88**,** 3872-3876.
PAN, Z. K., CHEN, L.-Y., COCHRANE, C. G. & ZURAW, B. L. 2000. fMet-Leu-Phe stimulates proinflammatory cytokine gene expression in human peripheral blood monocytes: the role of phosphatidylinositol 3-kinase. *The Journal of Immunology,* 164**,** 404-411.
PARKER, J. 1989. Errors and alternatives in reading the universal genetic code. *Microbiological reviews,* 53**,** 273.
PÉDELACQ, J.-D., CABANTOUS, S., TRAN, T., TERWILLIGER, T. C. & WALDO, G. S. 2006. Engineering and characterization of a superfolder green fluorescent protein. *Nature biotechnology,* 24**,** 79.
PEREZ-PINERA, P., KOCAK, D. D., VOCKLEY, C. M., ADLER, A. F., KABADI, A. M., POLSTEIN, L. R., THAKORE, P. I., GLASS, K. A., OUSTEROUT, D. G. & LEONG, K. W. 2013. RNA-guided gene activation by CRISPR-Cas9–based transcription factors. *Nature methods,* 10**,** 973.
PIATKOV, K. I., VU, T. T., HWANG, C.-S. & VARSHAVSKY, A. 2015. Formyl-methionine as a degradation signal at the N-termini of bacterial proteins. *Microbial Cell,* 2**,** 376.
RACKHAM, O. & CHIN, J. W. 2005. A network of orthogonal ribosome· mRNA pairs. *Nature chemical biology,* 1**,** 159.
RAVIKUMAR, A., ARRIETA, A. & LIU, C. C. 2014. An orthogonal DNA replication system in yeast. *Nature chemical biology,* 10**,** 175.
SCHMIDT, M. 2012. *Synthetic biology: industrial and environmental applications*, John Wiley & Sons.

SCHULMAN, L. H. & PELKA, H. 1985. In vitro conversion of a methionine to a glutamine-acceptor tRNA. *Biochemistry,* 24**,** 7309-7314.
SCHULMAN, L. H. & PELKA, H. 1988. Anticodon switching changes the identity of methionine and valine transfer RNAs. *Science,* 242**,** 765-768.
SCHULTZ, D. W. & YARUS, M. 1996. On malleability in the genetic code. *Journal of molecular evolution,* 42**,** 597-601.
SELMER, M., DUNHAM, C. M., MURPHY, F. V., WEIXLBAUMER, A., PETRY, S., KELLEY, A. C., WEIR, J. R. & RAMAKRISHNAN, V. 2006. Structure of the 70S ribosome complexed with mRNA and tRNA. *Science,* 313**,** 1935-1942.
SENKO, M. W., BEU, S. C. & MCLAFFERTYCOR, F. W. 1995. Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions. *Journal of the American Society for Mass Spectrometry,* 6**,** 229-233.
SHARP, P. A. 2009. The centrality of RNA. *Cell,* 136**,** 577-580.
SHARP, P. M. & LI, W.-H. 1987. The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic acids research,* 15**,** 1281-1295.
SIMONETTI, A., MARZI, S., MYASNIKOV, A. G., FABBRETTI, A., YUSUPOV, M., GUALERZI, C. O. & KLAHOLZ, B. P. 2008. Structure of the 30S translation initiation complex. *Nature,* 455**,** 416.
SUSSMAN, J. K., SIMONS, E. L. & SIMONS, R. W. 1996. Escherichia coli translation initiation factor 3 discriminates the initiation codon in vivo. *Molecular microbiology,* 21**,** 347-360.
TABOR, S. & RICHARDSON, C. C. 1985. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proceedings of the National Academy of Sciences,* 82**,** 1074-1078.
TEMME, K., ZHAO, D. & VOIGT, C. A. 2012. Refactoring the nitrogen fixation gene cluster from Klebsiella oxytoca. *Proceedings of the National Academy of Sciences,* 109**,** 7085-7090.
TOBIAS, J. W., SHRADER, T. E., ROCAP, G. & VARSHAVSKY, A. 1991. The N-end rule in bacteria. *Science,* 254**,** 1374-1377.
TOMITA, K., UEDA, T. & WATANABE, K. 1999. The presence of pseudouridine in the anticodon alters the genetic code: a possible mechanism for assignment of the AAA lysine codon as asparagine in echinoderm mitochondria. *Nucleic acids research,* 27**,** 1683-1689.
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. *Journal of Biological Chemistry,* 279**,** 8539-8546.
VARSHAVSKY, A. 2011. The N-end rule pathway and regulation by proteolysis. *Protein science,* 20**,** 1298-1345.
VARSHNEY, U., LEE, C. P., SEONG, B. L. & RAJBHANDARY, U. L. 1991. Mutants of initiator tRNA that function both as initiators and elongators. *Journal of Biological Chemistry,* 266**,** 18018-18024.
VARSHNEY, U. & RAJBHANDARY, U. L. 1990. Initiation of protein synthesis from a termination codon. *Proceedings of the National Academy of Sciences,* 87**,** 1586-1590.
VARSHNEY, U. & RAJBHANDARY, U. L. 1992. Role of methionine and formylation of initiator tRNA in initiation of protein synthesis in Escherichia coli. *Journal of bacteriology,* 174**,** 7819-7826.
VINCENT, R. M. 2017. Reprogramming the genetic code: defining a new start codon in Escherichia coli. *Masters of Research Macquarie University*.
VINCENT, R. M., WRIGHT, B. W. & JASCHKE, P. R. 2019. Measuring amber initiator tRNA orthogonality in a genomically recoded organism. *ACS synthetic biology,* 8**,** 675-685.
WANG, K., NEUMANN, H., PEAK-CHEW, S. Y. & CHIN, J. W. 2007. Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nature biotechnology,* 25**,** 770.

WANG, K., SACHDEVA, A., COX, D. J., WILF, N. M., LANG, K., WALLACE, S., MEHL, R. A. & CHIN, J. W. 2014. Optimized orthogonal translation of unnatural amino acids enables spontaneous protein double-labelling and FRET. *Nature chemistry,* 6**,** 393.

WANG, L., BROCK, A., HERBERICH, B. & SCHULTZ, P. G. 2001. Expanding the genetic code of Escherichia coli. *Science,* 292**,** 498-500.

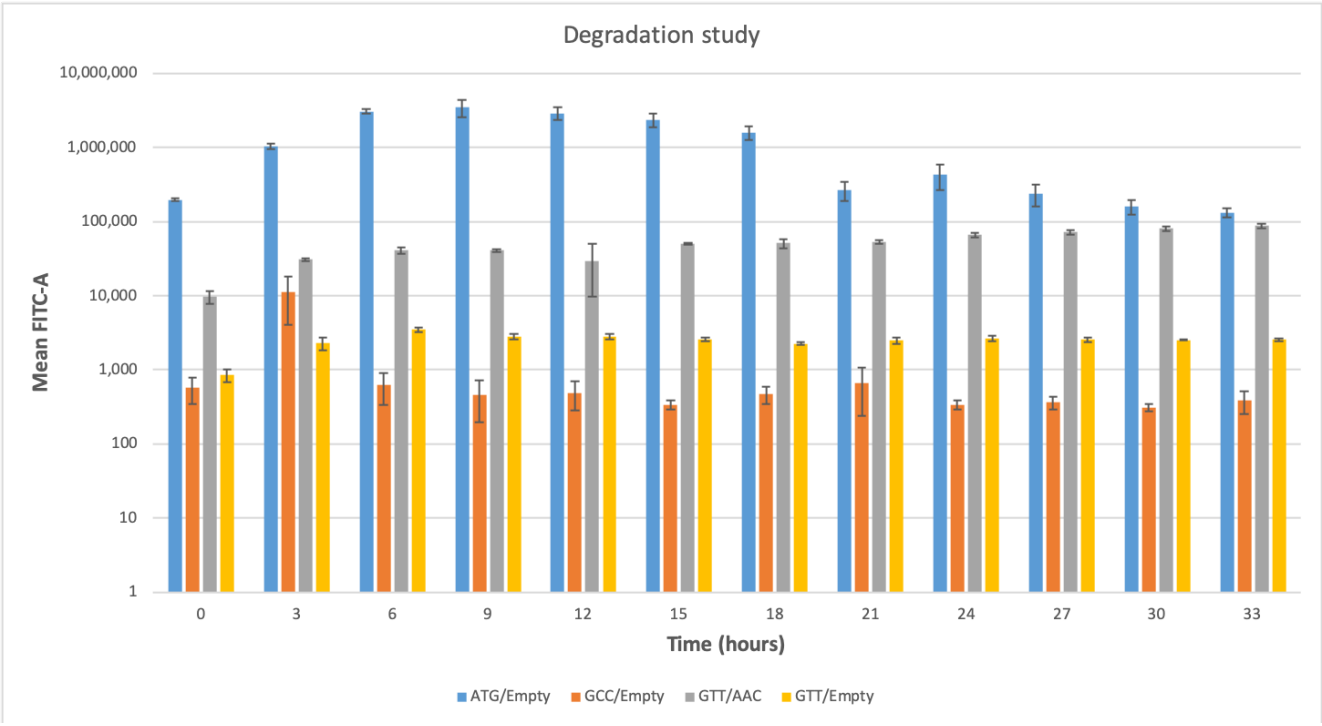
WHITAKER, W. R., LEE, H., ARKIN, A. P. & DUEBER, J. E. 2014. Avoidance of truncated proteins from unintended ribosome binding sites within heterologous protein coding sequences. *ACS synthetic biology,* 4**,** 249-257.

WU, G., YAN, Q., JONES, J. A., TANG, Y. J., FONG, S. S. & KOFFAS, M. A. 2016. Metabolic burden: cornerstones in synthetic biology and metabolic engineering applications. *Trends in biotechnology,* 34**,** 652-664.

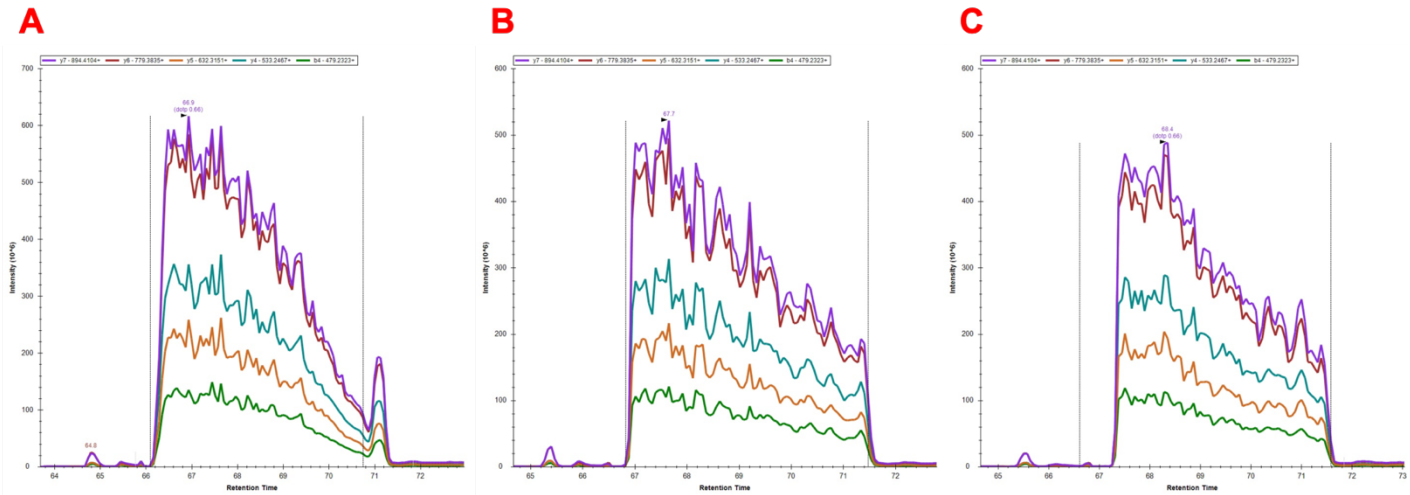
WU, X.-Q. & RAJBHANDARY, U. L. 1997. Effect of the amino acid attached to Escherichia coli initiator tRNA on its affinity for the initiation factor IF2 and on the IF2 dependence of its binding to the ribosome. *Journal of Biological Chemistry,* 272**,** 1891-1895.

YAMAMOTO, H., WITTEK, D., GUPTA, R., QIN, B., UEDA, T., KRAUSE, R., YAMAMOTO, K., ALBRECHT, R., PECH, M. & NIERHAUS, K. H. 2016. 70S-scanning initiation is a novel and frequent initiation mode of ribosomal translation in bacteria. *Proceedings of the National Academy of Sciences,* 113**,** E1180-E1189.

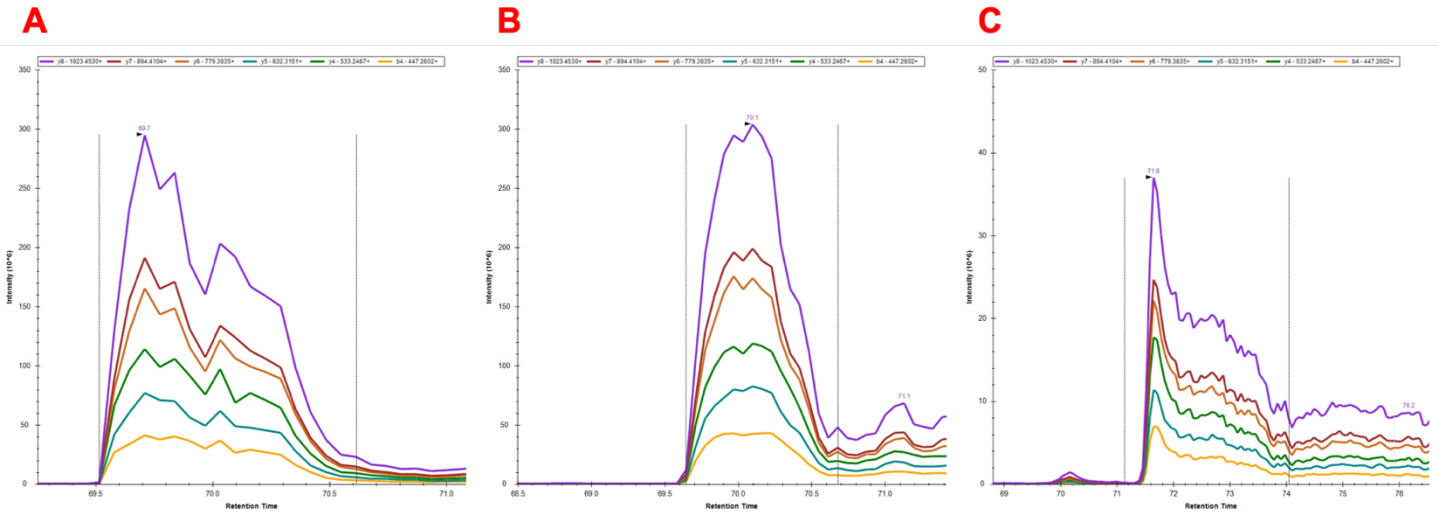
Supplementary material



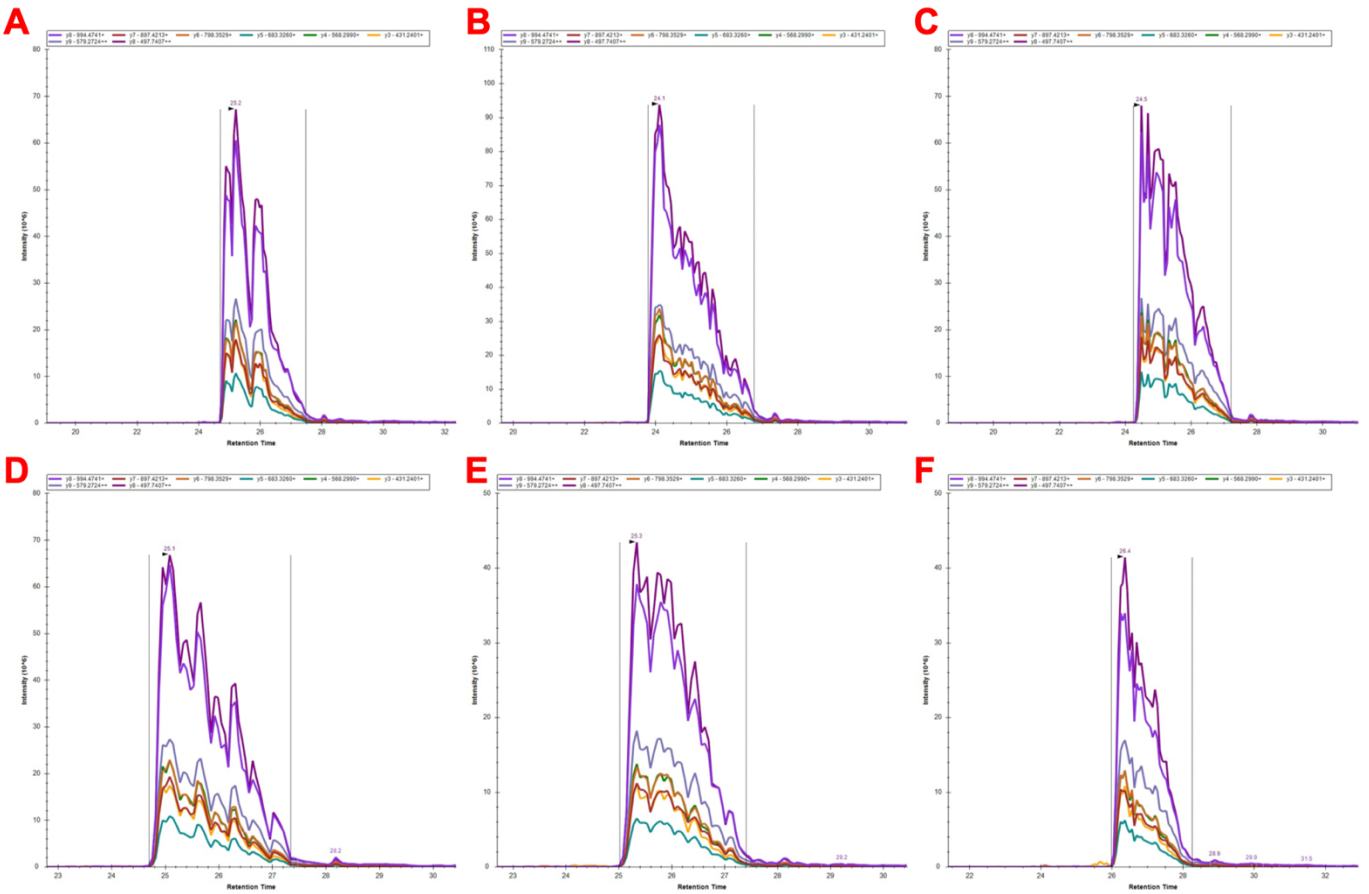
Supplementary Figure 1. Raw flow cytometry fluorescence data. Raw mean FITC-A fluorescence counts for four strains used in protein degradation study. *E. coli* strains with sfGFP (AUG) / i-tRNA^{CAU}, sfGFP (GCC) / i-tRNA^{CAU}, sfGFP (GUU) / i-tRNA^{AAC}, sfGFP (GUU) / i-tRNA^{CAU}.



Supplementary Figure 2. MS2 spectra of Methionine bearing N-terminal peptide in NanoLuc (AUG) / i-tRNA^{CAU} replicates. (A) Replicate 1. (B) Replicate 2. (C) Replicate 3. b-ions identified in all replicates identifying N-terminal methionine.



Supplementary Figure 3. MS2 spectra of Valine bearing N-terminal peptide in NanoLuc (GUU) / i-tRNA^{AAC} replicates. (A) Replicate 1. (B) Replicate 2. (C) Replicate 3. b-ions identified in all replicates identifying N-terminal valine.



Supplementary Figure 4. MS2 spectra of internal peptide in NanoLuc (GUU) / i-tRNA^{AAC} replicates and NanoLuc (AUG) / i-tRNA^{CAU} replicates. (A) AUG-Replicate 1. (B) AUG-Replicate 2. (C)AUG- Replicate 3. (D) GUU-Replicate 1. (E) GUU-Replicate 2. (F) GUU-Replicate 3. Internal peptide sequence VVYPVDDH identified in each condition.