Investigating Genetic Code Plasticity in Transient Expression Systems

Towards alternative genetic codes:
development of bacterial circuits for the robust and
efficient detection of transient sense-to-sense codon reassignment

Yan-Kay Ho

This thesis is submitted for the degree of
Doctor of Philosophy in Synthetic Biology and Biotechnology

Institute of Structural and Molecular Biology
University College London

As supervised by *Prof. Vitor Pinheiro* and *Prof. John Ward*,
with *Prof. Joanne Santini* as thesis chair.

08 August 2024

Declaration of Authorship

I, Yan-Kay Ho, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Signed: _____

Dated: 08 August 2024

"The kind of control you're attempting simply is... it's not possible. If there's one thing the history of evolution has taught us, it's that life will not be contained. Life breaks free, it expands to new territories, and crashes through barriers, painfully, maybe even dangerously [...]
I'm simply saying that life [pause] finds a way."

-– Dr. Ian Malcolm (played by Jeff Goldblum)
Jurassic Park, 1993

"Just because someone stumbles and loses their path, doesn't mean they're lost forever. Sometimes, we all need a little help."

-– Charles Xavier (played by James McAvoy) and
Professor X (played by Sir Patrick Stewart)
X-Men: Days of Future Past, 2014

"Two steps forward, one step back, is still one step forward."

-– Detective Rosalita "Rosa" Diaz (played by Stephanie Beatriz)
"He Said, She Said" (S6E8), Brooklyn Nine-Nine, 2019

Abstract

The natural genetic code, bar some exceptions, is universal. It dictates a specific and unambiguous set of rules regarding the assignment between RNA triplet (genetic information) and amino acid (protein product), which is enforced by transfer RNAs (tRNAs) and aminoacyl-tRNA synthetases (aaRSs). The latter ensures that amino acids are specifically charged to their cognate tRNAs, whilst the former establishes the physical bridge between amino acid and the associated messenger RNA (mRNA) triplet. However, these links can be rewritten by engineering the identity elements of tRNAs and/or by altering the substrate specificity of aaRSs to produce a modified genetic code. Alternative viable genetic codes, with reassigned codons, could be used to create enhanced biocontainment systems that are unable to contaminate the environment with engineered traits.

Leucine-to-histidine reassignment (using the mRNA sequence as the frame of reference) can be accomplished by altering the anticodon region of *E. coli* histidyl-tRNA (*Eco*tRNA^{His}) such that it base-pairs with leucine codons, whilst maintaining compatibility with its cognate histidyl-tRNA synthetase (*Eco*HisRS). Equivalent modifications to *E. coli* leucyl-tRNA (*Eco*tRNA^{Leu}) should generate the complementary histidine-to-leucine reassignment. Such sense-to-sense codon reassignments are expected to produce a significant degree of toxicity *in vivo*, making it difficult to detect reassignments in heterologously-expressed histidine/leucine-recoded reporters. By incorporating serine recombinase phage integrase PhiC31 (ΦC31) with these codon-modified reporters, a biological toggle circuit platform was developed that produces a secondary non-toxic and permanent response to augment any initial transient signals of codon reassignment.

I demonstrate that the developed platform is functional, and that leucine-to-histidine reassignment can be robustly detected. The genetic toggle switch assay, the first-of-its-kind, can be adapted to other codon reassignments and therefore represents a feasible route towards testing and characterising alternative sense-to-sense codon reassignments. Despite significant optimisation, the counter histidine-to-leucine reassignment, which would create the first viable orthogonal genetic code using the native translational machinery, was not successful in the time-frame of the project.

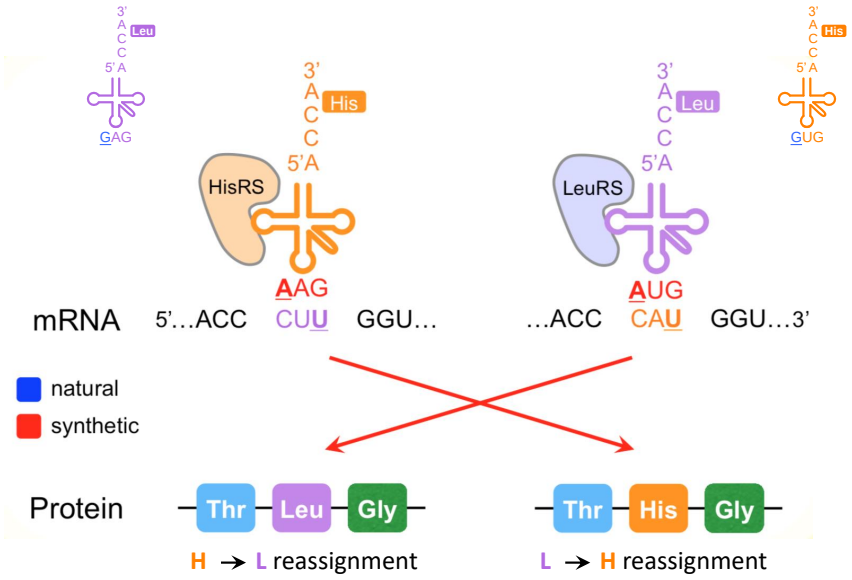


Figure 1: Implementing alternative sense-to-sense codon reassignments

In the natural system: histidyl-tRNA synthetase (HisRS) would charge histidyl-tRNA ($tRNA^{His}$) with histidine; the histidylated/aminoacylated $tRNA^{His}_{GUG}$ (His-$tRNA^{His}_{GUG}$) would recognise and bind the histidine codon (e.g. CAU) on an mRNA sequence and incorporate histidine into the nascent polypeptide chain (i.e. His-to-His assignment).
In the proposed synthetic system: HisRS would charge an *anticodon-modified* $tRNA^{His}$ with histidine, but as the anticodon for this $tRNA^{His}$ has been changed (from GUG) to AAG*, His-$tRNA^{His}_{AAG}$ would bind the *leucine* codon (i.e. CUU) on an mRNA sequence and instead incorporate histidine into the nascent polypeptide chain (i.e. Leu-to-His reassignment).
Equivalent modifications to the anticodon for $tRNA^{Leu}$ (from GAG to AUG*; $tRNA^{Leu}_{AUG}$) should provide the complementary His-to-Leu reassignment at the *histidine* (wobble-)codon CAU.

*Underlined nucleotides in blue show natural wobble base-pairing; underlined nucleotides in red show anticodon modification for complete Watson-Crick base-pairing between tRNA anticodon triplet and mRNA codon.

Impact Statement

Being able to investigate the plasticity of the genetic code through sense-to-sense codon reassignment provides a valuable approach to address fundamental questions about the process of translation in the central dogma of biology, investigate whether the universal decoding rules from RNA to protein was a historical "frozen accident", and explore if alternate coding systems could have future applications as additional biocontainment strategies in genetically modified organisms (GMOs).

For researchers interested in examining some of the key molecules used in translation, the Orphan Codon Invasion (OCI) system developed in Chapter 3 could be used to study the interactions between non-natural tRNAs and aaRSs pairings *in vivo*. In this research, this was to create anticodon-modified tRNAs that could be accepted by endogenous synthetases, and be co-opted into developing genetic code reassignments. By extension, changing the rule set for decoding the universal genetic code could shed light on whether the current natural code for all living organisms is immutable or, more likely, a result of a local optimum that was propagated throughout the tree of life. Moreover, the system developed could also be used to investigate other changes to the translation components: from analysis of tRNA structure in relation to function (e.g. through minimal tRNAs truncation experiments), to uncovering cross-species interactions of tRNA/aaRS sets, or for conducting co-evolution experiments to discover new aaRSs that can aminoacylate tRNAs and incorporate non-canonical amino acid analogues (i.e. added functionalities for genetic code expansion).

The effects of shifting from an endogenous/wild-type (WT) pair of tRNA/aaRS to that of an introduced homologous or mutated set can be further probed using the

genomic toggle Cre-Lox switch system initially developed in earlier work (data not shown), in a manner similar to developmental biology studies. The prerequisites for this approach, however, are that natural and synthetic tRNA/aaRS sets would need to be genomically integrated, and that both should be compatible with the host system. Alternatively, if it is unclear whether the synthetic set would present problems for the host, the combination of both the OCI system and the Genetic Toggle Switch (GTS) (a new method for investigating Orphan Codon Invasion *in vivo* with reduced toxicity to the host organism, as described in Chapter 4) – the resulting OCI-GTS platform as presented at the end of this chapter offers a heterologously-expressed approach for implementing sense-to-sense codon reassignment (and other engineered systems) with minimal impact to the host. Coupled with the recently published approaches for whole genome recoding by the Church and Chin groups (Wang et al., 2009a; de la Torre, 2020; Wang et al., 2016), the (Orphan Codon Invasion)-(Genetic Toggle Switch) (OCI-GTS) system would be a vastly more viable, feasible, and robust method for generating a variety of alternate genetic codes that have not previously been conceived, nor considered possible.

Together, the Orphan Codon Invasion and Genetic Toggle Switch systems developed here provides the initial steps towards enhanced biosafety in the production of chemicals, materials, and products from biocontained genetically modified or recoded organisms (GMOs/GROs), without contaminating the natural world.

Acknowledgements

There are many people to whom I owe my heartfelt thanks, who have supported me through thick and thin, in all manner of ways, along this long and winding journey of a PhD. I don't think I know how to truly thank you all, but I will give it my best.

To Prof. Vitor B Pinheiro – you taught me all I know about being an open, independent, and collaborative researcher, and I carry your lessons with me every day. Hopefully, I have also passed on a fraction of these experiences to those who I shared a lab and team with since then. Prof. John Ward and Dr. Joanne Santini, your insights helped shape my formative research and writing years, and hopefully some of it made its way into this final thesis.

To Dr. Eszter Csibra, Dr. Leticia Torres, Dr. Chris Cozens, Dr. Pedro Tizei, Dr. Antje Krüger, Dr. Marleen Renders, Paola Handal Marquez, Ana Riesco Gomez-Manor, Barbara Steijl, Dr. Luba Prout – you guys really made it all worth it to be in (and out!) the lab together. I couldn't ask for better people to share in the trials and tribulations of a lab life, and that's not even saying anything about the friendship, knowledge, shared experiences, and joy that you each brought with you. Roxie Mironska, Dr. Hugo Sinclair, Dr. Kwasi Amoako Kwakwa, Dr. Charmian Dawson, and Dr. Henry Bayele – you made life in our shared office all the more lively and fun. I miss having chats with each and every one of you.

To Prof. Chris Barnes and my adopted lab: Dr. Alex Fedorec, Dr. Linda Dekker, Dr. Ke Yan Wen, Dr. Bez Karkaria, Dr. Jack Rutter – you took me in when our group dwindled, and you always made me feel welcome, thank you.

To those unnamed: TZ, MR; RH, LK, LB – when times were tough, you made me smile and lent a friendly ear. You'll never know much much I needed that.

To Dr. Hongorzul Davaapil, Dr. Emma Barton, Dr. Cam Watson, Dr. Eszter Csibra, and Dr. Mike Nicoll – how you dealt with my constant refrain of "I'm still working at it..." I'll never know. This time, I do hope the end is in sight.

To Dr. Charmian Dawson, Mariam, and Jannah – because you needed a disembodied voice in your life for the past three(?!) years; or was it I who needed it? Either ways, you really welcomed me into your lives, and provided me the support I needed at every turn, be it a chat, a laugh, listening, notetaking, proofreading, transcribing, scripting, questioning, improving. I think you name it, and it's been done. Oh yeah, ... editing.

Last, but far from least, I want to thank my family: Dad, Mum, YK, CK, Gammy, Einir, and Jamie – for your unwavering support, untold patience, constant presence, and words of encouragement. Thank you for always being here for me. I love you all.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 A General Overview

Our growing understanding of biology, together with recent advances in foundational technologies, such as large-scale DNA synthesis, is bringing us ever closer to being able to engineer biology: faster and further than we could have ever before, and without the limitation of working within just a single species at a time. Because biology has the potential to autonomously replicate and evolve, biological engineering raises an ethical question regarding the possibility of contaminating the environment, with obvious parallels to the introduction of invasive species. Thus, the idea of biocontainment emerges and there are numerous approaches to implement it – each with their advantages and limitations.

Minimising the possibility (and ideally eliminating the risk) of engineered organisms breaking containment is a key aspect of making genetic engineering safer: not just to us but to our current ecosystem. However, such a simple statement misses the complexity of the task. An engineered organism needs to be useful and efficient (for the task it was designed), and it needs those traits maintained during function. If escaped, the engineered organism can interact with the environment, it can lose genetic material, it can acquire genetic material, and its genetic material can evolve. Therefore, "ideal" biocontainment addresses all those processes (via a number of different mechanisms; Torres et al., 2016) without significantly compromising the value of the organism for its application.

To date, no single biocontainment strategy has delivered a solution to all the problems described, and only recently have strategies been proposed to hamper the exchange of genetic information with the environment. One such approach lies on engineering the quasi-universal genetic code to develop an alternative: a viable genetic code that is efficient, but that remains incompatible with the natural one. Under those circumstances, the possibility of exchanging genetic material with the environment is removed, while requiring no changes in the cellular machinery itself – therefore maintaining the efficiency of the engineered organism. This would be referred to as sense-to-sense codon reassignment. To implement such reassignment, the specific and unambiguous set of rules that dictate the linkage between the genetic information encoded in ribonucleic acid (RNA) triplet codons and the amino acids of protein products, as enforced by transfer RNAs (tRNAs) and aminoacyl-tRNA synthetases (aaRSs), must be rewritten.

In practice, this entails changing the correlation between tRNA anticodon and charged/aminoacylated-tRNA, while evading natural error-checking processes that may be in place to monitor that interaction. In principle, this can be achieved by engineering a tRNA to recognise a non-cognate (sense) codon by modifying their recognition or "identity" elements, while synthetases can be evolved to pair with and aminoacylate these alternate substrates, i.e. engineering the aaRS to recognise the "new" tRNA elements.

Döring and Marlière (1998) first investigated deviations from the natural genetic code with mutant tRNAs that incorporate cysteine (Cys) in place of isoleucine (Ile) and methionine (Met) codons in a heterologously-expressed "recoded" gene within *E. coli*. For clarity, in this thesis, the term "recoded" refers to the non-/synonymous replacement of codons that would produce functionally-expressing proteins under an alternate coding system, but would be considered missense mutations in the quasi-universal genetic code as recognised by natural organisms.

Whilst successful in demonstrating transient and incomplete isoleucine-to-cysteine reassignment in one position of the thymidylate synthase (ThyA) protein in *E. coli*, this was only part way towards establishing a functional alternative genetic

code. The original isoleucine tRNA isoacceptors ($tRNA^{Ile}_{\underline{G}AU}$ and $tRNA^{Ile}_{\underline{C'}AU}$) were still present alongside the new isoleucine-to-cysteine synthetic $tRNA^{Cys}_{\underline{N}AU}$ variant, and the complementary cysteine-to-isoleucine (reverse reassignment for a fully orthogonal code) was not implemented. Nonetheless, it was the first – and crucial – demonstration that such sense-to-sense codon reassignment was viable.

In the absence of cellular regulatory mechanisms to identify mischarged tRNAs, incomplete reassignment leads to two populations of amino acids charged to a single anticodon – in the case of Döring and Marlière (1998), the charging of $tRNA_{\underline{G}AU}$ with Ile or Cys. Whilst sufficient for a functional ThyA to be synthesised, it was anticipated that Cys would also be incorporated in other cellular proteins in positions where Ile is expected. Therefore, a potential challenge to implementing sense-to-sense codon reassignments *in vivo* would be the degree of toxicity these reassignments may confer to the host. On a small scale (e.g. with low concentration or short expression time of these codon reassignment tRNAs), a reduction in fitness of the host in this manner may be tolerated as it demonstrates an initial proof-of-concept. However, as a long-term goal of demonstrating sense-to-sense codon reassignment as a viable route for orthogonal genetic codes in genetically modified or recoded organisms (GMOs/GROs) without deleterious effects, all genes within the host (and expression of any exogenous DNA) should be recoded according to this alternative system.

While there are now feasible, yet still non-trivial, approaches to create fully recoded genomes (Wang et al., 2009a; de la Torre, 2020; Wang et al., 2016), these were not available at the time my research took place, therefore alternative strategies needed to be developed to bypass cellular toxicity, whilst demonstrating bilateral sense-to-sense codon reassignment in a living organism.

Instead, the approach taken to mitigate toxicity to host (due to gene mis-expression) was to adopt recombinase enzymes as genetic switches (e.g. St-Pierre et al., 2013). Briefly, the principle behind the approach was to use a recombinase as a reporter for codon reassignment. If codon reassignment leads to host toxicity, limiting the exposure to the altered code minimises toxicity, but also limits the

possibility of detecting reporter activity. By using a recombinase as a (primary) reporter, even transient activation of the codon reassignment would lead to active recombinase synthesis and a permanent change in the host genome. The expected result was to develop a platform to study codon reassignment without being hampered by the toxicity of those reassignments.

In this introduction, I first discuss the natural genetic code, the key protein translation machinery (specifically tRNA and aaRS) involved in its implementation, and certain natural variations found within the quasi-universal code. I then discuss how synthetic genetic codes have been developed, predominantly in pursuit of research into genetic code expansion, and how they could be further developed (at least in this body of work) for use in recoding and genetic code reassignment. This is followed by a brief summary of the potential implications of research into genetic code reassignment, both theoretical (increasing understanding of tRNA and aaRS biology) and applied (protein engineering and biocontainment strategies). Finally, I outline the research that has been carried out during the course of this thesis in developing a platform of genetic circuits for the robust and efficient detection of transient sense-to-sense codon reassignment in *E. coli*.

1.2 What is the Genetic Code?

It is well-established that every living organism on Earth adheres to rules of the quasi-universal genetic code. Yet, there have been, and are still, minor differences in the decoding of the code, and it is not clear whether those initial near-universal rules reflected a physical constraint or a "frozen accident" early on in evolution. Like in all living things, the genetic code within is always open to change – as long as the change brings sufficient advantages that outweigh the fitness costs necessary to adapt and maintain the novel variations in the complex environmental and ecological niche with which it finds itself. It is this potential mutability that led us to investigate just how malleable the genetic code is, and whether we can modify the code in ways not yet seen before. First, however, we should start with how the code is implemented today, before exploring its past and the different aspects in which the code (and its associated components) have since evolved.

Implementation of the genetic code, through the translation of an mRNA to a polypeptide, involves multiple processes (see Figure 1.1). Briefly, the genetic code is established by aaRSs (also known as synthetases) that catalyse the chemical conjugation of amino acids (AAs) and tRNAs. This conjugation is highly specific, with synthetases capable of recognising cognate amino acids (see Section 1.2.1.1) and tRNAs (see Section 1.2.1.2).

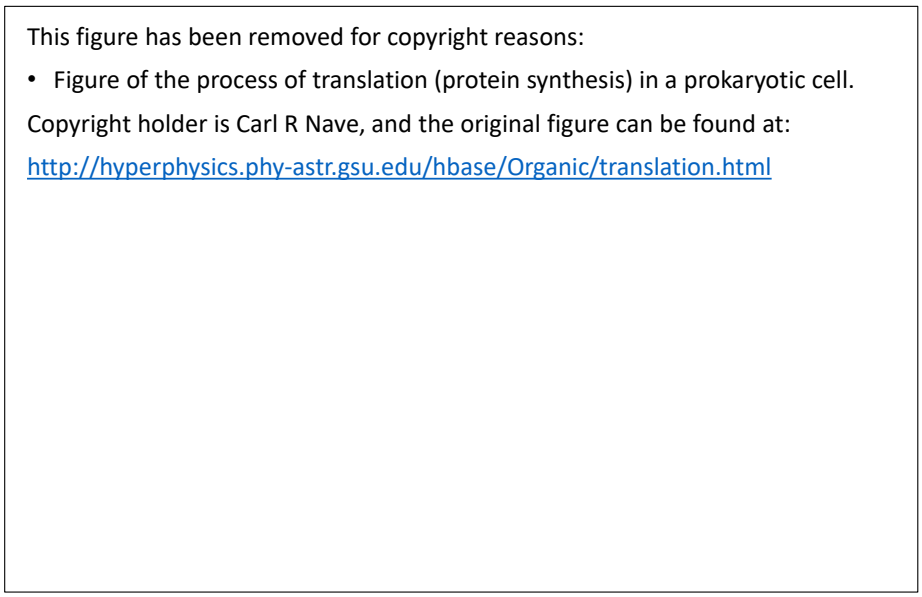


Figure 1.1: The process of translation (protein synthesis) in a prokaryotic cell

Protein synthesis is carried out within ribosomes, inside the cell. Aminoacylated tRNAs enter the ribosome, where the tRNA anticodon sequence can base-pair with the codons on the mRNA transcript, and transfer its amino acid to the nascent polypeptide chain. Once transferred, the deacylated tRNA exits the ribosome, where it can be charged again by an amino acid activated aminoacyl-tRNA synthetase. Figure from `http://hyperphysics.phy-astr.gsu.edu/hbase/Organic/translation.html` (last accessed: 06-Jan-2017).

Each tRNA is intrinsically associated to a single amino acid and the result is that there is a strong correlation between tRNA anticodon and amino acid identity. Thus, knowing the mRNA codon sequence, it is possible to predict the anticodon that will bind the mRNA, and hence which amino acid will be incorporated in that position by the ribosome – that is the genetic code.

AaRSs charge amino acids to tRNAs (a process known as aminoacylation) and therefore are the enzymes that establish the genetic code. The code is ancient,

conserved, and highly specific, which result from aaRSs having specificity towards their tRNA and amino acid substrates – and this can be species-specific. Theoretically, 20 synthetases should be necessary and sufficient to translate the genetic code, given that there are only 20 proteinogenic, canonical amino acids (cAAs). However, nature is more complex.

For example, some organisms can survive with less than the full set of aaRSs. In *E. coli*, glutamine (Gln) is charged directly to $tRNA^{Gln}$ by glutaminyl-tRNA synthetase (GlnRS). However, all archaebacteria, the majority of eubacteria, and most eukaryotic organelles lack the gene encoding GlnRS (Meng et al., 2022); instead, glutamine is incorporated into proteins by non-discriminating glutamyl-tRNA synthetases (GluRS) through a two-step reaction: first GluRS aminoacylates $tRNA^{Gln}$ with glutamate (Glu), before converting Glu-$tRNA^{Gln}$ to Gln-$tRNA^{Gln}$ via a tRNA-dependent amido-transferase (Glu-AdT; Hadd and Perona, 2014; Perona, 2013).

The complexity lies not just with the synthetases but also with the mRNA codons and tRNA anticodons that make up the genetic code itself. While there are only 20 cAAs, the nucleotide triplet basis of the code means that there are 64 ($= 4^3$) sequence variants for which to accommodate the 20 proteinogenic amino acids. This additional combinatorial space is filled by the need for stop codons, by tRNA isoacceptors (groups of tRNA that recognise the same amino acid but a different codon; Young and Schultz, 2010), and by tRNA wobble base-pairing (non-Watson-Crick base-pairing between tRNA anticodon and mRNA codon; Ran and Higgs, 2010; Allner and Nilsson, 2011).

Despite the high conservation and the complexity of how the genetic code is established, it is possible to engineer it. A number of groups over the last 20 years have engineered the genetic code successfully: through identification of species-specific tRNA/aaRS combinations (Zeng et al., 2014; Kwon et al., 2003; Mukai et al., 2015), and by engineering the aaRSs (Ho et al., 2016) or the tRNAs (Döring and Marlière, 1998; Pezo et al., 2004). The focus in the field has been to primarily expand the code by introducing new chemical functionality to the 20 cAAs (Liu et al., 2010; Young and Schultz, 2010) as will be discussed later in Section 1.3.1,

but some progress has also been made in altering the core of the genetic code itself (Biddle et al., 2016) as we will see in Section 1.3.2. Before going further, we should understand more about the components required for translation of the genetic code.

1.2.1 Key Players in Translation

1.2.1.1 Aminoacyl-tRNA Synthetases

Synthetases are ancient proteins and thought to be linked to the genetic code from its emergence. The main function of all aaRSs is to catalyse a two-step reaction for charging an amino acid to the 3'-terminus of a tRNA. The first step is amino acid activation by ATP, forming an aaRS-bound aminoacyl-adenylate (aa-AMP) intermediate. In the second step, this activated amino acid is transferred to the 3'-terminal nucleoside (conventionally, adenosine 76 [A76]) of a tRNA substrate, producing a charged/aminoacylated tRNA (aa-tRNA) molecule (Figure 1.2, black central path; Ibba et al., 2005; Ibba and Soll, 2000).

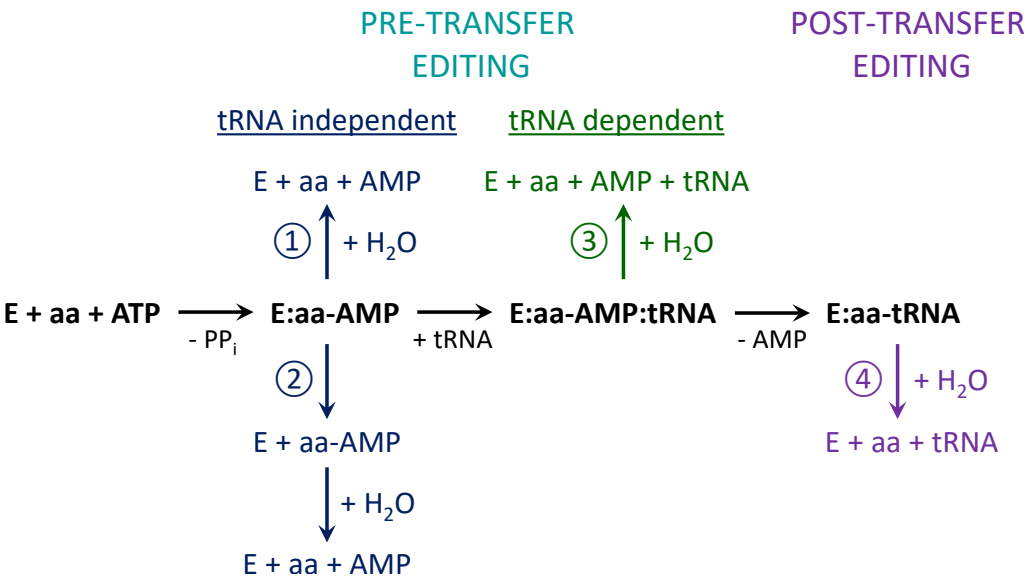


Figure 1.2: A schematic of the two-step reaction for amino acid activation and aminoacylation of tRNAs, and the aaRS editing pathways associated to correct for mis-acetylations

The central (black) pathway represents amino acid activation, tRNA binding, and aminoacylation of (non-)cognate amino acid to tRNA. The other coloured pathways represent the editing mechanisms for non-cognate amino acids. Pre-transfer editing of a non-cognate aminoacyl-adenylate (i.e. aaRS activated with non-cognate amino acid) includes: pathway 2, enhanced dissociation through solution hydrolysis; pathways 1 and 3, through enzymatic hydrolysis. Pathways 1 and 2 are tRNA-independent, whereas pathway 3 is tRNA-dependent. Post-transfer editing is shown in pathway 4, where a mischarged tRNA can be deacylated. Figure adapted from Dulic et al. (2010).

Whilst the two-step mechanism for amino acid activation and tRNA aminoacylation is conserved, the 20 aaRSs (corresponding to the 20 amino acids) are thought to have diverged from two ancestral enzymes, prototypical Class I and Class II synthetases – a division that is observed to this day in aaRSs (Carter and Duax, 2002; Carter and Wolfenden, 2015). Each class, with characteristic structures and differences in their active site architecture, can be further divided into three subgroups, primarily based on synthetase sequence similarity but also with respect to the properties of their amino acid substrates (Figure 1.3): (i) mainly aliphatic or sulphur-containing hydrophobics; (ii) with carboxyl (COOH) side chains or amidated (NH_2) derivatives; (iii) aromatics (Schimmel, 2008; Ribas de Pouplana and Schimmel, 2001).

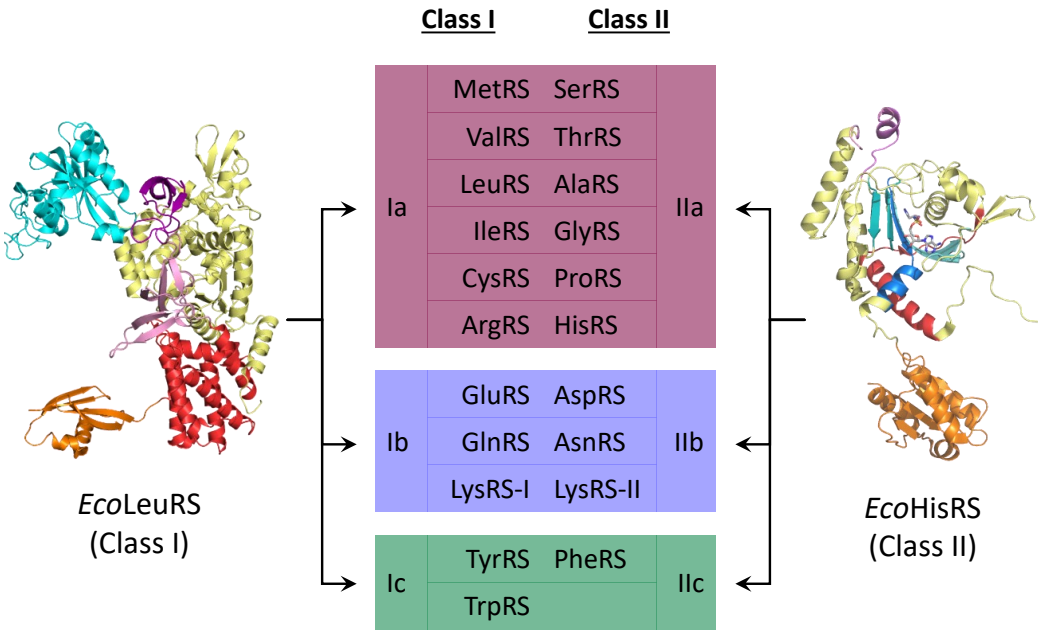


Figure 1.3: Categorisation of Class I and Class II aminoacyl-tRNA synthetases (aaRSs)

Each aaRS class can be further divided into three subdivisions, and the classes are thought to have derived from an ancient single-domain protein, which subsequently diversified into the 20 synthetases we see today. Leucyl-tRNA synthe50 tase (LeuRS) is an example of a Class I enzyme (left), whereas histidyl-tRNA synthetase is a Class II (right). Adapted from Ribas de Pouplana and Schimmel (2001). Crystal structures from *E. coli* proteins: LeuRS (PDB: 4AQ7) and HisRS (PDB: 2EL9).

For Class I enzymes, such as leucyl-tRNA synthetase (LeuRS), the catalytic core consists of a Rossmann fold, with conserved "HIGH" and "KMSKS" sequence motifs (Figure 1.4, left) that are involved in both steps of the tRNA charging reac-

tion. The KMSKS motif is involved in stabilising the transition state of amino acid activation, while the HIGH motif establishes hydrogen bonds with ATP for aa-AMP formation (Schmitt et al., 1995).

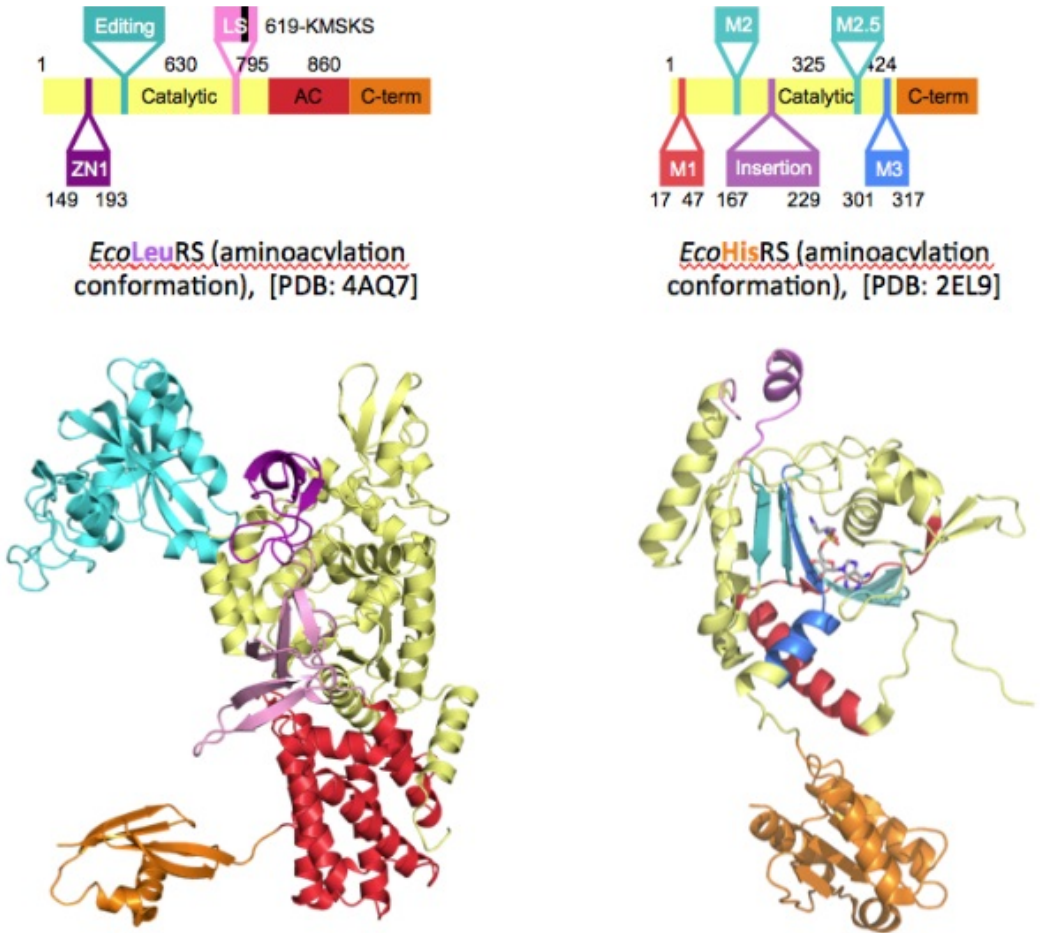


Figure 1.4: Annotated crystal structure domains within *E. coli* LeuRS and HisRS

The key domains and structural architecture of *Eco*LeuRS (PDB: 4AQ7) is shown on the left and the equivalent for *Eco*HisRS (PDB: 2EL9) are on the right. Residue numbers indicate domain boundaries and colours are matched to show comparable domains. Domains include the catalytic (yellow), zinc (ZN1; dark purple), editing (cyan), leucine-specific (LS, pink) with associated KMSKS motif (black), anticodon binding (AC; red), C-terminal (orange), insertion (light purple), and various unnamed motifs (M#; light red, cyan, blue). PDB structures annotated by YKH; adapted from Palencia et al. (2012) for *Eco*LeuRS (compare also to Figure 1.5), and Arnez et al. (1995) for *Eco*HisRS.

Conversely, Class II synthetases have a large cavity on one side of a seven-stranded antiparallel β-structure, with three conserved loop motifs making up a diffuse active site (Figure 1.4, right). In histidyl-tRNA synthetase (HisRS), motif 1 (M1) assists in aaRS homo-dimerisation interactions, and the orientation of motifs 2 and 3 (M2, M3)‘. The latter two help with contacting the tRNA acceptor arm

directly and with ATP-binding for amino acid activation, respectively (Ibba et al., 2005; Smith and Hartman, 2015).

Despite these architectural deviations, there is some commonality in the functions of the different domains within and between both aaRS classes (Figure 1.4). For example, the catalytic domain in both can recognise and bind with the 3' acceptor stem of tRNAs for aminoacylation. Similarly, the C-terminal domain, at least for class I synthetases, is primarily used in tRNA anticodon recognition (Boniecki and Martinis, 2012; Smirnova et al., 2000). The equivalent domain in class II aaRSs has a more disordered structure and so their function is less clear, but one proposed function is to provide additional stabilising structures for tRNA acceptor stem binding (Smith and Hartman, 2015). Moreover, some aaRSs in both classes have acquired extra domains to further facilitate differential binding of enzyme to amino acid or tRNA, and to edit incorrect interactions to their requisite substrates (Ibba et al., 2005).

Given the importance of its function, the catalytic domain dominates the majority of the aaRS structure, however mistakes do occur, both in amino acid activation and tRNA aminoacylation. The average translational error rate *in vivo* is estimated at $\sim 10^{-4}$ across bacterial (e.g. *Escherichia coli* [*Eco*], *Thermus thermophilus* [*Tth*]) to eukaryotic (e.g. *Saccharomyces cerevisiae* [*Sce*], humans [*Hsa*]) systems (Ibba and Soll, 2000; Reynolds et al., 2010; Jakubowski and Goldman, 1992). The consequence of misacylation in translation is cellular toxicity since the incorporation of incorrect amino acids can lead to misfolded and/or truncated proteins, and therefore non-functional waste products. Reviews by Schimmel and colleagues (2008) further highlight the enhanced deleterious effects of misacylating mammalian aaRSs and the causal association with (in some cases heritable) diseases in autoimmunity, angiogenesis, and apoptosis. The premise is that non-lethal mutations in ancillary (C-/N-terminal or editing) domains, while may reduce correction of misacylation, usually do not overtly affect aminoacylation functionality overall. As these mutations are not rectified, misfolded protein by-products can accumulate and, over time, become targets of the autoimmune system and/or for apoptosis (Guo et al.,

2010; Schimmel, 2008).

Given the high penalty for misacylation, synthetases are subject to strong selective pressure to recognise both cognate amino acid and cognate tRNA, and to develop stringent editing strategies to discriminate against, and correct, mischarged tRNAs. To help mitigate mistranslation, most aaRSs possess additional domains with editing mechanisms for correcting mistakes pre- and/or post-transfer of aa-AMP to tRNA.

Synthetases can initiate pre-transfer editing of misactivated aa-AMP either in a tRNA-independent or -dependent manner (Figure 1.2, pathways 1 and 2 (blue), or 3 (green) respectively). Pathway 2 differs from 1 and 3 as the synthetase first releases the aa-AMP and relies on the relative instability of the adenylate in aqueous solution for its hydrolysis; cleaving in the latter two depends on the synthetase (pathway 1) and tRNA binding (pathway 3). The location for pre-transfer editing was only recently demonstrated to be localised to the aminoacylation site, owing to findings that many aaRS of both classes show weak tRNA-dependent and -independent editing (Dulic et al., 2014, 2010; Ling et al., 2012; Hati et al., 2006), even in those that naturally lack editing domains (e.g. SerRS or ProRS) (Gruic-Sovulj et al., 2007; Splan et al., 2008). This suggests that editing occurs locally at the site of amino acid transfer to tRNA and potentially can be implemented soon after aminoacylation. However, much of the pre-transfer editing mechanisms are still not well understood (Martinis and Boniecki, 2010).

Post-transfer editing mechanisms clear incorrect amino acids from mischarged tRNAs (Figure 1.2, purple pathway 4) and synthetases often adopt the use of an editing domain. Together, the aminoacylation and editing sites make up what is known as the “double-sieve” model (Figure 1.5). The aminoacylation core acts as a coarse sieve to exclude larger, non-cognate amino acids being charged to the tRNA. Those that pass this initial checkpoint, both correctly charged and unedited misacylated tRNAs, are subsequently transferred to a finer sieve at the editing domain, which selectively binds and hydrolyses the smaller and/or isosteric near-cognates that passed the first sieve (Fersht and Kaethner, 1976; Mursinna et al., 2004). Some

suggest that the finer editing domain discriminates near-cognates even further: first by chemistry, before exclusion by size and steric differences (Hussain et al., 2010; Moras, 2010). For a select group of class I synthetases (LeuRS, IleRS, ValRS), this hydrolytic editing core is located in a later-acquired connective polypeptide 1 (CP1) domain, inserted in the aminoacylation catalytic site but which protrudes away from this domain (with 25∼35 Å separation distance) (Ling et al., 2009; Dulic et al., 2010). Although post-transfer editing mechanisms do exist for some class II synthetases, their structure and area of activation is also yet to be determined (Hati et al., 2006; Dock-Bregeon et al., 2000).

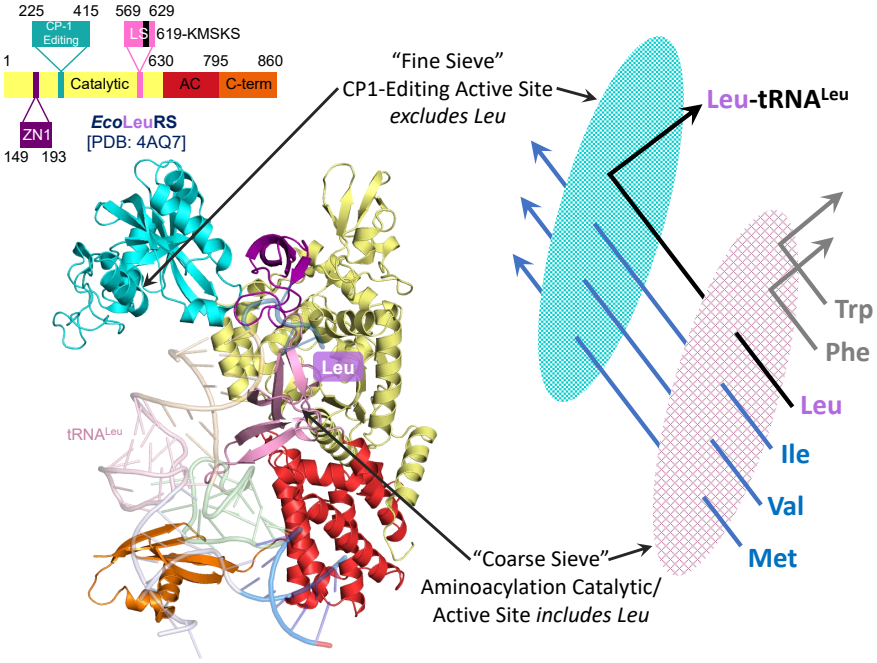


Figure 1.5: Crystal structure (left) and "double-sieve" schematic (right) in *Eco*LeuRS

The domains are the leucyl-specific domain (purple); Rossmann fold aminoacylation active site (made up of 7 parallel β-strands) (pink); connective polypeptide 1, CP1 editing domain (blue). The double-sieve diagram shows the exclusion of larger non-cognate amino acids (Trp, Phe) at the aminoacylation active site, but admittance of smaller near-cognates (Ile, Val), and the subsequent acceptance of tRNAs charged with near-cognate amino acids for hydrolysis. Adapted from Mursinna et al. (2004). Compare also to the *Eco*LeuRS in Figure 1.4.

As yet, there is no general rule as to whether aaRSs edit via a pre- or post-transfer mechanism, or whether the same mechanisms are conserved across different species. But more often than not, both proofreading processes can co-exist in the same enzyme with one pathway predominating over the other. Post-transfer

editing acts as a fail-safe that may activate if the former is compromised and, in this way, the redundancy of the error-checking methods maintains fidelity (Martinis and Boniecki, 2010). In terms of changing the genetic code, it is therefore rational that the catalytic and/or editing domain(s) would be the prospective targets for engineering in order to subvert the identity and function of aaRSs.

1.2.1.2 Transfer RNAs

Like aaRSs, tRNAs are ancient molecules, yet these short RNA adaptors (typically 76∼90 nucleotides (nt) in length) are responsible for providing the direct link between mRNA and the amino acid chain. The secondary structure of tRNAs can be visualised easiest as a cloverleaf, while an inverted L-shape best represents its tertiary structure (Figure 1.6). There are five key features of the tRNA: acceptor stem, D-arm, anticodon-arm, variable loop, and TΨC-arm (Giegé et al., 2012). Of particular interest are the acceptor stem with the necessary N73 discriminator and the 3'-CCA$_{\text{-OH}}$ terminus for tRNA aminoacylation/amino acid attachment, and the anticodon-arm with its crucial anticodon triplet recognition sequence (positions N34-N36) that base-pairs with mRNA codons.

The diversity of tRNA genes is not insubstantial: some can fold and be fully functional for protein synthesis directly post-transcription (e.g. *in vitro* transcription of tRNA$^{\text{Leu}}$; Tocchini-Valentini et al., 2000; Giegé et al., 2014). However, most need to first undergo some form of processing. The processing pathway required depends not only on the tRNA itself, but also its species of origin. Only the initial step, transcription of tRNA precursors (pre-tRNAs) by RNA polymerases, appears to be conserved (Sweetser et al., 1987). The majority of prokaryotic and eukaryotic pre-tRNAs will subsequently trim the leading 5'- and trailing 3'-sequences flanking the pre-tRNA but the order of their removal, and the nucleases required, can vary depending on species, tRNA, presence or absence of the 3'-CCA motif and any introns (Figure 1.7; Hopper and Phizicky, 2003; Hartmann et al., 2009).

Another key processing step, especially for all eukaryotic tRNA genes, is for tRNA nucleotidyl-transferases to append the 3'-CCA terminus (Chan and Lowe, 2009; Hartmann et al., 2009). In some instances, various bases within these mature

tRNAs can also have post-transcriptional modifications (PTMs) to give further functions, including providing structural stability and identity elements for recognition by aaRSs (Figure 1.6, Figure 1.7; Kirchner and Ignatova, 2015; Chopra and Reader, 2015; Hopper and Phizicky, 2003; Giegé et al., 2014). The addition of dihydrouridine (D17) in the D-loop, thymidine (T54) and pseudouridine (Ψ55) in the T-arm, and methylation of bases or ribose sugars, are a few examples of PTMs (Machnicka et al., 2014; Giegé et al., 2014; Björk et al., 1987). Although uncommon, there are a few cases in *E. coli* where tRNA modifications affect identity, all of which are located at the anticodon positions and/or adjacent nucleotide N37. For example, re-

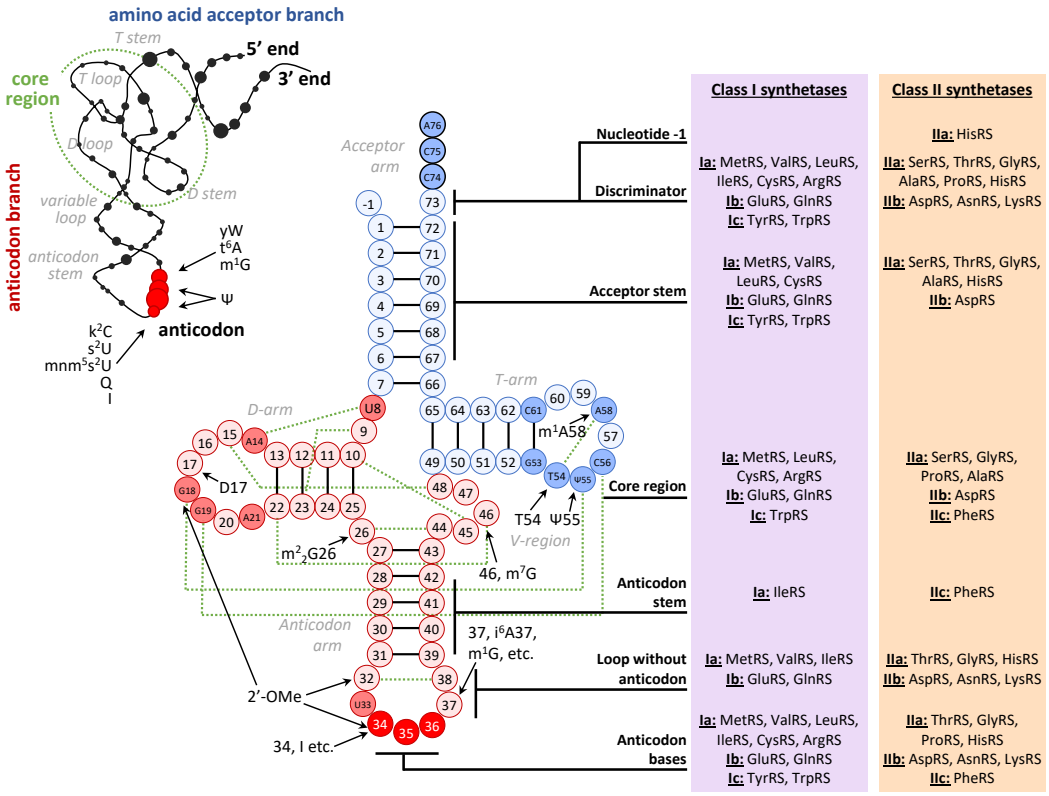


Figure 1.6: Secondary "cloverleaf" and tertiary "L-shape" (inset) structure of tRNA with locations of known identity elements recognised by Class I and Class I aaRSs

The conventional numbering system is used for the tRNA: the anticodon is always N34-N-36; discriminator at N73; CCA 3'-terminal group at positions 74-76. Constant nucleotides are explicitly indicated. Some tRNAs possess a 4-24 nt variable loop region; type I tRNAs have a small loop (4-5 nt); type II have longer sequences. The * symbol indicates Watson-Crick base-pairing; dotted green lines indicate other pairings important for L-shaped architecture. Location of identity elements to each class of synthetases are highlighted. Abbreviations: U, uridine; C, cytosine; A, adenosine; G, guanosine; T, thymidine; Ψ, psudoridine; k^2C, lysidine; s^2U, 2-thiouridine; mnm^5s^2U, 5-methylaminomethyl-2-thiouridine; Q, queuosine; I, inosine; m1G, 1-methylguanosine; t^6A, N6-threonylcarbamoyladenosine; yW, wybutosine. Adapted from Giegé et al. (2014).

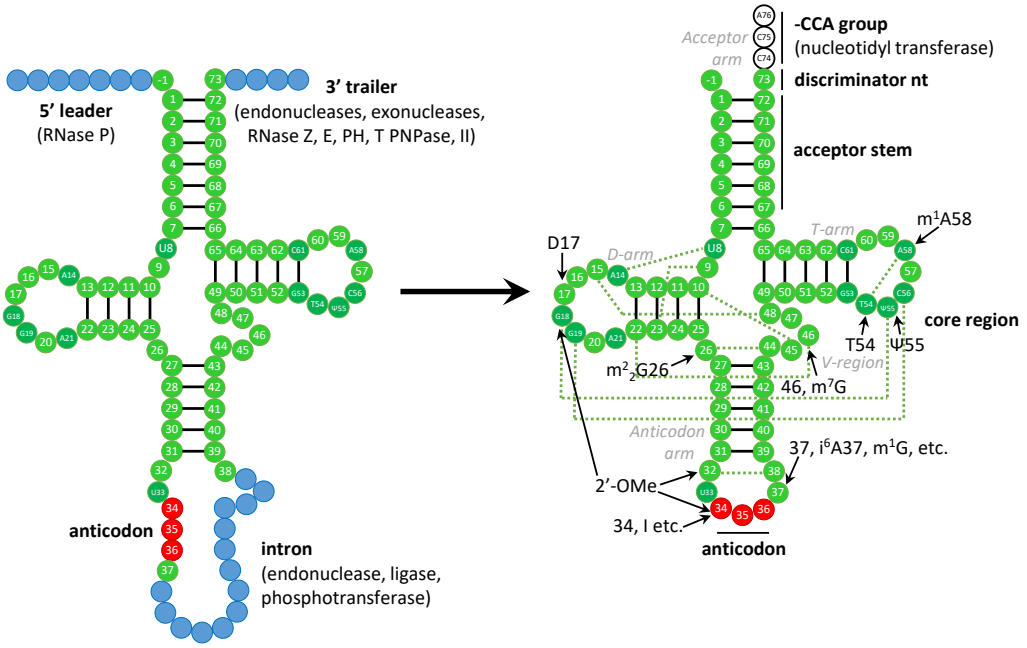


Figure 1.7: Schematic of both tRNA precursor (pre-tRNA; left) and mature tRNA (right)

The schematic also decpicts some post-translational modifications highlighted, both before and after tRNA processing. Each nucleotide is represented by a filled circle: mature tRNA (green); anticodon (red); 3'-CCA group (white); 5'-leading sequence; RNase Z, E, PH, T, PNPase, II (a mix of exo- and endo-nucleases) cleaves the 3'-trailing sequence; nucleotidyl-transferases append the 3'-CCA terinal group if it was not in the original transcript; splicing endonucleases, ligases, and phospho-transferases are responsible for removing introns. Adapted from Hopper and Phizicky (2003).

placement of C34 in $tRNA^{Met}_{(CAU)}$ with k^2C34 (lysidine), results in isoleucylation by IleRS and incorporation of isoleucine on codon $Ile_{(AUA)}$, instead of methionine at $Met_{(AUG)}$ (Giegé et al., 2014).

For the *E. coli* tRNAs that do not require PTMs, sequence variation, tertiary folding conformations, and tRNA/aaRS interactions, are critical for correct tRNA/aaRS specificity. Each tRNA can be defined by a set of identity elements (usually a group of single or base-paired nucleotides): determinants enable positive recognition and specific aminoacylation by cognate aaRS, whilst anti-determinants represent negative features that correctly hinders interactions between tRNAs and non-cognate synthetases. The variable loop in $tRNA^{Ser}$ (where the importance lies in its length more than sequence), for example, demonstrates both determinant and anti-determinant properties, as it simultaneously acts as an identity signal to SerRS, but also obstructs all non-cognate aaRSs, especially closely-related LeuRS, from interacting with it (Himeno et al., 1997).

Certain elements, like the N73 discriminator and 3'-CCA$_{\text{-OH}}$ acceptor, can be crucial to the identity of a tRNA, as synthetases use these sites for aminoacylation (Chambers, 1971; Crothers et al., 1972). Conservation of these tRNA-defining features across prokaryotes and eukaryotes, such as A73 in tRNA$^{\text{Leu}}$ and G3•U70 in tRNA$^{\text{Ala}}$ (Giegé et al., 1998; Francklyn and Schimmel, 1990; Naganuma et al., 2014), can provide an estimate of the identity strength. However, this may not hold true in all species: G-1 (and to a lesser extent, C73 in prokaryotes, or A73 in eukaryotes) is both necessary and sufficient for tRNA$^{\text{His}}$ identity, with the exception of α-proteobacteria *Caulobacter crescentus* (Nameki et al., 1995; Giegé et al., 1998). This species relies instead on the anticodon, A73, and U72 for recognition and aminoacylation by its native HisRS (Jackman and Phizicky, 2006; Yuan et al., 2011). Conversely, although the *E. coli* HisRS (*Eco*HisRS) C-terminal domain does contact the *Eco*tRNA$^{\text{His}}$ anticodon, this triplet may not be directly required, or at least plays only a minor role in identity (Arnez et al., 1995; Jackman and Phizicky, 2006).

Like the discriminator and acceptor stem, the anticodon is an often-used identity element as it holds a key role in direct sequence pairing to mRNA codons (Kisselev, 1985). Only the *E. coli* tRNAs for histidine, alanine, serine, and leucine are unusual in that the anticodon is not a (strong) identity element for their respective synthetases (Figure 1.8; Giegé et al., 1998; Ho et al., 2016). This is further supported by the exploitation of these tRNAs as suppressor tRNAs (Figure 1.9; Yan et al., 1996; Tocchini-Valentini et al., 2000; Tian et al., 2013).

Despite the diversity of identity elements and the stringent requirement for tRNA/aaRS specificity to maintain a genetic code *in vivo*, there is a degree of redundancy at the other end of the code: tRNA to mRNA. This is exemplified in the degeneracy of the genetic code, whether through tRNA isoacceptors (families of tRNA that are charged with the same amino acid but target a different codon) (Young and Schultz, 2010) or through two codons being targeted by the same tRNA (tRNA wobble base-pairing between the 3'-base of the mRNA codon and the 5'-base of the tRNA anticodon).

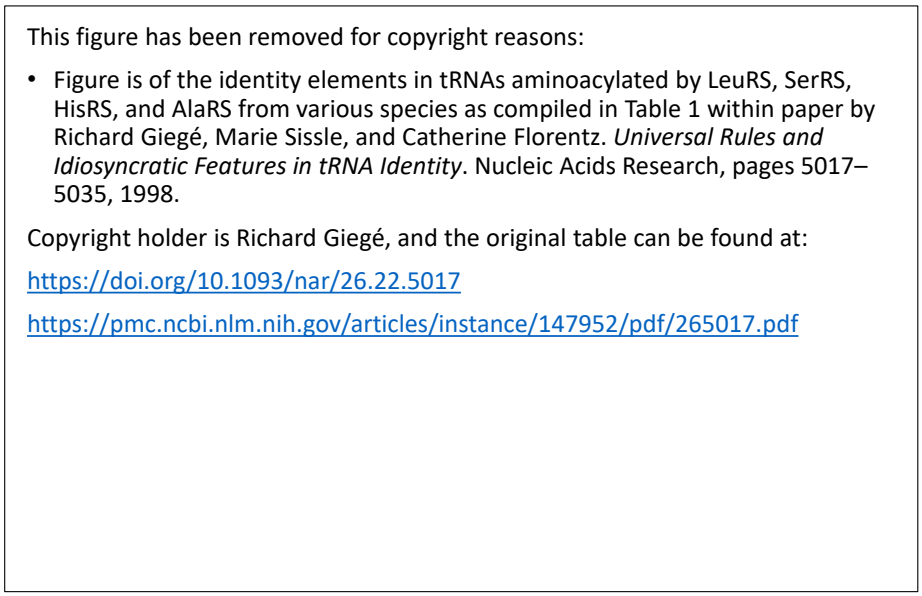


Figure 1.8: Identity elements in tRNAs aminoacylated by LeuRS, SerRS, HisRS, and AlaRS in various species

Identity elements for (1) *E. coli*, (2) *S. cerevisiae*, and (3) other organisms, are classified according to their location in the (a) amino acid accepting stem, (b) anticodon region, or (c) other tRNA domains. Identity nucleotides identified: *in vitro* (bold), *in vivo* (italics), or with both approaches (normal script). Numbering of residues is according to Sprinzl and Vassilenko (2005). For base pairs, (:) denotes Watson-Crick (WC) base-pairing, (•) non-WC pairs, and (–) tertiary pairs. R, purine (A or G); Y, pyrimidine (C or T or U). Leucine, serine, and alanine identities do not rely on anticodon positions. The role of the histidine anticodon as an identity element is conflicted. Figure from Giegé et al. (1998).

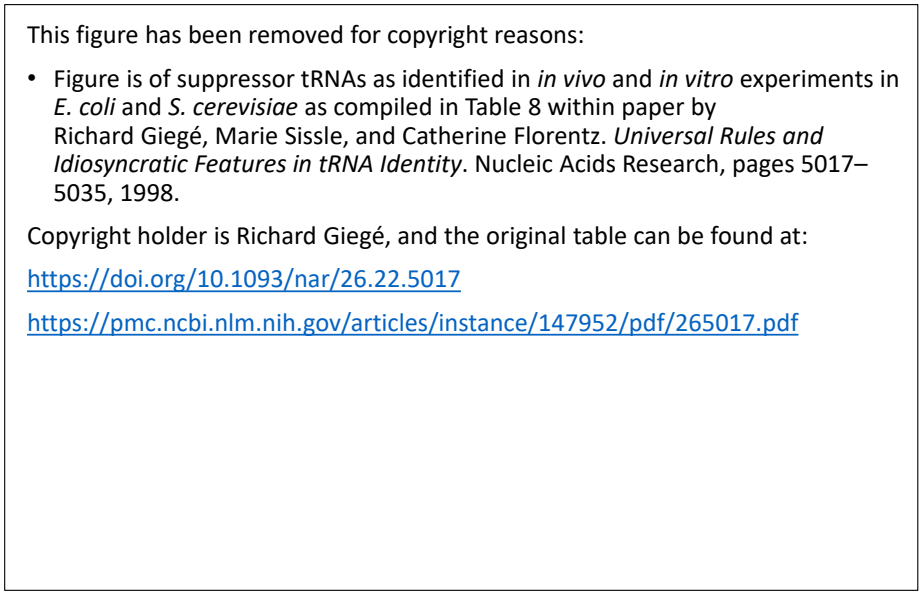


Figure 1.9: Examples of suppressor tRNAs

These are tRNAs that have either been engineered or are rationally designed *in vitro* transcribed chimeric tRNAs, that are capable of acquiring additional identities to their primary assignment, thus can be aminoacylated by non-cognate synthetases. Figure from Giegé et al. (1998).

In addition, the combinatorial space of identity elements is not fully populated and additional tRNAs and aaRS can be introduced that do not interact with the natural sets (i.e. introduced aaRSs cannot charge native tRNAs and introduced tRNAs cannot be charged by native aaRSs), also described as orthogonal tRNA/aaRS pairs. These pairs can, and have, been exploited to change the genetic code with the tRNA generally targeting a stop codon (akin to suppressor tRNAs; Figure 1.9). The result has been the very successful expansion (and, to a limited degree, reassignment) of the genetic code, as described by a number of groups (Krishnakumar and Ling, 2014; Wang and Schultz, 2001; Wang et al., 2001; Chin et al., 2003; Anderson et al., 2004; Liu and Schultz, 2010; Biddle et al., 2016).

Although those efforts increased the chemical diversity potentially available to biology, they fundamentally do not change the topology of the genetic code at its core, i.e. they expand but do not rewrite the genetic code.

1.3 Can the Genetic Code be Changed?

1.3.1 Genetic Code Expansion

1.3.1.1 Natural Genetic Code Expansion

Genetic code expansion is generally the direction most consider when contemplating how the genetic code can be modified. Given that there are only 20 canonical amino acids (cAAs) and (if considering just triplet nucleotides) a sequence space of 64 codons, there is sufficient room to introduce into proteins a whole host of novel, non-canonical amino acids (ncAAs) with new chemical, physical, and biological properties (Wang et al., 2001, 2009c).

Needless to say, nature has already beaten us to the punch with the atypical – but naturally proteinogenic – supplementary amino acids 21 and 22, selencysteine and pyrrolysine (Osawa et al., 1992; Krzycki, 2005; Lukashenko, 2010). Selenocysteine (Sec/U) is an analogue of cysteine, with a selenol group taking the place of the usual sulfur-containing thiol group in the sidechain, and it is present in a number of proteins (particularly those involved in redox activities, like glutathione peroxidases; Johansson et al., 2005) within eukarya and prokarya (both bacteria and archaea), but it is not universal in all organisms. In the presence of selenium, selenocysteine can be incorporated into the opal stop codon UGA through the use of a **SECIS** element in the mRNA; the ***se***leno***c***ysteine ***i***nsertion ***s***equence is an RNA sequence of approximately 60 nt that forms a stem-loop structure. In bacteria, the method with which selenocysteine is incorporated differs from the standard tRNAs[1] in that the $tRNA^{Sec}$ is initially charged/mis-acetylated with serine by the seryl-tRNA synthetase (SerRS), before being converted into $Sec\text{-}tRNA^{Sec}$ by selenocysteine synthase (SelA; Itoh et al., 2013). Similarly, whilst EF-Tu (elongation factor thermo unstable) is the prokaryotic component responsible for delivering a standard aminoacylated tRNA to the ribosome for incorporation into a polypeptide

[1] Although, it draws some parallels with how glutamine (Gln) is charged to $tRNA^{Gln}$ in archaebacteria (and the majority of eubacteria and in most eukaryotic organelles), as discussed earlier; that is, through an intermediary step with non-discriminating glutamyl-tRNA synthetases (GluRS). GluRS first aminoacylates $tRNA^{Gln}$ with glutamate (Glu) to give $Glu\text{-}tRNA^{Gln}$, before converting the glutamate to glutamine to produce $Gln\text{-}tRNA^{Gln}$ (Hadd and Perona, 2014; Perona, 2013).

chain, SelB, an alternative translational EF is required for selenoproteins.

Pyrrolysine (Pyl/O) is a lysine derivative that is currently only known to be used in some methanogenic prokaryotes, specifically the archaebacteria *Methanosarcina barkeri* (*Mb*; Srinivasan et al., 2002). Like selenocysteine, pyrrolysine is also incorporated into a stop codon, albeit the amber stop UAG (Ibba and Söll, 2004; Herring et al., 2007), however, unlike selenocysteine, incorporation of pyrrolysine during protein synthesis occurs in the same manner as the standard amino acids. The process requires only the *pylT* and *pylS* genes, encoding the $tRNA^{Pyl}_{(CAU)}$ and the Class II pyrrolysyl-tRNA synthetase (PylRS) respectively.

1.3.1.2 Synthetic Genetic Code Expansion through tRNA/Aminoacyl-tRNA Synthetase Engineering

In terms of extending into synthetic genetic code expansion: this requires the targeting of a triplet codon not otherwise required by the host cell for canonical translation/translation of natural proteins, and the development of a mechanism for charging a tRNA with ncAAs. As the entire palette of 64 codons is typically in use in model organisms of interest to bioengineers, sophisticated strategies need to be developed to wrest control of a target codon; most frequently by using an under-used stop codon, like UAG.

Orthogonal tRNA/aaRS engineering is the most prominent method to date for ncAA incorporation. There are a few key domains that can be targeted in enzyme engineering: for example, the active sites in aaRSs for amino acid activation, and for the aminoacylation to tRNAs. For synthetases with editing domains, the relevant residues could be mutated to provide another level of specificity for discriminatory binding and cleaving acceptable or unacceptable amino acids and/or aa-tRNAs. Depending on the required tRNA/aaRS pairing, directed evolution of the synthetases may be required to recognise and bind to alternative/engineered tRNA structures (discussed later). If an aaRS has the potential for multiple substrates, rational design or targeted evolution could predispose it towards the required substrates, either through better fit (e.g. recognition, conformation, binding).

Similarly, whilst aaRSs may need to be modified to accommodate for alternate

ncAAs, tRNAs may also require adaptation for higher affinity to its foreign amino acid cargo. These changes would predominantly centre on the tRNA acceptor stem, either on the single-stranded (3') or double-stranded (5'-3') region of the extended acceptor arm, as this is the esterification site for the amino acid. Of equal importance for genetic code modification is the anticodon stem-loop region of the tRNA; without both the amino acid acceptor stem and the anticodon, there would be no direct link between reading the genetic information (mRNA) and production of the resultant protein (polypeptide chain). Assuming that the ncAA has been correctly charged to the tRNA, the next significant consideration is the anticodon binding partner, the codon. Earlier examples of selenocysteine and pyrrolysine incorporation saw these aa-tRNAs base-pair with opal UGA and amber UAG stop codons respectively: unsurprisingly, many synthetic ncAAs have followed this pattern in being directed by these codons. As mentioned earlier, such tRNAs that target the amber, opal, and ochre codons are collectively known as stop codon or nonsense suppressors, as they prevent the incorporation of stop codons into the protein. More generally, these are known as tRNA suppressors as they can be used to suppress the identity of a given tRNA (and tRNA-like molecules, i.e. stop codons and release factors at a given compatible codon.

The earliest examples of tRNA/aaRS pairs that allow for ncAA incorporation into *E. coli* via stop codons were developed from amber (UAG) suppressors (Figure 1.9), such as *Mj*tRNA$^{\text{Tyr}}_{(\text{CUA})}$/*Mj*TyrRS from *Methanocaldococcus jannaschii* (*Mj*; Chatterjee et al., 2013; Rauch et al., 2016). As it was, since their discovery, it was also found that the orthogonal *Mb*tRNA$^{\text{Pyl}}$/*Mb*PylRS pair (and an equivalent pair derived from *M. mazei*) could be evolved to incorporate a range of ncAAs at the amber stop codon in *E. coli* (Neumann et al., 2010a; Yanagisawa et al., 2008). A later PylRS variant, iodo-phenylalanyl-tRNA synthetase (IFRS), further enhanced its substrate promiscuity to encompass a chemical library of 313 ncAAs (Guo et al., 2014), whilst work by the Schultz and Chatterjee groups spread the use of these synthetase variants beyond *E. coli* to enable ncAA incorporation in yeast and mammalian species as well (Liu and Schultz, 2010; Young and Schultz, 2010; Xiao et al.,

2013; Italia et al., 2017). (Zeng et al., 2014)

The use of stop codons for genetic code expansion, however, is challenging: whilst it was possible to efficiently encode many distinct ncAAs, a key limitation with nonsense suppression, as noted by the Chin group at the time (particularly when only amber suppression had been established), was that only a single type of ncAA can be incorporated at a time due to the lack of "blank" codons with which to introduce multiple novel amino acids (Neumann et al., 2010b). Even if multiple orthogonal tRNA/aaRS nonsense suppressors can be used simultaneously to direct ncAAs, only two of the three stop codons can realistically be used; the third must remain a stop codon. Given that the ochre stop codon UAA in *E. coli* is most frequently used (with 2,765 instances, compared to compared to 1,232 UGA instances for opal and 321 UAG for amber; Ho et al., 2016), in the absence of wholesale recoding of the host genome, it makes sense to direct novel tRNA/aaRS pairs to the two rarer stop codons as a way to minimise translational read-through and accidental mis-acetylations. Chatterjee et al. (2013), however, managed to test these limits and demonstrated ncAA incorporation at both amber (UAG) and ochre (UAA) in a double-mutant GFP expression plasmid encoding GFP-3TAG-151TAA, co-transformed with pUltra (encoding *Mb*PylRS^{opt} and tRNA$^{PylU25C}_{(UUA)}$; ochre-suppressing pyrrolysyl system) and pEVOL (encoding a *p*-azidophenylalanine (pAzF)-specific *Mj*TyrRS and tRNA$^{Tyr}_{(CUA)}$; amber-suppressing *Mj*Tyr system) suppressor plasmids into strain BL21(DE3).

1.3.1.3 Expansion into Rare and Novel Sense Codons

Given that there have been examples of genetic code expansion using novel amino acids into each of the stop codons, the natural extension would be to explore whether the sense codons[2] could also become potential target sites for incorporating ncAA. The challenge with targeting the wider codon space is that sense codons occur far more frequently than stop codons, hence, there is a significantly greater chance of causing cellular/host toxicity due to ncAA mis-incorporation, especially if the se-

[2]Arguably, depending on the types of new amino acids being introduced, this line of research strays close to genetic code reassignment (discussed later in Section 1.3.2).

lected codon has not been "freed-up" to allow for a new substrate to be introduced. Moreover, the task of wholesale genomic recoding to ensure that a codon would be fully available for adopting a non-cognate amino acid is non-trivial to say the least. Mukai et al. (2015) attempted to mitigate the potential problems of producing an admix population of cognate- and ncAA-incorporation at existing codons by targeting the rarest sense codon, $Arg_{(AGG)}$ (1,491 instances; Ho et al., 2016) in the *E. coli* genome. Another benefit of targeting arginine was that there are five other arginine codons (with their associated arginyl-tRNAs) that could potentially ameliorate the deleterious effects of (mis-)incorporating a ncAA (i.e. missense suppression) at $Arg_{(AGG)}$ positions. By creating an AGG-reading $tRNA^{Pyl}_{(CCU)}$[3] and engineering a new variant of the PylRS, known as HarRS, to recognise two new substrates L-homoarginine and L-N(6)-(1-iminoethyl)lysine (L-NIL) and the new $Arg_{(AGG)}$ codon, these ncAAs were expected to be incorporated to a degree. Note that one advantage of using a versatile pairing like $tRNA^{Pyl}$/PylRS was that the anticodon does not appear to be a (strong) determinant for PylRS recognition and is malleable enough to conform to different stop codon variants. As such, it was anticipated that $tRNA^{Pyl}$ would be relatively amenable to anticodon modifications that direct it towards the selected sense codon(s) for ncAA incorporation. Furthermore, Mukai et al. (2015) were able to synonymously replace some or all $Arg_{(AGG)}$ codons in the essential genes for their assays, but the majority of the $Arg_{(AGG)}$ codons remained in the genome. Instead, to mitigate the effects of the mis-incorporation, their genetic code expansion assays were allayed by use of a temperature sensitivity dependency that could only be rescued by correctly translating $Arg_{(AGG)}$ to L-homoarginine or L-NIL.

Ho et al. (2016) achieved a similar result by using the IFRS synthetase variant to incorporate 3-iodo-l-phenylalanine (3-I-Phe) at a number of more commonly-used codons in serine ($Ser_{(AGU)}$, $Ser_{(AGC)}$, $Ser_{(UCG)}$) and leucine ($Leu_{(CUG)}$) codons in a heterologously-expressed sfGFP within wild-type *E. coli*. Although there were

[3]It should be noted that earlier attempts had been made by Krishnakumar et al. (2013) and Zeng et al. (2014) to use PylRS for sense codon reassignment involving the arginine codons AGA, AGG and CGG, but with no established success.

varying degrees of success with incorporation of 3-I-Phe at these more frequently-used sense codon sites, it also demonstrated the possibility that well-established sense codons could become targets for genetic code expansion.

As opposed to working within the confines of the known genetic code, an alternative approach taken by the Schultz and Chin groups was to explore a quadruplet codon sequence space and engineering a new set of translation machinery for code expansion. The direction the Schultz group took was to disassociate competition of triplet- and quadruplet-codon recognition by full genome recoding to remove all relevant instances of competing aaRS, tRNA, and release factors (RF) that might prevent quadruplet codon suppression (Magliery et al., 2001; Anderson et al., 2001). Chin's group, on the other hand, delved down the route of ribosome engineering. Specifically, they looked to enhance ribosome processing of quadruplet codon suppressors through the development of orthogonal ribosomes (Rackham and Chin, 2005b,a; Neumann et al., 2010b). O'Donoghue et al. (2012) later developed upon the $tRNA^{Pyl}_{(UCCU)}$ variant to bind to the 4 nt AGGA codon and incorporate in the ncAA N6-(tert-butoxycarbonyl)-l-lysine (BocK), albeit with limited success: they were unable to determine the underlying reasons for the low basal level of frame-shift suppression from endogenous Arg-$tRNA^{Arg}$ on BocK-$tRNA^{Pyl}_{(UCCU)}$ incorporation. It was postulated that the 4 nt anticodon-codon binding (i.e. quadruplet "frame-shift" suppression) was accepted, but that the ribosome likely favoured catalysing aa-tRNAs with the native tri-nucleotide anticodon conformation over a tetra-nucleotide anticodon variant, which was why the BocK-$tRNA^{Pyl}_{(CCU)}$ isoacceptor had more success.

That is not to say that tRNAs with tetra-nucleotide anticodons could not be preferred, simply that the interacting components, like the ribosome, also need to be made compatible; after all, quadruplet (and five-, six-, or higher-base) codons would greatly open up the sequence space for ncAA incorporation at unique positions (Anderson et al., 2001). For example, the synthetically evolved orthogonal ribosome (ribo-Q1) that Chin and colleagues developed was able to decode a series of quadruplet codons as well as the amber stop, by accommodating aa-tRNAs with

tetra-nucleotide anticodons and by using the *Mj*TyrRS/tRNA$^{Tyr}_{(CUA)}$ and *Mb*PylRS and tRNA$^{Pyl}_{(CUA)}$ (Neumann et al., 2010b). The authors produced and screened 11 saturation mutagenesis libraries in the 16S ribosomal RNA of ribo-X (a previously evolved ribosome that was able to efficiently decode at the amber stop codon), covering 127 nt within 12 Å of a tRNA bound in the decoding centre, and recovered four ribo-Q1-4 able to restore function to a chloramphenicol acetyl transferase with an in-frame AAGA 146 quadruplet codon, with *Eco*SerRS/tRNA$^{Ser}_{(UCUU)}$. However, further incorporation of ^{35}S-cysteine into a protein that contains no cysteine and into a dual luciferase system found that there is little difference in the decoding translation fidelity between triplet- and quadruplet-decoding, and their evolved, and wild-type ribosomes.

One potential method to improve on the fidelity and rate of the quadruplet-decoding orthogonal ribosome would be select such a ribosome by directed evolution for efficient quadruplet decoding activity. In order to achieve effective selection, the bipartite ribosomal subunits needed to be physically coupled to avoid interaction with host subunits. Using a circular permutation approach to find proximal linkages between the 16S and 23S RNAs the groups of Jewett and Mankin (Orelle et al., 2015) engineered a tethered ribosome (Ribo-T) that maintained functionality without hindering the required subunit movements for translation. Their Ribo-T (only) catalysed at a rate of ∼45% relative to wild-type ribosomes. A further development on this (oRibo-T) was developed with the intention of engineering a ribosome capable of polymerising ncAAs and with any backbone-modified analogues. In recent works, the Chin group automated the design of not only orthogonal ribosomes, but also orthogonal aaRSs, tRNAs and mRNA operons such that it was possible to create a 68-codon, 24-amino acid genetic code that was able to efficiently incorporate four unique ncAAs in response to distinct quadruplet codons within a single recoded green fluorescent protein (GFP) sequence (Dunkelmann et al., 2021).

Schultz and others took a more radical approach for reducing competition between triplet- and quadruplet-decoding protein synthesis with ncAAs: that of full genome reprogramming. They argued that the low nonsense suppression efficiency

is a result of competing recognition of the first three bases in the alternate tRNAs by endogenous tRNAs or RF. Thus, they demonstrated that these competing elements could be (systematically) eliminated from the host genome to make way/the space in the current genetic code table for the alternate variants. By using MAGE to replace all 321 instances of the amber (UAG) stop codon with ochre (UAA) in *E. coli*, as well as by deleting RF1, they freed up all amber (UAG) codons to make way for engineered $tRNA_{(CUA)}$ to use UAG for ncAA coding (Wang et al., 2009b; Lajoie et al., 2013; Chatterjee et al., 2014; Zheng et al., 2016).

Church and colleagues took this genome-wide/scale recoding a step further and used MAGE to completely abolish 13 rare sense codons in the *E. coli* genome and replaced them with 42 highly-expressed essential genes (Lajoie et al., 2013). This effectively freed up 13 of the 64 codons on the natural code for ncAAs, whilst still retaining all 20 proteogenic cAAs, in a functional *E. coli* host.

1.3.2 Genetic Code Reassignment

Much of what has been achieved in this field to date has focused on the incorporation of ncAAs into an extended codon space: either by engineering orthogonal tRNA/aaRS pairs that are nonsense suppressors, or by engineering orthogonal pairs that are able to access the larger sequence space that is provided by quadruplet-decoding. However, a lesser-used approach for expansion is genetic code reassignment. Work by Mukai et al. (2015) and Liu et al. (2010) provided the first efforts towards this by characterising ncAA incorporation at rare sense codon Arg(AGG). Naturally, the limitation of this approach is that incorporation occurs at already occupied codon spaces, however, if genome-wide reassignment can free up 13 rare sense codons (Lajoie et al., 2013), then the only remaining requirement is the development of orthogonal tRNA/aaRS pairs that can be directed to those blank codons. While 13 new pairs would not be sufficient to cover the 200+ ncAAs now available (Liu et al., 2010), no study has yet incorporated more than three distinct ncAAs in the same protein.

Genetic code modification, however, is not just limited to expansion; given that it is possible to produce sense-to-sense codon reassignments to incorporate unnat-

ural amino acids in place of rare natural amino acids, it must also be possible to engineer sense-to-sense codon reassignments between two different natural/canonical amino acids. This poses a different sort of challenge from ncAA incorporation, since there is no requirement to expand into quadruplet codon territory, or to evolve basic translational components like the ribosome.

Nature itself has already made short forays in this direction: some *Candida* species can change the LeuRS/$tRNA^{Leu}$ identity to incorporate serine where there should be leucine, predominantly through a change in the tRNA anticodon sequence. Beside this singular natural example, there have not been any other organism that has developed this identity switch. As discussed above, however, some tRNA/aaRS pair interactions do not depend on the anticodon loop in the tRNA, making it possible to change the anticodon of these tRNAs to give a sense-to-sense codon reassignment. Döring and Marlière (1998) were one of the first to perform such an identity switch study: Ile/Met reassignment to Cys. Specifically, there was no deletion of the endogenous tRNAs, instead the new anticodon-modified tRNAs were introduced that had perfect Watson-Crick base-pairing with the desired codon. Although it is assumed that there was some competition between endogenous wobble-pairing and the Watson-Crick base-paring from the anticodon-modified tRNAs, this simple change seemed sufficient to allow Ile-to-Cys incorporation.

Whilst this system demonstrated that it is possible for a form of codon competition/invasion to produce sense-to-sense reassignment, it also highlighted that due to competition, the products will be heterogeneous. Depending on the codon reassignment, this could potentially prove fatal to the host. Therefore, in the ideal codon reassignment scenario, the endogenous tRNA would need to be removed from the host genome, so that the new codon-reassigning tRNA has sole access to its newly adopted codon.

This strategy could be pushed further. If we use the term “invade” to define an instance in which an engineered tRNA is used to decode a natural sense codon for a novel target amino acid, the first step above constituted one invasion event, made

efficicent by the removal of the native tRNA. In a second step, the native tRNA that had been replaced in the first scenario could instead be used to invade a second codon, which again may require the removal of another native tRNA, and so on so forth. Wholesale reassigned genomes, such as those described earlier, would be invaluable for the development of such complex reassignments. We can take this strategy one step further: the tRNA that was replaced could theoretically be used to invade another codon either competing through wobble-base-pairing, or by taking on the identity of the freed rare codons from the *E. coli* strain produced by Church and colleagues (Lajoie et al., 2013).

1.4 The Original Plan to Rewrite the Genetic Code

The original intention had been to establish the plan that is described as follows. However, upon reflection, a seemingly more viable approach (as outlined in Section 1.5) was devised, and it is with this latter plan that the main research for this thesis was conducted. Aspects of the plan as laid out below, and the results therein, were incorporated into the design of the rest of the research within this thesis. Should there be a future opportunity to revisit this work, the details that follow within this section, and from the rest of this thesis, may prove useful.

1.4.1 Initial project aims

My initial project aims to establish platforms to test aaRS and tRNA function in *E. coli*, with a view towards developing a system for the transient reassignment of sense codons *in vivo*. Sense codons can gain new assignments through the use of tRNA/aaRS pairs that have been engineered with new identities. However, if a codon that was originally specific for leucine were to also code for histidine, translation would produce a heterogeneous mix of naturally- and alternately-coded proteins (Figure 1.10). Given that these amino acids have contrasting physical and biochemical properties, the structure and function of a naturally coded protein – under an altered code – will likely be disrupted. Therefore, in order to study genetic code reassignment, whilst minimising toxicity, we must develop a transient *in vivo* system, else be confined to *in vitro* systems. Once a proof-of-principle plat-

form has been established that can demonstrably reassign sense-to-sense codons in an *in vivo* setting, we can use the platform to systematically explore further reassignments, be it between other sense codons or for large-scale refactoring in whole organisms. As such, we will be able to take one step closer to synthesising semantically orthogonal, genetically engineered machines/organisms (GEMs) and, hence, provide containment of genetic information within GEMs for enhanced biosecurity and biosafety.

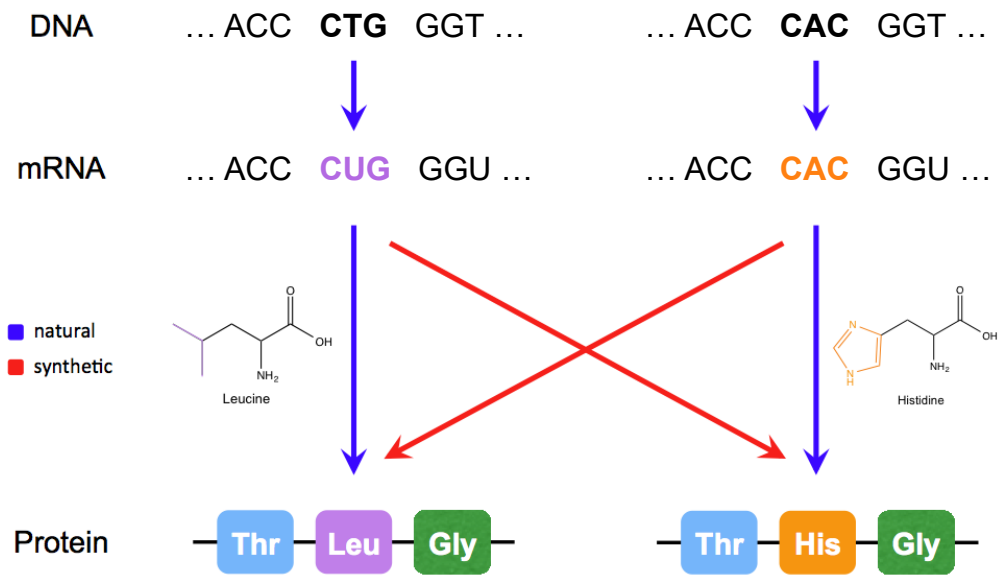


Figure 1.10: Representation of the proposed histidine-leucine alternative genetic code system

Representation of the proposed histidine-leucine transient sense codon reassignment system, which also demonstrates the possibility for toxicity to hosts, i.e. the translation of polysemeous codons Leu(CUG) (to incorporate a natural leucine, or histidine using the alternate translational machinery) and His(CAC) (naturally assigning histidine; alternate assignment for leucine) resulting in the production of heterogenous protein populations.

1.4.2 Sense codon reassignment strategies

Genetic code reassignment can be achieved using three different strategies: via codon invasion, through a toggle switch platform, or by engineering synthetases. The focus of codon invasion predominantly results in engineering tRNAs to provide a transient alternate identity at an already existing sense codon. The toggle switch system relies first on systematically knocking-out endogenous aaRSs and their associated tRNA isoacceptors to free up sense codons, before reintroducing (new) aaRS/tRNA reassignment pairs back into the *E. coli* host system. A subsequent step to this strategy is to develop an exogenous construct that, when reintroduced into the host, can flip between pairs of synthetases and/or tRNA(s) sets to investigate the aaRS/tRNA combinations that could be accepted by *E. coli*. The consequences of such a system can not only be useful in sense-to-sense codon reassignment, but would allow further exploration of tRNA and synthetase biology, as well as provide a platform for engineering orthogonality in these two key components of translation. Finally, reassignment can also be permitted through engineering synthetases to accept changes in substrate specificity, specifically towards accepting and charging alternate amino acids to its cognate tRNA substrates.

The specific approaches to how engineered tRNA/aaRS pairs can target codons for reassignment can depend on whether the target amino acid has degenerate codons, multiple isoacceptors, or orphan codons. The latter are complemented by tRNAs only through wobble base-pairing as there are no tRNAs that match orphans with Watson-Crick-pairing (Figure 1.11).

1.4.2.1 Sense-to-sense codon reassignment by codon invasion

Implementing sense-to-sense codon reassignment via codon invasion is fundamentally dependent on the ability to alter tRNA anticodons to match their target codons whilst maintaining the identity elements necessary for correct recognition and aminoacylation by cognate synthetases. That is, aaRS can still charge its cognate, albeit anticodon-modified, tRNA with its cognate amino acid, but the tRNA is able to base-pair with a different codon to the native set (e.g. LeuRS charging leucine to an anticodon-modified $tRNA^{Leu}$, which base-pairs with a histidine

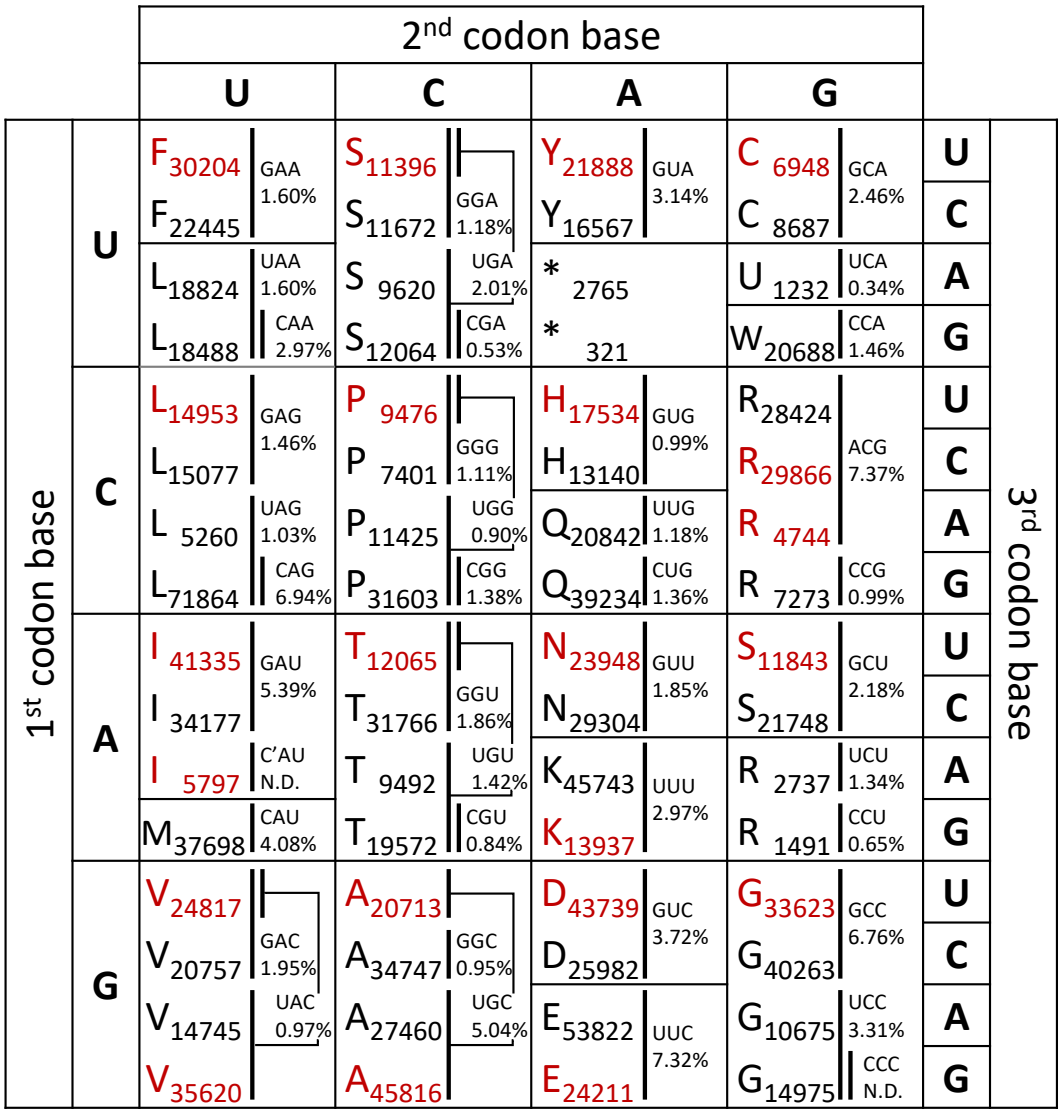


Figure 1.11: The genetic code table for *E. coli* with its associated tRNA isoacceptors

The genetic code annotated with the 46 decoding tRNA isoacceptors and their codon interactions (black vertical lines), intracellular tRNA isoacceptor abundances* (percentage next to black vertical lines), codon usage frequency** (number next to amino acid letter abbreviation), and orphan codons (red horizontal lines) for *E. coli*. Of the 64 codon variants: 22 are orphans (representing 17 of 20 amino acids); 3 code for stop codons Amber(UAG), Ochre(UAA), and Opal(UGA) (although the latter can also be reassigned to selenocysteine [U, Sec] in *E. coli*); 3 are singular codons (Met(AUG), Trp(UGG), Opal/Sec(UGA)). Adapted from Ho et al. (2016). *Isoacceptor abundances for each of the 64 codons were calculated using equation: isoacceptor abundance for a codon = (biochemically determined isoacceptor abundance) × (codon usage)/(total codon usage of all codons decoded by isoacceptor) (Ho et al., 2016; Lajoie et al., 2013; Dong et al., 1996).
**Codon usage values are based on NC_000913.2 (National Center for Biotechnology Information, 1 September 2011). N.D. indicates values not determined (Ho et al., 2016; Lajoie et al., 2013; Dong et al., 1996).

codon but transfers a leucine to the nascent peptide chain). Similar modifications will be required for the reverse assignment (e.g. HisRS aminoacylating histidine to anticodon-modified $tRNA^{His}$, which transfers histidine to the growing amino acid chain upon complementing a leucine codon), to complete the full sense-to-sense codon reassignments.

Orphan codons would be the prime targets for testing reassignment through codon invasion with anticodon-modified tRNAs. The rationale is that a tRNA that base-pairs with a codon through wobble pairing could be outcompeted if a new tRNA introduced into the system has perfect Watson-Crick-pairing with the codon; the latter would provide higher stability and stronger thermodynamic interactions with three base-pairs rather than two pairs and a weaker interaction (Kwon et al., 2003; Meroueh and Chow, 1999). As such, a tRNA with anticodon AAG could better complement orphan Leu(CUU), reducing the likelihood of wobble pairing from native *Eco*$tRNA^{Leu}_{(GAG)}$ (encoded by *leuU*), whilst simultaneously allowing sense-to-sense codon reassignment (Figure 1.12, bottom).

Although this strategy would not require knocking-out endogenous genes, the cell will likely suffer from toxicity as the polysemous Leu(CUU) codon will likely produce a heterogeneous population of protein products (i.e. naturally assigned proteins from native wobble pairing *Eco*$tRNA^{Leu}_{(GAG)}$, and alternately assigned products from Watson-Crick-pairing with the engineered $tRNA_{(AAG)}$). These issues could potentially be ameliorated by increasing the concentration of $tRNA_{(AAG)}$ to the exclusion of *Eco*$tRNA^{Leu}_{(GAG)}$ at Leu(CUU) and/or only transiently inducing the reassignment to limit accumulation of unrequired products. Moreover, the codon frequency of both Leu(CUU) and Leu(CUC) are low relative to the majority of the other leucine codons, as such, the chances of misincorporation – and hence increase in toxicity from miscoded protein production – of reassigned amino acids into essential proteins should also be low.

The process of invading an orphan His(CAU) follows the same method as with the Leu(CUU) orphan. There are only two histidine codons, His(CAU) and His(CAC), both of which only base-pair with *Eco*$tRNA^{His}_{(GUG)}$ (with wobble- and

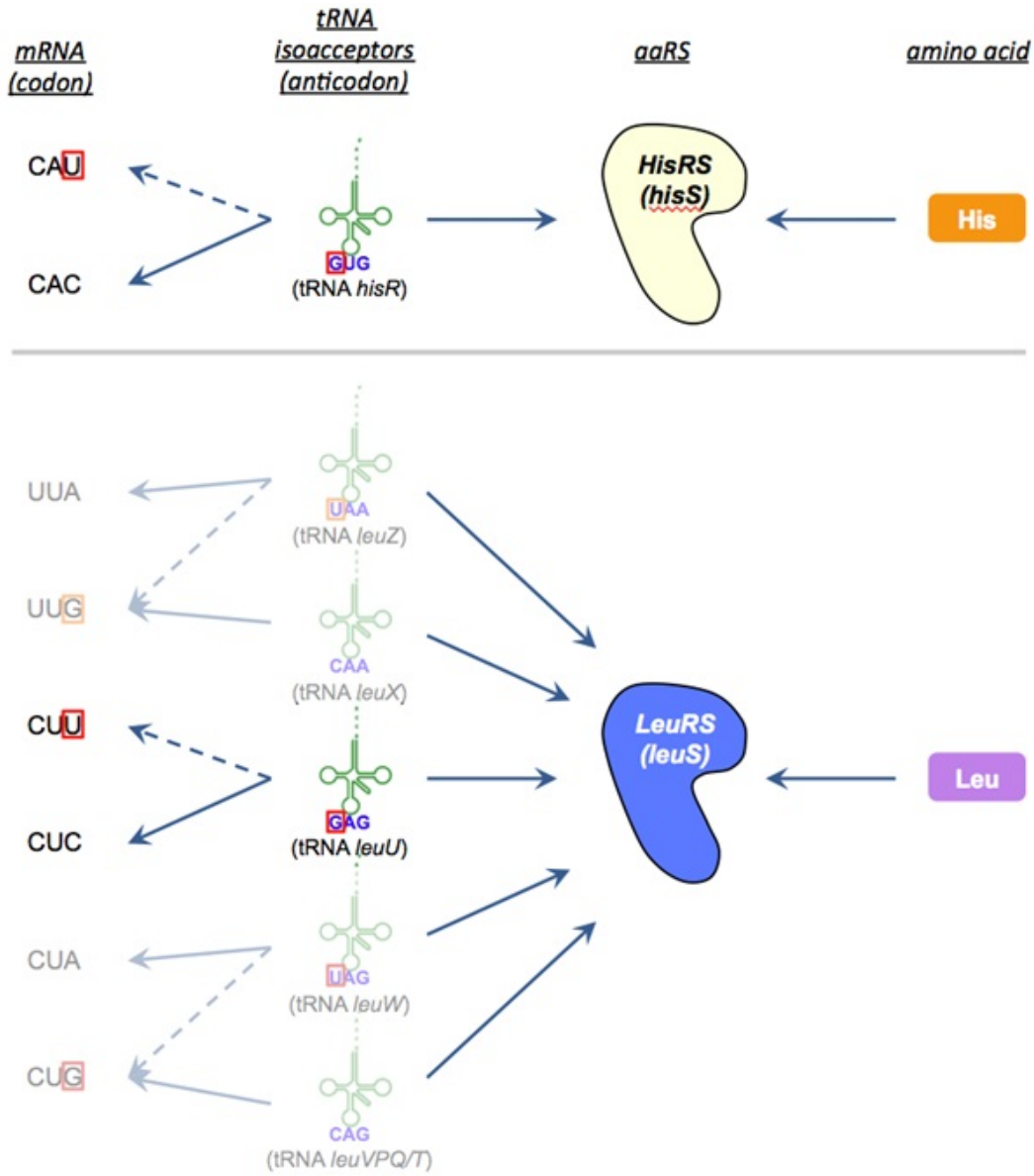


Figure 1.12: Mapping the codon-anticodon interactions for the translational machinery encoding histidine and leucine

Mapping the multiple many-to-many codon-anticodon, and the many-to-one tRNA-aaRS, interactions that are required for histidine (top) and leucine (bottom) assignment, following the decoding rules under the natural genetic code. Mapping interactions based from Ho et al. (2016).

Watson-Crick-pairing, respectively), which is only encoded by a single gene (*hisR*) (Figure 1.12, top). As there is only one isoacceptor for both histidine codons (both with comparable codon frequency), we assume that *hisR* is essential and that its tRNA transcript can only be charged by *Eco*HisRS (encoded solely by *hisS*). As with the engineered tRNA for invading the Leu(CUU) orphan codon, the engineered

tRNA for invading the orphan His(CAU) codon will have modified anticodon AUG – the perfect Watson-Crick base-pair complement.

In both codon invasion strategies, no particular amino acid was assigned to either $tRNA_{(AAG)}$ (the Watson-Crick match to orphan Leu(CUU)) or $tRNA_{(AUG)}$ (complement to orphan His(CAU)), because these tRNAs could potentially adopt any identity, provided their synthetases can aminoacylate them independently of their anticodon. Given that both native tRNAs possessed anticodon-independent identity elements for cognate synthetase recognition, we may assume that these engineered tRNAs (that have only been modified in the trinucleotide anticodon sequence) will retain this anticodon-independent recognition feature. As such, we could swap the identities of the two engineered tRNAs, i.e. $tRNA^{His}_{(AAG)}$ and $tRNA^{Leu}_{(AUG)}$ to test reassignment. The former would be charged with histidine but recognises codon Leu(CUU) (L→H), whilst the latter transfers a leucine to codon His(CAU) (H→L) (Figure 12). We can subsequently measure the efficiency of H→L and L→H, both independently and simultaneously, in reassigning singular and multiple codons in a recoded reporter system.

Once the engineered tRNAs have been introduced into the *E. coli* chassis, verification of reassignment activity and efficiency in a transient switching system can be assessed: a recoded fluorescent protein construct is the reporter of choice. The reporter is designed with an in-built requirement for leucine-histidine reassignment in order to be functional, which upon induction of the engineered tRNA/aaRS should incorporate histidine in place of leucine (L→H) and vice versa (H→L), thereby decoding a correctly expressing protein.

Specifically, sfGFP, with its fast maturation rates, clear visual phenotype, and low cellular burden, is ideal for providing a quick indicator of successful expression (Chudakov et al., 2010). Of particular interest for testing leucine-histidine reassignment is how mutations to the chromophore (residues 65-67) can affect the colour of the expressed fluorescence. In (sf)GFP, for example, the chromophore consists of S/T65-Y66-G67. However, a Y66H mutation can produce blue fluorescence (sfBFP) instead of green, and a further mutation, Y66L, renders the protein colourless

(Figure 1.13) (Chudakov et al., 2010). As such, we can recode the reporter to fluoresce either the blue (H66) or colourless (L66) variant, and test for a colorimetric shift upon inducing reassignment using the two sets engineered tRNA/aaRS pairs.

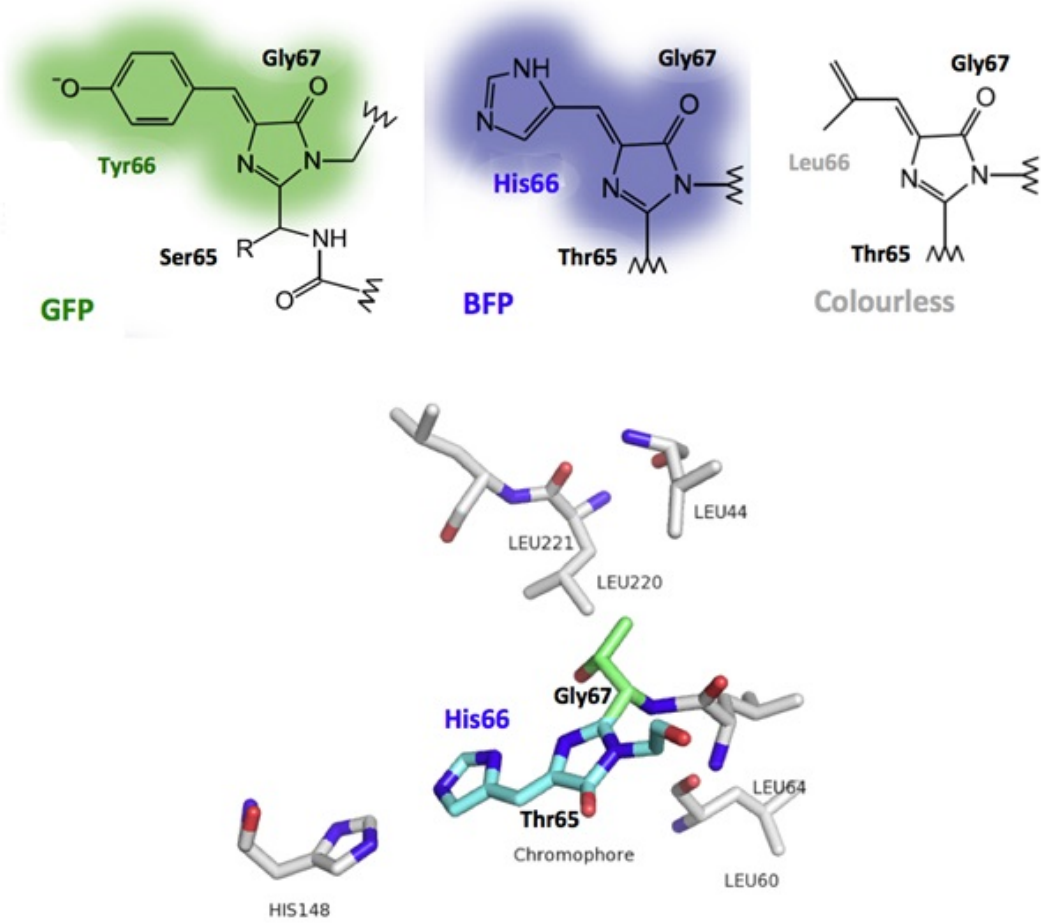


Figure 1.13: Representation of the chromophore in fluorescent proteins

Diagrammatic (top) and crystal structure (bottom) illustrating the chromophore (a cyclisation of residues 65-66) in green fluorescent protein (GFP) (top-left) and its colour-variant mutants, blue H66 (top-middle), and colourless L66 (top-right). Although this GFP variant shows a S65 residue, T65 would equally give the same phenotype. Adapted from Chudakov et al. (2010) Crystal structure of the expected sfBFP and its chromophore (T65-H66-G67) created as an amalgamation of BFP (PDB: 1BFP) and sfGFP (PDB: 2B3P) in order to show the identified leucine and histidine residues that could be reassigned using our *in vivo* transient leucine-histidine codon reassignment system.

1.4.2.2 The original aaRS/tRNA genetic toggle switch system

A parallel strategy for genetic code reassignment originates from the development of an *in vivo* aaRS and/or tRNA toggle switch system. In such a system, the aim was to knock-out genes coding for native synthetases and tRNA isoacceptors for

a given set of codons, reintroduce into *E. coli* an exogenous source of the native genes as part of a genetic toggle switch construct (to maintain host viability until a reassignment system is established: on by default, off when induced), before triggering the switch to test a different aaRS/tRNA set that is also encoded in the construct (Figure 1.14).

The degeneracy of the genetic code, existence of multiple isoacceptors (each of varying intracellular abundance levels, likely due to gene copy number variation), and species-dependent codon usage frequency, make it difficult to determine which codons, and associated tRNA isoacceptors, are essential for protein translation in a given species, and which could be repurposed (Figure 1.11; Ho et al., 2016).

While knocking-out LeuRS in *E. coli* (*Eco*LeuRS) may be straightforward (as it is encoded by only a single variant, *leuS*), the removal of its cognate isoacceptor, without compromising host viability, while feasible, may be more complex (Huang et al., 2012). This complexity is due to the uncertainty in ascertaining the essential tRNAs, given the mapping of many-to-many possible interactions between the five different *Eco*tRNA^{Leu} isoacceptors (transcribed from eight genes) to the six degenerate leucine codons (Figure 1.12, bottom; Thorbjarnardottir et al., 1985; Chan and Lowe, 2009). Isoacceptor abundance levels and codon frequency estimates may also help to clarify the relative importance of the various isoacceptors. For example, as the highly-abundant *Eco*tRNA$^{Leu}_{(CAG)}$, encoded by four of the eight *Eco*tRNA^{Leu} genes (*leuVPQ/T*), base-pairs with Leu(CUG) (which is incidentally the most commonly occurring codon in the *E. coli* genome with 71,864 instances; Sørensen et al., 2005; Ho et al., 2016), we would assume that this isoacceptor is essential for assigning leucine. However, the least abundant leucine isoacceptor, *Eco*tRNA$^{Leu}_{(UAG)}$ can also wobble pair with Leu(CUG) (as well as base-pair, with full Watson-Crick, to the rarest leucine codon, 5,260 instances), so we must also assume that, in the absence of *Eco*tRNA$^{Leu}_{(CAG)}$, *Eco*tRNA$^{Leu}_{(UAG)}$ may too be necessary.

In order to free a given codon for reassignment, Leu(CUG) for example, we would need to knock-out the five genes that encode endogenous *Eco*tRNA$^{Leu}_{(CAG)}$

and *Eco*tRNA$^{Leu}_{(UAG)}$ (*leuVPQ/T* and *leuW*, respectively). Whilst the removal of these isoacceptors may be sufficient to open up codon Leu(CUG) (and incidentally Leu(CUA) as well) for engineered tRNAs to pair with, it is not truly feasible for full codon reassignment as some leucine-encoding tRNAs (from genes *leuU,X,Z*) still remain. For this, we will need to continue to systematically and iteratively knock-out the remaining variants. Likewise, the complementing reassignment system - *Eco*HisRS (encoded by *hisS*) and its sole *Eco*tRNA*His* (*hisR*) - will also need to be deleted. Only then will the necessary environment for direct leucine-histidine codon reassignment be available.

Moreover, if the deletion of *leuVPQ/T,W* results in a non-viable host (i.e. insufficient leucine assignment from the remaining *leuU,X,Z* set), an exogenous supply of the removed *Eco*tRNA^{Leu} genes would need to be provided, at least until a leucine reassignment system is established.

A genetic toggle switch can be used for genetic code reassignment and it allows more severe interventions than simple codon invasion. A toggle switch would allow engineered tRNAs to target not only orphan codons but potentially all codons for a given amino acid.

In addition, genetic toggle switches can be used to probe synthetase and tRNA biology, testing the functional landscape of *E. coli* elements (whether aaRS or tRNA), measuring orthogonality of synthetases and tRNAs from other species, or replacing an *E. coli* aaRS/tRNA pair with an alternative system (e.g. switching *Eco*LeuRS for *Tth*LeuRS, with the potential to also introduce, or not introduce, *Tth*tRNA^{Leu} using a second toggle switch).

Engineered aaRS/tRNA (e.g. for L$\rightarrow$H) constructs, reintroduced into *E. coli* (e.g. with Δ*hisS,hisR*) will be toxic to the host. As such, a toggle switch system could be used to temporarily allay toxicity. The switch could be set-up with a functioning native *Eco*HisRS/*Eco*tRNA^{His} on one side to initially allow the host to survive and grow (i.e. under the natural code), but upon flipping the switch to investigate an altered genetic code using an engineered aaRS/tRNA set, the cell will likely mutate ore slowly perish.

While there are multiple methods to construct a toggle switch, our main design consists of a promoter (constitutive or inducible), flanked by an inverting pair of serine recombinase sequences (attB and attP), all of which sits between two oppositely-oriented components, e.g. *Eco*tRNA^{Leu} and *Tth*tRNA^{Leu} (Figure 14, orange boxes/-text); the promoter is originally arranged in the same direction as *Eco*tRNA^{Leu}. We can test this switch in an *E. coli* Δ*leuVPQ/T,W,U,X,Z* strain, where in its original state, the endogenous *Eco*LeuRS can use the exogenous *Eco*tRNA^{Leu} source for translation (phenotype: live *E. coli*). However, upon flipping the switch, i.e. inducing Bxb1 integrase activity, the promoter is inverted so that it allows only *Tth*tRNA^{Leu} transcription. As *Eco*LeuRS is unable to aminoacylate *Tth*tRNA^{Leu} (Soma and Himeno, 1998), translation with leucine will be prevented (phenotype: dead *E. coli*).

A similar system for testing the ability of *Eco*LeuRS and *Tth*LeuRS to aminoacylate endogenous *Eco*tRNA^{Leu} can also be constructed, again, with a reporter that provides a live/dead phenotype. However, we expect this system to show viable *E. coli* in both orientations of the switch as *Tth*LeuRS is known, at least *in vitro*, to be able to cross-aminoacylate *Eco*tRNA^{Leu} substrates (Kruger and Pinheiro, personal communication, December 2015; Krüger et al., 2015). In this instance, we can also incorporate a fluorescent protein fusion to *Tth*LeuRS, such that the live cells will fluoresce when the promoter is switched to the *Tth*LeuRS-FP direction (Figure 1.14, right-hand side of purple boxes/text).

Furthermore, the *in vivo* toggle switch system could also allow investigations into multiple combinations of switchable aaRSs and flippable tRNAs (Figure 1.14, purple boxes/text), where we can not only assess the engineered aaRS sets and tRNA pairs for implementing sense-to-sense codon reassignment, but gain more understanding into synthetase and tRNA biology as well. The latter two can be particularly useful in ascertaining the understudied rules that govern aaRS/tRNA (intra- and inter-specific) interactions across multiple species (Italia et al., 2017; Soma and Himeno, 1998; Weygand-Durasević et al., 1993). Subsequently, we may be able to use this *in vivo* toggle switch platform to evolve and investigate orthog-

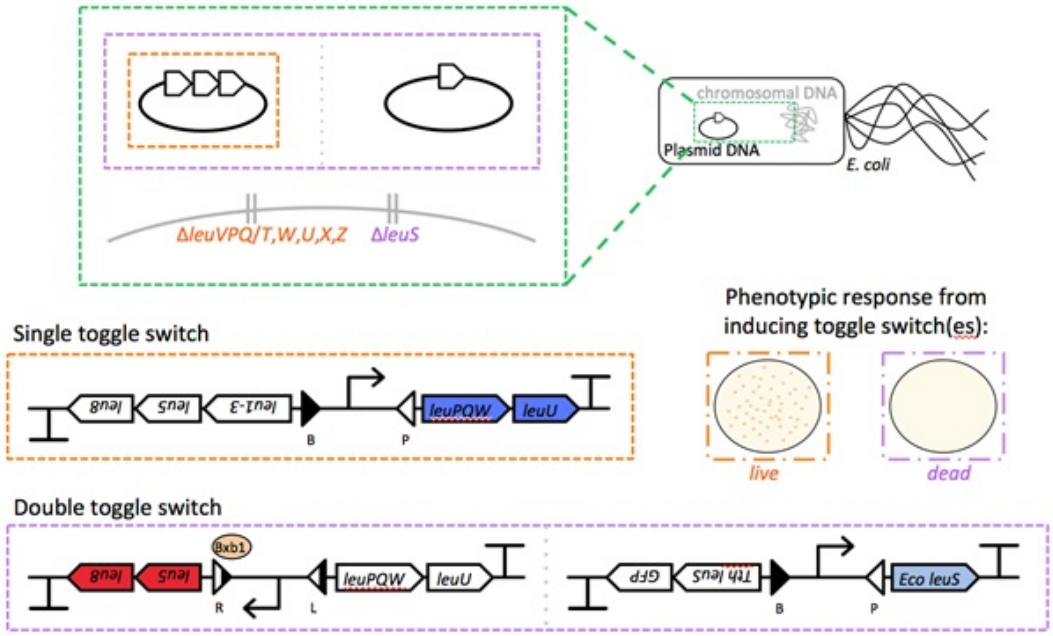


Figure 1.14: A schematic of the *in vivo* aaRS/tRNA genetic toggle switch system

Designs for the *in vivo* aaRS/tRNA genetic toggle switch system and how it can be used to measure orthogonal aaRS/tRNA systems. In the single toggle switch example (orange boxes/text), all genes for $tRNA^{Leu}$ isoacceptors have been deleted from the *E. coli* genome (Δ*leuVPQ/T,W,U,X,Z*) (the native LeuRS, encoded by leuS, has not been removed). We introduce into the cell a single toggle switch rescue plasmid that has genes for various *Eco*$tRNA^{Leu}$ (leuPQW,U) on one side (blue CDS) of the promoter (which is flanked by two inverted serine recombinase attB, attP sequences), and genes for *Tth*$tRNA^{Leu}$ (*leu1-3,5,8*) inverted on the other side. The endogenous *Eco*LeuRS can charge the exogenously supplied *Eco*$tRNA^{Leu}$ genes to allow cells to grow and survive. However, when Bxb1 serine integrase is induced, attB and attP irreversibly switch orientation (forming sites attR and attL), causing the promoter to be inverted such that it now promotes the transcription of the *Tth*$tRNA^{Leu}$ genes (red CDS; left side of the purple box at the bottom). Native *Eco*LeuRS cannot aminoacylate *Tth*$tRNA^{Leu}$, resulting in cell death. A double toggle switch system can be implemented (purple boxes), which allows the testing of four potential aaRS/tRNA combinations. If, as in this example, all endogenous genes for *Eco*LeuRS and *Eco*$tRNA^{Leu}$ was replaced with the double toggle switch, as *Eco*LeuRS cannot charge *Tth*$tRNA^{Leu}$, the phenotypic response would be cell death.

onal aaRS/tRNA pairs (such as intra-specific mutants) beyond those in the natural genetic code.

1.4.2.3 Synthetase engineering

The third strategy that can be pursued for genetic code reassignment is the engineering of synthetases to adopt alternate amino acids for aminoacylation. Thus, allowing an aaRS to charge a non-cognate amino acid to a cognate tRNA that will recognise its native codon, thereby creating an alternate assignment system. Moreover, given the potential of the previously described toggle switch platform in engineering/e-

volving orthogonality, if this system is successfully developed, we can use it to implement this third strategy for genetic code reassignment and direct the evolution of synthetases towards accepting alternate substrates (Bezerra et al., 2015).

1.4.3 A summary of the initial project aims

To summarise, the key aims of my project was to demonstrate that we can establish platforms capable of transiently reassigning sense codons *in vivo* by (i) invading orphan codons, and (ii) developing an aaRS/tRNA toggle switch platform, the latter of which can be further implemented in studying synthetase and tRNA biology, and in engineering orthogonality through directed evolution. Whilst codon invasion has been previously studied (Döring and Marlière, 1998; Kwon et al., 2003), this work would contribute greatly to the development of direct sense-to-sense codon reassignment *in vivo* (Döring and Marlière, 1998; Torres et al., 2016). The execution of reassignment for a colour-switching fluorescent protein is novel, and whilst we have not yet been able to fully demonstrate a histidine-leucine altered code, this platform has potential to provide further impact in the form of genetic informational containment. Finally, the toggle switch mechanism we are developing is unique and could prove to be an indispensable tool for furthering knowledge in the basic science of the genetic code translational machinery, as well as in providing secure semantic bio-containment of genetically engineered organisms.

1.5 Overview of this Thesis

This thesis summarises my efforts towards developing a robust and efficient platform of genetic circuits to introduce, detect, and assess the efficiency of sense-to-sense codon reassignment in *E. coli*.

The materials and methods used during the course of this research are detailed in Chapter 2. Further description of the oligonucleotides and genetic constructs created during the course of this work are also listed in Appendix A.

An initial approach for sense-to-sense codon reassignments focused on establishing a suitable host for codon reassignment research, where a natural tRNA and aaRS are provided *in trans* and can be easily engineered. Here, I used the lambda Red recombinase system to knock-out endogenous/genomic LeuRS and leucyl-tRNA ($tRNA^{Leu}$) genes in *E. coli*, followed by designs for implementing plasmid supplementation of rescue constructs. The primary rescue construct was designed as a bilateral switch, which on one side consisted of the native *Eco*LeuRS and one of the *Eco*$tRNA^{Leu}$ genes to maintain a functioning host (the initial "on" state), and an engineered tRNA/aaRS pair on the other side (initial "off" state) to test sense-to-sense codon reassignment. Upon expression of another heterologously-expressed gene, Cre recombinase (Cre) (and/or Bxb1 serine integrase (Bxb1)), the toggle switch would shift from expressing the native rescue *Eco*LeuRS/*Eco*$tRNA^{Leu}$ pair to the engineered codon reassignment set.

The initial plans for this approach was summarised earlier in Section 1.4, however, due to difficulties establishing this initial knock-out/rescue strategy for codon reassignment, I opted to pursue an alternative route. Instead of fully replacing the *E. coli* aaRS and tRNA, I decided to invade a natural codon using an anticodon-modified tRNAs: a strategy referred to as Orphan Codon Invasion (OCI) that mirrors the work previously carried out by Döring and Marlière (1998), and that is the focus of Chapter 3. As codon invasion would still be expected to be toxic to the host, the development of a reporter system that could minimise toxicity would be greatly beneficial to the detection of reassignment. The Genetic Toggle Switch (GTS) system devised would convert a transient codon invasion signal into a per-

manent reporter activation, and this is the focus of Chapter 4.

More specifically, in Chapter 3, anticodon-modified tRNAs and recoded reporter constructs were introduced as exogenous (plasmid) DNA into an unmodified host, allowing engineered tRNAs to compete with the endogenous tRNAs for the native aaRSs – ideally creating a population of alternative tRNAs that can lead to the synthesis of a stochastic protein, introducing different amino acids at the engineered anticodon. Whilst this approach would not negate cellular toxicity resulting from (mis-)expression of native genomic and engineered exogenous genes, it served as a proof-of-concept[4] that sense-to-sense codon reassignment was possible with synthetic tRNAs working in parallel with the existing natural translation machinery of the host. Chapter 3 also describes the design and generation of the first recoded reporters created to assess the viability of the anticodon-modified tRNAs for codon reassignment *in vivo*.

Knowing that codon invasion, due to mis-expression of other *E. coli* protein, would reduce host fitness and therefore mask successful invasion, I decided to look into strategies that could minimise that toxicity, primarily by limiting the exposure of the host to the modified genetic code. In Chapter 4, I adapt the core mechanism that was being developed in earlier work (data not shown) – the genetic switch to swap LeuRS/tRNA^{Leu} variants. Here, I pursued an approach where the successful expression of a functional recombinase reporter (which can only be achieved through sense-to-sense codon reassignment) triggers a genetic rearrangement to stop the codon invasion and to stably express a common reporter. Chapter 4 describes the design and assembly of such functional switches, which I term Genetic Toggle Switch.

The end of Chapter 4, specifically Section 4.2.7, details the incorporation of OCI and GTS into a viable *in vivo* platform for genetic code engineering. I demonstrate that codon reassignment can be achieved from leucine-to-histidine, and explain a means to which histidine-to-leucine may have been accomplished. The scripts used to create and analyse the flow cytometry data is documented in Ap-

[4] As well as a replication of prior research (Döring and Marlière, 1998), albeit with an alternative pair of sense codons/amino acids; histidine-leucine, instead of cysteine-isoleucine/methionine.

pendix B.

Finally, I make my closing discussions and conclusions in Chapter 5, and summarise both the benefit of the genetic circuit platform developed, and the impact of being able to use orthogonal genetic code systems in genetically modified and genetically recoded organisms.

Chapter 2

Materials and Methods

2.1 General Reagents and Materials

2.1.1 Bacterial Strains

Table 2.1 (page 45) lists the *E. coli* K-12 and B-strains used in this work respectively. Throughout the text, *E. coli* strains will be referred to by the strain names as listed here.

For the most part, an *E. coli* DH10β derivative was used as the cloning strain, whilst T7 Express *lysY/I*^{q} and BL21(DE3) were used as expression strains. DH5α_Z1 was used briefly as an expression strain in Chapter 4, and *E. coli* β1308 was used to replicate the ThyA reassignment assay by Döring and Marlière (1998).

Strain (Genotype)	Source	Application
K-12 strain derivatives		
NEB® 5-alpha Competent *E. coli* (High Efficiency) – K-12 derivative (*fhuA2Δ(argF-lacZ)U169 phoA glnV44 Φ80Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*)	NEB #C2987	Used for transforming a minority of plasmid constructs created. Efficient competent cells for subcloning.
DH5α_Z1 – K-12 derivative (*lacI^q, PN25-tetR, Sp^R, deoR, supE44, Δ(lacZYA-argFV169), Φ80 lacZΔM15, hsdR17(rK- mK+), recA1, endA1, gyrA96, thi-1, relA1*)	Provided by T. Folliard See also Expressys	Used to ensure minimal uninduced expression from pTetR/Apromoters, due to two genomic copies of *tetR* and *lacI* in the cellular genome.
NEB® 10-beta Electrocompetent *E. coli* 10β – K-12 derivative (*Δ(ara-leu) 7697 araD139 fhuA ΔlacX74 galK16 galE15 e14- Φ80ΔlacZΔM15 recA1 relA1 endA1 nupG rpsL (Str^R) rph spoT1 Δ(mrr-hsdRMS-mcrBC)*)	NEB #C3020	Used for cloning large plasmids and BACs.
***E. coli* β1308** MG1655 (*F$^-$λ$^-$*) derivative: *ΔthyA::erm$^+$*	Provided by R. Mutzel; V. Pezo	To replicate work by Döring and Marlière (1998)
B strain derivatives		
***E. coli* T7 Express *lysY/I^q* (High Efficiency)** – enhanced BL21 derivative (*MiniF lysY lacI^q(Cam^R) / fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10--Tet^S)2 [dcm] R(zgb-210::Tn10--Tet^S) endA1 Δ(mcrC-mrr) 114::IS10*)	NEB #C3013	Used for expressing constructs under pT7 control.
***E. coli* BL21(DE3)** – protease deficient B strain (*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*)	NEB #C2527	Used for co-transforming multiple plasmid constructs with various antibiotic resistances, and expressing constructs under pT7 control.
***E. coli* C41(DE3)**		Used for expressing constructs under pT7 control.

Table 2.1: List of bacterial K-12 and B strain derivatives used in this body of work

2.1.2 Parental Plasmids

Table 2.2 (page 46) lists the parental vectors/plasmids in the lab collection that were used in this work. All genetic constructs created were derived from this set. For plasmid sequence and maps, refer to Appendix A.

Plasmid	**Ori compatibility group**	**Copy number**	**Native antibiotic resistance marker**	**Promoter**
pUC19	ColE1/pMB1/ pBR322/pUC	100~300	AmpR, bla$^+$	pLac
pUC57 [P002]	ColE1/pMB1/ pBR322/pUC	100~300	KanR, kan$^+$	pLac
pET29a(+) [P003]	ColE1/pMB1/ pBR322/pUC	~20	KanR, kan$^+$	pT7lac
pSB1C3	ColE1/pMB1/ pBR322/pUC	100-300	AmpR, bla$^+$	N/A
pNGAL97 [P005]	ColE1/pMB1/ pBR322/pUC		AmpR, bla$^+$	pTetR/A
pBAD30	p15a	10~15	AmpR, bla$^+$	pBAD
pD881 [P008]	p15a	10~15	KanR, kan$^+$	prhaBAD
pCDF-Duet	CloDF13	20~40	SpcR, spt$^+$ StrR, stm$^+$	pT7lac
pKD46	pSC101/Rep101	~5	AmpR, bla$^+$	pBAD

Table 2.2: List of parental plasmids/vectors used in this body of work
Antibiotic resistance: + refers to resistant, - refers to sensitive to the associated antibiotic.

Co-transformation of plasmids was only carried out using plasmids with compatible (i.e. different compatibility groups) origins, to ensure plasmid stability.

2.1.3 Key Genetic Constructs

The key genetic constructs created and used in this body of work is as follows, with the majority being derived from parental vectors/plasmids from Section 2.1.2 Parental Plasmids. For completeness, Appendix A.1 (page 253) includes the context for each of the key constructs.

2.1.3.1 *E. coli* and *T. thermophilus* leucyl-tRNA synthetase

DNA and amino acid sequences for *Eco*LeuRS (*leuS*; 2,580 bp, 860 AA, 97 kDa) were obtained from NCBI (*E. coli* strain K-12 sub-strain MG1655 NC_000913.3, locus tag b0642 [FASTA]). Primers YKH007/YKH008 were used to PCR amplify the *leuS* sequence. The amplified product was purified and subcloned into [P003] `pET29a(+)` (4,852 bp) via `NdeI`/`HindIII` restriction cloning into the multiple cloning site (MCS), whilst simultaneously removing the upstream thrombin and S-tag sequences.

The gene sequence for *Tth*LeuRS (2,634 bp, 878 AA, 101 kDa) was also obtained from NCBI (*Thermus thermophilus (Tth)* NC_006461.1, locus tag TTH_RS00845 [FASTA]) was similarly obtained by post-doctoral researcher Dr. Antje Krüger through genomic DNA extraction and subcloning into a modified `pET29a(+)` vector.

Both *Eco*LeuRS and *Tth*LeuRS were used in the earlier strategy to implement a genomic knock-out and subsequent rescue with heterologous expression of the synthetases and tRNAs from plasmids.

2.1.3.2 *E. coli* histidyl-tRNA synthetase

DNA and amino acid sequences for *Eco*HisRS (*hisS*) were obtained from NCBI (*E. coli* strain K-12 sub-strain MG1655 NC_000913.3, locus tag b2514 [FASTA]). Designs were made using this sequence in the earlier knock-out/rescue strategy for sense-to-sense codon reassignment but this approach was subsequently discontinued in favour of the OCI-GTS system.

2.1.3.3 *E. coli* leucyl-tRNA and histidyl-tRNA

The engineered $tRNA^{Leu}$ and $tRNA^{His}$ for constructing the codon-reassigning anticodon-modified tRNAs of the Orphan Codon Invasion (OCI) system were based on the gene sequences *leuU* ($tRNA^{Leu}_{\underline{G}AG}$) and *hisR* ($tRNA^{His}_{\underline{G}UG}$) respectively. For the former, positions 34 and 35 of the anticodon were changed to $tRNA^{Leu}_{\underline{A}UG}$ (to have full Watson-Crick complementarity to histidine codon $His_{(CA\underline{T})}$), and changed to $tRNA^{His}_{\underline{A}AG}$ in the latter (to pair with leucine codon $Leu_{(CT\underline{T})}$). These design sequences were concatenated together and ordered as a gBlock® from IDT. Further details on the design (Section 3.1.3), creation (Section 3.2.1), and implementation (Section 3.2.5) of these engineered tRNAs are documented in Chapter 3.

2.1.3.4 Cre and lambda Red Recombinase Constructs

DNA and amino acid sequences for the Cre recombinase (a tyrosine recombinase[1]; X03453.1 [EMBL] [FASTA]; P06956) were obtained from the European Nucleotide Archive and UniProt respectively. Sequences matching the `EcoRI` and `HindIII` within the MCS of the [P007] `pBAD30` (4,919 bp) destination vector were added to each end of the 1,032 bp *cre* gene, and the resulting 1,128 bp *cre* construct (YKH036) was commercially synthesised (IDT gBlocks®, Belgium). Primer pairs YKH059/YKH060 and YKH057/YKH058 were designed to create an equivalent *cre* construct in vector [P008] `pD881`, whilst YKH076/YKH077 were used to insert a 9 bp (5'-TGAAATTCT-3') sequence upstream of the ribosome binding site (RBS) as a means to control leaky expression of the Cre recombinase.

To combine Cre with the lambda Red homologous recombination system, primers YKH093/YKH094 were designed, which would amplify the araC regulator, the araBAD promoter, and its three downstream lambda Red components (Gam, Beta, Exo) from the [P010] `pKD46` plasmid. This could be combined via Gibson Assembly with an `NheI`-linearised plasmid containing the rhaB-controlled *cre* recombinase. The sizeable (8,023 bp) resulting Cre/lambda Red plasmid was

[1] Alternative recombinases that were considered for use in this system included Bxb1 and the TP901 integrase. These were kindly provided by Tom Folliard as part of a 6,613 bp "Dual controller plasmid", where Bxb1 was under control of a pTetO promoter, TP901 was regulated by the araBAD promoter, and the plasmid had a ColE1 ori and chloramphenicol (Cam) resistance marker.

taken forth as the means to create the initial genomic integration/knock-out genetic switch for codon reassignment – the precursor of the final OCI-GTS system.

2.1.3.5 ΦC31 Recombinase Construct Variants

The DNA sequence for the *ϕc31* serine recombinase (1,866 bp, 621 AA, 69 kDa) was kindly provided in a [P003] `pET29a(+)` vector (totalling 7,080 bp) by then-PhD candidate, Hugo Villanueva[2]. Its gene sequence was taken as the base design for the *ϕc31* recombinase that was subsequently used in the prototypical Genetic Toggle Switch (GTS) circuit in Chapter 4. The final 3,269 bp [002] design construct (Figure 2.1) was commercially synthesised (Twist Biosciences, USA).

It should be noted that there were two approaches for recoding ΦC31 into an inactive variant. For making multi-site histidine-leucine recombinase recodings within the Inversion-sfGFP switch, the new Darwin Assembly method (Section 2.2.18) was deployed. Section 4.2.6.1 was dedicated to describing this approach, and the characterisation of the library of constructs. The second approach uses a more typical site-directed mutagenesis (SDM) approach with inverse polymerase chain reaction (iPCR) amplification to create either one or two point mutations at single sites within the recombinase in the Excision-tsPurple switch. The design rationale and construction of these ΦC31 recodings were detailed in Section 4.2.6.2.

2.1.3.6 Fluorescent and Chromoprotein Reporter Constructs

Fluorescent Proteins – the sfGFP and EBFP variants:

Fluorescent proteins, particularly superfolder green fluorescent protein (sfGFP), were used in the development of both OCI recoded reporters for demonstrating sense-to-sense codon reassignments (Chapter 3), as well as in the creation of a GTS circuit with a naturally-coded secondary/proxy reporter for amplifying the signal of histidine-leucine reassignment without adversely affecting the viability of the host organism (Chapter 4). The gene construct was created and kindly provided as plas-

[2]Hugo also provided Bxb1 (1,566 bp, 521 AA, 58.8 kDa) in `pET29a(+)` (totalling 6,780 bp). Whilst he was able to purify Bxb1, he did not have success in using it *in vitro*; *in vivo* expression was not tested. A collaborator, Dr. Andy Osborne, who was provided with the same constructs, was able to express and use the ΦC31 construct, hence, the GTS system was based on this recombinase.

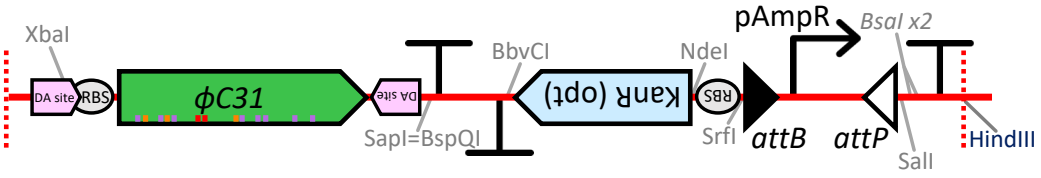


Figure 2.1: The design of the prototypical GTS ΦC31 recombinase circuit [002]

Chapter 4 discusses the role of the ΦC31 recombinase (in green) within the wider GTS circuit [002]. ΦC31 forms one half of this 3,269 bp GTS Inversion-kanamycin resistance (Kan^R) circuit that was commercially synthesised (Twist Biosciences, USA). The purple notches within the green *ϕc31* represents leucine codon sites that could be recoded, whilst orange notches show histidine sites. The two red notches identify two asparagines (at positions 200 and 204) that were recoded to $*Am_{(TAG)}$ stop codons. Notches are not positioned to scale. The other half of the circuit comprises of a Kan^R (in blue) construct. The only difference this presents from a typical operon is that the promoter is enclosed within *attB/attP* site-specific recombination sequences and that the promoter in this initial state is positioned facing the opposite direction to the rest of the Kan^R operon. Refer to Figure 4.9 for the full [002] design.

mid [203] (within a modified [P003] `pET29a(+)` vector; Cozens et al., 2012) by post-doctoral researcher, Dr. Chris Cozens.

In Chapter 3, alternative fluorescent variants of sfGFP – potential OCI codon reassignment reporters – were created through site-directed mutagenesis with the introduction of either a Y66H or Y66L substitution into plasmid [203]. For the two blue fluorescent variants, termed EBFP1.1, a reassign-able $Y66H_{(CAT)}$ variant was made with iPCR using primers YKH195/YKH201, whilst the synonymous but not reassign-able $Y66H_{(CAC)}$ was made with YKH165/YKH199. Similarly, non-fluorescent versions, termed noFP1.1, were produced using the full set of leucine codons. The primers used to make the synonymous leucine codon substitutions at residue 66 are listed in Table 2.3. After amplification, each product was treated with `DpnI/ExoI` (to remove the template plasmid and the single-stranded DNA (ssDNA) primers), phosphorylated with T4 DNA PNK (to enable ligation), and self-/re-ligated with T4 DNA ligase, before being transformed into both *E. coli* DH10β cloning and T7 Express *lysY/I^q* expression strains. A more detailed description of the generation of these sfGFP and EBFP1.1 variants, including the characterisation process is in Section 3.2.2.2.

Another set of blue fluorescent protein variants were also created from a known and established EBFP2 as obtained from FPbase (`https://www.fpbase.org`).

Proposed Point Mutation	SDM Primer ID	SDM Primer Sequence (5'->3')	Rv ID
Y66	YKH200	caccctgaccTATggcg	2
Y66His$_{(CA\underline{T})}$	YKH195	caccctgaccCATggcgttca	2
Y66His$_{(CAC)}$	YKH165	ccaccctgaccCACggcgttcagtg	1
Y66Leu$_{(CT\underline{T})}$	YKH176	caccctgaccCTTggcgttcagtg	2
Y66Leu$_{(CTC)}$	YKH188	ccaccctgaccCTCggcgttcagtg	1
Y66Leu$_{(CTA)}$	YKH189 | YKH190	ccaccctgaccCTAggcgttcagtg	1
Y66Leu$_{(CTG)}$	YKH190_v2	ccaccctgaccCTGggcgttcagt	1
Y66Leu$_{(TTA)}$	YKH196	caccctgaccTTAggcgttcagt	2
Y66Leu$_{(TTG)}$	YKH197	caccctgaccTTGggcgttcagt	2
SDM_Rv1	YKH199_Rv	TCACCAGGGTCGGCC	-
SDM_Rv2	YKH201_Rv	gTCACCAGGGTCGGC	-

Table 2.3: Primer designs for introducing point mutations in the sfGFP gene sequence

Paired with primer YKH199, these primers enable polymerase chain reaction (PCR) amplification to introduce the Y66, Y66H and Y66L mutations into the sfGFP sequence withn the [203] plasmid.

Not only was the sequence design codon-optimised for use in both *E. coli* and for re-assignment, additional flanking ends (including pT7 promoter, LacO operator, and RBS on the 5'-end, and T7 terminator on the 3'-end of the gene sequence) were added to make it suitable for expression and compatible for direct assembly into `pET29a(+)`. This 1,021 bp [025] construct was synthesised commercially (Twist Biosciences, USA). A further blue version, EBFP2.1 was made from EBFP2 by introducing the three mutations F99S, M153T, V163A, to bring it closer to that of sfGFP. Likewise, a non-fluorescent noFP2.1 followed on from EBFP2.1 with a simple H66L substitution.

A full summary of the green, blue, and non-fluorescent proteins that were synthesised and modified for use as OCI reassignment reporters are listed in Table 3.3 and are discussed in more detail in Section 3.2.2.2.

As for the use of sfGFP in the GTS system in Chapter 4: the *sfgfp* gene sequence from plasmid [071] (a replica of [203]) was simply subcloned into the basic Inversion switch circuit, as described in Section 4.2.4.1 (Figure 2.2).

Chromoproteins – TinselPurple:

TinselPurple is a vibrant non-fluorescent chromoprotein (Liljeruhm et al., 2018),

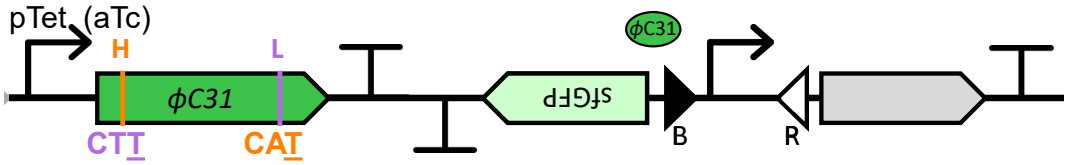


Figure 2.2: A schematic of the GTS Inversion-sfGFP circuit design

This design is one of the GTS circuits designed in Chapter 4. The primary reporter is the histidine-leucine recoded ΦC31 recombinase (in dark green), with the naturally-coded sfGFP (in light green) as the secondary/proxy reporter for augmenting the signal of ΦC31. The ΦC31 and sfGFP operons were designed in opposing directions to prevent accidental read-through. Refer to Figure 4.5 for the full implementation of this circuit.

sourced from the iGEM Parts Registry (BBa_K1033905, a composite part including the consensus RBS). The design of the tsPurple portion of the circuit was broadly that of a typical operon, except that it had a double-terminator sequence enclosed within two site-specific recombination sites (*attB/attP*) that interrupted the operon between the promoter and the RBS-tsPurple sequence (Figure 2.3).

To facilitate commercial synthesis of the construct, the design was split into two parts: a 415 bp switch fragment (including a 5'-segment of tsPurple) [032] was synthesised as a linear double-stranded DNA (dsDNA) gBlock® (IDT, Belgium), whilst the remaining 1,066 bp segment [022] was synthesised by Twist Biosciences (USA) and arrived in a pTwist_Kan_MC vector. The full construction is shown in Figure 4.18, and fully detailed in Section 4.2.5.1.

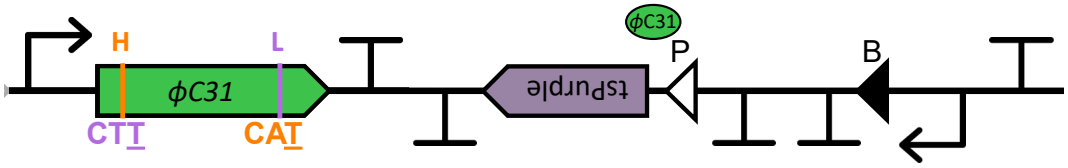


Figure 2.3: A schematic of the GTS Excision-tsPurple circuit design

This design is one of the GTS circuits designed in Chapter 4. The primary reporter is the histidine-leucine recoded ΦC31 recombinase (in green), with the naturally-coded TinselPurple (in purple) as the secondary/proxy reporter for augmenting the signal of ΦC31. The ΦC31 and TinselPurple operons were designed in opposing directions to prevent accidental read-through. Refer to Figure 4.5 for the full implementation of this circuit.

2.1.3.7 Periplasmic Secretor Constructs

Like the EBFP variants, two periplasmic secretor constructs were designed as re-coded reporters for use in the OCI system (Chapter 3). This was based on the two

major pathways used by bacteria to secrete proteins across the cytoplasmic membrane: the general *Sec*retion route (the Sec-pathway) that transports unstructured proteins (i.e. proteins in their unfolded conformation) across the transmembrane, and the *T*win-*a*rginine *t*ranslocation pathway (the Tat-pathway) that catalyses the translocation of folded proteins. One of the recoded reporters was formed of the signal sequences from OmpA (the initial 66 bp from NCBI *E. coli* NC_000913.3, locus tag b0957 [FASTA]) of the Sec-pathway, whilst the other was from SufI (the initial 81 bp from NCBI *E. coli* NC_000913.3, locus tag b3017 [FASTA]) of the Tat-pathway; both were coupled to the *lacZ*α sequence (225 bp, 75 AA; from 4-228 bp, residues 2-76 AA from NCBI *E. coli* NC_000913.3, locus tag b0344 [FASTA]).

Section 3.2.3.2 describes the design and construction of both the OmpA-LacZα construct [027] (438 bp) and the SufI-LacZα segment YKH388 (205 bp) in more depth, including how the latter could be made from the former once OmpA-LacZα had replaced the recoded reporter sfGFP (via Gibson Assembly) within the OCI construct. Construct [027] was commercially synthesised and subcloned into [P011] pTwist-Kan vector by Twist Biosciences (USA), whilst segment YKH388 (205 bp) was synthesised as a dsDNA gBlock® (IDT, Belgium).

2.1.4 Oligonucleotides

Refer to Appendix A.2 (page 253) for a list of oligonucleotides used in this body of work. Unless stated otherwise, all oligonucleotides were synthesised by IDT. Some larger genetic construct fragments were synthesised by GeneWiz and Twist Biosciences.

2.1.5 Enzymes

Unless otherwise stated, all enzymes used in cloning procedures were purchased from New England Biolabs (NEB), USA. dNTPs were purchased from Bioline, UK. Table 2.4 (page 54) lists all other enzymes used creating constructs for this work.

Enzyme	**Supplier**
General Enzymes	
Restriction endonucleases (see text)	NEB, Roche
Other Nucleases	
RNase A	Sigma Aldrich
Molecular Cloning Enzymes	
Shrimp alkaline phosphatase (SIP)	Promega
T4 DNA ligase	NEB
T4 polynucleotide kinase (T4 PNK)	NEB
Polymerases	
Q5 High-Fidelity Hot Start DNA polymerase (NEB #M0493)	NEB
KOD Xtreme™ Hot Start DNA polymerase (Novogen #71975)	Novogen
MyTaq DNA Polymerase (Bioline ##BIO-21113)	Bioline
OneTaq DNA Polymerase (NEB #M0481)	NEB
PCRBio Ultra Polymerase (#PB10.31)	PCR Biosystems

Table 2.4: List of enzymes used and their suppliers

2.1.6 Chemicals

General purpose chemicals and reagents were supplied by Fisher Scientific, Invitrogen, Sigma Aldrich, and Stratagene, unless otherwise stated. Table 2.5 (page 55) lists the general reagents provided by the aforementioned companies.

Reagent / Material	**Supplier**
Agarose (Molecular Grade)	Bioline
Ammonium persulphate (APS)	
Biotin	Sigma Aldrich
Bovine serum albumin (BSA)	NEB
Deoxyribonucleoside triphosphate set (PCR Grade; sodium salts, 100 mM each)	Roche Applied Science
1 kb DNA ladder	NEB, Invitrogen Life Technologies
Quickload 1 kb or 100 bp DNA ladders	NEB
Dihydrostreptomycin, streptomycin sulphate salt	Sigma Aldrich
Dynabeads(R) MyOne(TM) Streptavidin T1	Invitrogen Life Technologies
Filter paper	Munktell
GeneJet PCR purification and gel extraction kits	GeneJet
Minisart(R) syringe end filters (0.2 µm and 0.45 µm)	Sartorius, Fisher Scientific
Monarch PCR purification and gel extraction kits	NEB
β-Nicotinamide adenine dinucleotide hydrate (NAD)	Sigma Aldrich (#N7004)
NucAway(TM) spin columns	Ambion(R), Life Technologies
Prestained protein marker, broad range (11-245 kDa)	NEB
QIAQuick PCR purification and gel extraction kits	QIAGEN
Quick ligase kit)	NEB
Ribonucleoside triphosphate set (lithium salts, 100 mM each	Roche Applied Science
Streptavidin sepharose high performance	GE Healthcare
N,N,N',N'-tetramethylethylenediamine (TEMED)	

Table 2.5: List of chemical reagents used and their suppliers

2.1.7 Buffer, Solutions, and Media

Table 2.6 (page 56) lists the buffers, solutions, and Table 2.6 (page 57) lists media used in this work.

Name	Constituents
General Use Buffers	
Annealing buffer (10X) (molecular cloning)	100 mM Tris-HCl pH 7.5, 1 M sodium chloride, 2.5 mM EDTA
DNA loading buffer (6X)	50% (v/v) glycerol, 100 mM EDTA pH 8.0, 1% (w/v) SDS, 0.1% (w/v) bromophenol blue, 0.1% (w/v) xylene cyanol. VP: 50% glycerol with a pinch of bromophenol blue works well as 5x. 15% Ficoll 400, 10 mM EDTA, 0.1% (w/v) orange G or 0.1% (w/v) bromophenol blue.
Laemmli's buffer (2X)	40% (v/v) glycerol, 4% SDS, 125 mM Tris-Cl pH 6.8, 200 mM DTT, 0.01% (w/v) bromophenol blue
PAGE running buffer (1X)	50 mM Tris base, 384 mM glycine, 0.1% (v/v) SDS
Phosphate-buffered saline (PBS, 1X)	137 mM sodium chloride, 2.7 mM potassium chloride, 10 mM sodium phosphate dibasic, 1.5 mM monopotassium phosphate (pH 6.7)
PBST	PBS with 0.1% (v/v) Tween-20
Tris-Borate-EDTA (TBE) buffer (10X)	89 mM Tris base, 89 mM orthoboric acid, 5 mM Disodium EDTA
Solutions	
19:1 40% w/v acrylamide/bisacrylamide solution	Supplied by AccuGel, National Diagnostics
37.5:1 40% w/v acrylamide/bisacrylamide solution	Supplied by AccuGel, National Diagnostics
Coomassie Blue solution	45% (v/v) methanol, 10% (v/v) acetic acid, 0.2% (w/v) Coomassie Brilliant Blue R-250 (Sigma-Aldrich)

Table 2.6: List of buffers, solutions, and media used in this body of work

Name	Constituents
Culture Media	
2X TY medium	16 g tryptone, 10 g yeast extract, 5 g sodium chloride per litre
Lysogeny Broth (LB) aka. Luria-Bertani medium	10 g tryptone, 5 g yeast extract, 5 g sodium chloride per litre; antibiotics added where required (e.g. ampicillin at 50 µg/mL or kanamycin at 30 µg/mL
Super Optimal Broth (SOB)	10 g tryptone, 2.5 g yeast extract, 0.29 g sodium chloride, 0.093 g potassium chloride per 500 mL
SOC	see SOB + additives
Terrific Broth (TB)	12 g tryptone, 24 g yeast extract, 4 mL glycerol, 17 mM potassium dihydrogen phosphate, 72 mM dipotassium hydrogen phosphate per litre
LB-Miller	10 g tryptone, 5 g yeast extract, 10 g sodium chloride per litre
M9 Minimal medium	15 g monopotassium phosphate (KH_2PO_4), 64 g disodium phosphate (Na_2HPO_4), 2.5 g sodium chloride, 5 g ammonium chloride (NH_4Cl)

Table 2.6: (Continued) List of buffers, solutions, and media used in this body of work

2.1.8 Equipment

Table 2.7 (page 58) lists the equipment used in this work.

Equipment	Company
Analysers and Imagers	
FACSCaliburTM	BD Biosciences
LI-COR Odyssey Scanner	LI-COR
Nanodrop 1000 Spectrophotometer	Thermo Scientific
Typhoon TRIO Variable Mode Imager	GE Healthcare
Typhoon FLA 9500 scanner	GE Life Sciences
FLUOstar OPTIMA	BMG
Centrifuges	
Eppendorf 5810R Centrifuge	Eppendorf
OptimaTM MAX-E Benchtop ultracentrifuge	Beckman Coulter
OptimaTM L-90 K Preparative ultracentrifuge	Beckman Coulter
Tools	
Bio-Rad Powerpac 200 and 300	Bio-Rad
DynaMagTM-Spin Magnet	InvitrogenTM, Life Technologies
Thermocyclers	
C1000 Touch	Bio-Rad
Peqstar	PeqLab

Table 2.7: List of equipment used in this body of work

2.1.9 Lab Reagent/Equipment Suppliers

Table 2.8 (page 59) and lists the general reagents, materials, and equipment used in this work, and from which which companies they were supplied.

Company	**Products**
NEB, New England Biosciences	Enzymes, Monarch Nucleic Acid Purification kits, Miniprep kits
Promega	Enzymes
GeneJet	PCR, Gel Extraction, Miniprep purification kits
IDT, Integrated DNA Technologies	Oligonucleotide and GBlock dsDNA synthesis
Genewiz	dsDNA synthesis, cloned into commercial plasmids
Twist Biosciences	dsDNA synthesis, cloned into commercial plasmids
Sigma Aldrich	Chemical reagents
Bio-Rad	Thermocyclers
Thermo Fisher Scientific	ClipTips, falcon tubes
Peqlab	Gel electrophoresis kits
VWR	Chemical reagents
Merck Millipore	Amicon centrifucal filters
Expedeon	IstantBlue Coomassie Protein Stain
BMG	FLUORstar, Optima, Nanodrop
Alpha Laboratories	Agarose gel tanks
Corning Life Sciences	Vaccum filter flasks
Applichem PanReac	Chemical reagents
Invitrogen	DynaMagTM-Spin Magnet, Dynabeads MyOne Streptavidin C1 beads
Applied Biosystems, Life Technologies	SYBR Safe DNA Gel Stain
Beckman Coulter	Centrifuges
GE Healthcare	Typhoon TRIO Variable Mode Imager, Typhoon FLA 9500 scanner
Eppendorf	Eppendorf 5810R Centrifuge, tubes
Li-COR	LI-COR Odyssey Scanner
Roche	Enzymes
Qiagen	Miniprep kits, gel extraction kits, genome extraction kits
Brady	LabelMaker
Bioline	Enzymes

Table 2.8: List of companies for general reagents, materials, and equipment

2.2 Molecular biology techniques

2.2.1 Resuspending single/double-stranded oligonucleotides

Single-stranded oligonucleotides/primers and double-stranded gBlock®/DNA fragments that are ordered in a lyophilised state are resuspended in 10 mM Tris-HCl pH 8.0 before use, to a stock concentration of 100 µM and at least 10 ng/µL respectively (per the recommended guidelines from the suppliers[3]). Resuspended solutions are stored at -20 ºC for $\leq$1 year, or at -80 ºC for longer-term storage.

2.2.2 Single-stranded oligonucleotide annealing

To generate dsDNA fragments, complementary single-stranded oligonucleotides were annealed together either through a process of freeze-thaw and/or heat-cool (sequentially, if using both methods). Equimolar ratios of the oligo pairs were combined to a final concentration of 10 µM in either MilliQ water (H_2O) or 10 mM Tris-HCl pH 8.0 buffer.

With the freeze-thaw method, the oligo mixture is placed at -20 ºC, where the tube directly contacts ice, and left for 10 min to 1 h. Once frozen, the sample can be left at room temperature until the solution thaws. For the heat-cool annealing method, the oligo solution is heated to 95 ºC for 1-5 min, then cooled to room temperature (20 ºC) at a rate of 0.1 ºC per second.

2.2.3 Polymerase chain reaction (PCR)

PCR was used to amplify DNA fragments. For general amplification of DNA fragments (typically $\leq$3 kb), the Q5 High-Fidelity Hot Start DNA polymerase (NEB #M0493) (Q5 HiFi HS DNAP) was used. PCRBio Ultra Polymerase (PCR Biosystems #PB10.31) and KOD Xtreme™ Hot Start DNA polymerase (Novogen #71975) (KOD DNAP) were used for products that were not robustly amplified by Q5 HiFi HS or for amplification of DNA fragments >3 kb. For colony PCR (veri-

[3]IDT guidelines for resuspending oligos (`https://www.idtdna.com/pages/education/decoded/article/tips-for-resuspending-and-diluting-your-oligonucleotides`), gBlock® fragments (`https://www.idtdna.com/pages/education/decoded/article/tips-for-working-with-gblocks-gene-fragments`), and from Twist Bioscience for gene fragments and clonal genes (`https://www.twistbioscience.com/sites/default/files/resources/2021-08/Guidelines_DNA_Resuspension_30AUG21_Rev3.0.pdf`.

fying transformants prior to Sanger sequencing; typically for 0.1-kb DNA fragments) or when there is limited initial template, MyTaq (Bioline ##BIO-21113) (and OneTaq, NEB #M0481) was used. PCRs were run using the C1000 Touch (BioRad, USA) and Peqstar (Peqlab, Germany) thermocyclers. Primers for each reaction are listed in section A.2, and annealing temperatures (T_a) were selected according to DNAP used and melting temperature (T_m) defined when primers were designed using SnapGene software v3.0.3 (from GSL Biotech; available at `https://www.snapgene.com/`). Annealing temperatures were also compared to the values provided by the Melting Temperature Calculator available on the NEB website (`https://tmcalculator.neb.com`).

General DNA amplifications using Q5 HiFi HS DNAP were typically carried out in 50 µL reactions with the following reagents (final concentration or amounts): 1X Q5 HiFi HS reaction buffer, 0.2 mM dNTPs, 0.5 µM each primer, $\leq$1 ng template DNA, 0.02 U/µL Q5 HiFi HS DNAP, and the rest made up with MilliQ H_2O. Gradient PCRs were carried out in 10 µL reactions using the same reagents as described. Reaction conditions were as follows: initial denaturation at 98 ºC for 2 minutes; 25-35 cycles of denaturation at 98 ºC for 10 seconds, primer annealing temperature (T_a) at 50-72 ºC ($T_a = T_m + 3$) for 30 seconds, primer extension (i.e. elongation time (E_t)) at 72 ºC for 30 seconds to 2 minutes per kilobase (30"/kb if amiplicon is $\leq$3 kb, 1'00"/kb for 3-7 kb, 2'00"/kb for >7 kb or overnight reactions); final extension at 72 ºC for 5 minutes; held at 18 ºC once reaction is completed. Various additives were occasionally added to Q5 HS HiFi DNAP reactions that provided unexpected products and/or yield, including: formamide, combinatorial PCR enhancer solution (CES, made as a 5X stock with 2.7 M betaine, 6.7 mM DTT, 6.7% (v/v) DMSO, 55 µg/mL BSA, and used at 1X final concentration; stored at -20 ºC; Ralser et al., 2006), DMSO, or betaine.

Reactions using KOD Xtreme™ contained: 1X Xtreme buffer, 0.4 mM dNTPs, 0.3 µM each primer, $\leq$10ng template DNA, 0.02 U/µL KOD Xtreme HS DNAP, and was made up to a total reaction volume of 50 µL. Reaction conditions were as follows: initial denaturation at 94 ºC for 2 min; 20-25 cycles of denaturation at

98 ºC for 10 seconds, primer annealing at 50-68 ºC (T_a = lowest primer T_m) for 30 seconds, primer extension at 68 ºC for 1 minute per kilobase; final extension at 68 ºC for 5 minutes; held at 18 ºC once reaction is completed.

MyTaq HS DNAP reactions contained: 1X reaction buffer (includes dNTPs), 0.4 µM each primer, $\leq$1 ng template DNA, 0.1 U/µL MyTaq DNAP, and made up to a total reaction volume of 10 µL. Reaction conditions were as follows: initial denaturation at 95 ºC for 2 min; 30-40 cycles of denaturation at 95 ºC for 15 seconds, primer annealing at 50-72 ºC ($T_a = T_m$ – (2 to 5)) for 15 seconds, primer extension at 72 ºC for 15 seconds; final extension at 72 ºC for 5 min; held at 18 ºC once reaction is completed.

Where DNA bound to streptavidin-coated paramagnetic beads required amplification, 1 µL of resuspended bead slurry is used instead of purified template DNA.

Once amplified and verified via agarose gel, PCR products were digested with `ExoI` (3'→5' activity) and `DpnI` to remove primers and methylated template DNA respectively.

2.2.4 Agarose-TBE/LiOAc for DNA

DNA fragments (e.g. PCR and restriction enzyme digest products) were analysed in (non-denaturing) agarose gels using a horizontal electrophoresis system (Multi-SUB4, AlphaLabs; Cleaver Scientific). Agarose gels were prepared by dissolving (ultra-pure) agarose powder (0.5-3.0% (w/v), according to required fragment size separation) in either (i) 1X TBE buffer or (ii) 10 mM lithium acetate (LiOAc) solution. SYBR™Safe DNA gel stain (Invitrogen #S33102) was added to the gels at a final concentration of 4 µL/100 mL.

Prior to loading, 1/6 volume of 6X DNA loading buffer (NEB #B7024/5 purple/blue dyes; or a pre-made mix of 40% (v/v) glycerol, 0.025% (w/v) Bromophenol blue) was added to DNA samples. Depending on expected fragment sizes, the molecular weight (MW) ladders run alongside DNA samples as size markers include: 2-log/1 kb Plus MW (0.1-10 kb; NEB #N3200), 1 kb MW (0.5-10 kb; NEB #N3232), 100 bp MW (100-1,517 bp; NEB #N3231), low MW (25-766 bp; NEB #N3233).

Agarose-TBE gels were run in 1X TBE buffer at 100-150 V for 0.5-1 hour. Agarose-LiOAc gels were run in 10 mM LiOAc, limited to 260 V for 0.25-1 hour. On completion of electrophoresis, if staining appeared insufficient, gels were stained again in their respective buffer solutions with either SYBR™Safe (for dsDNA) or SYBR™Gold (Invitrogen #S11494; for all nucleic acids) for >10 min, covered and left on a gentle shaking platform. The DNA was visualised using either blue light (DR-89X transilluminator, Clare Chemical Research) or UV transillumination.

2.2.5 Gel extraction / DNA purification from agarose-TBE/LiOAc

Gel electrophoresis (on a low percentage agarose gel; typically 0.5% (w/v)) and subsequent gel extraction was used to purify DNA samples from mixed populations, be it from PCRs that produced significant by-products alongside the expected fragment or from restriction endonuclease digests. If detected, the required DNA fragment is excised from the gel under blue light illumination, minimising the amount of agarose carried over. The sample is then purified using the GeneJet Gel Extraction Purification Kit (Thermo Scientific) or Monarch Nucleic Acid Purification Kit according to manufacturer's instructions.

2.2.6 Non-denaturing polyacrylamide gel electrophoresis

Non-denaturing, or native polyacrylamide gels are used to visualise nucleic acids and/or proteins without causing the molecular structures to unfold. Native polyacrylamide gels consist of 1X TBE, 10-20% polyacrylamide (40% solution of 19:1 (w/v) acrylamide:bis-acrylamide), 0.1% (v/v) APS, 0.1% TEMED, made up to 15 mL or 60 mL depending on whether visualising on a small (10×10 cm^2) or large (20×20 cm^2) plate respectively.

2.2.7 Urea PAGE (for nucleic acids)

Urea-PAGE was used for visualising 10-200 nt nucleic acid fragments where denaturation of structure is desired. These gels were typically made to 8-20% polyacrylamide (depending on the expected size of nucleic acid) with 8 M urea in 1X TBE, 0.1% (v/v) APS, 0.1% TEMED, and casted homogenously onto vertical gel

systems (Peqlab).

An equal volume of urea-PAGE loading solution, consisting of 98% (v/v) formamide, 10 nM EDTA, 0.02% (w/v) Orange G dye was added to the samples, before being incubated at 95 ºC for 5 min, and loaded onto the gel. Gels were run in 1X TBE buffer, limited to 30 mA for 1.5-2 h.

If the nucleic acids were not fluorescently labelled, the gels were stained with SYBR^{TM}Safe (for dsDNA) or SYBR^{TM}Gold (for all nucleic acids), before being imaged on a Typhoon FLA 9500 scanner (GE Life Sciences).

2.2.8 SDS-PAGE (for protein electrophoresis)

Protein products from heterologous expression of synthetic constructs/DNA in *E. coli* were analysed in denaturing SDS-PAGE. Typically, the ratio of acrylamide:bis-acrylamide used in these gels were at 37.5:1 (w/v) rather than 19:1 (w/v), as the former permits gel formation with larger pores. The gels consisted of two layers: a bottom resolving gel made of 8-15% acrylamide:bis-acrylamide, 375 mM Tris-HCl pH 8.8, 0.1% (v/v) SDS, 0.1% (v/v) APS, 0.1% (v/v) TEMED; a top stacking gel of 4% acrylamide:bis-acrylamide, 125 mM Tris-HCl pH 6.8, 0.1% (v/v) SDS, 0.1% (v/v) APS, 0.1% (v/v) TEMED.

The resolving gel was poured first, and then overlaid with 2-butanol to ensure a clean horizontal interface with the stacking gel upon polymerisation. Once the resolving layer has set, the 2-butanol was washed away with dH_2O, before the stacking gel was poured on top and combs placed to cast the wells for the samples. Samples are mixed with equal volume of 2X Lamaeli SDS-PAGE loading buffer (100 mM Tris-HCl pH 6.8, 4% (w/v) SDS, 0.02% (w/v) Bromophenol Blue, 20% (v/v) glycerol, 20 mM TCEP), before being incubated at 95 ºC for 5 min, and loaded onto the gel. Gels were run with SDS-PAGE running buffer (25 mM Tris-HCl, 192 mM glycine, 0.1% (w/v) SDS pH 8.3), limited to 100V for 1.5-2 h. InstantBlue Coomassie Protein Stain (Expedon) was used to stain the gels for >30 min, then destained with dH_2O for >15 min, before imaging on a FLUOstar OPTIMA (BMG).

2.2.9 Nucleic acid purification

All other (agarose gel verified) PCR amplicons, and/or enzymatically-treated DNA fragments were purified using the GeneJet PCR Purification Kit (Thermo Scientific) or Monarch Nucleic Acid Purification Kit (NEB) according to manufacturer's instructions.

DNA fragments were sometimes additionally purified and/or concentrated through phenol/chloroform extraction and ethanol(/isopropanol) precipitation. A phenol:chloroform:isoamyl alcohol (25:24:1) solution was added to the DNA solution in a 1:1 ratio, mixed by vortexing briefly, before being centrifuged at 15,000 g RCF for 5-10 min at room temperature. The aqueous phase DNA sample (typically top layer; bottom layer contains organic solvents) was then transferred to a fresh tube and precipitation was carried out with a final concentration of $\geq$0.4 M ammonium acetate* (0.1 volume of sample), between 60-80% (v/v) ethanol or isopropanol (2-3 volume of sample), and 10-50 µg glycogen azure (1 µL of 10/20 mg/mL). Ammonium acetate wass typically used as it sublimates when the DNA pellet dries, however 0.3 M sodium acetate pH 5.2 is used if T4 PNK is to be used subsequently. Glycogen azure co-precipitates with DNA so was used to facilitate visual detection of the pellet. The precipitation solution is mixed by vortexing, then incubated either at -20 ºC for $\geq$1 h or at -80 ºC for 5 min, and subsequently centrifuged at 15,000 g RCF for 30 min at room temperature. The supernatant was discarded and the pellet washed at least once with 250 µL 70% (v/v) ice-cold isopropanol, and centrifuged at 15,000 g RCF for another 5 min at room temperature. The supernatant was discarded again, and the pellet was either air-dried at room temperature or $\leq$50 ºC. Once dried, the pellet was resuspended in 5-10 µL 10 mM Tris-HCl pH 8.0.

2.2.10 DNA quantification by spectrophotometry

The concentration of DNA solutions (e.g. from purified plasmids and PCR products) and their purity were quantified by absorbance measurements. Absorbance at 260 nm was used to estimate, whilst protein contamination was observed by evaluating the absorbance ratio 260/280, and solvent or guanidine contamination can be assessed with the 260/230 ratio. Majority of measurements were acquired using the

BMG Labtech SpectroStar Nano instrument with the LVis Plate accessory, which allowed for absorbance measurements from 2 µL droplets.

2.2.11 Phosphorylation of nucleic acids (oligos/PCR products)

For primers, 0.5-1 nmol of 100 µM oligo was added to 1X T4 DNA ligase buffer (containing 1 mM ATP), 10 U T4 polynucleotide kinase (T4 PNK), and made up to 50-100 µL for a final oligo concentration of 10 µM. For 5'-phosphorylation of PCR products (or pairs of annealed ssDNA, gBlock®, etc.), 500 ng of DNA (or $\leq$300 pmol of 5' termini) was added to 1X T4 DNA ligase buffer (containing 1 mM ATP), 10 U T4 PNK, 10 U DpnI (DpnI only for PCR products), and made up to 20-50 µL. These were incubated at 37 ºC for 40 min to 1 h, heat inactivated at 65 ºC for 20 min, and held at 18 ºC. Typically, phosphorylated primers were not subsequently purified, but phosphorylated PCR products tended to be purified with the GeneJet or Monarch PCR Purification Kits before undergoing ligation. Phosphorylation was unnecessary if the DNA fragment had undergone enzymatic restriction digestion.

2.2.12 Restriction endonuclease digestion of DNA

Restriction endonuclease enzymes were used for diagnostic/analytical digests to confirm that a given plasmid or PCR product had been produced as expected. The products of restriction were run on agarose gels to verify their identity.

For restriction endonuclease cloning, vectors (usually plasmids) and inserts (usually PCR products) were digested with restriction enzymes to produce compatible/complementary overhangs for ligation. Typically, vectors and inserts were digested in separate 10-20 µL reactions, containing: 1X Cutsmart (or NEBuffer 1.1, 2.1, 3.1 according to enzyme required) buffer, 200 ng to 1 µg DNA, and 10-20 U of enzyme(s). If the DNA fragment to be digested was the product of PCR amplification from methylated plasmid templates, these reactions also included 10 U DpnI to remove the any residual template. Most digestion reactions were incubated at 37 ºC (exceptions include Nt.BspQI at 50 ºC, and SmaI at 25 ºC) for 30 min to 2 h, before heat inactivation at 65 ºC or 80 ºC for 20 min. Products from restriction digests could be purified as described in previously, but were always purified if the enzyme

used could not be heat inactivated (e.g. Nb.BssSI, BamHI-HF, PstI-HF).

2.2.13 Dephosphorylation of nucleic acids (e.g. oligonucleotides/PCR products)

Alkaline phosphatases (e.g. Antarctic Phophatase, AnP; Shrimp Alkaline Phosphatase, rSAP; Calf Intestinal Phosphatase, CIP) were used to remove phosphates from 5'- (protruding, recessed, and blunt) and 3'-ends of ss-/ds-DNA/RNA. A 20 µL dephosphorylation reaction typically consisted of 500 ng of DNA (or $\leq$1 pmol of 5'-/3'-termini), with 1X AnP reaction buffer (or any NEB restriction enzyme buffer when supplemented with AnP reaction buffer to provide Zn^{2+} required for activity), and 5 U AnP. This was incubated at 37 ºC for 30 min, heat inactivated at 80 ºC for 2 min, and held at 18 ºC.

2.2.14 DNA ligation for restriction enzyme (sub-)cloning

Ligation reactions were typically carried out in a total volume of 10-20 µL, containing: 1X T4 DNA ligase buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM ATP, 10 mM DTT, pH 7.5 at 25 ºC), 200-400 U T4 DNA ligase (T4 DNA ligase), and a total of 100-200 ng DNA of both vector and insert(s). In classic restriction endonuclease cloning with only two parts, DNA was added in a vector:insert molar ratio of 1:3. The ligation mix was subsequently incubated at 20 ºC for 2 h, 16 ºC for 16 h, heat inactivated at 65 ºC for 10 min, and held at 18 ºC. For simple ligations, the reaction was stopped after 2 h to use directly for transformation; more complex or blunt-ligations were left overnight to in an attempt to increase ligation efficiency. A higher concentration of T4 DNA ligase ($\leq$800 U) was typically used for blunt-ligations or where fragments contained single/2 nt overhangs. Fast ($\leq$10 minute) ligations were achieved using a higher T4 DNA ligase concentration and/or with the addition of 3% PEG8000 (final concentration) crowding agent.

Once ligated, the mixture was stored at 4 ºC until it was ready to be transformed into *E. coli*.

2.2.15 Type IIS and Golden Gate Assembly

The Golden Gate Assembly method was used to combine complex DNA fragments and/or for simultaneous multi-insert assemblies. After verification, purification, and quantification, PCR amplicons were digested with specific Type IIS (TIIS) restriction endonucleases (e.g. BsaI, SapI, BbvI, BspMI) and either ligated sequentially, as described previously in Section 2.2.14 , or in a single-one pot reaction. With BsaI as the TIIS enzyme (recognition and cut site GGTCTC(1/5)), a 20 µL reaction mix typically consists of: 1X T4 DNA ligase buffer (or 1X CutSmart buffer supplemented with 1 mM ATP), 20 U BsaI-HF (with 10 U DpnI if an insert was amplified from a methylated plasmid), 200-400 U T4 DNA ligase, and a total of 100-200 ng of all the DNA inserts together at equimolar amounts (assuming each are of approximate lengths). The mix is incubated at 37 ºC for 1-2 h, and heat inactivated at 65 ºC (or 80 ºC if incubated also with DpnI) for 20 min. Other TIIS enzymes that were used, and the respective variations in the reaction formulation, were further specified within the results.

2.2.16 Gibson Assembly

Gibson assembly was another isothermal method typically used when assembling multiple fragments in a single-pot reaction. The reaction mix contains a 5' exonuclease to generate long overhangs, a DNA polymerase to extend 3'-end of annealed fragment pairs, and a DNA ligase to seal the nicks in the assembled DNA product. Like with TIIS cloning, the fragments for the Gibson method need to be synthesised or amplified with primers that give the overlapping assembly sequences. Once all DNA parts were prepared, they were assembled together with the NEBuilder® HiFi DNA Assembly Master Mix (NEB #E2621) as per the protocol described by the manufacturers. In both cloning methods, once ligated, the mixture can be stored at 4 ºC until it is ready to be transformed into *E. coli*.

2.2.17 Site-directed mutagenesis (SDM): inverse PCR (iPCR)

SDM Site-directed mutagenesis (SDM) was typically used to introduce point mutations in plasmids via inverse PCR; an amplification of the entire plasmid sequence

with a pair of primers that anneal adjacently around the site where a mutation was required. The mutation was added at the 5'-end of one of the primers and replaces the wild-type nucleotide(s) as a non-annealing sequence. PCR was performed using NEB Q5 HiFi Hot Start DNA polymerase or KOD Xtreme™, with up to 30 cycles. Once purified, the products were phosphorylated and re-circularised by blunt-ligation.

2.2.18 Darwin Assembly (DA): one-step multiple SDM incorporation

Darwin Assembly was used as a fast and efficient multi-site mutagenesis method (Cozens and Pinheiro, 2018). In brief, it consisted of five main steps: generation of a single-stranded plasmid; annealing of SDM and boundary oligos; isothermal assembly of the full fragment spanning the mutagenic oligos; recovery and purification of assembly for PCR amplification; sub-cloning into a new vector.

2.2.19 Sanger/DNA sequencing of plasmids

To verify that plasmids were constructed as designed, aliquots of purified plasmid constructs were sent to GATC Biotech AG, Germany for Sanger sequencing, as per the provider's instructions. Sequencing results were aligned to the constructs designed using SnapGene software v3.0.3 (from GSL Biotech; available at `https://www.snapgene.com/`) or visualised on 4Peaks to verify correct assembly of the required constructs. If the sequencing results were not as expected, and compromised downstream experiments, another colony was isolated from the transformants, and submitted for sequencing again.

2.3 Microbiology techniques

2.3.1 *E. coli* culture: preparation and storage

E. coli liquid/inoculation cultures were carried out in a range of media as listed in Table 2.6. Cultures were typically grown in Luria-Bertani (or lysogeny) broth (LB) (1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl) medium. For pro-

tein over-expression experiments, strains are grown in richer media such as: 2X TY (1.6% (w/v) tryptone, 1% (w/v) yeast extract, 0.5% (w/v) NaCl); SOC (2% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.48% (w/v) $MgSO_4$, 0.3603% (w/v) dextrose, 0.05% (w/v) NaCl, 0.0186% (w/v) KCl); TB (1.18% (w/v) tryptone, 2.36% (w/v) yeast extract, 0.94% (w/v) K_2HPO_4, 0.22% (w/v) KH_2PO_4). In all these cases, for solid medium, 1.5% (w/v) agar is added.

Conversely, a simple M9 minimal medium is also used for expression of toxic proteins, consisting of: 1X M9 salts (4.25% (w/v) $Na_2HPO_4.2H_2O$, 1.5% (w/v) KH_2PO_4, 0.5% (w/v) NH_4Cl , 0.25% (w/v) NaCl), 2 mM $MgSO_4.7H_2O$, 0.1 mM $CaCl_2.2H_2O$, 0.4% (w/v) glucose or fructose, and optionally 1X (1 mg/mL) BME vitamins. The M9 media was also supplemented with either 1 mM L-leucine for use with leucine-auxotroph 10β strain, or 1 mM L-arginine hydrochloride for the DH5α-family strains. To make solid M9 media, 2X M9 salts and 3% (w/v) agar were first autoclaved separately. Once both have cooled to 50-60 ºC, aseptically mix together the 2X M9 salts and 3% (w/v) agar, and then add the other reagents specified above for making the liquid M9 media.

Antibiotics were used to ensure selection of expected transformations and inoculums, including: ampicillin (Amp), with stock solution made at 100 mg/mL in 50% (v/v) ethanol, and 100 µg/mL working concentration; kanamycin (Kan), with stock solution at 50 mg/mL in dH_2O, and 50 µg/mL working concentration; chloramphenicol (Cam/Chl), with stock solution at 34 mg/mL in 100% (v/v) ethanol, and 10 or 34 µg/mL working concentrations; spectinomycin (Spc), with stock solution at 20 mg/mL in dH_2O, and 20 µg/mL working concentration; streptomycin (Str), with stock solution at 100 mg/mL in dH_2O, and 50 or 100 µg/mL working concentrations. All antibiotic stocks are filter-sterilised and stored at -20 ºC.

Media and glassware used for cultures were sterilised by autoclaving at 121 ºC for 20 min. All antibiotic-free media were kept at room temperature; antibiotic-containing media were kept at 4 ºC.

E. coli were generally cultured in plastic-/glass-ware in a 1:4 ratio of media to air, and grown at 30 ºC or 37 ºC at 250 RPM overnight (~16 h) until stationary phase.

Permanent stocks of the strains were made from the overnight cultures, resuspended in 15% (v/v) autoclaved glycerol, in a final volume of 1.2 mL, and were stored at -80 ºC.

2.3.2 Measurement of bacterial growth by spectrophotometry

Cell density of *E. coli* liquid cultures is estimated by measuring light scattering in spectrophotometer cuvettes with 1 cm path length, using the SpectroStar Nano (BMG Labtech, UK) instrument. Absorbance measurements are taken at 600 nm, where cell concentration of Luria-Bertani grown *E. coli* with OD600 of 0.1 is 2×10^7 cells/ml (`https://bionumbers.hms.harvard.edu/bionumber.aspx?id=100985&ver=14&trm=od600+escherichia+coli&org=`). If OD600 > 1, the sample was diluted ten-fold in its respective medium before being measured again; the resulting value is multiplied by 10 to produce an estimate of the original cell density of the culture. All measurements were compared against its respective culture medium blank (i.e. medium with no bacterial growth).

2.3.3 Preparation of electro-competent cells for transformation

A liquid culture (10 mL LB in a 50 mL centrifuge tube) of the selected *E. coli* strain was started from its glycerol stock (stored at -80 ºC), typically with no antibiotics added, and incubated at 37 ºC, 250 RPM overnight (16 h). Upon saturation, the culture was used to seed (typically 1-2.5%, approximately equivalent to OD600 = 0.02-0.05) an over-day culture of 200-400 mL LB (with no antibiotics) in a 2 L conical/shake flask, incubated at 30 ºC or 37 ºC, 250 RPM. Once cell density reaches OD600 = 0.4, the culture was transferred to pre-chilled 500 mL centrifuge tubes and centrifuged at 4 ºC, 4000 RPM (= 3250g RCF), for 30 min. The supernatant was discard and the pellets are each resuspended/washed in 50 mL ice-cold, filter-sterilised 1 mM HEPES pH 7.0 buffer. These 50 mL centrifuge tubes were centrifuged again at 4 ºC, 4000 RPM (= 3250g RCF), for 10 min, supernatant again discarded. This time, the pellets from two 50 mL centrifuge tubes were combined together and resuspended in a total of 50 mL 1 mM HEPES, before centrifuging as per the conditions previously. This resuspending, centrifuging/pelleting, pooling

and resuspending process was repeated until the cell pellet was in a single tube of 50 mL 1 mM HEPES, at which point it was pelleted once more and all supernatant was removed as much as possible. This cell pellet (from an initial 200 mL culture) was finally resuspended in a total volume of 2 mL 1 mM HEPES mixed with 15% (v/v) filter-sterilised glycerol. The resulting mixture of electro-competent cells were divided into 50 µL aliquots, immediately flash-frozen in dry ice, and subsequently stored at -80 ºC.

2.3.4 Transformation of DNA into (electro-)competent cells

Plasmids were transformed into electro-competent *E. coli* using electroporation as this has greater transformation efficiency over the use of chemically-competent cells (BioRad, USA). Up to 100 ng ($\leq$5 µL) of ligated DNA (or 1-10 ng of purified plasmid) was mixed with 50 µL of electro-competent cells thawed on ice, and pipetted into pre-chilled electroporation cuvettes with a 0.2 cm gap between its electrodes. If necessary, phenol/chloroform extraction and ethanol/isopropanol precipitation was first performed on the DNA solution to minimise salt carry-over during transformation. Electroporation was implemented using a Gene Pulser II electroporator (BioRad, USA) at 2.5 kV, 200 Ω and 25 µF ; the DNA-cell solution was optimally pulsed at 4.80-5.20 ms. Directly after electroporation, 450 µL LB (or 2X TY) was added to the cuvette, and incubated at 37 ºC for 1 h (ideally between 30-90 min) for recovery. After this incubation period, cells were plated on LB (or 2X TY or M9) agar containing the required antibiotic for selection of transformants, and incubated again at 37 ºC overnight (~16 h).

2.3.5 DNA plasmid preparation and isolation

From transformation plates, individual colony-forming units (CFUs, colonies) were isolated and usually (i) transferred to a secondary storage/spotting LB agar plate, (ii) used to seed an liquid culture overnight for plasmid multiplication, and (iii) used as DNA template in colony PCRs as an initial means to verify transformation of required construct/plasmid. The remaining culture was centrifuged to pellet the cells, which is subsequently lysed, and plasmid DNA was purified using the

GeneJet/Monarch Miniprep Purification Kit. If the original culture was 3 mL, plasmid purification was performed as per manufacturer's recommendations, with the exception that the sample was eluted with 30 µL 10 mM Tris-HCl pH 8.8. If the starting culture was 10 mL, the buffer volumes were increased as follows: 350 µL Resuspension buffer, 350 µL Lysis buffer, 450 µL Neutralisation buffer; Wash buffer remained as recommended, but elution was with 30-50 µL 10 mM Tris-HCl pH 8.8. The key change in increasing amount of lysis buffer used is an attempt to increase the amount of plasmid DNA purified from the cell pellet. Once plasmids were purified, the DNA concentration and purity were checked by spectrophotometry as described previously and/or by visualisation on agarose gel (typically after analytical digestion). If necessary, plasmids were sent for Sanger sequencing.

Chapter 3

OCI: Heterologously-Expressed Reassignment

"Developing the Orphan Codon Invasion System to Implement Heterologous Expression of an Alternate Genetic Code In Vivo"

3.1 Orphan Codon Invasion

In this chapter, I explain the key concepts that form the basic theory behind my Orphan Codon Invasion (OCI) strategy for sense-to-sense codon reassignment. I first focus on my plan to exploit the wobble base-pair hypothesis to introduce non-cognate amino acids at what I term "orphan" codons (Section 3.1.1), before moving on to define "codon invasion" via the exemplary work of Döring and Marlière (1998) in engineering tRNA sequences for unidirectional *in vivo* codon reassignment (isoleucine-to-cysteine and methionine-to-cysteine; Section 3.1.2). Having established this, I then introduce how my OCI approach sought to modify tRNA anticodons to invade orphan codons as a means of implementing bi-directional sense-to-sense codon reassignment (Section 3.1.3).

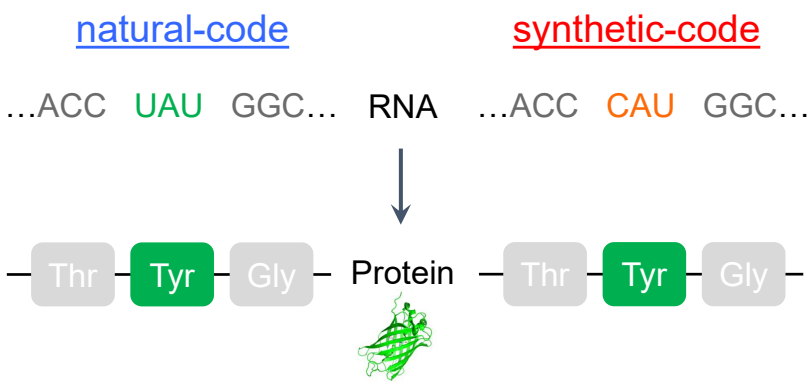
Analogous to Döring and Marlière (1998), I explore here how to engineer tRNA to create genetic codon ambiguity *in vivo* (Section 3.2.1). Such an engineering step is the crucial intermediate towards sense-to-sense codon reassignment: the ambiguity created by the engineered tRNA can be resolved by removal of the

original assignment – so that the engineered tRNA becomes the only assignment. Repeating this procedure with another set of engineered tRNA leads to an organism operating under a different genetic code. Codon reassignment of tRNAs also requires the development of suitable reporter assays (Section 3.2.2, Section 3.2.3, Section 3.2.4) to monitor and quantify the invasion process (Section 3.2.5).

Briefly, in order to design a viable codon invasion strategy, I first lay out the rationale for selecting which amino acids (and their associated tRNAs, cognate aaRSs, and complementary orphan codons) could be used for non-synonymous sense-to-sense codon reassignment. I place particular emphasis on identifying a tRNA pair whose cognate aaRSs do not recognise the anticodons, thereby reducing the chances of these endogenous aaRSs rejecting my alternative (orphan codon pairing) tRNAs when their anticodons are engineered. Moreover, of the candidate tRNA pairs selected for anticodon modifications, importance is placed on pairs that, when swapped to pair with alternative orphan codons, would lead to significant changes in the resulting proteins' physico-chemical properties, making it compatible with fast and robust assays for determining codon reassignment.

With regards to reporter assays for assessing sense-to-sense codon reassignment *in vivo*: I required a heterologously-expressed reporter construct that could rapidly output a strong, visual phenotype, with a robust and long-lasting signal (Section 3.1.4 and Section 3.1.5). The selected reporter(s) must contain the identified candidate amino acids for sense-to-sense codon reassignment as critical residues, where "mis"-translation (i.e. correct codon reassignment) would result in a discernible phenotypic response from the natural expression. To create this assay, the gene sequence of the reporter construct would first need to be recoded to conform to the proposed orthogonal genetic code assignment by changing the codon of a critical amino acid to that of a non-synonymous sense codon of an alternate amino acid (Figure 3.1). In this way, the resulting translated protein is rendered inactive under the natural genetic code system, yet becomes fully functional upon implementation of the OCI reassignment strategy.

The remaining sections of this chapter document the design and construction of



	UAU → Tyr [WT] CAU → His [WT]	CAU → Tyr [Recoded]
Naturally-coded organism	**Functional**	Non-functional Toxic (likely)
Synthetically-coded organism	Non-functional Toxic (likely)	**Functional**

Figure 3.1: The basic principle of the sense-to-sense codon reassignment reporter assay.

Under the natural genetic code, codon $\text{Tyr}_{(\text{UAU})}$ encodes tyrosine, whilst under the synthetic code the $\text{His}_{(\text{CAU})}$ codon has been recoded to accept tyrosine. In a naturally-coded organism, this change in the codon assignment will likely lead to the production of a non-functional protein that is toxic to the natural host. A synthetically-coded organism that recognises this alternate genetic code, however, will be able to correctly reassign the $\text{His}_{(\text{CAU})}$ codon and incorporate tyrosine, thereby producing a functional protein.

these novel codon-invading anticodon-modified tRNAs, and of the various recoded reporters necessary to assess reassignment (Section 3.1.6). The results of expressing these constructs together in *E. coli* are discussed, including the degree to which reassignment occurred, and any observable fitness costs to the host (Section 3.2). Finally, I review whether the OCI strategy, and the chosen reassignment pair, were sufficient in demonstrating sense-to-sense codon reassignment (Section 3.3), and what further steps may be necessary to make alternative genetic codes, and genetically recoded organisms, a reality.

The following flow diagram (Figure 3.2) should provide additional guidance to explain which sections in Chapter 3 were successful in taking the research forward, and which were abandoned owing to experimental issues.

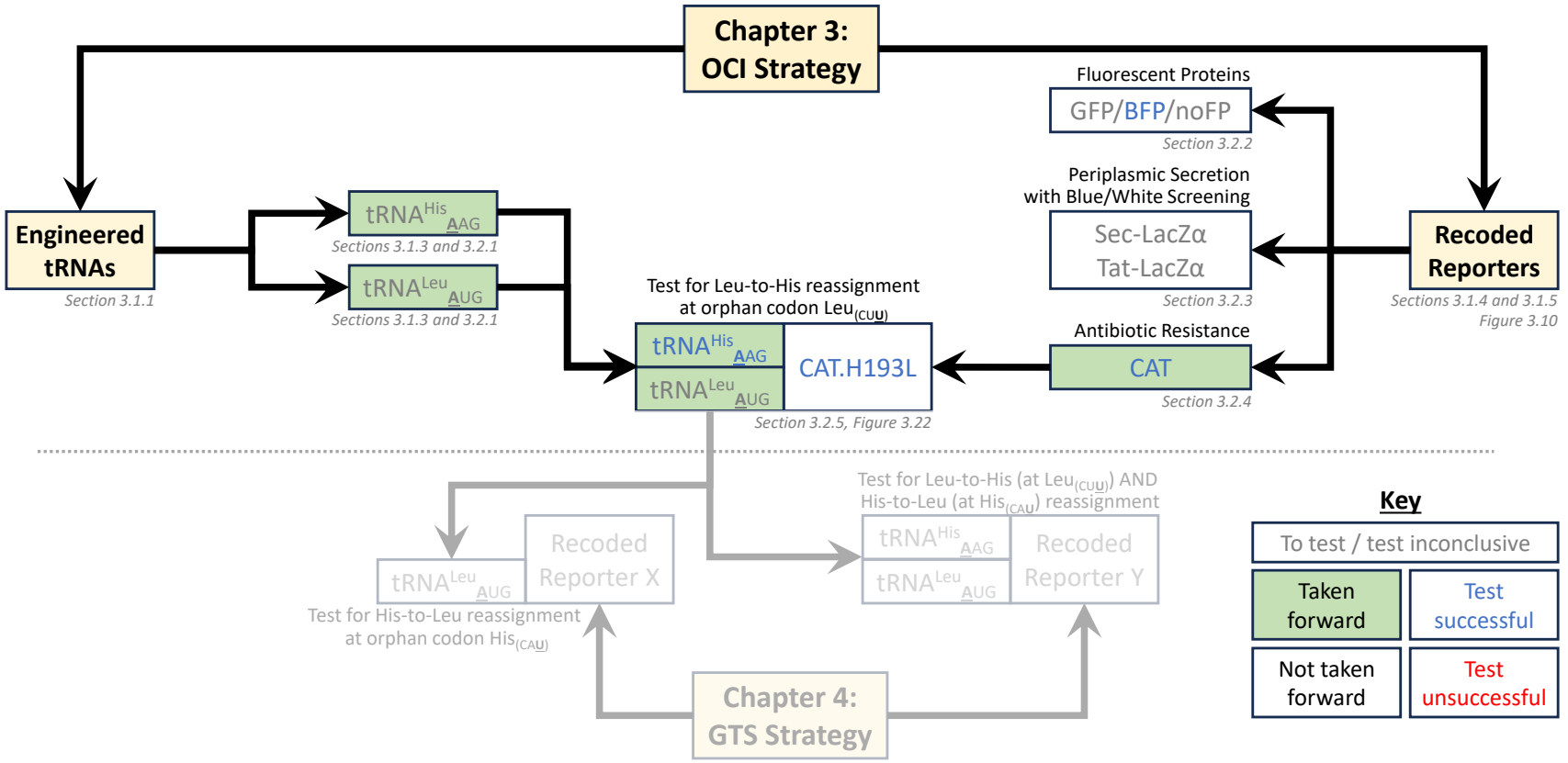


Figure 3.2: A summary flow diagram of the strategies that were successful whilst developing the OCI in Chapter 3

Elements of the OCI strategy that were taken forward are shaded in green, with experiments that were successful in demonstrating sense-to-sense codon reassignment in blue font. The associated sections relating to the work are shown in grey font below each box. Some elements were taken forward as they were essential to demonstrating reassignment but there had been no test yet available to assess its viability and/or the results of the initial tests were inconclusive. The faded section at the bottom of the figure relates to how the OCI strategy builds into the later GTS strategy of Chapter 4; its associated flow diagram is shown in Figure 4.2.

3.1.1 Genetic code redundancy, the wobble hypothesis, and orphan codons

The genetic code is filled with redundancies. Just as multiple codons can correspond to the same amino acid, multiple codons can be served by a single tRNA isoacceptor. This is why it is possible for *E. coli* to require only 46 tRNA isoacceptors, yet be able to adequately complement all 61(+1) sense codons that encode the 20(+1)[1] natural amino acids (Figure 3.3; Ho et al., 2016). Moreover, the number of tRNA isoacceptors and the sets of codons each pairs with can differ from one organism to the next. A good example of genetic code redundancy can be seen with serine (Ser/S) where, for *E. coli*, the *Eco*tRNA$^{\mathrm{Ser}}_{\underline{\mathbf{U}}\mathrm{GA}}$ can pair with codons Ser$_{(\mathrm{UC}\underline{\mathbf{U}})}$, Ser$_{(\mathrm{UC}\mathbf{A})}$, and Ser$_{(\mathrm{UC}\underline{\mathbf{G}})}$ (Figure 3.3, tRNA highlighted in pale green). The tRNAs are able to do so due to "wobble" base-pairing within the triplet codon/anticodon sequences (Crick, 1996; Roth, 2012); wobble-pairing is herein represented as underlined bases. In fact, 25 of the 46 *E. coli* tRNA isoacceptors (encompassing 17 of the 20 canonical amino acids; Figure 3.3, triplets highlighted in green) have at least one wobble partner[2], further demonstrating the extent of code redundancy. For certain tRNA/mRNA anticodon/codon pairings, this means that the first base (5'-end, tRNA nucleotide position 34) of the anticodon triplet has more steric freedom with its complementary pair, the third base (3'-end of the codon), to allow for these non-standard base-pairings (Figure 3.3, codons highlighted in grey or not at all/white). These wobble mismatches typically manifest as G·U, but the genetic code also includes a handful of other wobble-pairings, such as U·U (in Val/V, Ser/S, Pro/P, Thr/T, Ala/A), and C·A and A·A (in Arg/A). Furthermore, the non-standard G·U wobble has almost comparable thermodynamic stability (and nearly isomorphic structure) to the standard Watson-Crick pairings, which may explain why it is

[1] The "(+1)" refers to selenocysteine (Sec/U), a rare but naturally-occurring proteinogenic amino acid found in *E. coli* (and some other organisms within archaea, bacteria, and eukarya) that is an alternative translational recording of the opal stop codon UGA.

[2] The remaining tRNA isoacceptors (covering 11(+1) of the amino acids; Figure 3.3, highlighted in purple, blue, yellow) pair with their respective codons with full Watson-Crick complementarity (Curran, 1998). Note, nine of these isoacceptors are the only partners to their codons, with methionine (Met/M; one instance), tryptophan (Trp/W; one instance), glutamine (Gln/Q; two instances), and seleoncysteine (Sec/U; one instance) having no isoacceptor redundancy at all (purple highlights).

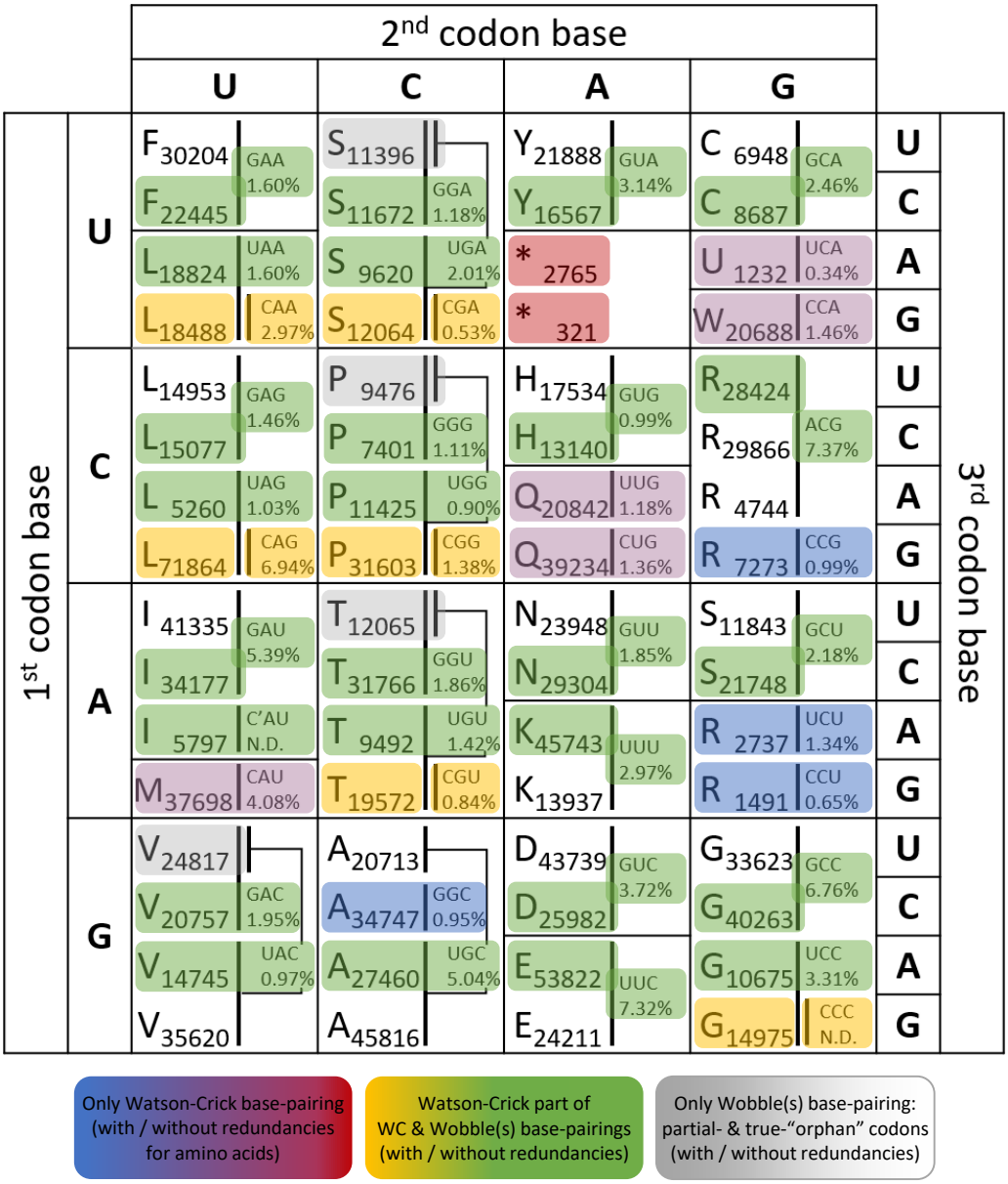


Figure 3.3: A tRNA-centric view of how the genetic code is implemented in *E. coli*.

Each of the 46 tRNA isoacceptors present in *E. coli* is involved in implementing the organism's genetic code. However, this assignment is not a one-to-one assignment and it operates with redundancies in a number of cases. This figure (adapted from Ho et al., 2016) shows the approximate number of encodings per cell for each specific codon (based on National Center for Biotechnology Information (NCBI) NC_000913.2, as referenced by Lajoie et al., 2013) and their respective tRNA isoacceptor abundances (Dong et al., 1996).
Colours denote whether the tRNA/codon has: only Watson-Crick (WC) one-to-one isoacceptor-to-codon associations, with (blue) and without (purple) redundancy for the amino acid; both Watson-Crick one-to-one (yellow) and one-to-many Wobble-pairing (green) associations, both with redundancies; one-to-many associations featuring only Wobble-pairing isoacceptors as partial- (grey; with redundancies) or true- "orphan" (white; without redundancies) codons.

the most common substitute for G·C and A·U (Varani and McClain, 2000). This is evident from a series of RNA denaturation experiments (on a 25 nt stem-loop structure based on *Eco*tRNA$^{Ala}_{GGC}$) demonstrated by Strazewski et al. (1999), where they showed that the mid-point melting temperature (T_m) of the Watson-Crick base-pairs were 4.5-12.4°C higher than that of the G·U wobble: G·C (T_m = 85.4°C), C·G (T_m = 84.0°C), U·A (T_m = 77.5°C), A·U (T_m = 76.3°C), versus wobble-pairing G·U (T_m = 73.0°C), and U·G (T_m = 70.5°C); standard error of all RNAs was 1.0°C. The melting temperatures of other aforementioned wobble-pairs were even lower, including: C·A (T_m = 63.0°C), U·U (T_m = 60.3°C), and A·A (T_m not determined). Hence it was determined that it could be possible to exploit these energetically weaker wobble-pairing bonds for the OCI strategy (Section 3.1.2).

Not only does wobble-pairing allow a single tRNA isoacceptor to complement multiple codons, multiple tRNA isoacceptors can serve the same codon – once more exemplifying the redundancy of the genetic code. In these many-to-one tRNA-to-codon associations, one isoacceptor pairs with full Watson-Crick complementarity to the codon, while the other isoacceptors are wobble-pair partners. But, there are also some codons where all their tRNA partners are wobble-pairs, such as with Pro$_{(CC\underline{U})}$ and its complements *Eco*tRNA$^{Pro}_{\underline{G}GG}$ (wobble G·U) and *Eco*tRNA$^{Pro}_{\underline{U}GG}$ (wobble U·U) (Figure 3.3, such codons are highlighted in grey). As they have no truly dedicated (Watson-Crick pairing) tRNA partner, I refer to these isolated wobble-pairing codons as "orphan" codons. These orphans can be further defined as partial- or true- orphan codons by the number of wobble-pairing isoacceptors that serve them. Specifically, codons that are only complemented by a single wobble-pairing isoacceptor are what I term "true orphans": such is the case with His$_{(CA\underline{U})}$, where its only partner is *Eco*tRNA$^{His}_{\underline{G}UG}$ (wobble G·U), for example. Even then, *Eco*tRNA$^{His}_{\underline{G}UG}$ only moonlights as a decoder of the His$_{(CA\underline{U})}$ codon; it has a more stable, fully Watson-Crick complementing codon partner in His$_{(CA\underline{C})}$. The consequence of having multiple codon partners, where one is a more favourable match, means that the rate at which the orphan codon could be decoded would theoretically be reduced. I can take advantage of this weaker wobble-pairing property

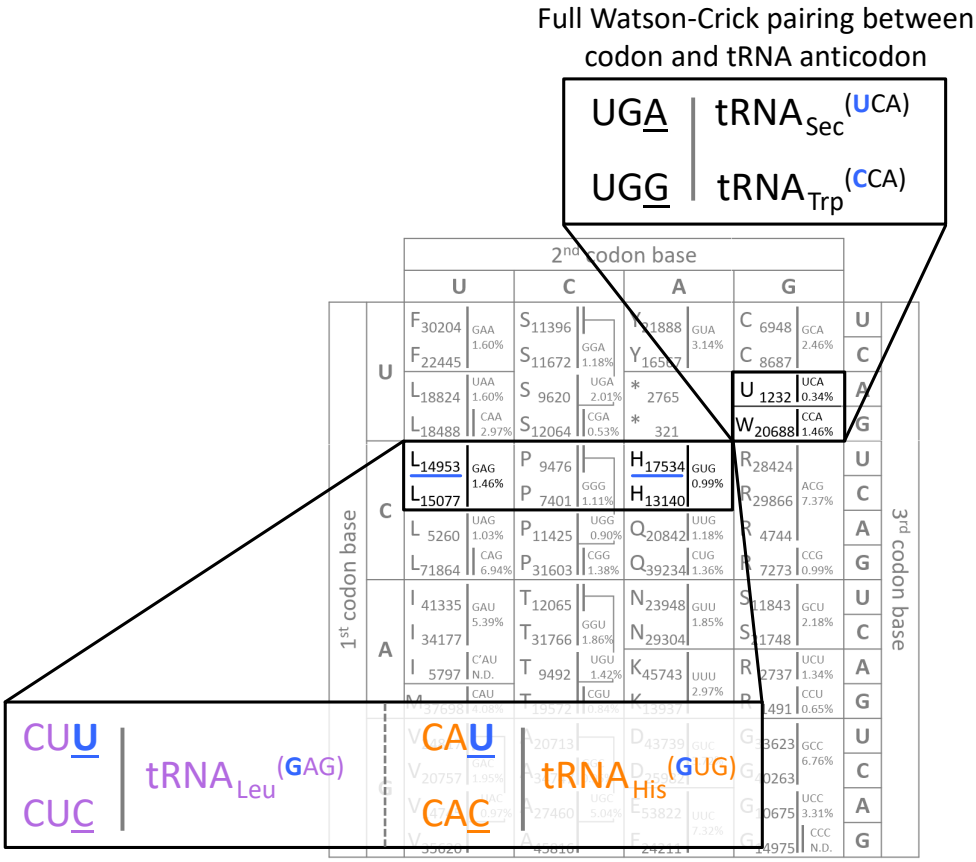


Figure 3.4: Examples of Watson-Crick and wobble codon-anticodon base-pairings from the *E. coli* genome

Demonstrating that the majority of amino acid codon assignments in *E. coli* have orphan codons, i.e. any codon with no tRNA isoacceptor that complements it with full Watson-Crick base-pairing. Only $\text{tRNA}^{\text{Leu}}_{\underline{\text{G}}\text{AG}}$ base-pairs with leucine codons CUU and CUC, however, this tRNA wobble base-pairs with orphan $\text{Leu}_{(\text{CU}\underline{\text{U}})}$ (blue nucleotides) and fully complements codon $\text{Leu}_{(\text{CU}\mathbf{C})}$. Similarly, $\text{tRNA}^{\text{His}}_{\underline{\text{G}}\text{UG}}$ wobble-pairs with orphan $\text{His}_{(\text{CA}\underline{\text{U}})}$, whereas it has full Watson-Crick pairing with codon $\text{His}_{(\text{CAC})}$. In contrast, only three assignments $\text{Met}_{(\text{AUG})}$, $\text{Trp}_{(\text{UUG})}$, and stop codon opal/$\text{Sec}_{(\text{UGA})}$ do not have orphan codons as there is only one codon for these amino acids, and each possess their own dedicated tRNA isoacceptors. Figure adapted from Ho et al. (2016).

by introducing engineered tRNA isoacceptors that are a better complement to orphan codons, and thereby allow the implementation of alternate tRNA-codon assignments for sense-to-sense codon reassignment.

All in all, there are 17(+1)[3] such true orphan codons, accounting for 15 of the 20 natural amino acids (Figure 3.3, codons not highlighted/white) within the *E. coli* genome that could potentially be induced to accept new tRNA isoacceptors. Despite having a range of opportunities for reassignment, orphan codons (i.e. codons that can only be accessed by tRNAs through wobble-pairing) do not have a significant impact in translation, because under cellular conditions, and in the absence of competing tRNAs, the wobble-pairing still operates efficiently. In addition, a given organism's codon frequency has also likely evolved towards a balance point where any loss of incorporation efficiency can be matched by the rarity of orphan codons. In *E. coli*, the rarer $\text{Arg}_{(\text{CG}\underline{\textbf{A}})}$ (codon usage per cell: 4,744) may well be an example of such evolutionary balance given its $\text{tRNA}^{\text{Arg}}_{\underline{\textbf{A}}\text{CG}}$ partner has more abundantly-occurring complements in $\text{Arg}_{(\text{CG}\underline{\textbf{C}})}$ (codon usage per cell: 29,866) and $\text{Arg}_{(\text{CG}\textbf{U})}$ (codon usage per cell: 28,424) (Figure 3.3; Ho et al., 2016).

It is the rarity of these orphan codons, and their predisposition for being supplanted by an alternate tRNA isoacceptor, that will prove to be critical for my research in creating a genetically recoded organism (GRO): one with a distinctly alternate genetic code assignment to the natural. Specifically, orphan codons present opportunities to introduce alternate tRNAs with anticodon modifications that fully complement the codon with three Watson-Crick base-pairs and, as such, have the possibility of contending with the native wobble-pairing tRNAs for the targeted codons. Furthermore, despite having higher thermodynamic stability than other non-standard wobble-pairs, the G·U wobble bond is still not as stable as a standard Watson-Crick pair (Varani and McClain, 2000), which suggests that an anticodon-modified tRNA could theoretically out-compete a native wobble-pairing tRNA and "invade" an orphan codon.

Whilst not specifically targeting orphan codons, Döring and Marlière (1998)

[3]This "(+1)" refers to isoleucine codon $\text{Ile}_{(\text{AU}\underline{\textbf{A}})}$, which has a single, but unusual, $\text{tRNA}^{\text{Ile}}_{\underline{\textbf{C'}}\text{AU}}$ isoacceptor partner.

Amino Acid	**Codon**	**Anticodon (natural; for $tRNA^{Ile}_{\underline{N}AU}$ & $tRNA^{Met}_{\underline{N}AU}$)**	**Anticodon (synth.; for $tRNA^{Cys}_{\underline{N}AU}$)**
Isoleucine, Ile/I	AU**U**	**G**AU (wobble)	**A**AU (WC)
	AU**C**	(WC)	**G**AU (WC)
	AU**A**	**C'**AU (wobble)	**U**AU (WC)
Methionine, Met/M	AU**G**	**C**AU (WC)	**C**AU (WC)
Cysteine, Cys/C	UG**U**	**G**CA (wobble)	unchanged
	UG**C**	(WC)	

Table 3.1: Comparison of the natural and anticodon-modified tRNAs used by Döring and Marlière (1998) for reassigning isoleucine and methionine to cysteine.

In bold and underlined are nucleotides of the codon and anticodon that wobble-pair together. All codons and anticodons are in the 5'-to-3' orientation.

did pioneer the research to investigate what I term "codon invasion" for sense codon reassignment. They proposed introducing four variants of anticodon-modified $tRNA^{Cys}_{\underline{N}AU}$ (where N = A/T/G/C) into *E. coli* to assess their ability to base-pair, with full Watson-Crick complementarity, to all AU**N** codons ($Ile_{(AU\underline{U})}$, $Ile_{(AU\underline{C})}$, $Ile_{(AU\underline{A})}$, and $Met_{(AU\underline{G})}$; Table 3.1), and demonstrate isoleucine-to-cysteine (and methionine-to-cysteine) sense codon reassignment in a heterologously-expressed recoded reporter, ThyA. ThyA is an enzyme that catalyses the conversion of deoxyuridine monophosphate (dUMP) to deoxythymidine monophosphate (dTMP), the latter of which is an essential precursor for DNA biosynthesis. Critical residue C146 in the essential *thyA* gene was mutated to either isoleucine or methionine (*thyA*.C146I/M[4] respectively), rendering a non-functional/missense protein product upon translation under the natural system. Furthermore, the endogenous *thyA* copies were deleted from the host genome, producing a *thyA*-deficient that could be used as a live/dead selection platform. That is, if the isoleucine-to-cysteine and methionine-to-cysteine reassignments are successful, the host would be alive, thereby suggesting that the anticodon-modified *Eco*$tRNA^{Cys}_{\underline{N}AU}$ variants were transcribed and able to reassign the recoded *thyA*.C146I/M reporter, ensuring the continued survival of the *E. coli* host.

Their results demonstrated that whilst all non-native anticodon-modified $tRNA^{Cys}_{\underline{N}AU}$ isoacceptors were able to compete with the natural *Eco*$tRNA^{Ile}$ and

[4]Where a gene has been recoded for reassignment, the mutations will be written like this, i.e. [gene name].[residue mutated at position # from X1 to X2]. If multiple mutations are incorporated into the recoded gene, they will be separated by further periods.

*Eco*tRNA^{Met} for the codons, and correctly reassign a recoded *thyA*.C146I/M to recover ThyA activity, the efficiencies of reassignment varied: average *thyA*-deficient *E. coli* growth rates, when expressing each tRNA$^{Cys}_{\underline{N}AU}$ variant, ranged from 0.280 ± 0.019 to 0.533 ± 0.024 (mean ± standard-deviation). Moreover, using tRNA$^{Cys}_{\underline{U}AU}$, Döring and Marlière (1998) was able to show the genetic stability of this codon invasion: the *thyA*-deficient *E. coli* were able to maintain the isoleucine-to-cysteine "mis"-coding for over 530 generations, with no reversion of the Ile$_{(AU\underline{A})}$ codon to Cys$_{(UGU)}$ or Cys$_{(UGC)}$. Consequently, this early research provided the principle foundations on which I built my OCI strategy.

3.1.2 Identifying amino acid pairs for reassignment

Given that Döring and Marlière (1998) had shown a set of tRNA$^{Cys}_{\underline{N}AU}$ that were able to invade all isoleucine and methionine codons alike (including orphan codons Ile$_{(AU\underline{U})}$ and Ile$_{(AU\underline{A})}$), I hypothesised that it would be possible to design any number of alternative tRNAs with anticodons modified to better-complement the 17(+1) true orphans previously identified (Figure 3.3, codons not highlighted/white). This next section details the process for identifying which of these orphan codons would be more amenable for codon invasion as a means to implement and investigate bidirectional sense-to-sense codon reassignments.

As with Döring and Marlière (1998), the anticodon-modified tRNAs for my OCI strategy need to function *in vivo* with the host's endogenous aaRSs. While the orphan codons are the "destination" of the engineered tRNA, the "source" of the engineered tRNAs is restricted to aaRSs that do not use (or do not strongly discriminate against) the tRNA anticodon as an identity (i.e. recognition) element. Of the 20 natural aaRSs, five are known to fit that criteria (further detailed below).

A second issue arises when both natural (but containing wobble-pairing) and engineered tRNAs are introduced. The two tRNA isoacceptors are expected to compete in translation, creating ambiguity at each incorporation – also known as statistical proteins (Woese, 1965; Berg and Brandl, 2021; Pezo et al., 2004; Barbieri, 2019). The more an engineered tRNA can out-compete the natural wobble tRNA, the closer the system is towards codon reassignment. However, competition for in-

corporation will not be limited to the reassignment reporter system and, as a result, multiple host proteins are expected to also be affected. Where the reporter will be active due to the codon invasion by the anticodon-modified tRNAs, most host proteins are expected to be inactivated by the invasion, implying that the invasion is likely to be toxic. Wholesale genome engineering (as later achieved by the Chin group; see Wang et al., 2016; Fredens et al., 2019; Robertson et al., 2021) can be used to remove the orphan codon from the genome, thus removing toxicity, but that approach is not feasible in the time-frame and resources of a single PhD. Instead, I opted to create a transient reassignment system, using the reporter to convert a transient signal into a stable output – the Genetic Toggle Switch (GTS), later described in Chapter 4.

Criterion 1: tRNA anticodon is not a required identity element for aaRS

In fitting with this first requirement, analysis of the literature highlighted a subset of five *E. coli* tRNA/aaRS pairs where the anticodon was reported to not be a major determinant for the respective synthetases (Figure 3.5; Giegé et al., 1998; Yan and Francklyn, 1994; Yan et al., 1996; Giegé et al., 2014; Tocchini-Valentini et al., 2000; Rosen and Musier-Forsyth, 2004). As modifications to this part of the tRNA should still allow viable interactions with the native aaRS, these five candidates could potentially be engineered for sense-to-sense codon reassignment. The corresponding amino acids are proline, alanine, serine, leucine, and histidine.

It should be noted that whilst the review by Giegé et al. (2014, adapted in Figure 3.5) displays many of the proposed key determinants for tRNA/aaRS pairs within *E. coli*, the relative significance of each element (which can have species-specific variations also) is still under debate. As such, it was used as a guideline as to what had previously been explored. For instance, in *Eco*tRNA^{Cys}, the 1-72 base-pair and U73 discriminator base on the acceptor stem are one of the strongest recognition elements for *Eco*CysRS (Ming et al., 2002; Hauenstein et al., 2004; Liu et al., 2012), in that if a specific motif is missing to insert into the Rossmann fold of this class I bacterial aaRS, these docking-defective synthetases would not be able to discriminate and pair with the tRNA. The same mechanism of recognition appears

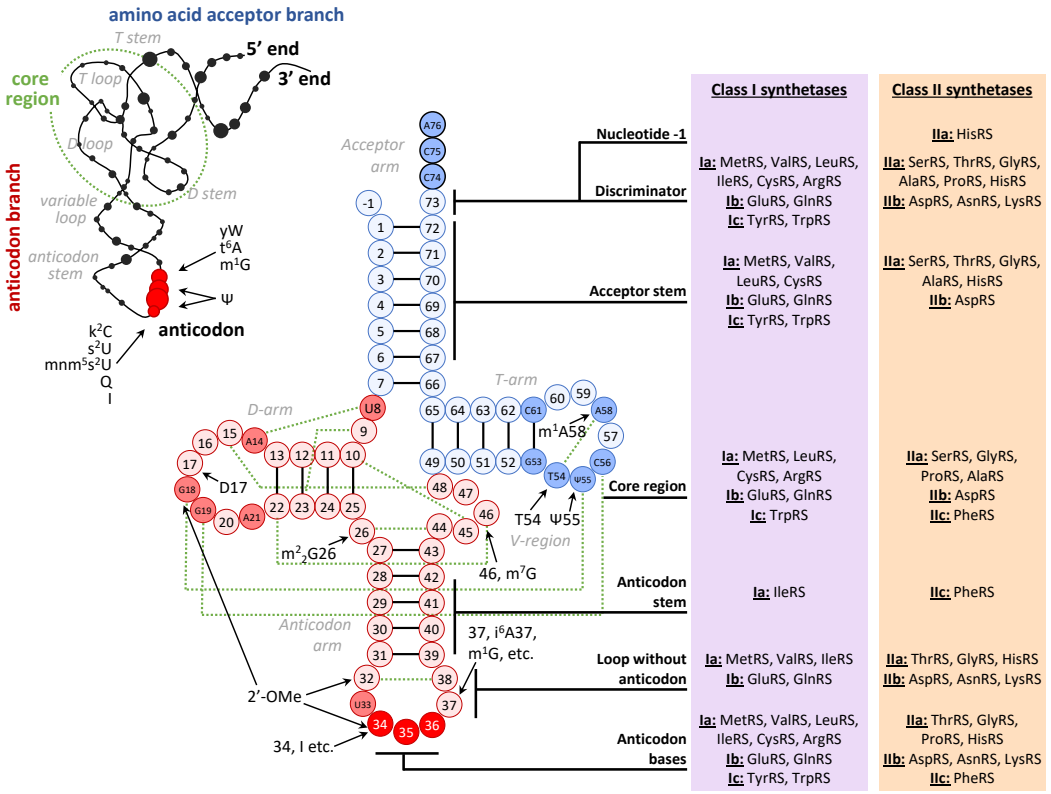


Figure 3.5: Clover-leaf folding structure of tRNA and its 3D L-shaped organisation

Clover-leaf folding structure of tRNA and its 3D L-shaped organisation, including a compilation of regions which have known identity determinants, regardless of whether these are major or minor elements for recognition and/or aminoacylation. It is noted that nearly all amino acids (bar alanine and serine) listed show that the anticodon is an identity determinant, which is logical as each anticodon represents a clear linkage between tRNA and mRNA, suggesting that the cognate synthetases may also use this feature for identification. Figure adapted from Giegé et al. (2014).

to be critical also for many other bacterial class I and for three class II synthetases, as well as the bacterial EF-Tu. Hence as Döring and Marlière (1998) only modified the anticodon (an alleged discriminator, albeit not critical identity element), and not the acceptor stem, they were still able to charge cysteine to the anticodon-modified $\text{tRNA}^{\text{Cys}}_{\underline{\text{N}}\text{AU}}$. Moreover, it is possible that the ThyA selection reporter used was such a necessary component to *E. coli* viability that even if *Eco*$\text{tRNA}^{\text{Cys}}_{\underline{\text{N}}\text{AU}}$/*Eco*CysRS was not highly compatible, it was sufficient for the ones that survived selection. Later discussions with Marlière (personal communication, 13-July-2019) confirmed also

that if the anticodon-modified *Eco*tRNA$^{Cys}_{\underline{N}AU}$ isoacceptors were sufficiently over-expressed, native *Eco*CysRS could aminoacylate them as, by chance, there would likely be enough "accidental" interactions. After which, between the necessity for the host to express ThyA, and with the minimal changes to the engineered tRNAs, there would unlikely be additional major issues with uptake by the ribosome and other translation factors to encode cysteine at recoded isoleucine/methionine AU**N** codons in the *thyA*.C146I/M reporter sequence (Döring and Marlière, 1998).

Regardless, whilst these additional factors could be considered when optimising my proposed OCI strategy for sense-to-sense codon reassignment, the initial approach remained: modify the tRNA with weak anticodon determinants of the five aforementioned amino acids.

Criterion 2: sufficient physico-biochemical differences between amino acid reassignment pair to enable selection

Having narrowed down the potential amino acids that could be compatible with the proposed anticodon-modified tRNA method for Orphan Codon Invasion (to proline, alanine, serine, leucine, and histidine), the next step was to select which pairing(s) from the five would be most pragmatic for characterising sense-to-sense codon reassignment *in vivo*.

In nature, it is well-documented that some reassignments between related amino acids may be tolerated (i.e. still produce functioning proteins despite being non-synonymous substitutions) in what is known as statistical proteins as this form of redundancy confers a degree of adaptive or selective advantage (Woese, 1965; Berg and Brandl, 2021). A key example of such natural occurrences is the Leu/Ser ambiguity as seen in *Candida albicans* and certain other CTG-clade fungi (Santos et al., 1993, 1996; Ji et al., 2016; Butler et al., 2009; Bezerra et al., 2015). In these organisms, the otherwise universal Leu$_{(CUG)}$ codon can be naturally translated as serine. Such ambiguity in protein translation arose initially from random mutations in the anticodon loop of a native *Cal*tRNA^{Ser}, producing an atypical *Cal*tRNA$^{Ser}_{(\underline{C}AG)}$ (further reviewed in Barbieri, 2019; Pezo et al., 2004). Subsequent co-evolution between this anticodon-modified tRNA variant and both its cognate seryl- (*Cal*SerRS)

and non-cognate leucyl- (*Cal*LeuRS) synthetases resulted in *Cal*tRNA$^{Ser}_{(\underline{C}AG)}$ being recognised and naturally aminoacylated with approximately 97% bias towards serine and 3% to leucine (Rocha et al., 2011; Bezerra et al., 2013; Mateus et al., 2013; Simões et al., 2016). *C. albicans* has even been engineered to accept what should be more intolerable ambiguity, with up to 28% leucine mis-incorporation (albeit with significant impact in cell morphology, though seemingly not on growth rate; Gomes et al., 2007). Despite the leucine-to-serine mis-incorporation afflicting 4168 (~67%) Leu$_{(CU\mathbf{G})}$-encoded proteins in *C. albicans*, this species has been able to survive the natural ambiguity with minimal impact on growth rate and fitness as it has, over time, synonymously replaced the once 98% prevalence of Leu$_{(CU\mathbf{G})}$ with other leucine codons (namely, Leu$_{(UA\mathbf{A})}$ and Leu$_{(UA\mathbf{G})}$). Moreover, remaining instances of Leu$_{(CU\mathbf{G})}$ (or any new introduction of this codon) tend to be localised in non-conserved positions or where serine residues would be expected, thus anticipating serine (mis-)incorporation. This natural occurrence of large-scale genome recoding, in tandem with the emergence of the unusual *Cal*tRNA$^{Ser}_{(\underline{C}AG)}$/(*Cal*SerRS and *Cal*LeuRS) pairing, is the reason behind the Leu/Ser ambiguity in *C. albicans*.

Whilst this could be considered codon reassignment (albeit not necessarily by design), performing such substitutions between amino acids with similar properties would make it difficult to determine whether my proposed OCI system is working as expected. As such, I needed reassignment pairs to have as great a difference as possible in terms of physico-biochemical properties, so that any reporter protein would be able to provide two distinct and detectable states of expression: one under the natural system, the other under the synthetic reassignment code. Specifically, the selection assay planned for verifying reassignment uses a heterologously-expressed reporter that has been engineered with missense or nonsense point mutations (i.e. "recoded") with a seemingly loss-of-function phenotype when decoded under the natural genetic code, but "regains" its function under the reassignment system. Having a distinct pattern of expression for the two states provides an easily measurable output that can be used as proxy for codon reassignment.

A basic approach for identifying pairs of amino acids that have significant

physico-chemical differences was to look at amino acid substitution tables like the BLOcks SUbstitution Matrices (BLOSUM). These statistical measures for estimating the protein similarity and/or divergence use log-odds score tables to estimate the chances of observing biologically meaningful amino acid replacements within a sequence (Henikoff and Henikoff, 1992). Briefly, scores range from -4 to +3 for amino acid substitutions, where: a score of zero indicates that the likelihood of substitution between a given pair of amino acids is as expected by chance; positive values suggests that a given amino acid is more likely to be used as a replacement; negative values implies unlikely substitutions. Typically, more positive scores correlate with greater physico-chemical similarities between amino acids that, even as a substitute, should not greatly affect protein function. Hence, to meet the second criteria, and find pairs of amino acids with significantly different side chain properties that, when swapped around, are more likely to impact the structure, interactions, and function of a reporter protein, I focused on pairs with the most negative BLOSUM scores.

As an initial reference point, BLOSUM62 (the standard alignment database matrix for comparing mid-range related protein sequences) was used to identify highly contrasting substitutions between the five prospective amino acids that fitted the first criteria for reassignment (Figure 3.6). BLOSUM62 scores of -2 or less were taken as the arbitrary cut-off threshold for defining disruptive pairwise substitutions for the reassignment selection assays. This gave the following five substitutions (scores): histidine-leucine (-3), leucine-proline (-3), leucine-serine (-2), histidine-alanine (-2), and histidine-proline (-2). In most of these pairs, one tends towards low reactivity (alanine, leucine, proline), whilst the other is typically involved in catalysis (histidine, serine). Replacing hydrophobic, non-polar residues with neutral-hydrophilic, polar residues (or vice versa) seemed most likely to be disruptive to a protein, hence the decision to design reassignment between one of the former three aliphatic amino acids with one of the latter two.

Being small, serine can be found both on the protein surface or buried within its interior. As a surface residue, its size enables it to assist in tight turns, much like proline. Serine's slightly polar nature means that it often resides in the func-

tional centres of enzymes as its hydroxyl group enables it to hydrogen-bond with other (usually polar) residues (Betts and Russell, 2003). Specifically, serine is generally involved in the active, catalytic site of hydrolases (e.g. protease, lipase, phosphatase) or transferases (e.g. protein kinase), as the nucleophilic element of
2310 the classic DHS (acid-base-nucleophile) catalytic triad motif (Warshel et al., 1989; Ekici et al., 2008). Histidine's involvement in these kind of covalent catalysis re-

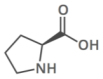
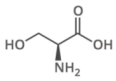
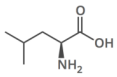
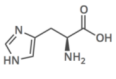
	PROLINE *CCT, CCC, CCA, CCG*	**ALANINE** *GCT, GCC, GCA, GCG*	**SERINE** *TCT, TCC, TCA, TCG, AGT, AGC*	**LEUCINE** *CTT, CTC, CTA, CTG, TTA, TTG*	**HISTIDINE** *CAT, CAC*
					
Blosum62	Pro	Ala	Ser	Leu	His
Pro	7	-1	-1	-3	-2
Ala	-1	4	1	-1	-2
Ser	-1	1	4	-2	-1
Leu	-3	-1	-2	4	-3
His	-2	-2	-1	-3	8
Volume, Å³	Small 108-117	Tiny 60-90	Tiny 60-90	Large 162-174	Medium 138-154
MW, g.mol⁻¹ (-H2O)	115.13 (97.12)	89.10 (71.08)	105.09 (87.08)	131.18 (113.16)	155.16 (137.14)
Structure	Aliphatic (pyrrolidine)	Aliphatic		Aliphatic	Aromatic (imidazole)
Polarity	Non-polar	Non-polar	Polar	Non-polar	Polar
Hydro-phobicity	Hydrophobic - Neutral	Hydrophobic	Neutral - (Hydrophilic)	V. Hydrophobic	Neutral - Hydrophilic
Charge	0	0	0 (non-acidic)	0	+ve (basic, non-acidic)
Other properties	2º amine; Rigid, solvent exposed (usu.)		Hydroxylic; Nucleophilic		pK_a=6.04 at physiological pH

Figure 3.6: Blosum62 matrix and physico-chemical properties of potential amino acids suitable for reassignment

The five potential amino acids that have associated tRNAs that do not use the anticodon region as identity/recognition elements by their cognate aaRSs include: proline (Pro), alanine (Ala), serine (Ser), leucine (Leu), and histidine (His). Note that the values are mirrored along the diagonal line from top-left to bottom-right, therefore only the values below the diagonal in the table are shown.

actions is typically as the base, but sometimes as a weak acid when its imidazole side chain is unprotonated. Because of these biochemical properties, a replacement of either of these catalytic residues by alanine, leucine, or proline (and vice versa) would likely be significantly detrimental to protein function. However, leucine-serine was discounted for reassignment as this substitution pair had already been previously demonstrated, both synthetically (*in vitro* in human and *in vivo* in *E. coli* respectively; Breitschopf et al., 1995; Ho et al., 2016) and naturally (as previously described above in *C. albicans*; Santos et al., 1993). Moreover, given that nature had already provided an elegant demonstration of leucine-to-serine reassignment *in vivo*, as well as having achieved whole genome recoding to further consolidate this alternative assignment, demonstrating an different reassignment pair would be a more compelling demonstration of orthogonal coding systems.

With regards to the aliphatic amino acids, proline seemed a good candidate. It has a distinctive physical property in that its pyrrolidine side chain connects with the protein backbone twice to form a five-membered ring (as an imino acid) in a rigid conformation. Correspondingly, proline is often located at very tight turns and kinks on the protein surface and is generally unable to adopt more main-chain configurations as seen in most other amino acids. As such, it is not usually substituted for other amino acids, nor can others easily take its place, as exemplified by the negative substitution scores seen in the BLOSUM62 matrix compared with the other reassignment amino acid candidates (e.g. leucine-proline at -3 and histidine-proline at -2; Figure 3.6). Furthermore, despite its unique physical structure, its relatively inert biochemical properties means that it is very rarely involved in the active or binding sites of a protein (Betts and Russell, 2003). Hence, for the purposes of reassignment, the histidine-proline pair would likely be the more striking study as their differences span both physical and chemical features (unlike the biochemical similarities of the leucine-proline pair).

In contrast to proline, histidine has many distinctive physico-chemical properties that could be useful in the reassignment selection assay. Having a pK_a (6.04~6.5) near to the physiological pH of *E. coli* means that it can act as both a

proton donor and acceptor to change the aromatic imidazole side chain from neutral(/basic) to positively-charged (more acidic) respectively. Its amphoteric nature also allows histidine to be buried deep in protein cores, or be solvent-exposed on protein surfaces. In addition, the aromatic ring can participate in stacking interactions; as a result, histidine is frequently found in the active catalytic triad motifs (like the DHS/SHH/etc. as described previously; Warshel et al., 1989; Ekici et al., 2008) and binding sites of proteins (Betts and Russell, 2003).

Hence, to reassign histidine would appear to be relatively straightforward: the amino acid is naturally incorporated by the same $\text{tRNA}^{\text{His}}_{\underline{\mathbf{G}}\text{UG}}$ isoacceptor at only two codons, $\text{His}_{(\text{CA}\underline{\mathbf{U}})}$ (an orphan, via wobble-pairing) and $\text{His}_{(\text{CAC})}$ (with Watson-Crick complementarity), and a new prolyl-tRNA could be designed with its anticodon modified to $\underline{\mathbf{A}}$UG (i.e. $\text{tRNA}^{\text{Pro}}_{\underline{\text{A}}\text{UG}}$) to fully complement the histidine orphan and achieve histidine-to-proline reassignment. However, one slight reservation in designating proline as the other reassignment candidate was that it has multiple redundancies for codon complementation: the four proline codons ($\text{Pro}_{(\text{CCN})}$) are each served by at least one of the three different tRNA^{Pro} isoacceptors. Crucially, the only orphan codon $\text{Pro}_{(\text{CC}\underline{\mathbf{U}})}$ is wobble-paired with two different tRNA variants, $\text{tRNA}^{\text{Pro}}_{\underline{\text{G}}\text{GG}}$ and $\text{tRNA}^{\text{Pro}}_{\underline{\text{U}}\text{GG}}$, which means that even if a newly introduced anticodon-modified tRNA ($\text{tRNA}^{\text{Xaa}}_{\underline{\text{A}}\text{GG}}$ with full Watson-Crick complementarity to the codon) is able to be aminoacylated by the native aaRS to a sufficient degree, it would be competing with two other natural (albeit wobble-pairing) tRNA variants for the same orphan $\text{Pro}_{(\text{CC}\underline{\mathbf{U}})}$ codon. That is, its chances of implementing codon reassignment would be much reduced compared to an fully-compatible anticodon-modified tRNA that is competing for a true orphan codon, which in turn is naturally only partially-complemented by one other tRNA variant.

This leaves two remaining candidates that could be coupled with histidine to demonstrate the sense-to-sense codon reassignment selection assays: alanine and leucine. Both have aliphatic side chains, are hydrophobic, neutrally-charged, and non-polar. Alanine, in contrast with the physical and biochemical properties of the other aliphatic candidate proline described earlier, is one of the most chemically

inert amino acids, and hence is hardly ever directly involved in protein function. Instead, it is highly-enriched/over-represented in low complexity regions (LCRs) within protein sequences in bacteria, phages, and eukaryotes. These LCRs are generally very low in amino acid diversity and were once considered "junk" or neutral linker regions between protein domains (Ntountoumi et al., 2019). In addition, its short methyl side chain contributes to its tiny size, and allows it to mimic the secondary structure preferences of many of the other amino acids if used as a replacement. As a result, alanine is often used in "scanning mutagenesis" (alanine scanning), a site-directed mutagenesis technique to determine the bioactivity of a given residue in the function or structure of a protein in question by randomly replacing residues with alanine. In this respect, alanine could be a good candidate to pair with histidine to assess our proposed reassignment system. With regards to leucine: unlike alanine, it tends to be under-represented in these LCRs due to its increased hydrophobicity and bulkier branched-chain form, but its other similar biochemical properties sometimes results in its use as an alternative in scanning mutagenesis if a larger amino acid is needed to conserve the size of mutated residues. Hence, like alanine, leucine would also contrast well with histidine for codon reassignment.

As such, the final consideration was based on the associated orphan codons that would be used for invasion with anticodon-modified tRNAs, and the relative impact their reassignment may have on the host. Alanine has two orphan codons, $\text{Ala}_{(\text{GC}\underline{\mathbf{U}})}$ and $\text{Ala}_{(\text{GC}\underline{\mathbf{G}})}$, which are only served by $\text{tRNA}^{\text{Ala}}_{(\underline{\mathbf{U}}\text{GG})}$ through wobble base-pairing. This tRNA in itself also has its own dedicated Watson-Crick target, $\text{Ala}_{(\text{GC}\mathbf{A})}$: stretching it across three highly used codons in the *E. coli* genome. On the one hand, introducing an engineered tRNA that is dedicated only to either of these orphan codons should make it easier to implement codon reassignment, especially compared with the $\text{tRNA}^{\text{Ala}}_{(\underline{\mathbf{U}}\text{GG})}$ that has a three codon targets. However, alanine codons have very high representation across the *E. coli* genome such that ambiguous translation under both natural and synthetic codes at an alanine orphan may also be excessively detrimental to host fitness. Similarly, whilst leucine codons have the highest usage overall for *E. coli*, this amino acid is represented by six codons and

five tRNA isoacceptors, arguably providing it higher tolerance/acceptance for introduced perturbations from anticodon-modified tRNAs. Though of its three wobble-pairing codons, only one is a true orphan; the only complement to orphan $\text{Leu}_{(\text{CU}\underline{\textbf{U}})}$ is $\text{tRNA}^{\text{Leu}}_{\underline{\textbf{G}}\text{AG}}$, which itself has one other (Watson-Crick) target $\text{Leu}_{(\text{CU}\textbf{C})}$. Given that this $\text{Leu}_{(\text{CU}\underline{\textbf{U}})}$ orphan is less commonly used in *E. coli* than the alanine codons (so may incur less fitness penalties to the host), and has less complex interactions to contend with than with the proline codons, the histidine-leucine reassignment pair was considered to be the most suitable compromise for validating sense-to-sense codon reassignment.

3.1.3 Designing anticodon-modified tRNAs, compatible with endogenous *E. coli* aaRSs, to implement reassignment with the Orphan Codon Invasion system

Having established histidine and leucine as the amino acids to test the OCI reassignment system *in vivo*, and the respective orphan codons ($\text{His}_{(\text{CA}\underline{\textbf{U}})}$ and $\text{Leu}_{(\text{CU}\underline{\textbf{U}})}$) to invade, the next step was to design the anticodon-modified tRNAs to implement the reassignments. The requirements of these anticodon-modified tRNAs were as follows: to engineer the anticodon triplet (and adjacent bases, if necessary) to have Watson-Crick complementarity with a non-cognate orphan codon reassignment target, yet be able to be recognised and aminoacylated by its otherwise cognate aaRS. Specifically, for a leucine-to-histidine reassignment using orphan codon $\text{Leu}_{(\text{CU}\underline{\textbf{U}})}$, we would introduce anticodon-modified $\text{tRNA}^{\text{His}}_{\underline{\textbf{A}}\text{AG}}$, and for a histidine-to-leucine reassignment at orphan $\text{His}_{(\text{CA}\underline{\textbf{U}})}$, a new isoacceptor $\text{tRNA}^{\text{Leu}}_{\underline{\textbf{A}}\text{UG}}$ would be used. Both these tRNAs were designed based on pre-existing tRNA sequences so as to reduce the possibility of tRNA/aaRS incompatibility.

To minimise the risk of rejection of engineered $\text{tRNA}^{\text{His}}_{\underline{\textbf{A}}\text{AG}}$ by host *Eco*HisRS (genomically encoded by the single-copy *hisS* gene; Chaliotis et al., 2017), the only native histidyl-tRNA ($\text{tRNA}^{\text{His}}_{\underline{\textbf{G}}\text{UG}}$, gene *hisR*; Figure 3.7, top) was used as the base sequence. Here, the first two bases (positions 34 and 35) of the anticodon were changed from $\underline{\textbf{G}}\textbf{UG}$ to $\textbf{AA}\textbf{G}$, to fully match orphan $\text{Leu}_{(\text{CU}\underline{\textbf{U}})}$ for leucine-

to-histidine reassignment. This alternate match, $\text{tRNA}^{\text{His}}_{\underline{\text{A}}\text{AG}}$, for $\text{Leu}_{(\text{CU}\underline{\text{U}})}$ was also defined as *hisR(leu)*, and as it was to be heterologously expressed within the host, it would be competing with the endogenous leucyl-tRNA, *Eco*$\text{tRNA}^{\text{Leu}}_{\underline{\text{G}}\text{AG}}$, for this orphan codon too.

Deciding on the source sequence for the proposed histidine-to-leucine $\text{tRNA}^{\text{Leu}}_{\underline{\text{A}}\text{UG}}$ engineered variant was a little more complex as there is a larger se-

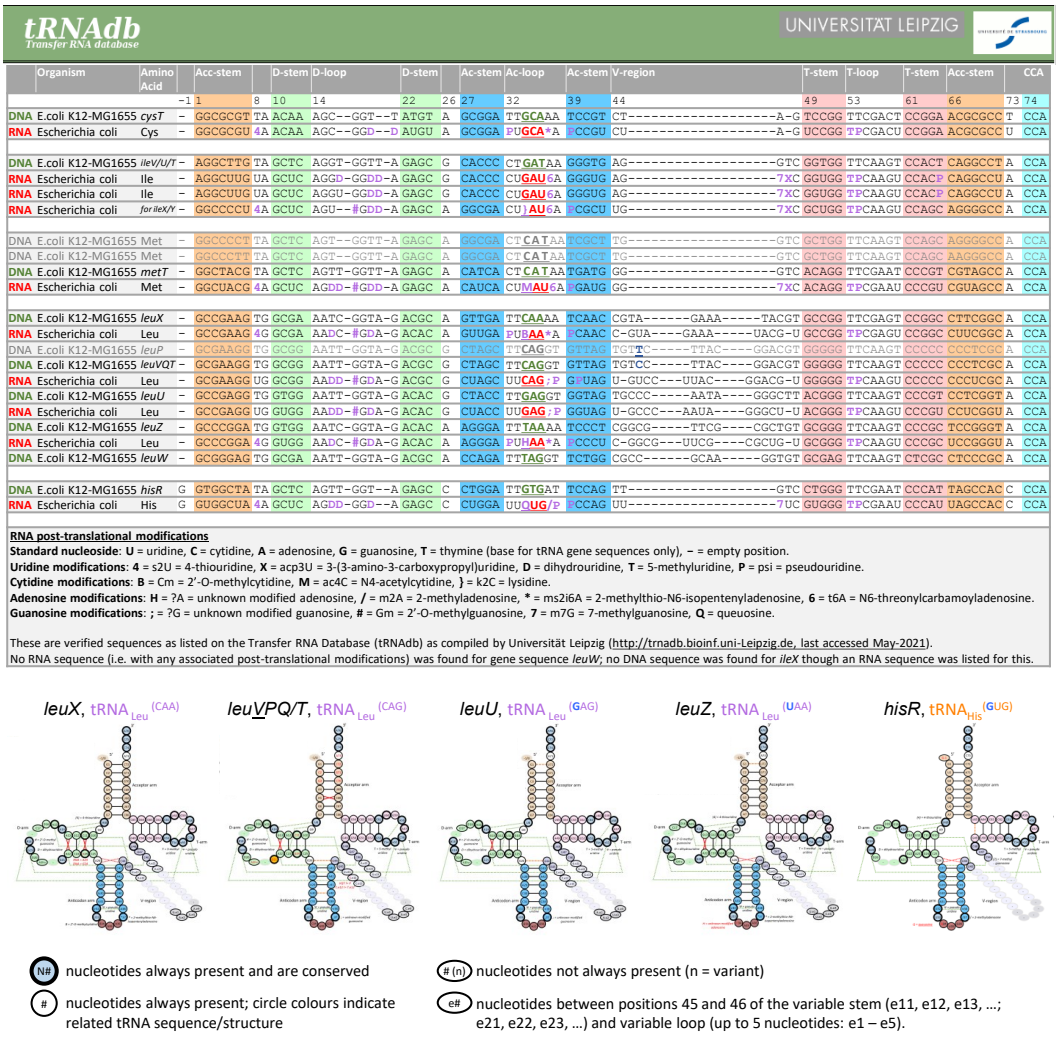


Figure 3.7: Gene and tRNA sequences associated with amino acids: cysteine, isoleucine, methionine, leucine, and histidine.

The top table is of the tRNA gene/DNA sequences for glsecol K-12 strain and the associated processed tRNA sequences for general *E. coli*, as a compilation of results from the Transfer RNA database (tRNAdb, `http://trna.bioinf.uni-leipzig.de`; Jühling et al., 2009; Sprinzl and Vassilenko, 2005): a curated, experimentally-validated and sequence-verified tRNA database that includes information on post-translational modifications. The different *Eco*tRNA^{His} and *Eco*tRNA^{Leu} isoacceptors and their DNA sequences are shown below; adapted from (Fung et al., 2014).

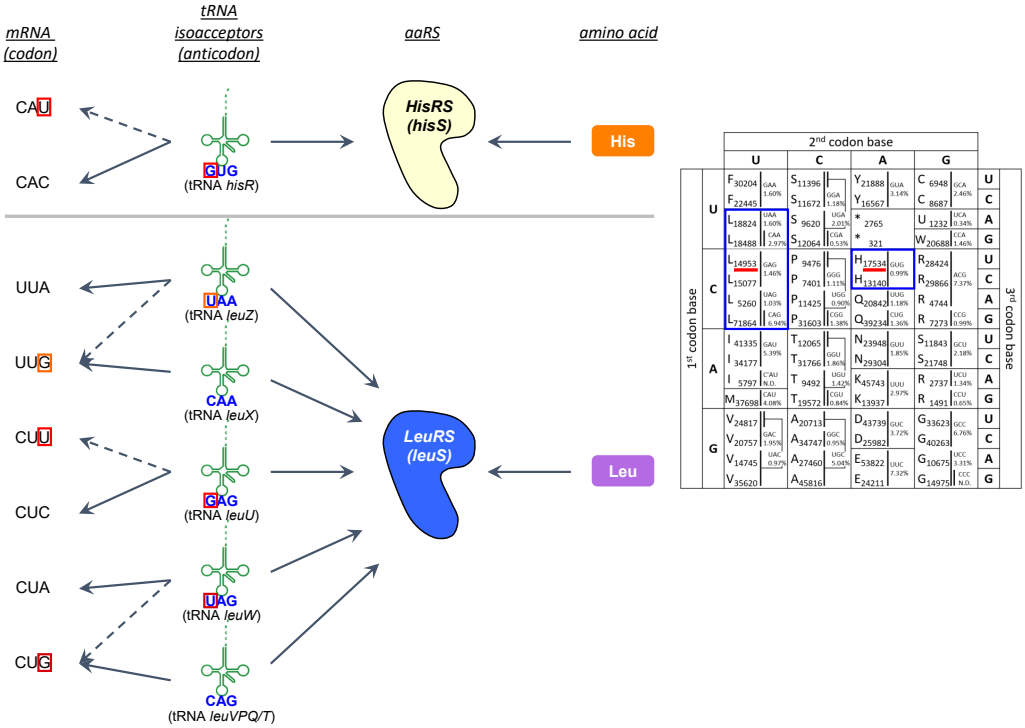


Figure 3.8: Association between the histidine and leucine amino acids with their codons, tRNAs (and respective anticodons), and cognate aaRSs

The wobble base-pairing between the codons and their respective tRNA anticodons are highlighted by the red/orange boxes, with the key orphan codons underlined in red within the blue boxes.

lection from which to choose. As mentioned previously, there are five tRNA^{Leu} isoacceptors, encoded by eight gene sequences (six of which are unique), that complement the six leucine codons (Figure 3.8). Unlike *Eco*HisRS, which has a one-to-one interaction with its $\text{tRNA}^{\text{His}}_{\underline{\mathbf{G}}\text{UG}}$, the same *Eco*LeuRS (from single genomic copy gene, *leuS*) is able to recognise and aminoacylate all five tRNA^{Leu} isoacceptors despite differences in their sequences (Figure 3.7, Figure 3.8). That is not to say that there is equal preference for each isoacceptor: *in vitro* experiments by Fung et al. (2014) shows quite a range in affinity and turnover rate estimates from the most favoured *Eco*LeuRS isoacceptor, $\textit{Eco}\text{tRNA}^{\text{Leu}}_{\text{CAG}}$ (genes: *leuVPQ,T*; $K_m = 2.0 \pm 0.4$ μM, $k_{cat} = 0.100 \pm 0.003$ min^{-1} respectively) to the least, $\textit{Eco}\text{tRNA}^{\text{Leu}}_{\underline{\mathbf{G}}\text{AG}}$ (gene: *leuU*; $K_m = 5.4 \pm 3.2$ μM, $k_{cat} = 0.013 \pm 0.002$ min^{-1} respectively). Although possessing lower affinity and slower turnover to *Eco*LeuRS, at least *in vitro*, it still seemed more prudent to use $\textit{Eco}\text{tRNA}^{\text{Leu}}_{\underline{\mathbf{G}}\text{AG}}$ as the foundation

sequence, and simply make the two base changes from **G**A**G** to **AU**G, to fully match the orphan codon $\text{His}_{(\text{CA}\underline{\text{U}})}$ for implementing histidine-to-leucine reassignment. It may be that by changing the tRNA as little as possible, the host may still be able to transcribe and apply any *in vivo* modifications to the engineered $\text{tRNA}^{\text{Leu}}_{\underline{\text{A}}\text{UG}}$ (defined also as *leuU(his)*) to ensure comparable expression levels as the endogenous *Eco*$\text{tRNA}^{\text{Leu}}_{\underline{\text{G}}\text{AG}}$ (Grosjean et al., 1998). Moreover, even if there is reduced interaction between *Eco*LeuRS and *Eco*$\text{tRNA}^{\text{Leu}}_{\underline{\text{G}}\text{AG}}$ (and, by extension, expected further diminished activity with synthetic $\text{tRNA}^{\text{Leu}}_{\underline{\text{A}}\text{UG}}$), this may serve to buffer the expected toxicity from mis-translation that is expected to arise from ambiguous decoding from both native and reassignment tRNAs – *Eco*$\text{tRNA}^{\text{His}}_{\underline{\text{G}}\text{UG}}$ and $\text{tRNA}^{\text{Leu}}_{\underline{\text{A}}\text{UG}}$ respectively– for the same orphan $\text{His}_{(\text{CA}\underline{\text{U}})}$ codon.

A summary of the two proposed anticodon-modified tRNAs and their interaction with their associated codons – both the natural (in blue) and alternate (in red) code – is shown in Figure 3.9.

If deemed necessary, the bases around the anticodon-modified tRNAs could be changed and, similarly, directed evolution to co-evolve the engineered tRNAs to be accepted by the native *Eco*LeuRS and *Eco*HisRS was considered (discussed later in Section 3.3).

3.1.4 Recoding leucine and histidine codons in selection reporters to establish robust reassignment assays

In parallel with identifying the target codons and the appropriate tRNAs to engineer for Orphan Codon Invasion, a suitable reporter selection system is needed to complete the histidine-leucine reassignment assay. The heterologously-expressed reporter needs to have fast expression that leaves a strong and long-lasting signal. It should remain stable and be continually expressed in *E. coli* without greatly impacting host fitness, whether grown in liquid or solid media. The signal output should ideally be visual, or have a quantifiable phenotype that could be used as a proxy for assessing reassignment. It is only when anticodon-modified tRNAs are successfully transcribed and aminoacylated by the host aaRSs to correctly reassign the recoded reporter that a functionally-restored protein should be expressed.

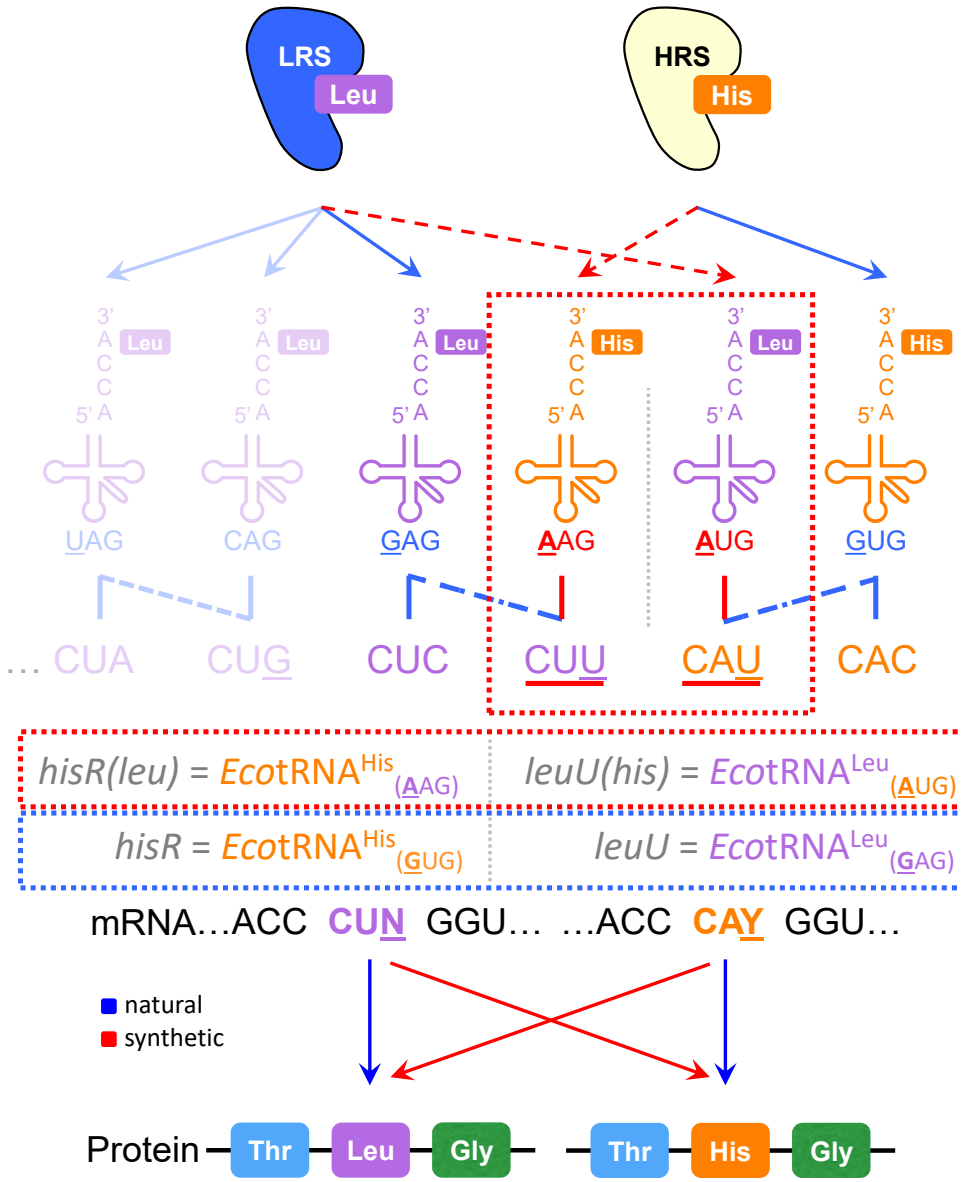


Figure 3.9: Proposed modifications to the anticodon of $tRNA^{His}$ and $tRNA^{Leu}$ for invasion of orphan codons $\text{Leu}_{(\text{CU}\underline{\text{U}})}$ and $\text{His}_{(\text{CA}\underline{\text{U}})}$ respectively

Native/natural anticodon sequences and their base-pairing interactions are shown in blue, whereas the proposed anticodon-modified tRNAs are in red. Those highlighted in the dotted red boxes are the anticodon-modified tRNAs that should enable leucine-to-histidine($Eco\text{tRNA}^{\text{His}}_{\underline{\text{A}}\text{AG}}$) and histidine-to-leucine ($Eco\text{tRNA}^{\text{Leu}}_{\underline{\text{A}}\text{UG}}$) codon reassignment. The tRNAs and codons that relate to leucine are shown in purple, and those relating to histidine are in orange.

To test a leucine-to-histidine reassignment, for example, the ideal reporter would need to have at least one critical histidine that would result in an inactive

protein (i.e. not displaying the desired phenotype; "negative phenotype") when recoded to another sequence, specifically to orphan Leu$_{(CT\underline{T})}$. However, when the engineered leucine-to-histidine reassigning *Eco*tRNA$^{His}_{\underline{A}AG}$ is transcribed alongside this inactive reporter, reassignment under the alternative code should take place, resulting in a "regain-of-function" of the recoded reporter (i.e. presenting the desired "positive phenotype"; Figure 3.10c). The same principle applies for testing histidine-to-leucine reassignment, except that essential leucine codons are recoded to orphan His$_{(CA\underline{T})}$, which are expected to be rescued with synthetic *Eco*tRNA$^{Leu}_{\underline{A}UG}$.

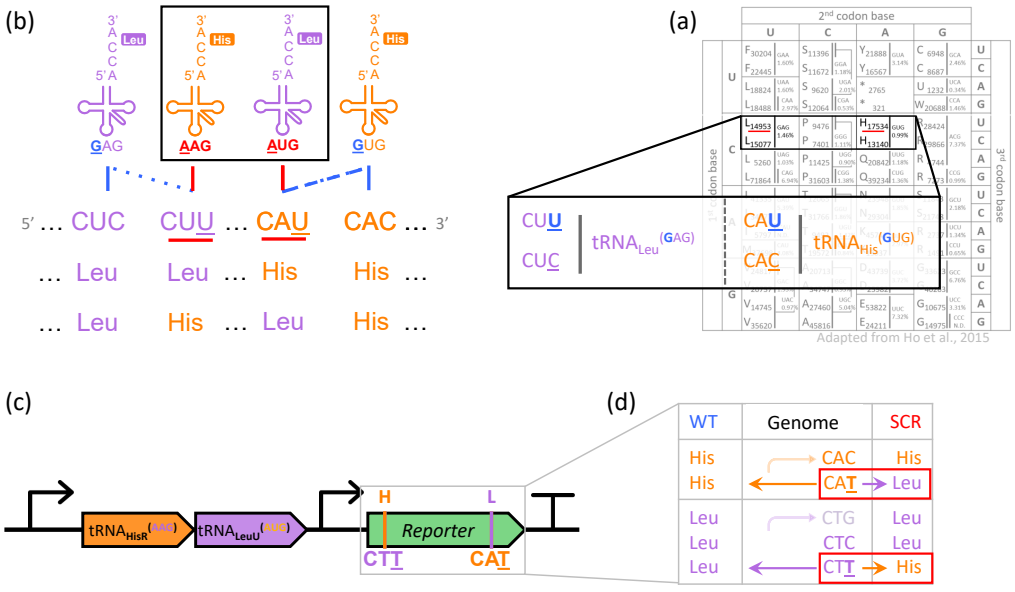


Figure 3.10: A summary of the proposed anticodon-modified tRNAs OCI system for establishing the histidine-leucine codon reassignments

(a) Highlights the target tRNAs that will be the basis for creating the anticodon-modified tRNAs for histidine-leucine codon reassignments; (b) shows the proposed changes to the anticodons, and the likely admix incorporation of leucine and histidine at the orphan Leu$_{(CU\underline{U})}$ and His$_{(CA\underline{U})}$ codon sites; (c) is a schematic of the OCI system, which includes both the anticodon-modified tRNAs and a recoded reporter; (d) is a table showing that the recoded reporter should have all sites where histidine should be incorporated to be recoded to His$_{(CAC)}$, whilst where leucine should be Incorporated to be recoded to Leu$_{(CTC)}$ or Leu$_{(CTG)}$. The sites (both specific sites selected for reassignment in the recoded reporter, but also all other instances within the unmodified host genome) where there are codons His$_{(CA\underline{T})}$ and Leu$_{(CT\underline{T})}$ will likely have a mixture of histidine or leucine, depending on whether the mRNA is decoded by the natural or synthetic tRNAs. WT = wild-type/natural decoding; SCR = sense-to-sense codon reassignment (i.e. histidine-leucine reassignment).

Whilst recoding essential leucines and histidines is necessary for the regain-of-function assay of the reporter protein, it was equally important to recode the "non-essential" $\text{Leu}_{(\text{CT}\underline{\textbf{T}})}$ and $\text{His}_{(\text{CA}\underline{\textbf{T}})}$ codons that have not been designated for reassignment, and which should remain leucine and histidine respectively. Recoding these other residues to synonymous codons ensures that there are no inadvertent reassignments that may confound the results, and the expected recovery of the inactive reporters. At best, ambiguous decoding of these other leucines and histidines still recovers reporter function, but its apparent activity may be reduced, which could be misconstrued as additional toxicity from ambiguous tRNA translation. At worst, it could result in the reporter remaining inactive despite successfully reassigning the designated, essential codon(s), i.e. a false negative phenotype. Hence, histidine residues that should not be reassigned were recoded to the only other histidine codon, $\text{His}_{(\text{CAC})}$, and the non-critical leucine residues were recoded to *E. coli*'s most preferred synonymous codon, $\text{Leu}_{(\text{CTG})}$ (Figure 3.10d)[5]. Once the engineered tRNAs are introduced *in vivo*, the $\text{His}_{(\text{CAC})}$ codon, being a perfect Watson-Crick match to the $\text{tRNA}^{\text{His}}_{\underline{\textbf{G}}\text{UG}}$ will remain translated as histidine, whilst orphan codon $\text{His}_{(\text{CA}\underline{\textbf{U}})}$ will be statistically translated as histidine (naturally) or leucine (reassignment), depending on the efficiency of invasion by the engineered $\text{tRNA}^{\text{Leu}}_{\underline{\textbf{A}}\text{UG}}$ (Figure 3.10b).

Note that these ambiguities affect not only the heterologous reporter, but also that of all the proteins expressed by the host genome. This invariably infers a significant fitness cost to the host, and likely will reduce the signal of correct reassignment from the recoded reporter system. As such, the resulting population of proteins produced would likely have a mix of both functional and non-functional activity.

Given the challenge *in vivo* sense-to-sense codon reassignment already poses, reassignment was planned only at essential reporter residues, however, if histidine-leucine reassignment can be achieved in the small-scale OCI system, it would imply that alternate genetic codes could be implemented upon wholesale recoding of an

[5]It would probably have been equally sufficient to recode the non-essential leucines to any of the other four synonymous codons, such as $\text{Leu}_{(\text{CTC})}$, but as the leucine-to-histidine $\text{tRNA}^{\text{His}}_{\underline{\textbf{A}}\text{AG}}$ was based off the same sequence as the natural complement, $\text{tRNA}^{\text{His}}_{\underline{\textbf{G}}\text{UG}}$, it seemed safer to use a different leucine codon altogether to avoid any unexpected reassignments, i.e. should $\text{tRNA}^{\text{His}}_{\underline{\textbf{A}}\text{AG}}$ be able to also reassign $\text{Leu}_{(\text{CTC})}$, as well as the $\text{Leu}_{(\text{CTT})}$ partner it was designed for.

organism's genome – therein begins another strategy for semantic biocontainment of genetically modified organisms.

3.1.5 Selecting reporters for the reassignment assays

Several sets of reporter systems were examined concurrently to explore a strategy that would be most accurate in detecting reassignment. In general, it seemed more practical to attempt detection from inactive-to-active protein expression (i.e. regain-of-function) as natural variation in *in vivo* expression rates and initiation may make it harder to determine when loss of signal will occur if using the reverse, active-to-inactive, detection method.

With this in mind, fluorescent reporters (and other markers that express coloured pigments) seemed an ideal first approach. Such proteins tend to produce visually robust phenotypes that can be quickly and directly observed in *E. coli*, without the need for downstream purification for quantification. An adaptation upon this would be to measure the efficiency of periplasmic secretors that have been coupled to an efficient colorimetric system (such as, split fluorescent proteins or α-complementation of β-galactosidase (*lacZ* for blue-white screening).

Whilst colorimetric outputs are an easy detection system, without some form of selection pressure, there was the possibility that the host would expel the newly introduced (not yet genomically-incorporated) anticodon-modified tRNAs, as they would likely pose too much of a fitness cost to the host. To that end, an alternative plan was to direct our search towards catalytic enzymes conferring antibiotic resistance as an appropriate reassignment reporter. It would be easy to apply a selection pressure (i.e. antibiotics) and use a live/dead screening of the *E. coli* to observe whether reassignment has occurred.

Figure 3.11 summarises the reporters that were explored as potential systems for the OCI reassignment assays. Although it was possible to investigate the ability of each for implementing leucine-to-histidine or histidine-to-leucine reassignments, the nature of the reporters' characteristics, and in the interest of gaining definitive signals of sense-to-sense codon reassignment, predisposed the assays to be used for assessing leucine-to-histidine rather than the alternative. Exploring an assay for the

histidine-to-leucine reassignment will be addressed in greater depth in Chapter 4.

3.1.6 Implementing the Orphan Codon Invasion system

Thus far, we have discussed the key elements involved in setting up sense-to-sense codon reassignment in an *in vivo* system. We have identified a pair of amino acids, leucine and histidine, with orphan codons (i.e. is only complemented by natural tRNAs through wobble base-pairing) open to invasion from novel tRNAs with anti-codons modified to have full Watson-Crick complementarity to the codon sequences ($Leu_{(CU\underline{U})}$ and $His_{(CA\underline{U})}$). We have also identified the parameters for designing the recoded reporters to characterise histidine-leucine reassignment. The reassign-able codons have been defined ($Leu_{(CT\underline{T})}$ and $His_{(CA\underline{T})}$), as have the codons that should not be reassigned (i.e. synonymous recoding of all non-critical residues to $His_{(CAC)}$ and $Leu_{(CTG)}$). Finally, we have focused the search for reporter function towards enzymes that produce fluorescence, coloured pigments, or have measurable catalytic activity in the host.

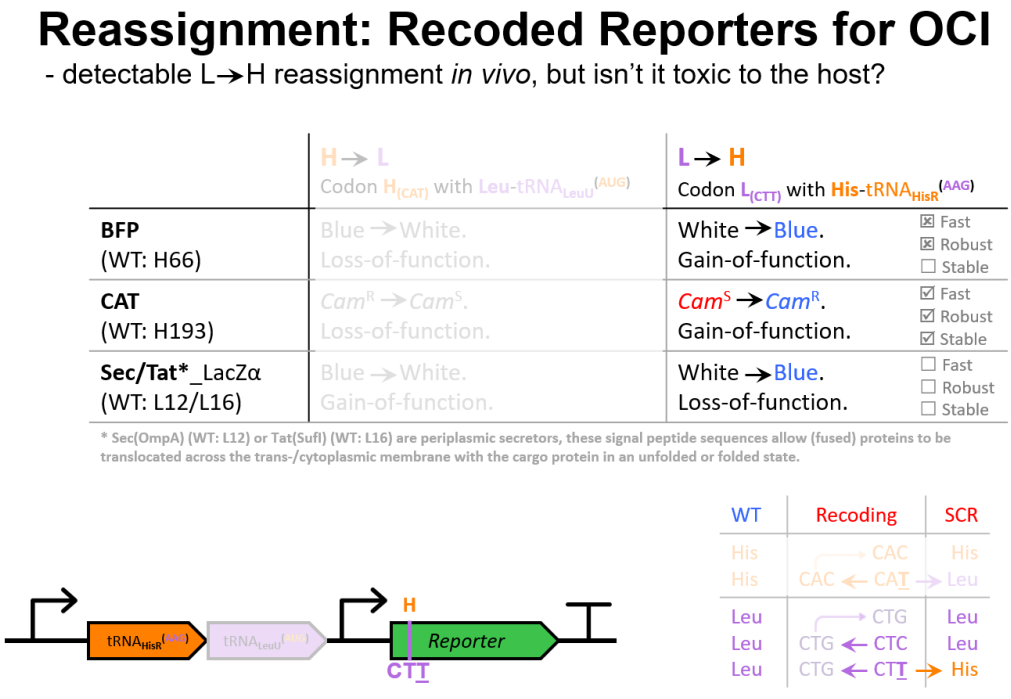


Figure 3.11: Proposed reporters to be recoded for the reassignment assays

To be explored in Section 3.2.2 (fluorescent proteins), Section 3.2.3 (secretion tags coupled to LacZα), and Section 3.2.4 (chloramphenicol acetyl-transferase). Note, these assays would only be able to demonstrate leucine-to-histidine reassignment.

Together, the anticodon-modified tRNA and the recoded reporters make up the Orphan Codon Invasion system with which I will endeavour to demonstrate the feasibility of implementing histidine-leucine reassignment *in vivo*.

3.2 Results and Discussion to OCI

To validate viable sense-to-sense codon reassignment with the Orphan Codon Invasion system, it was necessarily to assemble multiple components, which include demonstrating that the engineered tRNA genes can be heterologously expressed in the host, as well as creating inactive recoded reporters that exhibit a measureable signal of selection upon successful reassignment.

Here, I begin by discussing the construction of the anticodon-modified tRNAs, before moving to the assembly of the recoded reporters. For each recoded reporter, the expected positive and negative phenotypes will first be verified and, once confirmed that the two states are sufficiently different, they will be expressed alongside the engineered tRNAs to test the reassignment assay. Successful reassignment should result in the inactivated reporter (i.e. negative phenotype) regaining its function, and once again displaying the desired positive phenotype.

3.2.1 Constructing anticodon-modified tRNAs for the Orphan Codon Invasion system

In Section 3.1.3, I reasoned that the probability of the proposed anticodon-modified tRNAs being accepted by host aaRSs may be increased if the synthetic sequences were based on pre-existing tRNA isoacceptors within the *E. coli* genome. To recap, the native *E. coli* tRNA gene *hisR* (*Eco*tRNA$^{\mathrm{His}}_{\underline{\mathbf{G}}\mathrm{UG}}$) became the foundation sequence for the proposed tRNA$^{\mathrm{His}}_{\underline{\mathbf{A}}\mathrm{AG}}$for the leucine-to-histidine reassignment invading orphan codon Leu$_{(\mathrm{CU}\underline{\mathbf{U}})}$, whilst the histidine-to-leucine tRNA$^{\mathrm{Leu}}_{\underline{\mathbf{A}}\mathrm{UG}}$variant was modelled from *leuU* (*Eco*tRNA$^{\mathrm{Leu}}_{\underline{\mathbf{G}}\mathrm{AG}}$) to target orphan codon His$_{(\mathrm{CA}\underline{\mathbf{U}})}$. Table 3.2 (and Figure 3.9) summarises these tRNAs and the codon-anticodon triplets, as they appear in the tRNA naturally, and the proposed anticodon change for the synthetic variants (LH, HL) in order to achieve sense-to-sense codon reassignment.

Foundational *Eco*tRNA gene	**Natural (codon-anticodon)**	**Synthetic, recoding (codon-anticodon)**
hisR	*hisR* *Eco*tRNA$^{His}_{GUG}$ (CAU-GUG)	*hisR::leuAAG* tRNA$^{His}_{AAG}$ (CUU-AAG)
leuU	*leuU* *Eco*tRNA$^{Leu}_{GAG}$ (CUU-GAG)	*leuU::hisAUG* tRNA$^{Leu}_{AUG}$ (CAU-AUG)

Table 3.2: *E. coli* tRNA genes (codon-anticodon) in the natural system and in the proposed synthetic sense-to-sense codon reassignment system

Native *E. coli* tRNA gene *hisR* (*Eco*tRNA$^{His}_{GUG}$) became the foundation sequence for the leucine-to-histidine reassigning tRNA$^{His}_{AAG}$, whilst the histidine-to-leucine tRNA$^{Leu}_{AUG}$variant was modelled on *leuU* (*Eco*tRNA$^{Leu}_{GAG}$). Note both codon and anticodon triplets are shown in the 5'-to-3' direction.

The gene sequences for *hisR* and *leuU* were initially acquired from the tRNAscan-SE Genomic tRNA Database (GtRNAdb; `http://gtrnadb.ucsc.edu/`; Chan and Lowe, 2016), under the tRNA gene predictions using the "*Escherichia coli str. K-12 substr. MG1655*" genome[6]. Although the GtRNAdb has more extensive records, it consists only of DNA prediction sequences, hence the necessary sequences were cross-referenced with those from tRNAdb (`http://trna.bioinf.uni-leipzig.de`; Jühling et al., 2009; Sprinzl and Vassilenko, 2005). tRNAdb compiles a more curated, experimentally-validated and sequence-verified tRNA gene subset, including *in vivo*[7] expression data of the mature, post-modification tRNA sequences. From these sequences, only minor modifications were made to the first two bases of the anticodons (Table 3.2, from blue to red) to

[6]Note that the GtRNAdb has four versions for *E. coli* strain K-12, substrain MG1655; the genome used was `http://gtrnadb.ucsc.edu/genomes/bacteria/Esch_coli_K_12_MG1655/`, which is from the "ASM584v2" assembly that was submitted by Univ. Wisconsin, 26-September-2013, `https://www.ncbi.nlm.nih.gov/assembly/GCA_000005845.2` (GenBank sequence U00096.3, `https://www.ncbi.nlm.nih.gov/nuccore/U00096`; Blattner et al., 1997). The relevant sequences were also cross-referenced with those from the "*Escherichia coli BL21DE3*" genome (of which there were two versions; Jeong et al., 2009), to compare whether the tRNA gene sequences differed between the K-12 and B-strains; no differences were found.
Sequences acquired at the time were from GtRNAdb Data Release 16 (September-2015), containing tRNA predictions using the beta version of tRNAscan-SE 2.0. Later work used in this thesis also referred to the subsequent Release 17 (December 2017) (archive: `http://gtrnadb.ucsc.edu/archives.html`; change log: `http://gtrnadb.ucsc.edu/change_log.txt`; Chan and Lowe, 2009, 2016). Whilst tRNA gene sequences were consistent across all versions of both strains, their genomic positions and tRNAscan-SE IDs differed between versions: any tRNA ID numbers mentioned refer to GtRNAdb Release 16 of the genomic sequence with GenBank ID U00096.3.

[7]Note that whilst there were gene sequences for the K-12 substrain, the post-modification tRNA sequences were specific to *E. coli* as a whole (Figure 3.7).

make the sequences, $tRNA^{His}_{\underline{A}AG}$ (*hisR::leu**AAG***) and $tRNA^{Leu}_{\underline{A}UG}$ (*leuU::his**AUG***). To form the full sense-to-sense codon reassignment system, the two sequences were constructed together as construct [201] (Figure 3.12).

As it was yet unclear whether the two engineered tRNA sequences would be able to be expressed *in vivo* as just the gene sequences alone, additional features were added to the final tRNA reassignment construct design [201] to mirror the *E. coli* genomic context and facilitate the natural tRNA processing as closely as possible. Firstly, each anticodon-modified tRNA was designed to include flanking 10~11 bp of the *E. coli* genomic context directly upstream and downstream of their source tRNA gene sequence, which I hypothesised should be sufficient to permit tRNA processing and maturation by ribozymes RNase P, E and T (or PH) (Mohanty and Kushner, 2019). For a similar reason, additional DNA sequences were included within the 395 bp tRNA construct, resulting in the following schema (Figure 3.12). Lastly, to incorporate this construct design into a heterologously-expressed plasmid, compatible homologous arms were added to the full sequence termini. At the 5'-end, the homologous sequence arm contained restriction sites `BsaI=SpeI` and `BamHI`, whilst at the 3'-end the sequence included `SapI`, the `t-lpp` lipoprotein hairpin terminator, `HindIII`, and `BsaI=EcoRI`. The final 395 bp anticodon-modified tRNA construct ([201], oligo YKH159) was ordered as a linear dsDNA from IDT.

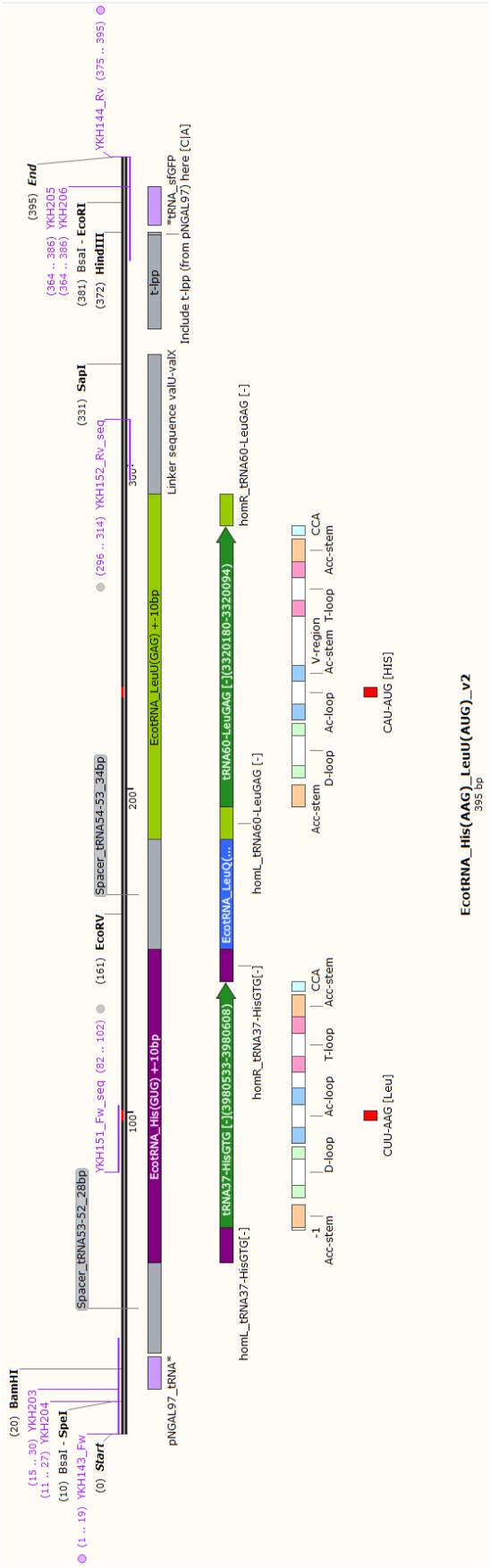


Figure 3.12: A schematic of the two anticodon-modified tRNAs sequences in construct [201] (395 bp)

The plasmid to which the engineered tRNAs construct [201] was designed to be cloned under pTetR/A control in a high-copy pNGAL97 vector [P005], with the intention that the associated recoded reassignment reporter construct would also be included within the same plasmid to complete the full OCI design (Figure 3.13). This would be achieved through a three-part `BsaI`-digestion/ligation assembly strategy. When assembled, the $P_{TetR/A}$ promoter would be present upstream of the whole engineered tRNAs reassignment construct [201], which would facilitate inducible transcriptional control (by anhydrotetracycline, aTc) of the reassignment system. Various restriction sites were also built into the construct design to facilitate any troubleshooting issues. For example, `BamHI` and `EcoRV` could be used to remove the $\text{tRNA}^{\text{His}}_{\underline{\mathbf{A}}\text{AG}}$ (*hisR::leu**AAG***) sequence, whilst `EcoRV` and `SapI` could remove the $\text{tRNA}^{\text{Leu}}_{\underline{\mathbf{A}}\text{UG}}$ (*leuU::his**AUG***) portion. If necessary, both engineered tRNAs sequences could be removed using `BamHI` and `SapI`.

Further downstream in the same plasmid would be the recoded reporter construct, which was planned with sfGFP in the first instance, but could be replaced with alternative reporters (using a combination of the following restriction enzymes: `SapI`, `HindIII`, or `EcoRI` with `XbaI` or `HindIII`). Like the synthetic tRNA operon, the recoded reporter would also be under inducible transcriptional control: this time with the P_{T7} promoter and *lacO* operator, and regulated by the strong consensus RBS sequence (AAGGAG). The design and construction of the recoded reporters is discussed further in Section 3.2.2, Section 3.2.3, and Section 3.2.4.

One other consideration was with regards to potential read-through transcription of either engineered reassignment tRNAs or recoded reporter from the other constructs' promoter. To mitigate this, and enable a degree of control over each portion of the OCI system built into this [202] plasmid design, the two operons were positioned in opposite orientation to one another (with the tRNA construct [201] placed in the same direction as other operons within the plasmid). Finally, should pNGAL97 [P005] prove to be an unsuitable vector (e.g. copy number too high causing significant toxicity to the host; pTetR/A not stringent enough) for this reassignment assay, there were additional design features in place to replace vari-

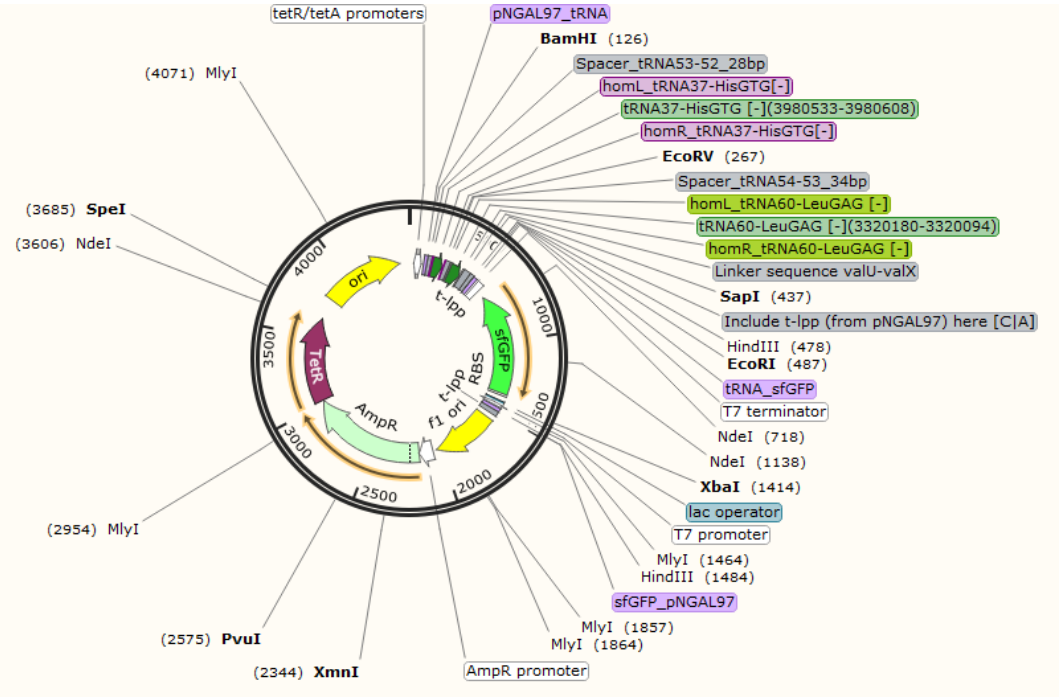


Figure 3.13: A schematic of the two anticodon-modified tRNAs sequences with the recoded sfGFP reporter, construct [202] (4,461 bp)

ous parts of the vector, else both engineered tRNAs and the recoded reporter could be excised together with `BamHI` and `XbaI` to be subcloned into an alternate vector background.

3.2.2 Assessing the feasibility of blue fluorescent protein for leucine-to-histidine reassignment

3.2.2.1 Fluorescent proteins display spectral diversity depending on chromophore structure

Since the discovery of GFP from the jellyfish *Aequorea victoria* in 1962, *av*GFP has become a well-characterised and much enhanced tool; one that has been used to develop a full spectrum of fluorescent proteins (FPs) that are now invaluable across many biological fields. The structure of *av*GFP (PDB: 1EMA), and most of its derivatives, is of a β-barrel formed of 11 anti-parallel β-sheets (with an approximate length of 40 Å, diameter of 30 Å; Figure 3.14), and with an α-helix threaded through the centre of the cylinder. Of particular importance, however, is a tripeptide

core, situated at positions 65-67 along this central axial α-helix, known as the chromophore. The chromophore is a molecule that absorbs light at a particular wavelength and emits the unabsorbed light at another wavelength, which can be detected as a specific colour output if emitted within the visible light range. In *av*GFP, the chromophore is made up of the three covalently-bonded residues S65-Y66-G67, which must follow a series of maturation steps before it produces the green fluorescence characteristic of it and its derivatives (i.e. within the 500-530 nm range[8]). The biosynthesis of the chromophore begins with a number of torsional adjustments (folding), where the carboxyl carbon of the amino acid at position 65 is positioned in close proximity to the amino nitrogen of the typically invariant G67. Nucleophilic attack from the former to the latter, followed by a dehydration step, forms an imidazoline heterocyclic ring system (cyclisation). Upon aerial oxidation of the other typically invariant residue Y66, the electron conjugation of the imidazoline ring system is extended to also include this aromatic substituent, resulting in the subsequent production of fluorescence (Figure 3.15 Tsien, 1998; Reid and Flynn, 1997; Prasher et al., 1992; Heim et al., 1994).

Whilst seemingly invariant for the GFPs, mutagenesis studies on the three residues involved in the chromophore (and its intramolecular interactions with nearby residues) has been shown to modify the excitation/emission maxima of protein fluorescence, resulting in the development of FPs that display a diverse spectral range (Figure 3.16). For example, introducing the Y66W and Y66H aromatic substitutions into *av*GFP shifts both excitation and emission towards shorter wavelengths, producing cyan (CFP, 436/485 nm) and blue (BFP, known as P4, 382/448 nm) fluorescent proteins respectively. This change is likely a result of the differences in electron availability between the azole and phenolic side chains (Heim et al., 1994; Tsien, 1998; Cubitt et al., 1998). As expected, mutations that do not contribute an aromatic group to the chromophore, such as Y66L, leads to non-fluorescent proteins. Though such deficiencies could be recovered with the

[8]Specifically, *av*GFP has a bimodal absorption band with excitation/emission maxima (λEx/λEm) at 395/508 nm (major peak) and at 475/503 nm (minor peak) (as reviewed by Pédelacq et al., 2006; Day and Davidson, 2009; Chudakov et al., 2010; Cubitt et al., 1998; Campbell and Davidson, 2010).

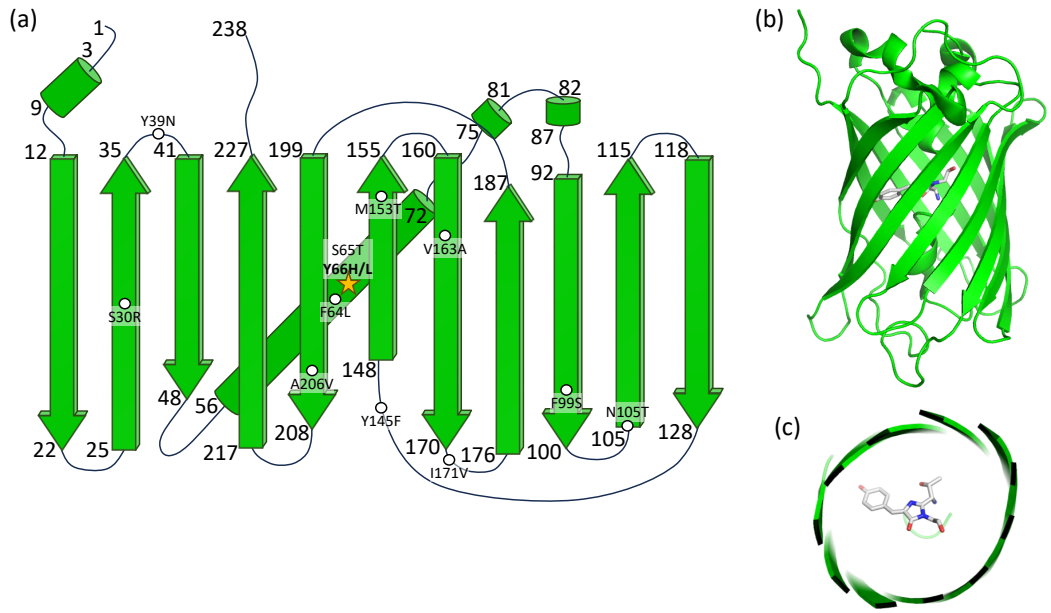


Figure 3.14: A schematic and crystal structure of sfGFP

(a) A topology map of sfGFP, adapted from Pédelacq et al. (2006); (b) the generalised structure of GFP (PDB: 1EMA), including (c) a view of the chromophore.

provision of an alternate aromatic compound, such as in the case of enhanced yellow fluorescent protein (EYFP), with the support of the phenol side chain from Y203 to the G65-Y66-G67 chromophore. In Section 3.2.2.2, I discuss both BFPs and non-fluorescent variants in greater depth as the critical Y66H/L changes within the chromophore makes these possible candidates for use as a recoded reassignment reporter in the OCI system.

In addition to establishing a wide palette of FPs, these mutagenic changes also affect protein folding and maturation efficiency of the chromophore. Key attributes that may be altered by these modifications include: protein (re-)folding rates, photostability, how well the molecule emits light in terms of the proportion of photons emitted to absorbed (i.e. its quantum yield, QY or ϕ), the quantity of light absorbed at a given wavelength by a fluorophore or chromophore of a given concentration (i.e. its molar extinction coefficient, EC or ε), and ultimately the molecular brightness of fluorescent expression as this is a product of QY and EC (Cubitt et al., 1998; Wall et al., 2014; Tsien, 1998). For example, the major peak for *av*GFP has ε = 25,000~30,000 M^{-1} cm^{-1} and ϕ = 0.79, whereas its

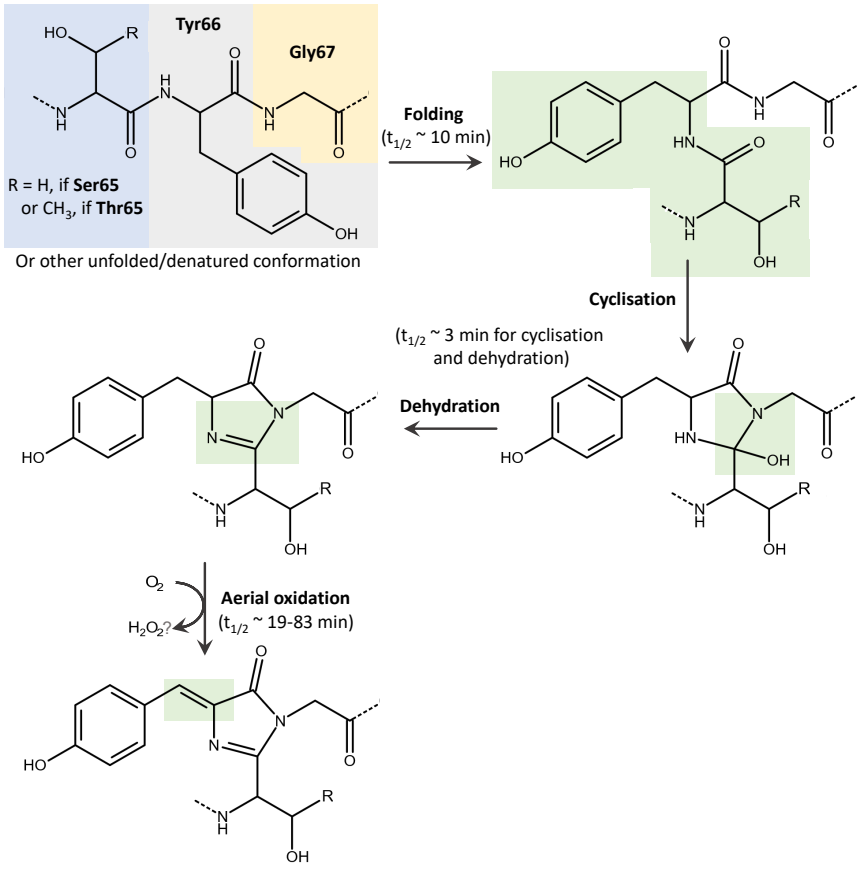


Figure 3.15: Intramolecular biosynthesis of GFP chromophore

Intramolecular biosynthesis of GFP chromophore as determined by Tsien (1998) and Reid and Flynn (1997), and as first isolated by Prasher et al. (1992), and characterised by Heim and colleagues (Heim et al., 1994; Heim and Tsien, 1996).

blue-fluorescing derivative (P4, i.e. *av*GFP.Y66H, has ε = 13,500 M^{-1} cm^{-1} and ϕ = 0.21, resulting in a seven-to-eight-fold reduction in brightness from just a single residue substitution (Patterson et al., 1997; Tsien, 1998; Heim and Tsien, 1996; Lambert, 2019, `https://www.fpbase.org/`, last accessed 01-Nov-2023). Consequently, commonly-used FPs will often have undergone various iterations of point mutagenesis before the desired characteristics are achieved.

The following work, however, focuses more on two specific FP variants: sfGFP and the EBFP family (Table 3.3). sfGFP is one of the brightest and most robust GFP derivatives. It was screened from a DNA shuffling library of folding reporter GFP (frGFP), which combines the F64L/S65T substitutions found in enhanced GFP (EGFP) (Cormack et al., 1996), the "cycle-3" (F99S/M153T/V163A) triple mutations from αGFP (Crameri et al., 1996), along with six further substitutions (S30R,

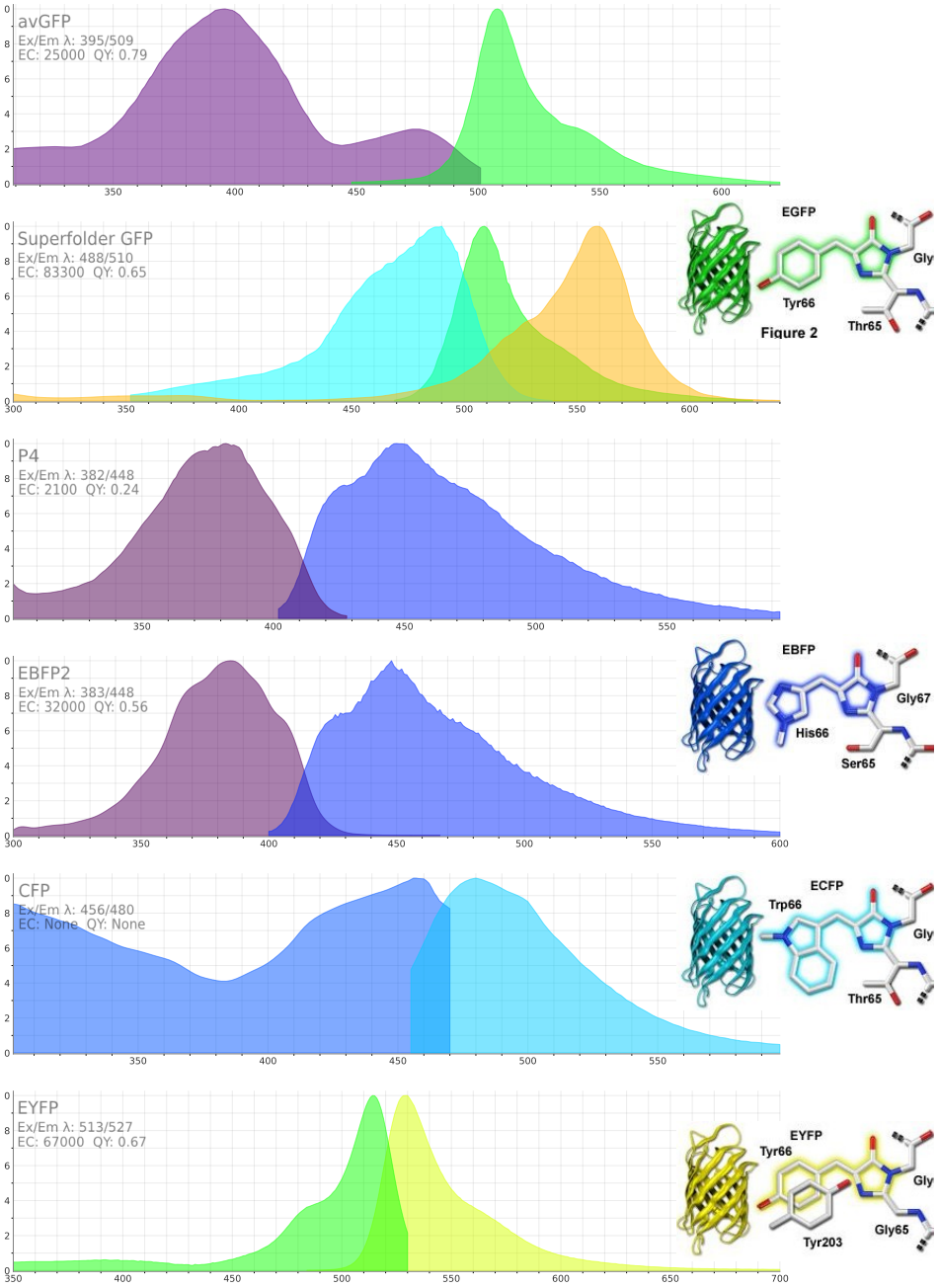


Figure 3.16: Spectral images of various fluorescent proteins with their excitation and emission peaks

Fluorescent proteins and their excitation/emission spectral images as acquired from `https://www.fpbase.org/` (Lambert, 2019). From top to bottom: avGFP (395/509 nm), sfGFP (488/510 nm), P4 (avGFP.Y66H; 382/448 nm), EBFP2 (383/448 nm), CFP (456/480 nm), EYFP (513/527 nm).

Y39N, N105T, Y145F, I171V, and A206V; Waldo et al., 1999; Pédelacq et al., 2006). The result of these modifications enables sfGFP to mature extremely quickly (with a refolding initial rate of 5.0E-1 s^{-1}, and maturation time of 13.6 min), have

increased thermo- and photo-stability (with a half-life of $t_{1/2}$ = 208.26 s; Cranfill et al., 1996), potentially reduce protein aggregation at high concentrations, as well as raise its EC (83,300 M^{-1} cm^{-1}) and overall brightness (54.15 $\times 10^3$, i.e. two-to-three fold brighter than *av*GFP; Pédelacq et al., 2006). All these properties combined make sfGFP an extremely effective reporter for developing a viable sense-to-sense codon reassignment detection system.

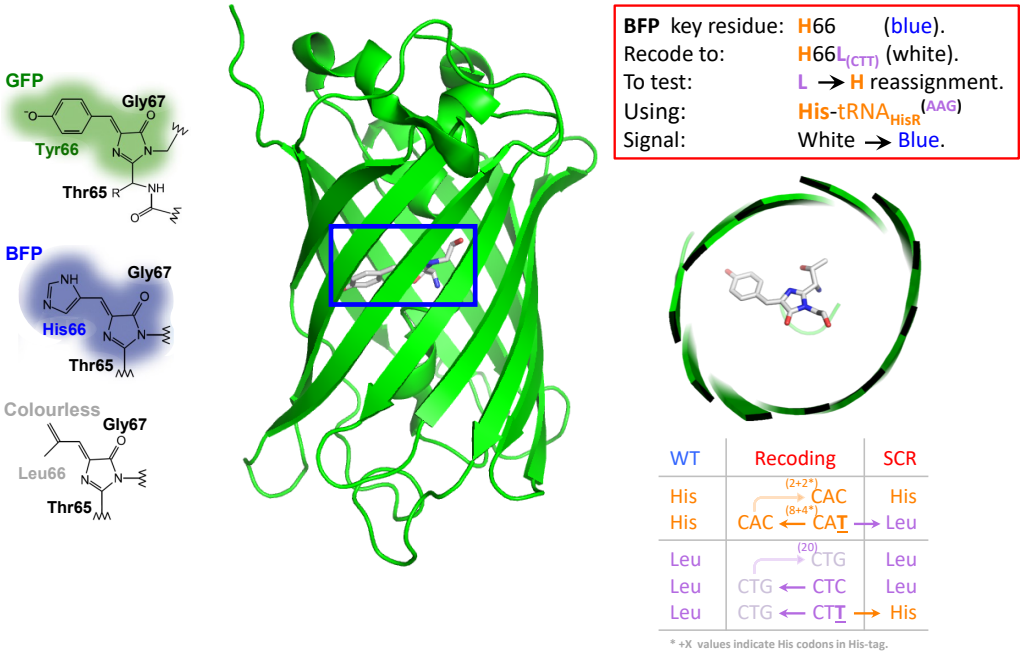


Figure 3.17: Proposals for adapting sfGFP as a recoded reporter for assaying leucine-to-histidine reassignment

Crystal structure of sfGFP with the chromophore highlighted (blue box). The chromophore molecules on the left show the residue changes in order to present as green, blue, and colourless. The key and summary table on the right explains the reassignment characterisation assay and the necessary histidine/leucine codons that were needed to be recoded in order to create the reporter for characterisation.

FP [PDB ID]	**Mutations**	**λEx/λEm, nm**	**EC, M^{-1} cm^{-1}**	**QY**	**Brightness $\times 10^3$ (wrt sfGFP)**	**Ref**
*av*GFP [1EMA]	(Chromophore: S65 / *Y66* / G67)	395/508 (major) 475/503 (minor)	25,000~30,000	0.79	19.75~23.70 (0.36~0.44)	
sfGFP [2B3P]	*av*GFP + M1_S2insV / (F64L / S65T) / (*Y66* / Y145F) / (F99S / M153T / V163A) / (S30R / Y39N / N105T / I171V / A206V)	485/510	83,300	0.65	54.15 (1.00)	
P4	*av*GFP + *Y66H*	382/448	2,100	0.24	0.50 (0.01)	
BFP [1BFP]	*av*GFP + (*Y66H* / Y145F)	381/445	14,000	0.38	5.32 (0.10)	
EBFP	*av*GFP + M1_S2insV / (F64L / S65T) / (*Y66H* / Y145F)	380/440	31,500	0.20	6.30 (0.12)	
EBFP1.1*	EBFP + (F99S / M153T / V163A) / (S30R / Y39N / N105T / I171V / A206V) OR sfGFP + **Y66H*	NR	NR	NR	NR (NR)	
EBFP1.2	EBFP + (**T65S** / S72A / N198S) / (S30R / Y39N / N105T / I171V / A206V)	379/446	41,000	0.45	18.45 (0.34)	
EBFP2	EBFP1.2 + (I128V / **V150I** / D155V / **V224R**)	383/448	32,000	0.56	17.92 (0.33)	
EBFP2.1*	EBFP2 + *(F99S / M153T / V163A)	NR	NR	NR	NR (NR)	
noFP1.1*	sfGFP + **Y66L*	NR	NR	NR	NR (NR)	
noFP2.1*	EBFP2.1* + **Y66L*	NR	NR	NR	NR (NR)	

Table 3.3: GFP derivatives and the point mutations required for GFP, sfGFP, EBFP, EBFP1.1, EBFP1.2; EBFP2.

The entries highlighted in green, blue, and grey were used or created (*) for use in this research, reflecting the colours emitted by the fluorescent proteins respectively (with the exception of grey, which represents no fluorescence). The residue/mutation within the chromophore that is responsible for the change in the colour fluorescence is also represented in their respective font colour. All other modified residues are listed, with key residues in italic, and those associated with reducing photobleaching are in bold (`https://www.fpbase.org/`, last accessed 01-Nov-2023; Lambert, 2019). NR = not reported.

By comparison, BFP reporters are a little less robust. The most modified, moderated[9] variant in Table 3.3, EBFP2, is only a third as bright as sfGFP. It takes about twice the time as sfGFP to mature (25.0 min), and photobleaches approximately four times faster (photo-stability half-life of $t_{1/2}$ = 55.0 s; Subach et al., 2008). Moreover, it requires ultraviolet (UV, approximately 380 nm) excitation, yet emits within the visible spectrum range (440~470 nm). The advantage of a BFP reporter for the OCI system, however, is in the critical histidine residue at position 66 in the chromophore. The significant change in properties that arise from single point mutations – that is, Y66H is the only necessary mutation to shift a GFP towards blue fluorescence, and the substitution of a non-aromatic residue, like leucine, into this position should result in noFP. This, therefore, makes sfGFP a suitable reporter for codon invasion (Figure 3.17). Hence, to investigate the potential of the Y66H/L reporter system, I decided to generate a series of BFP and noFP constructs to explore the limits and stability of this detection system.

3.2.2.2 Recoding fluorescent reporters: from green to blue to colourless

The first necessary step for ascertaining a BFP variant as a possible recoded reporter for the OCI system was to characterise the positive and negative controls for reassignment. The *sfgfp* gene within plasmid [203] (an in-lab/house plasmid, made from the [P003] `pET29a(+)` vector by post-doctoral researcher, Dr. Chris Cozens; Table 3.3) was used as the foundational sequence with which to introduce the Y66H and Y66L mutations, and make *ebfp1.1* (positive) and *nofp1.1* (negative) respectively. Inverse PCR with Q5 HiFi HS DNAP, using the site-directed mutagenesis (SDM) primers listed in Table 2.3, was used to generate all possible histidine and leucine codon substitutions at position 66 in *sfgfp*. Though no significant differences were expected between the synonymous substitutions, it seemed prudent to generate all variants in case of substantial codon usage preferences by the host.

As the *sfgfp* variants were each under control of a pT7, the T7 Ex-

[9] As published on FPbase (`https://www.fpbase.org/`, last accessed 01-Nov-2023; Lambert, 2019), the free and open-source, community-editable database for fluorescent proteins and their properties.

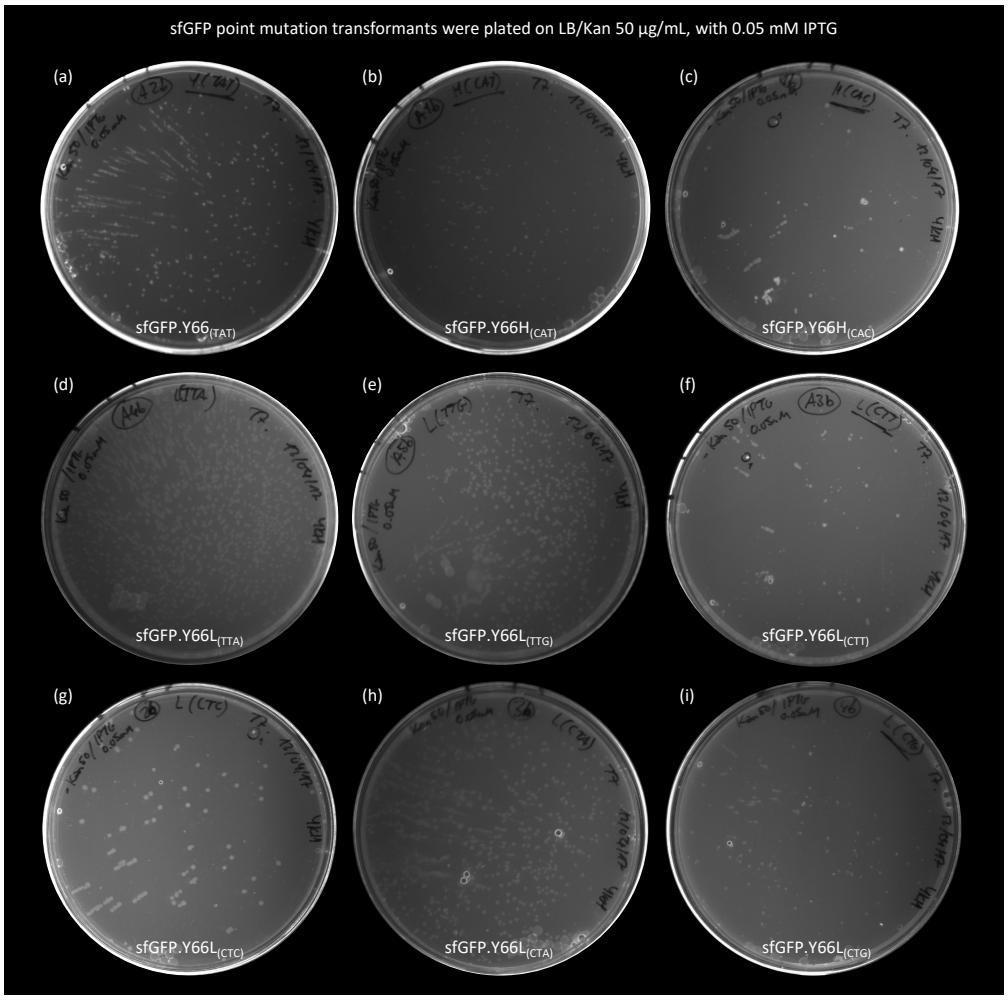


Figure 3.18: *E. coli* transformants of sfGFP point mutations that were used to create Y66H (EBFP1.1) and Y66L (noFP) variants

The complete set of recoded histidine (Y66H) and leucine (Y66L) codon variants for sfGFP were created and the transformants plated onto 50 μg/mL kanamycin in LB media plates that were supplemented with 0.05 mM IPTG in an attempt to screen for fluorescence/lack of fluorescence. The transformants into T7 Express *lysY/I^q* were visualised under UV light.
Plate (a) had the original unmodified sfGFP variant, which was visibly expressed green fluorescence (albeit not as clearly visualised here). Of note were plates (c) sfGFP.Y66His$_{(CAC)}$ (Watson-Crick complement, as opposed to the histidine wobble variant in (b)), (f) wobble variant sfGFP.Y66Leu$_{(CT\underline{\mathbf{T}})}$, and (g) Watson-Crick sfGFP.Y66Leu$_{(CTC)}$ variant (that is charged by the same tRNA isoacceptor as the one to Leu$_{(CT\underline{\mathbf{T}})}$ in (f)), which produced the lowest number of CFUs and also required an extra 24 h in order for any colonies to form; these plates were imaged a day after all other plates. No fluorescence in the recoded variants could be observed that were significantly different from one another; additional explanation in main text.

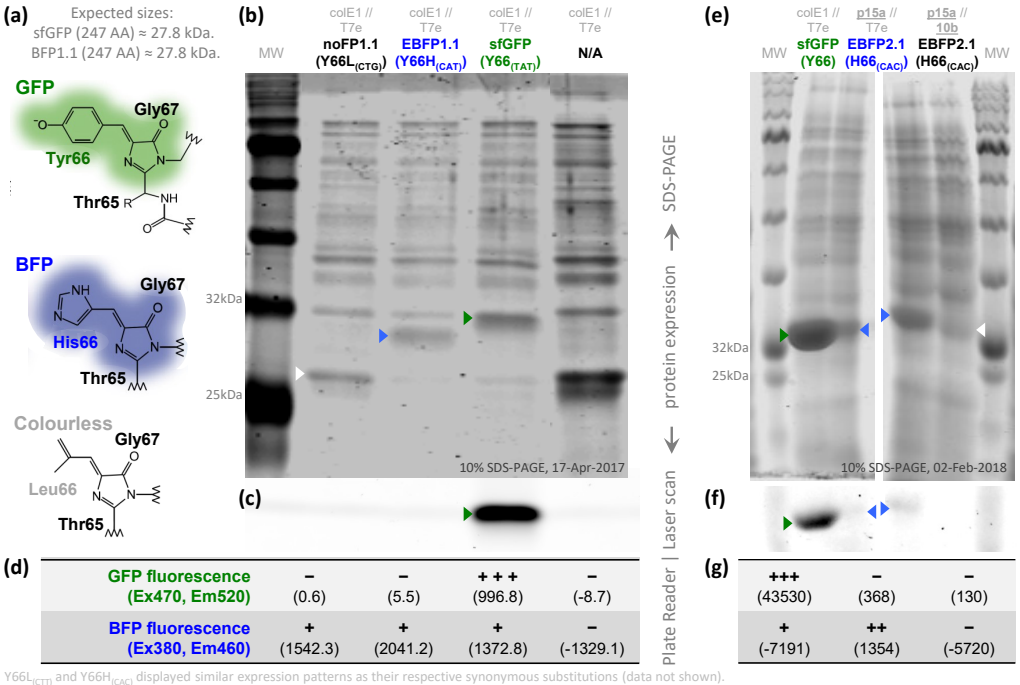
press *lysY/I^q* transformants were plated on 50 μg/mL kanamycin in LB media (LB Kan 50 μg/mL) agar plates, with and without 0.05 mM isopropyl-b β-D-thiogalactopyanoside (0.05 mM IPTG). The former were used as an initial check for fluorescence (Figure 3.18) , however, it was not possible to detect a discernible difference between blue and no fluorescence expression unaided (data not shown).

It should be noted, however, that the plates may not have been excited at the optimal wavelength (nor for long enough), even accounting for the fact that the two *ebfp1.1* variants have not previously been characterised before as BFPs. While the Bio-Rad GelDoc machine had a trans-UV excitation source, it was later understood that the plates were likely excited at 302 nm (default UV illumination provided as standard), and did not have the optional 365 nm UV lamp installed, which would have been closer to the expected ~380/440 nm peaks; excitation at ~300 nm would be at the tail-end of the expected range for a BFP (Figure 3.16). Although unlikely to be significant, an additional comment should be added that three variants (sfGFP.Y66His$_{(CAC)}$, sfGFP.Y66Leu$_{(CT\underline{T})}$, sfGFP.Y66Leu$_{(CTC)}$; Figure 3.18c,f,g) displayed slower growth, with fewer colony forming units (CFUs), and were imaged the following day. Otherwise, there did not seem to be any other noticeable differences. Hence, for the experiments moving forward, variant sfGFP.Y66His$_{(CA\underline{T})}$ (construct [204]) became known as EBFP1.1, whilst Y66Leu$_{(CTG)}$ (construct [205]) was termed noFP1.1.

As the initial visual inspection between heterologously-expressed EBFP1.1 and noFP1.1 proved to be inconclusive, a small-scale culture expression experiment was carried out on the different sfGFP recoded variants that were transformed into T7 Express *lysY/I^q*. In a 96-well plate, a 5% seed of each sample was grown in 200 μL 2xTY Kan 50 μg/mL media, and were induced with 0.5 mM IPTG (10x from prior plate-based assays) at 37 ºC for up to 5.5 h. Fluorescence (using the BMG LabTech FLUOstar OPTIMA microplate reader with filter sets at 470/520 nm for green and 380/460 nm for blue) and OD600nm readings were taken every hour from induction. Final fluorescent measurements were background subtracted (to take into account any auto-fluorescence) and normalised by OD600nm (to account

for any differences in experimental growth; Figure 3.19d). There seemed to be weak blue fluorescence detected in EBFP1.1 (2,041.2 a.u.), but it was only marginally higher than that of noFP1.1 (1,542.3 a.u.) and sfGFP (1,372.8 a.u.). As for green fluorescence, the sfGFP was the only one that showed observable expression (visually and by plate reader quantification) (996.8 a.u., compared to 5.5 a.u. for EBFP1.1). Fluorescence from the sfGFP sample was also confirmed using a GE Healthcare Life Sciences Typhoon FLA9500 biomolecular imager (filter set 473/510LP nm, excited with a blue LD laser, with an LPB emission filter; Figure 3.19c); no filter sets were available for detecting blue fluorescence.

A last attempt to visualise these variants *in vivo*, using fluorescence microscopy as kindly undertaken by Dr. Andrew Osborne, also failed to demonstrate any blue fluorescence (data not shown).



Y66L(CTT) and Y66H(CAC) displayed similar expression patterns as their respective synonymous substitutions (data not shown).
Negative Ex/Em values are due to subtraction of raw values from the fluorescence of the blank (2*TY media).

Figure 3.19: Protein expression of sfGFP point mutations, Y66H (EBFP1.1) and Y66L (noFP1.1), and EBFP2.1.

Selected Y66H and Y66L recoded variants were cultured and the proteins expressed before being visualised on SDS-polyacrylamide gel electrophoresis (SDS-PAGE). This appeared to produce more tangible differences, both in terms of slight shifts in sizes of the proteins expressed (b,e), and that seen when excited at 470 nm under both the Typhoon FLA 9500 scanner (c,f), and from end-point scans with FLUOstar OPTIMA plate reader (d,g).

Despite issues in detecting *in vivo* visual expression of the fluorescent pro-

teins, it was possible to observe a shift in soluble protein migration patterns between the three variants after protein extraction (Figure 3.19b). EBFP1.1 presented as a slightly smaller protein (~30 kDa, lane 3) than sfGFP (~32 kDa, lane 4), while the noFP1.1 variant was presumed to have migrated around the ~27 kDa mark (lane 2); though to some degree, the latter's migration pattern mirrored that of the negative control (the same `pET29a(+)` background vector, lane 5). Whilst unexpected, it should be noted that both EBFP1.1 and sfGFP migrated to a point that was larger than their predicted molecular weight (of ~27.8 kDa, including its 6xHis tag; observed also by Liu et al., 2019). Overall, although this *in vitro* approach was more definitive for differentiating between the three phenotypes, it required multiple additional steps where errors could be introduced or propagated, and was not a fast and readily-visual method of detection.

Given the lack of discernible fluorescence in both the EBFP1.1 and noFP1.1, it was unclear whether introducing the single Y66H point mutation in *sfgfp* to generate these were new variants would be sufficient in producing viable BFPs. Despite its less-modified relatives (e.g. EBFP) having, at best, approximately 12% brightness levels to sfGFP, it had been anticipated that it would at least reach an equivalent degree of brightness given that it had been designed with a set of mutations that had made sfGFP both a fast-folding and extremely bright protein (Table 3.3). Hence, I had optimistically made the assumptions that the resulting EBFP1.1 would have λEx/λEm wavelengths in the same range as the rest of the BFP family, and that its brightness would be within detectable limits. Figure 3.19b-d suggested that this was not the case.

In an attempt to overcome these unknown variables and make use of this key Y66H mutational property found in the *av*GFP/BFP family, it was determined that a better assessment would be to use an established variant, like EBFP2. This version has been shown to be at least a third in brightness relative to sfGFP, and with mutations (V150I and V224R) that reduced its photobleaching rates (full differences detailed in Table 3.3; Ai et al., 2007). The subsequent 1,021 bp [025] sequence that was sent for synthesis by Twist Biosciences was a codon-optimised iteration

of EBFP2, suitable for use in reassignment (i.e. all leucines recoded to $Leu_{(CTG)}$, and histidines to $His_{(CAC)}$), and designed with the same operon elements as the earlier [204] EBFP1.1 plasmid (i.e. with pT7, lacO, RBS, T7 terminator, and flanking termini homologous to a cloning site in `pET29a(+)`).

In advance of creating the noFP2.1 variant (which would be from a H66L substitution in EBFP2), the [026] EBFP2 plasmid was simply transformed into T7 Express *lysY/I^q* as the test-case for blue fluorescence, and into DH10β as the negative control. This latter strain has no T7 RNA polymerase to induce EBFP2 expression, so could be treated as the negative (no fluorescence) condition; the lack of blue fluorescence in this control was also confirmed in an earlier small-scale induction assay (data not shown). When expression was induced for 3 h and measured on a plate reader, there appeared to be blue (and possibly some green) fluorescence detected in the EBFP2 transformed into T7 Express *lysY/I^q*, but not in DH10β (Figure 3.19g). Cross-checking this with a 473/510LP nm laser scan of the soluble fraction of proteins that were run on an SDS-PAGE gel also indicated a shadow of fluorescence (presumably weak green emission) in the T7 expressed EBFP2 (Figure 3.19e-f, lanes 3 and 4, technical replicates), and again no detectable fluorescence in the DH10β negative control (lane 5). Whilst this scan had primarily been to reproduce the same green fluorescence presence/absence experiments as had been carried out with the earlier EBFP1.1 variant, after seeing expression in the plate reader it had seemed worth using this approach as additional confirmation for the T7 expressed EBFP2. It should be noted that this was likely green fluorescence, not blue, as the λEx/λEm spectra for EBFP2 suggests that the blue excitation range does not exceed 440 nm (Figure 3.16; Lambert, 2019; Ai et al., 2007).

It is probable that this weak green emission from the EBFP2 could still have been exploited for use as a leucine-to-histidine recoded reporter[10]. Realistically, to achieve the level of fluorescence that could be quantified by a plate reader from

[10]If, for example, instead of p15A, the EBFP2 was in a higher-copy colE1 plasmid, producing greater amounts of protein that could be detected above a no-fluorescence signal, ideally from *E. coli* in solid agar media. Had it been deemed more fruitful to continue optimising this reporter system, subcloning this variant into the colE1-containing `pET29a(+)` vector context (including the addition of LacI repression control) would have been one of the next steps; as would have been the creation of the noFP2.1 variant.

liquid cultures alone would have taken at least 18 h (data not shown), assuming also that there remained sufficient isopropyl-b β-D-thiogalactopyanoside (IPTG) induction to limit the loss of fluorescence from photobleaching. Moreover, the proteins had been assessed in controlled and optimised conditions thus far; if implemented with the sense-to-sense codon reassignment reporter assay as planned, which is expected to be toxic to the host, the system would likely fail to provide any indication of fluorescence and, by extension, reassignment. Given the slow, multi-step approach for quantifying fluorescence, the uncertainty over the photostability of the EBFP variants, and the lack of robustness in the assay, this fluorescence reporter was deemed impractical as a reassignment detection system.

3.2.3 Designing periplasmic secretors for histidine-to-leucine reassignment

3.2.3.1 Using periplasmic secretor fusions to re-distribute gene fragment cargoes

Bacterial periplasmic secretors was the next proposed system for establishing a recoded reporter for sense-to-sense codon reassignment. Signal peptide sequence OmpA and SufI of the Sec- and Tat- secretion systems respectively (Natale et al., 2008) are rich in leucine residues, which if recoded to histidine should prevent translocation of these signal sequences, and any associated cargo, across the transmembrane. By coupling the inactive, recoded secretors to the α-peptide fragment (lacZα) of the *lacZ* gene (β-galactosidase), the blue-white screening system could be used to detect reassignment.

The coupling of a signal sequence to LacZ was inspired by similar work by Beckwith and colleagues (Michaelis et al., 1983), where they were characterising the export of mutant alkaline phosphatase (*phoA*) signal sequences through phoA-lacZ fusions. The general principle of α-complementation of the *lacZ* gene is that the activity of a non-functional deletion mutant of LacZ (the ω-peptide; deletion of residues 11-41 of the N-terminus) can be rescued by a fragment bearing the deleted sequence (the α-peptide; formed of 3-90 residues of LacZ) (Ullmann et al.,

1967; Langley et al., 1975; Langley and Zabin, 1976). Typically for screening, the *lacZ*α sequence (encoding the first 59 residues) is carried on a plasmid (e.g. pUC19 and pBluescript) and this complements the ω-peptide expressed from a *lacZ*Δ*M15* *E. coli* strain (e.g. DH5α, JM109, XL1-Blue). X-gal, a colourless analogue of lactose, is used to detect the α-complementation and subsequent LacZ activity: the latter cleaves X-gal to form 5-bromo-4-chloro-indoxyl, which dimerises and oxidises to form 5,5'-dibromo-4,4'-dichloro-indigo, a bright blue insoluble pigment.

The *lac* operon (consisting of a P_{lac} promoter, *lacO* operator, *lacZYA* genes, and a terminator) is typically used in *E. coli* for the transport (*lacY*, β-galactoside permease) and metabolism (*lacZ*, β-galactosidase) of lactose. Whilst glucose is the preferred carbon source for most bacteria, when it is not available, LacZ cleaves lactose into the preferred glucose energy source and by-product galactose monosaccharide.

Due to the interacting elements surrounding the *lac* operon, various considerations need to be accounted for when using the LacZ α-complementation (blue-white) screening system. Mainly, media for growing the complementation assay strains should not include glucose, but IPTG can be added to enhance LacZ expression. Additionally, due to light sensitivity, agar plates containing X-gal should be stored in the dark.

An alternative design may be to use split-GFP (or other split fluorescent proteins) (Knapp et al., 2017; Romei and Boxer, 2019). Like LacZ, when GFP is split into GFP_{1-10} and GFP_{11}, and one of the fragments is removed, the activity is abolished. However, when the two parts are brought into close proximity again, activity can be restored. With this system, it would not be essential to use a *lacZ*Δ*M15* strain, nor the need to make specific media without glucose.

Like with the BFP/noFP reporters, it should be possible to test both leucine-to-histidine and histidine-to-leucine reassignment using the periplasmic reporters. However, if using LacZα as the cargo for detection, it may be more practical to assess the former reassignment, and visualise reassignment by the colour of CFUs on X-gal-containing agar plates. This assay would mean using the leucine-rich

(recoded with synonymous $\text{Leu}_{(\text{CTT})}$ codons for critical residues) signal sequences as wild-type, which should result in the export of the LacZα fragment out of the cytoplasm, thus the formation of white CFUs. Upon transcription of the anticodon-modified *Eco*$\text{tRNA}^{\text{His}}_{\underline{\mathbf{A}}\text{AG}}$, some leucine residues in the signal sequences should be reassigned to histidine, thus abolishing the ability to transport LacZα out of the cell. This would allow α-complementation to take place and the formation of some blue-pigmented colonies.

With leucine-to-histidine reassignment, we should see a loss-of-function of the signal sequences (corresponding white to blue CFUs). The hypothesis is that even if there were some blue CFUs prior to reassignment, the addition of the *Eco*$\text{tRNA}^{\text{His}}_{\underline{\mathbf{A}}\text{AG}}$ should produce a noticeable increase in blue CFU count. The concern with testing for histidine-to-leucine reassignment is that the rate at which the signal sequence is rescued (and the resulting export of LacZα from the cytoplasm) may not be fast enough, so that there would not be a significant increase in white CFUs over the pre-reassignment blue colonies. However, it is theoretically possible to test both reassignments with this system.

Similarly, if the smaller GFP_{11} fragment is coupled to the signal sequences, and the larger GFP_{1-10} is in the cytoplasm, it is still more prudent to assess leucine-to-histidine reassignment. The wild-type signal sequence would export GFP_{11} out to the periplasm, resulting in little to no GFP-expressing CFUs. Upon leucine-to-histidine reassignment, some signal sequences should lose activity and retain the smaller GFP fragment in the cytoplasm with the GFP_{1-10} part, which would give some GFP-expressing colonies. Determining no fluorescence from a background of GFP-expressing colonies seems trickier to assess but, if necessary, possible to attempt.

3.2.3.2 Designing and constructing signal sequences OmpA (Sec) and SufI (Tat) with LacZa

For the OmpA_LacZα reporter, two 438 bp sequences [027] and [027b] were ordered from, synthesised, and subcloned into [P011] pTwist-Kan vector by Twist Biosciences: one had the codon $\text{Leu}_{(\text{CTT})}$ at residue position 12 in the OmpA sig-

nal sequence [028], the other (non-functional variant) had codon $His_{(CAT)}$ [028b] (Figure 3.20).

The design of the SufI_LacZα reporter was similar to OmpA_LacZα, the only difference being the 81 bp SufI signal sequence replacing the 66 bp OmpA sequence (Figure 3.21). A 205 bp SufI_LacZα truncated fragment was ordered as a GBlock from IDT ([206] aka YKH388), which was to be used as is it was.

However, problems with the construction and transformation into a suitable *E. coli* strain (data not shown) meant that this reporter system was also discontinued.

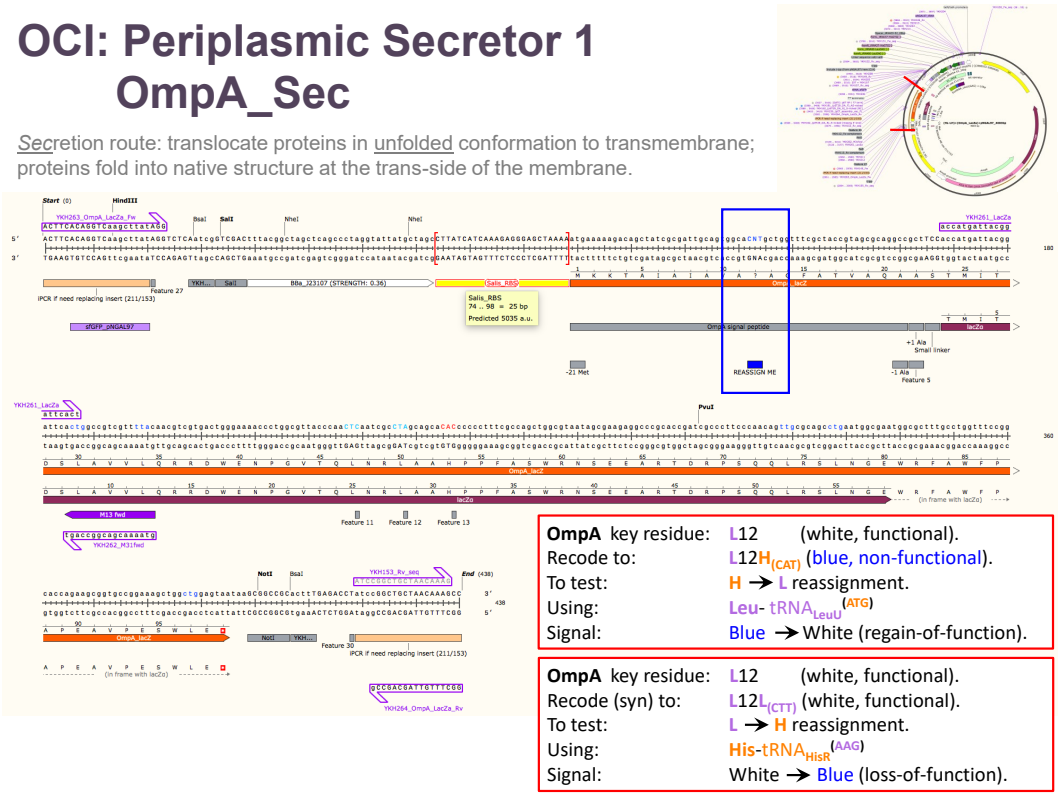


Figure 3.20: The OmpA_Sec periplasmic secretor recoded reporter design constructs [027] and [027b]

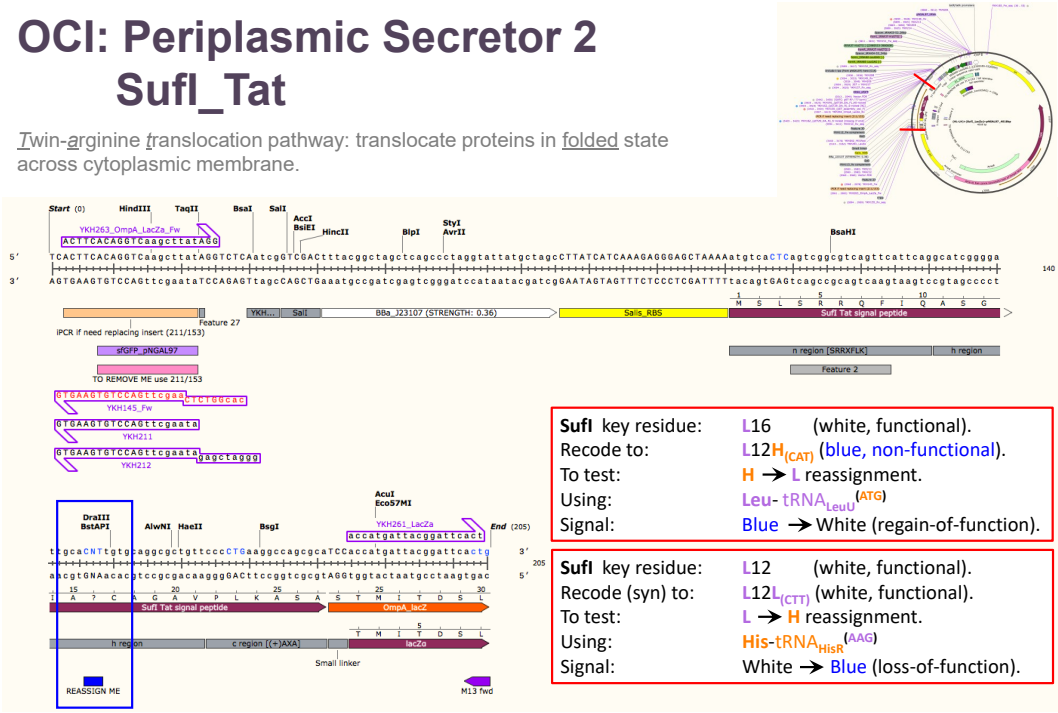


Figure 3.21: The SufI_Tat periplasmic secretor recoded reporter design construct [206]

3.2.4 Constructing and characterising chloramphenicol acetyl-transferase as an OCI reporter for leucine-to-histidine reassignment

3.2.4.1 Retaining critical chloramphenicol acetyl-transferase activity for recoding

CAT was the final OCI reporter that had been identified as a potential candidate for use in the reassignment assay. It was developed in parallel with the earlier reporter systems and, given the issues in creating suitable recoded reporters from both the BFP derivatives (Section 3.2.2.2) and the periplasmic secretors (Section 3.2.3.2), it was essential to proceed with any gains following this approach. Unlike the previous two heterologous expression systems, selective pressure could be applied to the host to retain the recoded CAT construct (i.e. with the introduction of chloramphenicol). Under such environmental conditions, *cat* might be considered an essential gene for survival and, as demonstrated by Döring and Marlière (1998) with ThyA, *E. coli* could be coerced into non-natural (isoleucine-to-cysteine and methionine-to-cysteine) codon reassignments given the relevant tools (over-expressed anticodon-modified tRNAs) and impetus (absence of any other pathway to generate the necessary dTMP precursor for DNA biosynthesis) to do so.

One particular key to CAT activity, as identified by Kleanthous et al. (1985), was residue H193 (Leslie et al., 1988). Through a number of mutagenesis experiments, Lewendon et al. (1994, 1988) demonstrated that H193A/Q/D/Y substitutions completely inhibited CAT. The fact that such a range of substitutions were able to abolish CAT activity so effectively prompted the hypothesis that recoding the H193 residue to the non-reactive, aliphatic leucine (a substitution pair that has been predicted to be relatively disruptive; Figure 3.6) should also be equally likely to suppress its catalytic activity. A *cat*.H193L would, therefore, provide another opportunity to create a leucine-to-histidine reassignment reporter.

3.2.4.2 Designing and constructing an inactive CAT recoded reporter

Analysis of the *cat* gene from [P011] pSB1C3 revealed that the sequence had a total of 12 histidine codons (five $His_{(CAC)}$, seven $His_{(CA\underline{T})}$ including H193), and 13 leucines (six $Leu_{(CTG)}$, one $Leu_{(CT\underline{T})}$, one $Leu_{(CTC)}$, one $Leu_{(CTA)}$, three $Leu_{(TTA)}$). To create a viable reporter for reassignment, the seven non-$Leu_{(CTG)}$ codons would need to be recoded to $Leu_{(CTG)}$, and the $His_{(CA\underline{T})}$ codons to $His_{(CAC)}$. As for the critical H193, that would be recoded to $Leu_{(CT\underline{T})}$, to create the initial (theoretically) loss-of-function mutant – sensitive to chloramphenicol, Cam^S – for testing leucine-to-histidine reassignment.

An issue with creating chloramphenicol acetyl-transferases.H193$Leu_{(CT\underline{T})}$ in pSB1C3 is that another selection marker (encoding for a different antibiotic resistance) would be needed to recover the inactive chloramphenicol acetyl-transferases gene. As it was, colleague Dr. Chris Cozens, had previously modified the pSB1C3 to include an additional *bla* gene (conferring ampicillin- and carbenicillin resistance), resulting with plasmid [207] pSB1C3A2. As a demonstration of the Darwin Assembly method (Section 2.2.18; Cozens et al., 2012), Chris mutated the seven non-$Leu_{(CTG)}$ codons in the *cat* gene to $Leu_{(CTG)}$, creating plasmid [208] pSB1C3A2_CTG.

To make this [208] compatible for leucine-to-histidine reassignment, however, I performed a further round of Darwin Assembly, changing six of the seven $His_{(CA\underline{T})}$ codons to $His_{(CAC)}$ and either leaving the seventh – the critical H193 residue – unchanged as $His_{(CA\underline{T})}$ to confirm it is still active despite the synonymous codon changes (plasmid [209]), or recoding it to $Leu_{(CT\underline{T})}$ to generate the necessary reporter for testing leucine-to-histidine reassignment (plasmid [209b]). To create the former construct [209], the inner oligos used were YKH269, YKH270, YKH271, YKH272, and YKH273; for the latter [209b], YKH269, YKH270, YKH271, YKH278, and YKH279 were used.

Modified plasmids were transformed into DH10β strains, and plated on LB Amp (100 µg/mL) to recover successful transformants. Individual colonies were

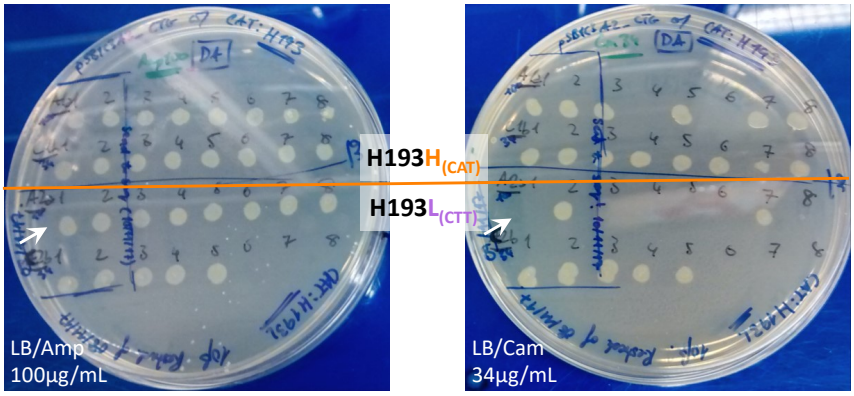
picked and grown in LB Chl (34 µg/mL) or LB Amp (100 µg/mL) to confirm CAT sensitivity of the CAT.H193L$_{(CTT)}$ mutants; the former to verify the loss of Cam^{R}, the latter to again recover the inactive *cat*-bearing plasmids (Figure 3.22a). The recovered plasmids were purified and sent for Sanger sequencing (Figure 3.22).

3.2.5 Combining and characterising the CAT.H193L inactive, recoded reporter with anticodon-modified *Eco*tRNA$^{His}_{\underline{A}AG}$ and *Eco*tRNA$^{Leu}_{\underline{A}UG}$ to test reassignment and characterise effect on host

Having verified correct recoding of the Cam^{S} sense-to-sense codon reassignment variant CAT.H193L$_{(CTT)}$ and the Cam^{R} positive control CAT.H193L$_{(CTT)}$ in plasmid [207] `pSB1C3A2`, both constructs were each subcloned in place of the *sfGFP* reporter that was created in Section 3.2.2.2. These plasmids were subsequently transformed into DH10β strains and plated on LB Chl (34 µg/mL), with and without aTc to induce engineered tRNA transcription. Even without induction of the tRNAs, there was sufficient basal transcription to allow reassignment and recovery of the inactive CAT.H193L$_{(CTT)}$ (Figure 3.22).

As all the OCI reporters assessed thus far allow only the characterisation of leucine-to-histidine reassignment, realistically only the *Eco*tRNA$^{His}_{\underline{A}AG}$ anticodon-modified variant is necessary in the constructions. To demonstrate this is the case, the gene sequence for *Eco*tRNA$^{His}_{\underline{A}AG}$ and *Eco*tRNA$^{Leu}_{\underline{A}UG}$ were each independently deleted from the earlier plasmid constructions. As before, these plasmids were transformed into DH10β strains and plated on LB Chl (34 µg/mL). As shown in Figure 3.23, with the positive control CAT.H193 variant, the deletion of each tRNA does not impact the CAT activity, and *E. coli* growth does not appear impaired. For the sense-to-sense codon reassignment variant CAT.H193L$_{(CTT)}$, however, deletion of the essential *Eco*tRNA$^{His}_{\underline{A}AG}$ resulted in no growth, whereas there was CAT activity (albeit with hampered strain growth) with the deletion of the as-yet expendable *Eco*tRNA$^{Leu}_{\underline{A}UG}$ (but retention of the necessary *Eco*tRNA$^{His}_{\underline{A}AG}$).

Whilst it was reassuring to see successful leucine-to-histidine reassignment



CAT key residue: H193 (*Cam*^{R}; live).
Recode to: H193L$_{(CTT)}$ (*Cam*^{S}; dead).
To test: L → H reassignment.
Using: His-tRNA$_{HisR}^{(AAG)}$
Signal: *Cam*^{S} → *Cam*^{R}.

WT	Recoding	SCR
His	(5) CAC	His
His	CAC ← (6) CAT →	Leu
Leu	(6+7*) CTG	Leu
Leu	CTG ← CTC	Leu
Leu	CTG ← CTT →	His

* C.Cozens previously recoded 7 Leu codons (3 TTA, 2 CTT, 1 CTC, 1 CTA) to WT (CTG).

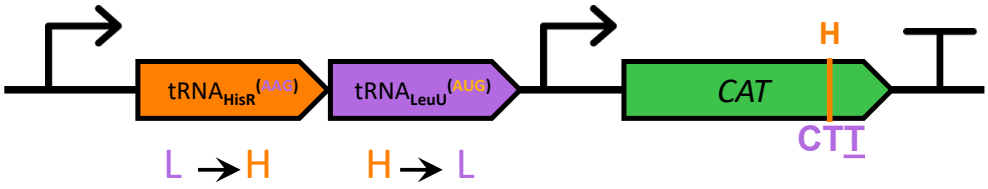
(a) Biological replicates of DH10β carrying plasmids containing either CAT.H193H or CAT.H193L, with rows 1 and 3 taken from transformants that were originally plated on 100 μg/mL ampicillin in LB media, and rows 2 and 4 from transformations previously plated on 34 μg/mL chloramphenicol in LB media

	Codon triplet sequence for:					pSB1C3A2_CTG_CAT				
Codon	WT	Optimised	Primers for DA SDMs	pSB1C3	pSB1C3A2_CTG	A1a1_ catH193_AC -275	C1b1_ catH193_AC -275	A2a1_ catH193L_AnoC -275	A2a2_ catH193L_AC -275	C2b1_ catH193L_AC -275
H17	caT	caC		CAT	CAT	CAC	CAC	CAC	CAT	CAC
H21	caT	caC	YK: YKH269_H17caC_H21caC	CAT	CAT	CAC	CAC	CAC	CAT	CAC
L39	ctg	CTG	-	ctg	ctg	ctg	ctg	ctg	ctg	ctg
L45	TtA	CtG	CC: CAT_L45CTG	tta	CTG	CTG	CTG	CTG	CTG	CTG
H53	cac	cac	-	cac	cac	cac	cac	cac	cac	cac
H61	cac	cac	-	cac	cac	cac	cac	cac	cac	cac
L63	ctT	ctG	CC: CAT_Leu63CTG	ctt	CTG	CTG	CTG	CTG	CTG	CTG
L66	ctg	ctg	-	ctg	ctg	ctg	ctg	ctg	ctg	ctg
H70	caT	caC	YK: YKH270_H70caC	CAT	CAT	CAC	CAC	CAC	CAT	CAC
L82	ctg	ctg	-	ctg	ctg	ctg	ctg	ctg	ctg	ctg
H89	cac	cac	-	cac	cac	cac	cac	cac	cac	cac
H96	caT	caC	YK: YKH271_H96caC	CAT	CAT	CAC	CAC	CAC	CAT	CAC
L105	ctC	ctG	CC: CAT_Leu105CTG // CAT_L105CTGv2	CTC	CTG	CTG	CTG	CTG	CTG	CTG
H110	cac	cac	-	cac	cac	cac	cac	cac	cac	cac
L117	ctA	ctG	CC: CAT_L117CTG	CTA	CTG	CTG	CTG	CTG	CTG	CTG
H118	cac	cac	-	cac	cac	cac	cac	cac	cac	cac
L131	ctg	ctg	-	ctg	ctg	ctg	ctg	ctg	ctg	ctg
L158	TtA	CtG	CC: CAT_L158CTG	TTA	CTG	CTG	CTG	CTG	CTG	CTG
L184	ctg	ctg	-	ctg	ctg	ctg	ctg	ctg	ctg	ctg
L187	ctg	ctg	-	ctg	ctg	ctg	ctg	ctg	ctg	ctg
H192	caT	caC	YK for L: YKH278	CAT	CAT	CAC	CAC	CAC	CAT	CAT
H193	**cAt**	**L(cTt) / H(cat)**	YK for H: YKH272	CAT	CAT	CAT	CAT	CTT	CAT	CAT
H200	caT	caC	YK for L: YKH279 YK for H: YKH273 (would have overlapped with YKH278, so needed to make YKH279 to prevent this).	CAT	CAT	CAC	CAC	CAC	CAT	CAT
L205	ctT	ctG		CTT	CTG	CTG	CTG	CTG	CTG	CTG
L208	TtA	CtG	CC: CAT_L205+8CTG [CAT_L205+8CTG___RC]	TTA	CTG	CTG	CTG	CTG	CTG	CTG
A147	gcc	gcc		gcc	gcc	acc	gcc	gcc	gcc	gcc

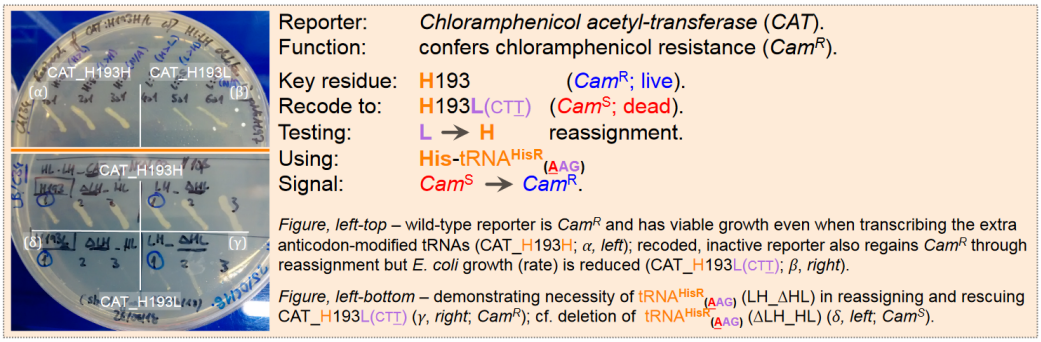
(b) Sanger sequencing table of selected colonies from Figure 3.22a, with the two columns in topped with yellow boxes as the correct Darwin Assemblies of the CAT.H193H and CAT.H193L constructs respectively

Figure 3.22: Screening for correctly Darwin Assembly recoded CAT.H193 and CAT.H193L reporters (constructs [209] and [209b] respectively)

	H → L Codon $H_{(CAT)}$ with $tRNA_{LeuU}^{(AUG)}$	L → H Codon $L_{(CTT)}$ with $tRNA_{HisR}^{(AAG)}$
H193H(CAT)	*Non-informative.* ***Live (WT, Cam^R)***. *If H>L, becomes Cam^S,* unlikely to be selected for.	*Non-informative.* ***Live (WT, Cam^R)***. *No $L_{(CTT)}$ codons present;* $tRNA_{HisR}^{(AAG)}$ unnecessary.
H193L(CTT)	***Dead (Cam^S)***. *No $H_{(CAT)}$ codons present;* $tRNA_{LeuU}^{(AUG)}$ unnecessary.	***Live (Cam^R)***. *Residue 193$L_{(CTT)}$ reassigned back to H by $tRNA_{HisR}^{(AAG)}$.*



(a) A summary table of CAT.H193/L(CTT), with both and each synthetic tRNA



(b) Results from deletion of either engineered $tRNA^{His}_{\underline{A}AG}$ (*hisR::leu**AA**G*) or $tRNA^{Leu}_{\underline{A}UG}$ (*leuU::his**A**UG*) to show that $tRNA^{His}_{\underline{A}AG}$ was both necessary and sufficient for recovery of the Cam^R phenotype

Figure 3.23: The full OCI reporter system with the CAT.H193L recoded reporter and the engineered tRNAs reassignment assay

by *Eco*tRNA$^{His}_{\underline{A}AG}$, it was clear that there was some level of toxicity to the host – endpoint visualisation of colony formation appeared to show less growth from those containing CAT.H193L compared to the wild-type. One hypothesis as to why CAT activity was recovered in the CAT.H193L$_{(CTT)}$ inactive variant is that the *cat* gene was too essential to the host when the selective media included chloramphenicol. The necessity to maintain the *cat* gene was such that even though engineered *Eco*tRNA$^{His}_{\underline{A}AG}$ and *Eco*tRNA$^{Leu}_{\underline{A}UG}$ were causing mis-translation in other genomic genes in the host, strains that retained these tRNAs (albeit as slow-growing and unhealthy specimens) would survive.

3.3 Conclusion to OCI

In summary, at this stage, I have explained the rationale for the choice and design of the anticodon-modified tRNAs, *Eco*tRNA$^{His}_{\underline{A}AG}$ and *Eco*tRNA$^{Leu}_{\underline{A}UG}$, to demonstrate leucine-to-histidine and histidine-to-leucine reassignment respectively. These tRNAs were designed to target orphan codons that do not have a dedicated tRNA isoacceptor that forms full Watson-Crick base-pairs between the triplet tRNA anticodon and the mRNA codon; instead orphans are served by tRNAs that have two nucleotides with Watson-Crick complementarity and one wobble base-pairing. As such, tRNAs with anticodons modified to have the full Watson-Crick complementarity to the triplet sequence of the orphan codons should be able to compete with the native tRNAs, and allow sense-to-sense codon reassignment. The anticodon-modified tRNAs were successfully constructed in two different plasmid backgrounds: pSB1C3 and pNGAL97. The orphan codons selected as targets for reassignment were Leu$_{(CTT)}$ and His$_{(CAC)}$.

Three reporter systems were designed and assessed for viability as robust, detectable reassignment in *E. coli*: fluorescent proteins, periplasmic secretor signal peptide sequences fused to a secondary reporter, and chloramphenicol resistance. To test reassignment activity of the engineered tRNAs *in vivo*, visually detectable reporters with critical histidine and/or leucine residues were identified. Active and

inactive controls of the reporters were constructed and characterised; inactive controls were created by recoding critical leucine codons to $His_{(CAT)}$ and essential histidine residues to $Leu_{(CTT)}$. The reporters were also optimised for reassignment by mutating non-essential leucine codons to $Leu_{(CTG)}$ and histidine codons to $His_{(CAC)}$; these codons theoretically would not be targets for reassignment. By mutating non-critical leucines and histidines to synonymous but non-reassign-able codons, the engineered tRNAs should only reassign the orphan wobble codons – corresponding to residues vital to reporter activity – that have been specifically marked for reassignment ($His_{(CAT)}$ and $Leu_{(CTT)}$).

The fluorescent reporters of choice were BFP derivatives, which had a single crucial residue at position 66: H66 produced blue fluorescence, whilst L66 abolished fluorescent activity. Theoretically, this reporter would have been ideal to demonstrate leucine-to-histidine reassignment, however, in practice it had too weak a signal and was too slow and cumbersome for detection. Fortunately, chloramphenicol acetyl-transferase, with essential residue H193, proved not only to have clear phenotypic distinction (Cam^R, live *E. coli*, as CAT.H193; Cam^S, dead *E. coli*, as $CAT.H193L_{(CTT)}$), but also fast-acting and long-lasting. The *cat* variants were successfully recoded, and subcloned into pSB1C3A2 and tRNAs(LH-HL)_CAT_pNGAL97; the latter of which was used to successfully demonstrate leucine-to-histidine reassignment.

One clear conclusion presented here is that it is possible to show leucine-to-histidine reassignment in *E. coli*. Anticodon-modified *Eco*$tRNA^{His}_{\underline{A}AG}$ was compatible with the native *Eco*HisRS to reassign and restore activity to the recoded $CAT.H193L_{(CTT)}$ reporter *in vivo*, which in turn proved robust enough to be easily detected.

In order to demonstrate reversible sense-to-sense reassignment and the possibility for the creation of alternative genetic codes, orthogonal to nature, the next step is to find a viable reporter for histidine-to-leucine reassignment *in vivo*.

A final issue that has not yet been addressed here is the effect of the engineered tRNAs on the host. As described in the beginning, there is no knock-out of the host

tRNAs, which means that native wobble-pairing *Eco*tRNA$^{\text{Leu}}_{\underline{\mathbf{G}}\text{AG}}$ (gene *leuU*) would still correctly assign leucine to orphan codon Leu$_{(\text{CTT})}$, whilst anticodon-modified *Eco*tRNA$^{\text{His}}_{\underline{\mathbf{A}}\text{AG}}$ will incorporate histidine in the same position. Likewise, native wobble-pairing *Eco*tRNA$^{\text{His}}_{\underline{\mathbf{G}}\text{UG}}$, (gene *hisR*) would correctly assign histidine to orphan codon His$_{(\text{CAT})}$, whilst anticodon-modified *Eco*tRNA$^{\text{Leu}}_{\underline{\mathbf{A}}\text{UG}}$, would add leucine. This level of mis-translation to genes that have not been recoded to accept alternative reassignments is detrimental to host health, as shown with the CAT.H193L$_{(\text{CTT})}$ reporter. Therefore, another objective to address now is to find a way to reduce the level of toxicity to the host, whilst being able to robustly detect the histidine-leucine reassignments.

The resolution to both the histidine-to-leucine reassignment reporter and mitigation of host toxicity are addressed in the following Chapter 4: Genetic Toggle Switch.

Chapter 4

GTS: Exploring Tolerable Alternative Codes

"The Genetic Toggle Switch System – an Approach to Limit the Toxicity Incurred to a Host that Possesses both the Natural and an Alternative Genetic Code"

4.1 Introduction to GTS

4.1.1 Potential Strategies to Deal with Reassignment-Related Toxicity to Host

The previous chapter (Chapter 3) introduced the idea of Orphan Codon Invasion (OCI), where anticodon-modified tRNAs, $tRNA^{His}_{\underline{A}AG}$ and $tRNA^{Leu}_{\underline{A}UG}$, were designed for use in various recoded reporter systems to implement leucine-to-histidine and histidine-to-leucine reassignment *in vivo* at orphan codons $Leu_{(CT\underline{T})}$ and $His_{(CA\underline{T})}$ respectively. Using a chloramphenicol acetyl-transferase gene that had been inactivated through recoding (*cat*.$H193L_{(CTT)}$), it was possible to reassign residue $H193L_{(CTT)}$ back to histidine, resulting in the recovery of CAT activity and rescue of the host.The reverse histidine-to-leucine reassignment, however, could not yet be demonstrated.

This chapter details the development of the Genetic Toggle Switch (GTS), a

complementary strategy to the OCI system previously described, with the goal of mitigating the toxicity caused by the co-expression of engineered tRNAs in an otherwise natural system. One ideal solution to resolving this host toxicity problem would be for wholesale recoding of the *E. coli* genome to accept the new codon assignments, which includes also the removal of all associated tRNA translators, whilst incorporating the anticodon-modified tRNAs into the system (likely via genomic integration, but feasible as heterologously-expressed plasmids). That is, recoding all $\text{Leu}_{(\text{CT}\underline{\textbf{T}})}$ and $\text{Leu}_{(\text{CTC})}$ codons to $\text{His}_{(\text{CA}\underline{\textbf{T}})}$ (as a new leucine-encoding site, complemented by anticodon-modified $\text{tRNA}^{\text{Leu}}_{\underline{\textbf{A}}\text{UG}}$), and all histidine codons $\text{His}_{(\text{CAY})}$ to $\text{Leu}_{(\text{CT}\underline{\textbf{T}})}$ (as the new histidine-encoding codon, with the engineered $\text{tRNA}^{\text{His}}_{\underline{\textbf{A}}\text{AG}}$ counterpart), as well as deleting all the formerly cognate/wild-type but now mis-translating tRNAs (i.e. *leuU*, *Eco*$\text{tRNA}^{\text{Leu}}_{\underline{\textbf{G}}\text{AG}}$; *hisR*, *Eco*$\text{tRNA}^{\text{His}}_{\underline{\textbf{G}}\text{UG}}$; Figure 4.1). This implementation would not be unlike the natural transitional process that evolved in *C. albicans* for accepting leucine-to-serine reassignment, except that the *C. albicans* genome had, over time, reduced its reliance on the leucine/serine-encoding $\text{Leu}_{(\text{CUG})}$ codons in order to diminish the adverse effects of ambiguous codon assignment (Santos et al., 1993, 1996). Nonetheless, for implementing my alternative histidine-leucine code, unless whole genome recoding could be established in advance in a second organism, there would still be toxicity problems with proteins being mis-translated in any intermediate host by the anticodon-modified tRNAs. However, as recently demonstrated by Chin and colleagues (Fredens et al., 2019), recoding is possible if executed via specific intermediates: they recoded an *E. coli* with a reduced genetic code, thus freeing the cell from its need for $\text{tRNA}^{\text{Ser}}_{\text{CGA}}$ (*serU*) and $\text{tRNA}^{\text{Ser}}_{\text{UGA}}$ (*serT*), and their respective $\text{Ser}_{(\text{TCG})}$ and $\text{Ser}_{(\text{TCA})}$ codons. They additionally removed release factor 1 (*prfA*), thereby freeing stop codon TAG as well. It was then possible to re-introduce the removed codons with engineered tRNAs, thus enabling a codon-reassigned organism to be made. At the time that my research was carried out (starting from 2014), these genome-recoded strains had not yet been created, and the large-scale DNA synthesis and host engineering tools (Wang et al., 2016; Robertson et al., 2021) were still under development. In

the absence of such resources for wholesale recoding, an alternative solution – the transient Genetic Toggle Switch system (this work) – was necessary.

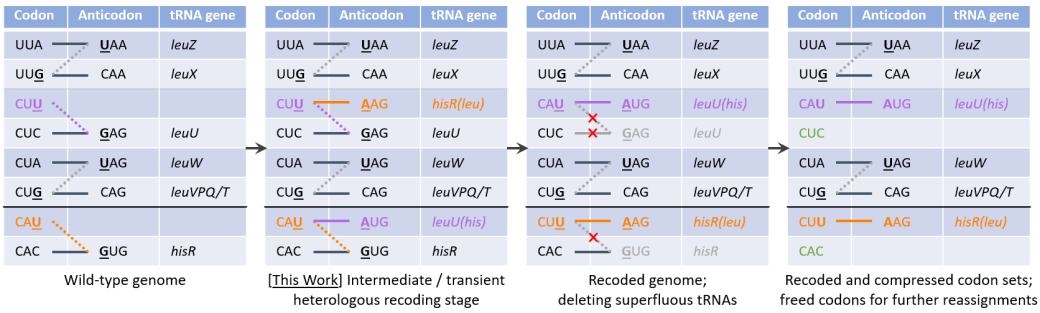


Figure 4.1: Proposed wholesale recoding of the *E. coli* to accept an alternative histidine-leucine genetic code

From the natural genetic code (far left box), to the alternative genetic code with the proposed histidine-leucine admix natural-synthetic code (second from left box; this work), to the proposed steps in the future for a for fully histidine-leucine synthetic genetic code (right two boxes).

As it is not possible to negate the toxicity to host entirely, the alternative strategy is to minimise it. This can be achieved by lowering the copy number of plasmids bearing the engineered tRNAs, reducing induction strength and/or limiting heterogenous transcription activity to a short time-frame, after which the tRNAs would cease to be transcribed. Lower, or transient, expression levels of anticodon-modified tRNAs risk precluding detection of codon reassignment in a reporter system – a system that must tolerate mis-incorporation during normal growth, and amplify the reassignment signal in a detectable and non-lethal manner. In short, I needed to develop a circuit that could take a transient burst of synthetic tRNA activity and output a permanent, non-toxic response, in order to reasonably detect sense-to-sense codon reassignment without incapacitating the host.

The fundamental principle for resolving this problem was inspired by research from the Endy lab in refining biological control through recombinase-based Boolean logic gates (Bonnet et al., 2013). Essentially, genetic switches made of site-specific recombinases could be used to turn on (or keep off) a downstream output, given the correct combination of input transcription signals. The feasibility of this approach stems from the efficiency and sensitivity of recombinases in that expression of only a few copies would be sufficient to activate the one-way switch, and catalyse a cascade of downstream enzymatic activity designed to produce an in-

nocuous, but amplified, response. The resulting system, therefore, should be able to convert and enhance analogue signals, like a weakly/variably-expressing synthetic tRNA, into an easily detectable digital/binary output (such as fluorescence/non-fluorescence, antibiotic resistance/sensitivity, or a live/dead phenotype). Doing so minimises the detrimental mis-translation effects on the host (owing from decoding multiple genetic codes), whilst providing a definitive indicator that reassignment has been realised.

In short, if the anticodon-modified tRNAs signals are able to implement reassignment in a histidine-leucine recoded recombinase reporter to a sufficient degree, this recombinase switch could be used to turn on (or off) a secondary downstream reporter system and provide another cleaner, augmented response as a proxy for successful codon reassignment.

Besides acting as a layer of signal amplification, the recombinase switch also presents an additional opportunity to uncover recoded reporters for testing reassignment under the OCI system. Finding a loss-of-function recombinase, where crucial histidine(s) have been recoded to leucine(s), which can subsequently be rescued upon introduction of the anticodon-modified tRNAs, could be used to further verify that the previously demonstrated leucine-to-histidine reassignment is replicable. Perhaps more importantly for validating the full system: recovering activity in a non-functional variant, resulting from recoding critical leucine(s) to histidine(s), would provide the first evidence that the reverse histidine-to-leucine reassignment is also possible.

The next step in this chapter, therefore, was to identify a suitable recombinase that is sensitive, precise, and efficient enough to perform the necessary unidirectional DNA rearrangement to act as a genetic on/off toggle switch (GTS) and amplify the signal of the anticodon-modified tRNAs in a non-toxic manner *in vivo*. From there, the compatibility of the recombinase with the OCI approach would be investigated by creating another recoded reporter for generating the full histidine-leucine codon reassignment system.

Like the flow diagram summary for Chapter 3 (Figure 3.2), the following figure

(Figure 4.2) for Chapter 4 should provide additional clarity as to the direction that I took to demonstrate sense-to-sense codon reassignment. Whilst there were additional routes that I could have explored further, there was not enough time during the experimental phase of the thesis to investigate these to completeness. Regardless, the GTS strategy developed a serine recombinase switch platform that was able to show leucine-to-histidine with yet another recoded reporter, and had the potential to demonstrate alternative reassignment systems in future extensions on this research.

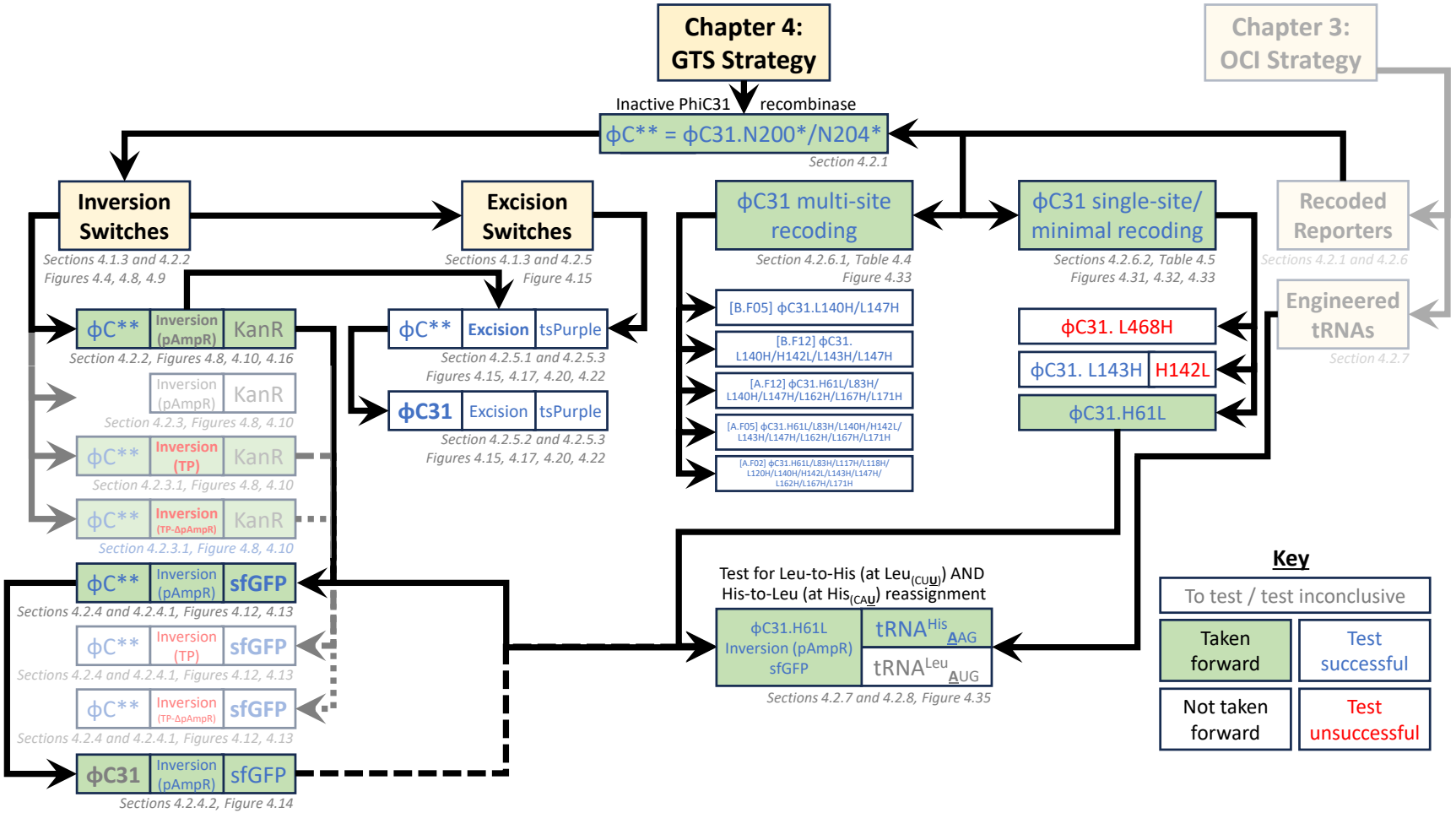


Figure 4.2: A summary flow diagram of the strategies that were successful whilst developing the GTS and proposed OCI-GTS in Chapter 4

Elements of the GTS strategy that were taken forward are shaded in green, with experiments that were successful in demonstrating sense-to-sense codon reassignment in blue font. Tests that were unsuccessful are in red font, and those that were not yet tested/inconclusive in grey font. The associated sections relating to the work are also shown in grey font below each box. Some elements were taken forward for the OCI-GTS system as they were essential to demonstrating reassignment but there had been no test yet available to assess its viability and/or the results of the initial tests were inconclusive. The faded section to the right of the figure relates to how the OCI strategy of Chapter 3 built into the GTS strategy of this chapter. The associated flow diagram from OCI can be found in Figure 3.2.

4.1.2 Recombinases and Site-Specific Recombination

Site-specific recombinases can be broadly separated into either the tyrosine- (Tyr) or serine- (Ser) recombinase families. This classification is based on amino acid sequence homologies and respective mechanistic relatedness when implementing DNA recombination. Depending on which amino acid (Tyr or Ser) within the active/catalytic sites that carry out the nucleophilic attack to instigate DNA strand exchange, determines which family the recombinase is designated. Although tyrosine and serine recombinases provide the same functional outcomes, they are an example of convergent evolution in that the conserved (catalytic and binding) domains between the families display different protein sequences, structures, and reaction mechanisms (Figure 4.3). A key difference, however, is that recombination with tyrosine recombinases can be reversed, whereas serine recombinases typically have a uni-directional mode of activity. As one aim of the GTS is to develop an irreversible switch for signal amplification, it was the serine recombinases that became the primary focus for this research. Hereon in, the term "recombinase" refers to

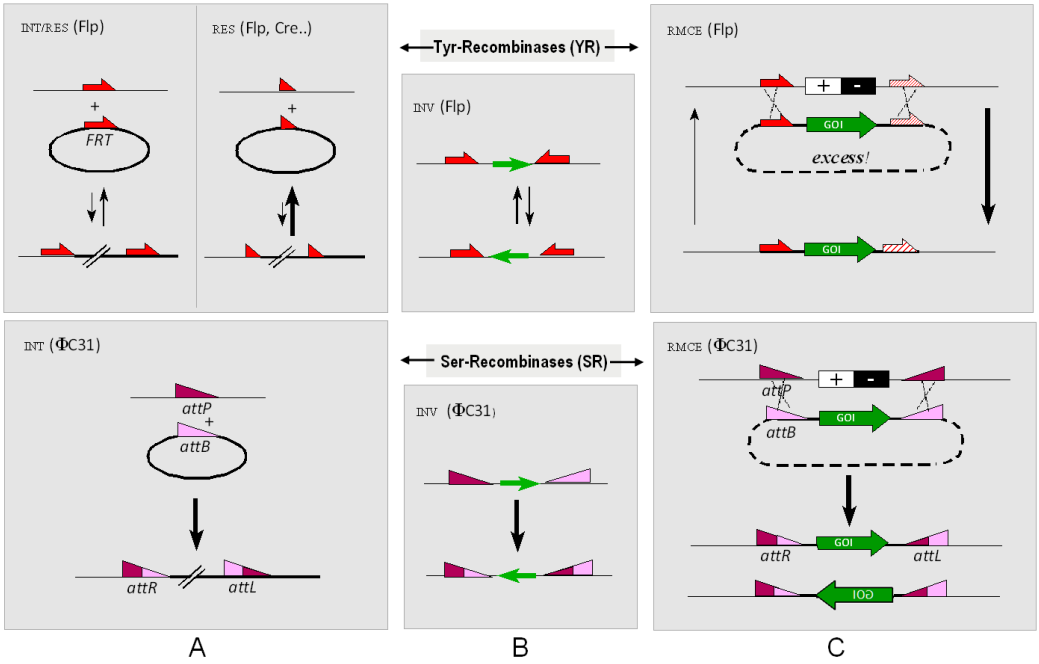


Figure 4.3: A summary of Tyr vs Ser recombinase crossover/reaction mechanisms

A range of reactions that can be implemented with tyrosine and serine recombinases: from integration (INT) to inversion (INV). Figure by Juergen Bode, CC BY-SA 3.0, `https://commons.wikimedia.org/w/index.php?curid=21206197` (last accessed: 07-Oct-2022).

those of the serine family, unless otherwise indicated.

As the name implies, site-specific recombinases enable a precise form of genetic recombination, allowing the exchange of DNA segments at specially defined regions. Recombinase activity typically produces insertions (including translocation and cassette exchange), deletions, and inversions at targeted sites: these sites tend to be regions that share a degree of sequence homology. More specifically, in site-specific recombination, target (or recognition) sites consist of two partially palindromic motifs, ranging from 30-200 nucleotides (nt) in length, to which recombinases may bind. Within each motif, there is a dinucleotide cross-over sequence where an active and bound recombinase complex can catalyse the breakage, exchange, and re-ligation of the recognition sites, resulting in the rearrangement of the targeted DNA sequence.

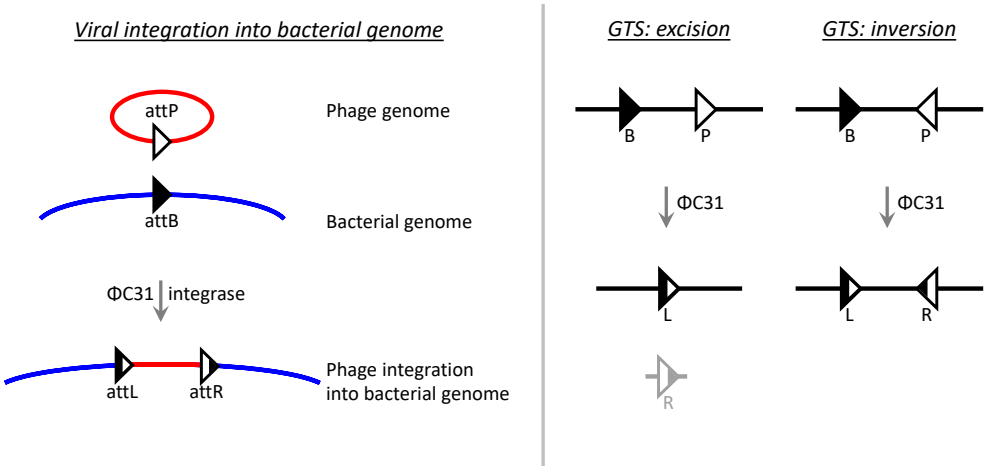


Figure 4.4: Site-specific recombinases for insertion, excision, and inversion

A generalised example of phage integrase activity (left), and the proposed activity that could be used as part the GTS strategy using the ΦC31 recombinase.

In nature, site-specific recombinases are widely used by lysogenic phages to facilitate integration into the host's genome. Often described as integrases, these enzymes bind to a specific attachment site within phage DNA (*attP*) and to a homologous site in the bacterial genome (*attB*). There, they mediate integration: cleaving the DNA sequences in the core region of the *attB* and *attP* sites to allow the phage genome to be inserted into the bacterial chromosome. Once integrated, the hybrid halves of the *attB*/*attP* sites (now *attL* and *attR*; Figure 4.4, left) are re-ligated, seal-

ing together the two genomes.

Whilst the natural system has been primarily used for integrating DNA fragments, recombinases have been repurposed as natural "molecular machines/tools" for genome and genetic editing and engineering: going beyond DNA insertion, into excision and inversion (Figure 4.4, right). Such precise modification of DNA sequences relies on the location and orientation of the *attB* and *attP* sequences. If both attachment sites are within the same sequence, and are in *cis* orientation (facing the same direction), activation of the recombinase will lead to the removal of the sequence in between the *attB* and *attP* sites. If the co-localised attachment sites are oriented in *trans* (facing opposite one another), this would result in an inversion of the intervening sequence. By taking advantage of these properties, different switch designs could be created to enhance an incoming signal following the expression of engineered tRNAs. For example, if a promoter is placed between *attB* and *attP* sites that are in *trans* and, in its original state, the promoter faces away from a downstream reporter, activation of recombinase should invert the enclosed promoter and allow expression of the downstream reporter. Similarly, *cis*-oriented attachment sites could be used to drop-out a feature that prevents expression of a secondary reporter downstream, thereby acting as an off-to-on switch. Both of these alternate functions form the crux of the proposed GTS designs (Section 4.2.3 and Section 4.2.4 for inversion; Section 4.2.5 for excision), and is the focus of the remainder of this chapter.

4.1.3 A Recombinase-based Solution for Mitigating Host Toxicity during Sense-to-Sense Codon Reassignment

Although the OCI system (with its heterologously-encoded anticodon-modified tRNAs and recoded reporter) was able to demonstrate leucine-to-histidine reassignment, the long-term expression of this alternative reassignment, alongside the natural genetic code, was detrimental to the host (Chapter 3). Moreover, it was not clear if the impact in growth was due to the metabolic burden of continually expressing extraneous reporters and orthogonal tRNAs, toxicity caused by genetic code ambiguity, or a combination of the two. Whilst it may have been possible to assess the

burden of heterologous gene expression *in vivo* by using a real-time fluorescence-based "capacity"/metabolic monitoring system, as proposed by Ceroni et al. (2015), the method required genomic integration of a constitutively-expressed GFP operon. Given prior difficulties in integrating new constructs into the bacterial genome (data not shown) and the high likelihood that the GTS system would also be using fluorescence (or a similar visual indicator) as a proxy signal of OCI activation, I sought an alternative design strategy to decouple the effects of anticodon-modified tRNAs expression from that of cellular burden so as to adequately detect the proposed non-natural codon reassignments without critically compromising the host.

Given the simplicity of phage integration and the efficiency of recombinases (Groth and Calos, 2004) as well as previous demonstration that they can be used as part of biological circuits (Bonnet et al., 2013; Roquet et al., 2016), harnessing these enzymes as reporters for codon reassignment brings significant advantages. I reasoned that recombinase expression (under the modified genetic code) could be used as a transient reporter that, once activated, would lead to a permanent genetic alteration, enabling the expression of a non-toxic reporter operating within the natural genetic code. This would establish a Genetic Toggle Switch (GTS) system.

With a sufficiently responsive recombinase, only a low level – or even a transient burst – of anticodon-modified tRNA activity would be required to trigger this one-way genetic switch. By separating expression of the primary recoded reporter from the secondary (naturally-coded) reporter[1], the latter would be able to persist as an indicator of reassignment without accruing additional toxicity from the engineered tRNAs designed for the former. In this way, the OCI-GTS systems together would be able to demonstrate a transient toxic reassignment signal, whilst delivering a permanent non-toxic response.

The recombinase and the secondary reporters are only two parts of the GTS system; the remaining key element is the switch itself. From Section 4.1.2 (and

[1]The main plan was for the secondary reporter to be naturally-coded, but there were designs for both reporters to function under the alternate histidine-leucine code, had host toxicity been less problematic. Whilst this runs counter to the rationale for having a non-toxic secondary reporter to indicate the presence of sense-to-sense codon reassignment, it could have demonstrated multi-target codon reassignment across more than one recoded reporter.

specified further in Figure 4.5), we know that the orientation of recombination recognition sequences, *attB* and *attP*, around a segment of DNA can result in the excision of the internal DNA sequence by site-specific recombinases if the recognition sites face the same direction, or inversion if they oppose one another. With these properties in mind, two general GTS designs were proposed.

The GTS excision design consists of a set of terminators between similarly-oriented *attB* and *attP* sites that block an upstream inducible promoter from initiating transcription of its downstream reporter; in Figure 4.5 these are represented by pTrp and chromoprotein TinselPurple (tsP, tsPurple) respectively. Upon activation of a recombinase, the terminating sequences would be removed, resulting in unhindered expression of the chromoprotein.

By contrast, the prototypical GTS inversion constructs consists of a pair of inverted *attB* and *attP* sequences flanking a (constitutive) promoter[2] that is initially facing away from its downstream reporter. Upon expression of a recombinase, the promoter would be flipped around so that it is oriented in the same direction as its reporter, allowing transcription to occur. The example in Figure 4.5 uses pAmpR as the promoter and sfGFP as the reporter, but it was designed so that other sequences/markers could be similarly used.

In terms of the choice of recombinase, one such site-specific recombinase that has the irreversible, unidirectional recombination activity necessary for both excision and inversion GTS designs is ΦC31 (Merrick et al., 2018). This recombinase doubles as a new recoded reporter target for anticodon-modified tRNAs in the OCI codon reassignment strategy, as well as the primary reporter in the GTS excision or inversion constructs. Thus, when working in combination with each other, these two strategies generate the full OCI-GTS system (further explored in Chapter 4).

Having defined the prototypical excision and inversion GTS design circuits, and selected the necessary recombinase, the remaining chapter sections will discuss in turn: (i) the designing and recoding of the ΦC31 switch construct (Sec-

[2]Subsequent derivations of the GTS inversion designs saw the introduction of other genetic elements/parts (including RBSs and terminators) within the internal DNA segment between the *attB* and *attP* sites.

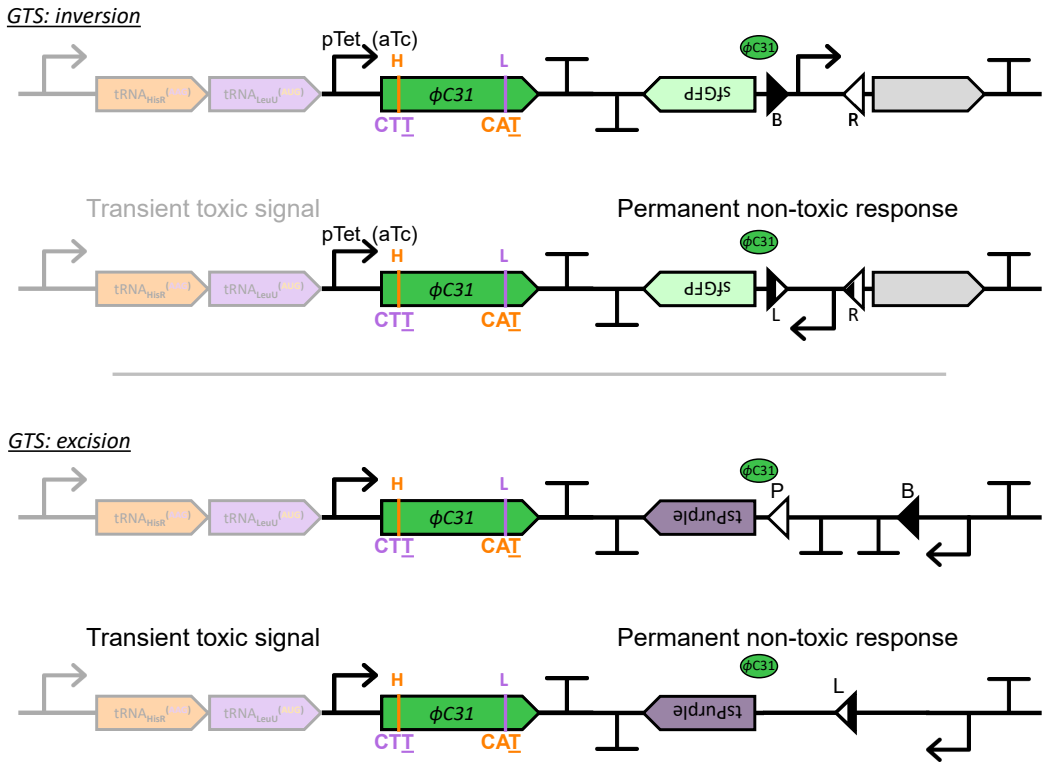


Figure 4.5: A generalised schematic of the Inversion and the Excision ΦC31 Genetic Toggle Switch circuits

Top: an earlier design of the GTS inversion construct proposed the use of dual secondary reporters: the promoter would be initially aligned to transcribe one coding sequence (e.g. red fluorescent protein (RFP); represented by the grey coding sequence (CDS) arrow-box), and upon successful recombination that inverts the promoter, the sfGFP could subsequently be expressed. sense-to-sense codon reassignment success could then be measured by the ratiometric change between the fluorescent proteins. However, as an initial design, this would likely have provided more challenges in both construction and characterisation/analysis, hence a simplified design, using just one secondary reporter, was implemented. Bottom: a proposal for the GTS excision construct follows a similar use of the ΦC31 recombinase, but in this case, the intention is to drop out a set of intervening terminators to allow expression of the *tsP* secondary/proxy reporter.

tion 4.2.1), including finding critical leucine residues to demonstrate the histidine-to-leucine reassignment that could not previously be investigated; (ii) characterising and troubleshooting of the inversion (Section 4.2.3, Section 4.2.4) and excision (Section 4.2.5) circuits, including the development of subsequent switch design variants; (iii) combining the recoded recombinase primary OCI reporter with the secondary GTS excisions or inversion reporter constructs into an improved OCI-GTS system for assessing reassignment *in vivo* (Section 4.2.6).

4.2 Results and Discussion to GTS

4.2.1 Characterising active and inactive ΦC31

As with the previous OCI reporters, it was necessary to first analyse the sequence and structure of ΦC31 to ascertain which leucines and histidines (if any) are crucial to function. Once established, it should be possible to create recoded, non-functional recombinase reporters that can be used alongside the anticodon-modified tRNAs (from the OCI system) in rescue/re-gain-of-function assays to characterise histidine-leucine codon reassignment. In particular, it was imperative to find essential leucines that, when recoded to histidine, would produce inactive ΦC31 variants suitable for investigating the yet unproven reverse histidine-to-leucine reassignment. If no essential leucine or histidine can be found, it could also be possible to create an inactive ΦC31 reporter by simply recoding all, or at least multiple, instances of histidine and leucine codons – in effect, wide-scale disruption rather than purposeful targeting of vital residues. This would greatly increase the number of recoded codons that the engineered tRNAs must correct, thereby increasing stringency in selection for recovering recombinase functionality and demonstrating a successful reassignment.

A preliminary literature review of mutagenesis studies revealed several leucines within ΦC31 that are conserved across members of the serine recombinase family, and in regions that are known to be involved in recombination activity or that have high DNA binding affinity to the *attB*/*attP* recognition sequences (Liu et al., 2010; Rowley et al., 2008; Rutherford and Van Duyne, 2014; Rutherford et al., 2013). Most of the studies stemmed from alanine-scanning mutagenesis experiments, although leucines were occasionally used when it is necessary to retain the size of a mutated residue but with a substitution that would still be considered relatively inert. Liu et al. (2010), for example, demonstrated that L143A[3] fully

[3]There is a discrepancy between the residue numbering from the ΦC31 sequences in the literature and that of the sequence from the PDB structure (ID: 4BQQ) of the recombinase, where literature numbering = PDB number + 8. In this thesis, I endeavour to use the literature numbering, however experimental designs and computational analysis were originally based on the PDB numbering and this may be reflected in these results. Where relevant, I will note if the results have used the PDB numbering, if the literature numbering has not been used. For reference, Table 4.1 displays both

abolished recombination activity *in vivo* and *in vitro*, as well as partially reduced *attB* binding affinity. Their research also highlighted other point mutations (Table 4.1c, L350A and L468A among them), which had varying degrees of success in diminishing DNA binding affinity, however, these substitutions alone did not completely inhibit ΦC31 functionality. Of the handful of systematic studies that discuss the effects of a wide range of substitutions at specific critical sites on recombination activity (like Smith and colleagues who used random site-directed mutagenesis to make a library of 11 E449 variants in order to reveal the mechanism for reversing the directional control in ΦC31; Rowley et al., 2008), none of the leucines of putative importance have been experimentally mutated to histidine for characterisation, nor have any critical histidines been uncovered to have warranted further study.

numbering schemes.

	PDB numbering	**Literature numbering (PDB # +8)**	**Structural Region**	**Affects function?** (Impact: * = minor, ** = moderate, *** = major)	**Ref**
(a)	N192*	N200*	β7	*** LOF	VP/AK?
	N196*	N204*	β7	*** LOF in DNA binding domain?	VP/AK?
(b)	H53	H61	β2	**	PyMOL
	L75	L83	αB	Not reported	PyMOL
	L109	L117	αD2	Not reported	PyMOL
	L110	L118	αD2	Not reported	PyMOL
	L112	L120	αD2	Not reported	PyMOL
	L132	L140	αE1	*	PyMOL
	H134	H142	αE1	*	PyMOL
	L135	L143H	αE1	***	PyMOL, Lit: L143A[a]
	L139	L147	αE1	Not reported	PyMOL
	L154	L162		Not reported	PyMOL
	L159	L167	αE2	Not reported	PyMOL
	L163	L171	αE2	Not reported	PyMOL
	L460	L468		* L468[a]/P[b] only partially abolishes activity as a single point mutation	PyMOL, Lit: L468A[a]/P[b]
(c)		R18A, ***L143A***, I141A, E153A, I432A, V571A		*? "great decrease in DNA binding affinity"	[a]
		G182A/F183A, C374A, C376A/G377A, Y393A, V566A		**? "completely lost their ability to bind to the specific target DNA attB as compared with wild-type proteins"	[a]

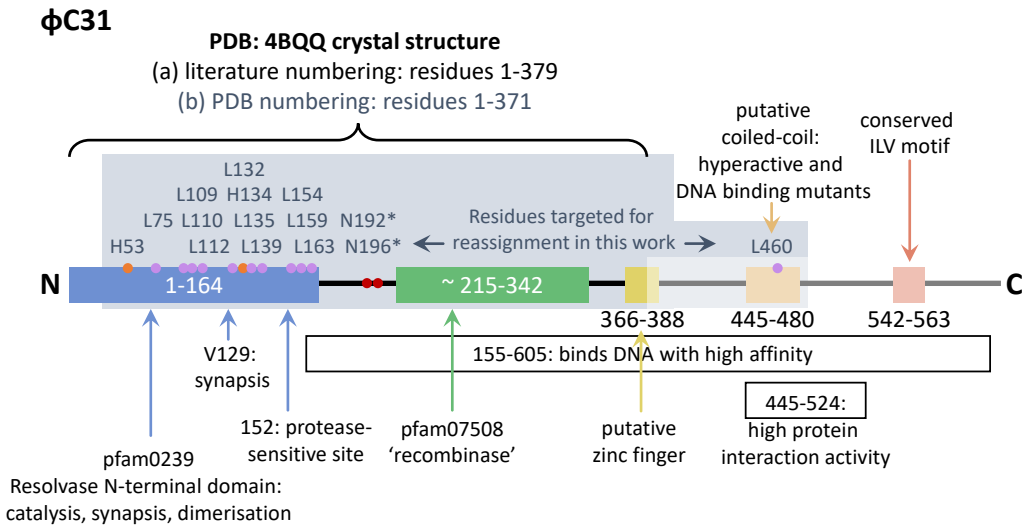
Table 4.1: Key candidate residues for recoding and potential reassignment in the ΦC31 recombinase

(a) Mutations that were included in the first ΦC31 construct, alongside the standard OCI codon optimisation (where all leucine and histidine codons in the DNA sequence of ΦC31 were initially recoded to $Leu_{(CTG)}$ and $His_{(CAC)}$ respectively). (b) Candidate mutations identified from PyMOL analysis of ΦC31 (PDB: 4BQQ) where point mutation recoding leucine-to-histidine (Leu-to-His) or histidine-to-leucine (His-to-Leu) may render the protein inactive; in preparation for reassignment to reactivate function. Identified for testing experimentally. (c) Mutations highlighted from literature that were shown to affect and/or abolish ΦC31 function. NR = not reported. [a]Liu et al., 2010; [b]Rowley et al., 2008.

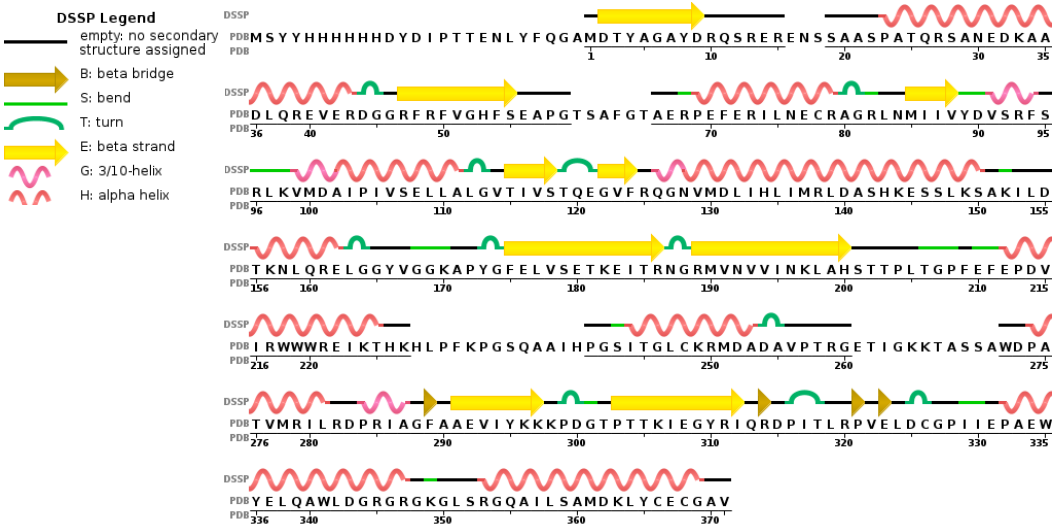
In an attempt to find other leucines and any histidines that, when recoded to histidine and leucine respectively, could produce a functionally-impaired recoded recombinase, an *in silico* examination of the ΦC31 crystal structure (PDB: 4BQQ) was carried out. With PyMOL, it was possible to display how specific point mutations may change the structure and its intramolecular interactions away from the initial model. Sites where histidine-leucine substitutions were predicted to result in a greater probability of increased physical steric hindrance were, therefore, selected as targets for creating recoded recombinases; the hypothesis being that such replacements would result in a reduction or full loss of functional activity.

It should be noted that one limitation to this computational approach was the lack of available structural data on most of the C-terminal domain of ΦC31 (Figure 4.6; presumably truncated during crystallisation). Of the 613 residue-long protein, the most complete model available in the literature (PDB: 4BQQ) accounts for only residues 1-379, which encompasses the catalytic resolvase N-terminal domain (pfam0239, residues 1-164), the putative "recombinase" (pfam07508, residues ~215-342), and most of the zinc finger domains (residues 366-388) (Figure 4.6a; Smith et al., 2010). This, therefore, precluded the opportunity to model L468H (as identified by Liu et al., 2010), which resides in a putative coiled-coil domain (residues 445-480) associated with hyperactive DNA binding mutants, and from investigating mutations to the conserved ILV-rich motif (residues 542-563). Although Smith and colleagues also noted that a good portion of ΦC31 binds DNA with high affinity at residues 155-605, and highlighted residues 445-524 as being an area of high protein interaction activity, these areas could not be fully examined via *in silico* methods available at the time.

Additionally, the mutational predictions conducted using PyMOL would only be able to provide an approximation of the structural effects on local interactions as the model is static. That is, when introducing in potential substitutions, PyMOL is not able to relax the structure of the protein model, alter the backbone, repack neighbouring side chains out of the way, or provide a difference in Gibbs free energy (as a means to assess the destabilising effect of the mutation on the structure). In effect,



(a) Conserved motifs in ΦC31 as adapted from Smith et al. (2010)



(b) An overview of the secondary structures in ΦC31 (RCSB-PDB: 4BQQ ChainA)

Figure 4.6: An overview of the key domains and secondary structures within ΦC31

no further information could be derived regarding local protein dynamics beyond that of putative van der Waals overlaps to imply potential steric clashes in the existing fixed local structure. Whilst at the time, it may have been possible to use other protein design and structure prediction software, like FoldX and Rosetta, to deter-

mine the energetic and alternative conformation effects of various point mutations on the surrounding side chains and overall protein stability/function, the ultimate aim was to find a recoded recombinase with loss-of-function activity that can be used to test sense-to-sense codon reassignment. As such, despite the limitations of using PyMOL predictions, this approach could be sufficient in giving a quick and visual first impression of whether a given mutation would cause clashes in a critical region within a protein; after which the model would be experimentally validated to confirm whether the predicted intramolecular clashes truly exist and are capable of abolishing recombinase function as required.

In total, 10 histidines and 43 leucines within ΦC31 sequence were identified, though the 4BQQ model only had structural data on five of the histidines and 26 leucines (not including the three leucines identified earlier from literature) from which to model the mutations that could potentially create inactive histidine-leucine recoded reporters. At the time of review, all previous examples of codon reassignment in synthetic/heterologous genes had been limited to one recoded codon per gene (Döring and Marlière, 1998). It was surmised that the fewer number of recoded leucines and histidines required to inactivate ΦC31 (ideally one of each) would help maximise the chances of the anticodon-modified tRNAs in rescuing function upon successful histidine-leucine reassignment. Hence, the initial search space for the *in silico* mutagenesis investigations was primarily focused on the N-terminal catalytic domain of the recombinase: the hypothesis being that the most essential residues for recombinase activity would be located in this region, and recoding key leucines and histidines here may abolish function.

Having further refined the search criteria, the five histidine residues (**H61**, H142, H151, H208, **H234**) previously identified and a reduced set of 14 leucines (L45, **L83**, L91, L105, **L117**, **L118**, **L120**, **L140**, **L143**, **L147**, L156, **L162**, **L167**, **L171**) were selected to simulate whether substituting histidine to leucine and vice versa could create an inactive ΦC31 for use as a viable recoded reporter. Note, underlined residues match key residues previously identified from literature, whilst those in bold highlight residues that PyMOL Mutagenesis Wizard simulations pre-

dict would likely clash with their neighbours.

The wild-type structure and the effects of the two most likely-occurring side chain orientations (rotamers) for each recoded residue simulation is further displayed in Figure 4.7 and Figure 4.8. The green and red lines/disks in the simulations indicated pairwise overlap of atomic van der Waals radii (i.e. steric clashes): atoms that were almost in contact or slightly overlapping one another were illustrated in green, whereas rotamers that were predicted to result in significant clashes with nearby residues were represented by large red disks and lines. As listed in Table 4.1b and shown in Figure 4.7, the final shortlist of 13 residues selected for introduction into ΦC31 were: H61L (in beta sheet 2, β2); L83H (in alpha helix B, αB); L117H, L118H, L120H (αD2); L140L, H142L, L143H, L147H (αE1); L162H (unassigned structure), L167H, L171H (αE2); L468H (unassigned structure). Their selection was based on qualitative assessment of mutations that produced relatively more clashes, the predicted likelihood/frequency of the rotamers occurring, and the relative proximity of the candidates to one another (which was expected to simplify the recoding efforts).

Given both the literature and the computational modelling predicted that mutating L143 would be sufficiently disruptive to ΦC31, this became the prime candidate for recoding to histidine; with the resulting, potentially inactive, ϕ*c31*.L143H to be used as the first demonstration of histidine-to-leucine reassignment. Similarly, as L140 and L147 are positioned on the same side of the αE1 helix as L143, it was possible that recoding all three leucines to histidines would cause clashes that could be additionally destabilising to the recombinase. Residue H142 also sits close to these αE1 helical leucines, and a ϕ*c31*.H142L mutation could be another opportunity to demonstrate leucine-to-histidine reassignment, however, its *in silico* simulations did not suggest any problematic intramolecular interactions with such a recoding. That is not to say that this would be the case once recoded, but it seemed prudent to also explore recoding another histidine residue, such as H61 in beta sheet β2, to further validate the reverse reassignment. To this end, oligos YKH297 (for $H_{cAC}142L_{cTT}$), YKH296 (for $L_{cTG}143H_{cAT}$), YKH298 (for a composite of the two, $H_{cAC}142L_{cTT}$

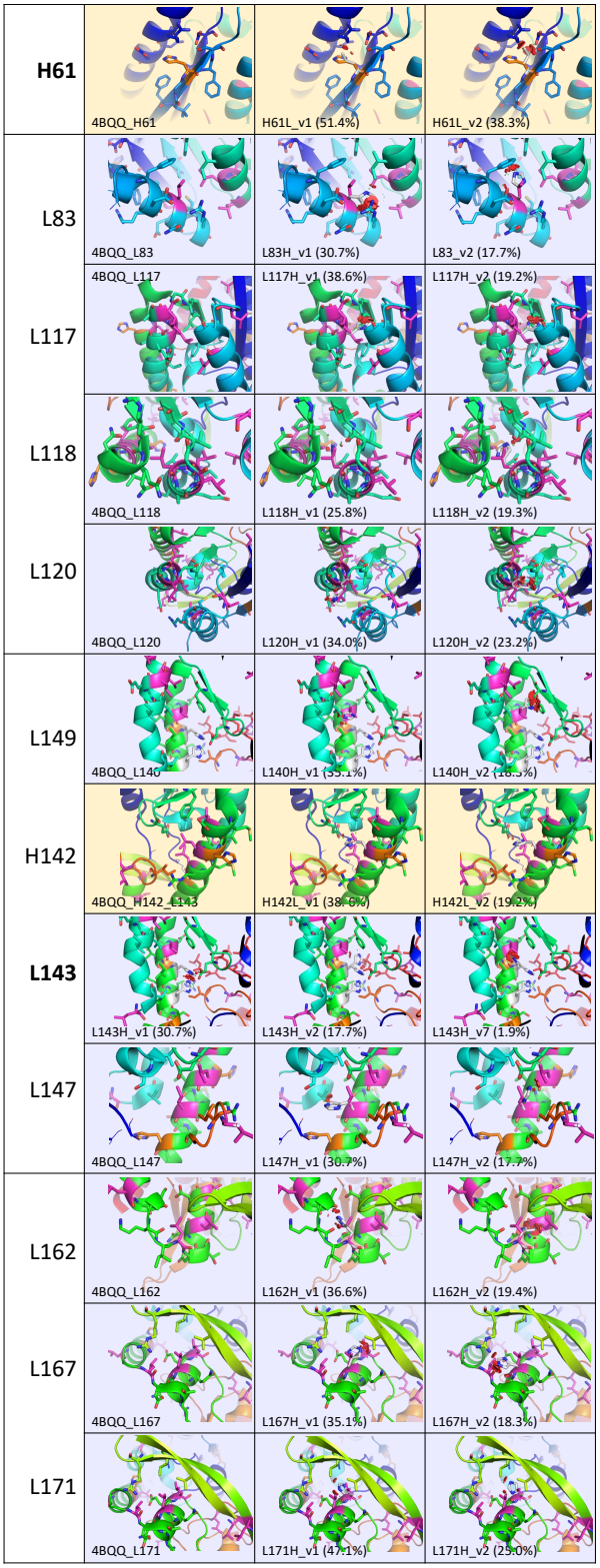


Figure 4.7: PyMOL mutagenesis modelling of key histidines and leucines in ΦC31

Histidine residues are in orange, leucines as purple, and grey where mutations have been simulated. Clashes are seen as red circular disks; percentage values show the likelihood of these side-chain orientations occurring. Expected key residues are in bold. The rest of the ΦC31 structure has a blue-to-red gradient colour showing the direction from the N- to the C-terminal.

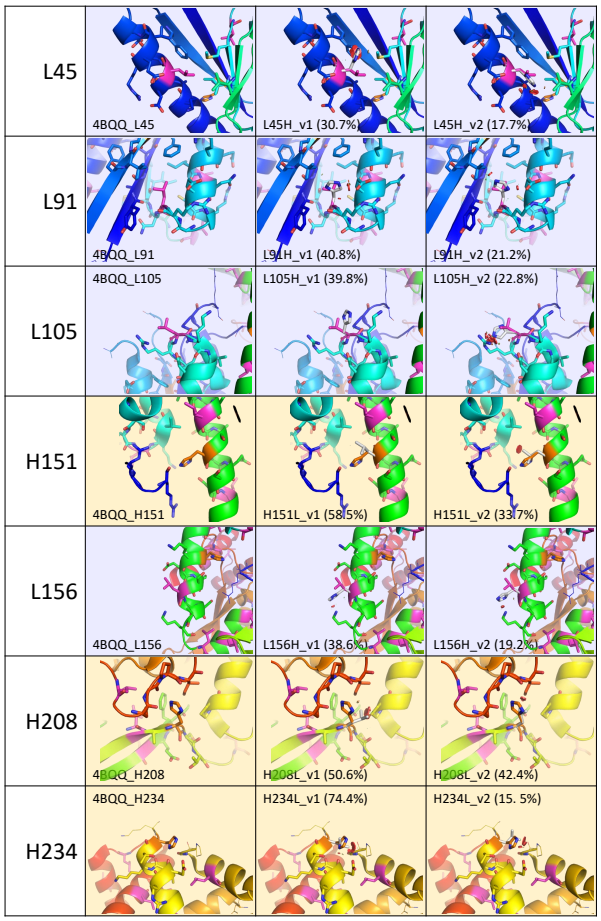


Figure 4.8: PyMOL mutagenesis modelling of other histidines and leucines in ΦC31

It is expected that the recoding the residues highlighted in Figure 4.7 should produce a viable non-functional ΦC31, however, should this fail, the residues shown here would considered. Histidine residues are in orange, leucines as purple, and grey where mutations have been simulated. Clashes are seen as red circular disks; percentage values show the likelihood of these side-chain orientations occurring. Expected key residues are in bold. The rest of the ΦC31 structure has a blue-to-red gradient colour showing the direction from the N- to the C-terminal.

and $L_{cTG}143H_{cAT}$), and YKH314 (for recoding all four αE1 helical candidates) were designed for site-directed mutagenesis (see Table A.2b in Appendix A.3 for Darwin Assembly oligo sequences).

In the event that more residues are needed to be recoded to create an inactive ΦC31, additional oligos spanning the shortlisted candidates identified in Table 4.1b were also created (Table A.2c). Moreover, these oligos were designed such that they could be introduced in a combinatorial manner (singularly or as multiple point mutations) to generate a library of recoded ΦC31 reporters, and thereby increase the chance of producing a non-functional variant. Section 4.2.6 discusses the im-

plementation of this one-pot, multi-site Darwin Assembly mutagenesis method and presents the ΦC31 mutants generated from this approach.

Another benefit of constructing a library of multi-site recoded recombinases is that it would extend beyond all previous sense-to-sense codon reassignment research, and present the opportunity to assess the extent to which anticodon-modified tRNAs are able to reassign and rescue function to reporters that have a number of recoded residues. That is, to investigate whether engineered tRNAs from the OCI strategy (Chapter 3) can recover not just one recoded codon within the inactivated ΦC31, but whether they could be transcribed at a sufficient level as to demonstrate bi-directional codon reassignment – in effect, a proto-alternative code – to a dozen or more differently recoded codons within the same gene, without severely incapacitating its naturally-coded host.

With this plan in hand for creating an inactive recoded recombinase, we turn next to the design and construction of the basic Genetic Toggle Switch circuit (Section 4.2.2), in which ΦC31 plays a central role in mitigating host toxicity so as to permit detection of successful codon reassignment. The GTS consists of two primary designs: the inversion (Section 4.2.3, Section 4.2.4) and excision (Section 4.2.5) constructs. Once optimised, the recombinases from these circuits are replaced with the site-directed mutagenesis designs (as detailed in this section) to create a viable but inactive ΦC31 recoded reporter (Section 4.2.6). Finally, this realised GTS circuits is coupled with the anticodon-modified tRNAs from the OCI to form the OCI-GTS system (Section 4.2.7), thereby producing a transient *in vivo* expression system for implementing alternative genetic codes.

4.2.2 Constructing the basic GTS circuit design

The earliest GTS construct was designed by Vitor Pinheiro, with the aim to be used in both my PhD project and a complementary reassignment project by post-doctoral researcher, Dr. Antje Krüger. Common to both projects was the use of a codon-optimised *ϕc31* gene and a simple recombinase-based inversion switch that consisted of a promoter inserted between opposing recombination recognition sequences *attB* and *attP* (the segment between the black and white triangles in Fig-

ure 4.9; construct `Recombinase_circuit_VP.dna` [002]). The promoter sequence, *pAmpR* (a regulatory element taken from an ampicillin resistance *bla* operon of the pNGAL97 [P005] plasmid), was designed to be oriented in the opposite direction to its downstream secondary reporter construct, with the intention that it would be irreversibly inverted upon activation by ΦC31. The subsequent rearrangement of the *attB*/*attP* switch and its embedded promoter sequence would therefore align *pAmpR* in the same direction as the intended reporter, thus allowing transcription of this operon (Figure 4.9).

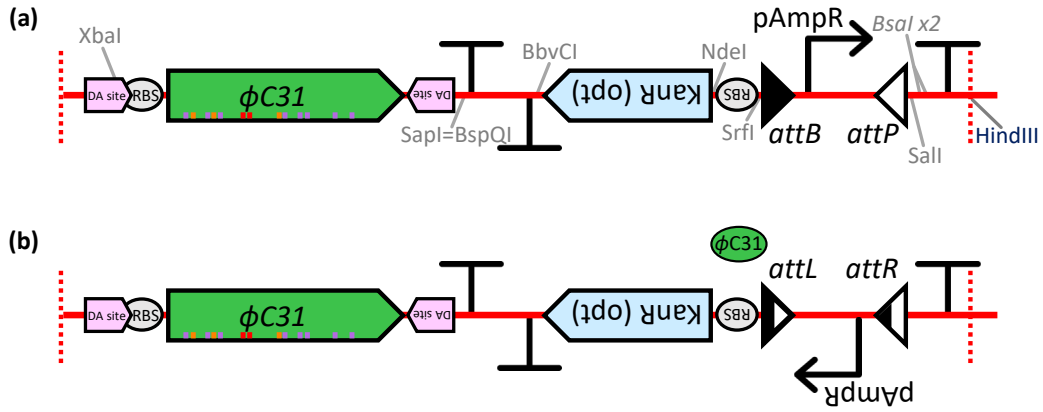


Figure 4.9: Recombinase circuit v1: [002] `Recombinase_circuit_VP.dna` (3,269bp)

The initial/prototypical GTS circuit design, featuring ΦC31, the inversion switch, and Kan^R as the secondary reporter system. (a) naïve state; (b) activated stated. Note that the tRNA reassignment constructs are not yet linked to the GTS circuit.

In this prototypical GTS design, the secondary reporter that was coupled to the switch was a codon-optimised *kanR* operon (for antibiotic resistance to kanamycin). The design flanking *kanR* had been additionally designed so that most of this reporter system could be replaced by another reporter of choice (e.g. an *sfgfp* operon, as in Figure 4.5), via restriction digest or Gibson assembly, by using the specifically incorporated (`NdeI` or `SrfI`)/(`BbvCI` or `SapI=BspQI`) restriction sites that flank the secondary reporter construct in [002] (Figure 4.9).

As with designing the OCI reporters, codon optimisation for both *ϕc31* and *kanR* genes consisted of synonymous replacement of all leucine and histidine codons to $Leu_{(CTG)}$ and $His_{(CAC)}$ respectively (if they had not already been marked to be orphan codons for reassignment). Like before, the rationale was to remove

any leucine and histidine wobble codons that could be ambiguously (re)assigned when using the OCI anticodon-modified tRNAs (i.e. through the heterologously-expressed alternate genetic code embedded within a naturally-coded organism), and to check that the gene was still functional despite the optimisation. Once confirmed that the synthesised gene constructs work as designed, Darwin Assembly was used to create point mutations at candidate inactivating sites in *ϕc31* (i.e. changing crucial histidines and leucines to orphan wobble codons $\text{Leu}_{(\text{CT}\underline{\textbf{T}})}$ and $\text{His}_{(\text{CA}\underline{\textbf{T}})}$ respectively) as identified earlier in Section 4.2.1 and Table 4.1a,b, and subsequently implemented in Section 4.2.6. As with the other OCI reporters, the recoding should initially render the reporter non-functional under the natural system, but with the addition of the alternatively coding $\text{tRNA}^{\text{His}}_{\underline{\textbf{A}}\text{AG}}$ (for leucine-to-histidine reassignment at wobble codon $\text{Leu}_{(\text{CU}\underline{\textbf{U}})}$) and $\text{tRNA}^{\text{Leu}}_{\underline{\textbf{A}}\text{UG}}$ (for histidine-to-leucine at wobble codon $\text{His}_{(\text{CA}\underline{\textbf{U}})}$), the sense-to-sense codon reassignment system should allow *ϕc31* to be reassigned back to produce a functioning protein.

To facilitate engineering these loss-of-function mutations, the [002] prototypical GTS construct was designed with two Darwin priming regions that flank the *ϕc31* sequence, as well as a `(Nt/Nb.)BbvCI` recognition site that would be necessary in implementing Darwin Assembly (Figure 4.9, Section 4.2.6). Moreover, not only did the upstream sequence to *ϕc31* include an RBS within the Darwin priming site, this entire segment was designed to match (and be subcloned via Gibson Assembly or with restriction cloning into) a region within vector pNGAL97 [P005] downstream of a *pTetR/A* promoter to help regulate its expression (Figure 4.10).

Unlike the typical OCI reporter designs, the codon-optimised *ϕc31* gene ordered for synthesis was already a non-functional variant – not for histidine-leucine reassignment, but with amber/stop codon suppression. The changes that rendered *ϕc31* inactive were the introduction of TAG amber stop codons ($*\text{Am}_{(\text{TAG})}$) in place of asparagine codons at positions 200 and 204 of the protein (represented by the two red markers within the green *ϕc31* gene sequence in Figure 4.9 and subsequent figures; listed in Table 4.1a). The term *ϕc31_*Am200_*Am204* (shortened to *ϕc***) is used to distinguish between the stop-inactivated variant and the active *ϕc31* ver-

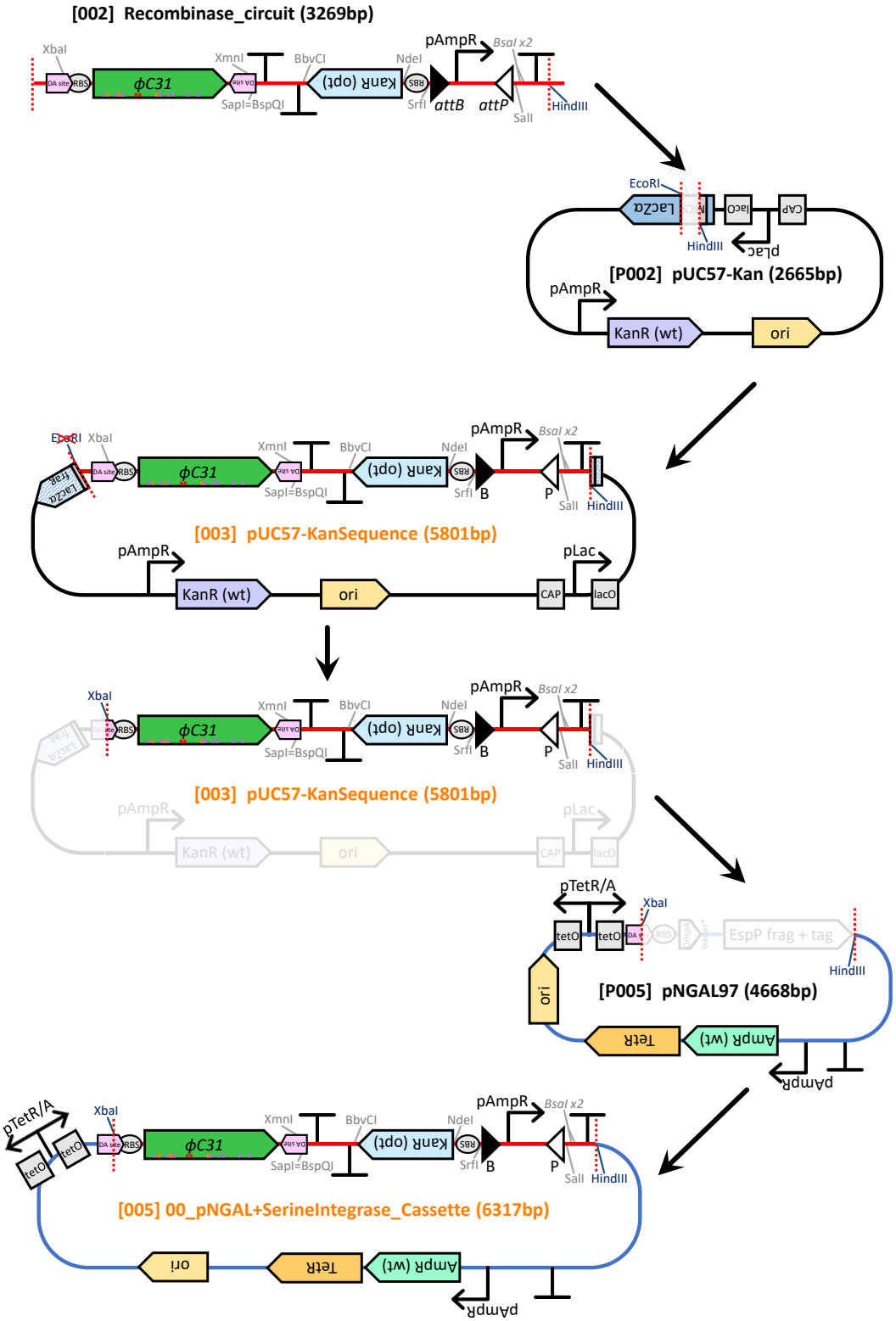


Figure 4.10: The prototypical design for all GTS constructs – from synthesis of the recombinase circuit [002] to assembly into the characterisation plasmid [005]

Figure 4.10: The prototypical design for all GTS constructs – from synthesis of the recombinase circuit [002] to assembly into the characterisation plasmid [005]

To create construct [003], the full GTS inversion prototype (`Recombinase_circuit_VP.dna` [002], 3,269 bp), synthesised by Genewiz, was subcloned into their standard vector pUC57-Kan [P002] (2,665 bp), between restriction sites EcoRI/HindIII, resulting in the 5,801 bp `pUC57-KanSequence(2665bp)_est-seq_5801bp.dna` [003b] plasmid.
For construct [005], then designed [002] inversion construct was further subcloned into the pNGAL97 [P005] plasmid, creating a 6,317 bp plasmid: `20170511_AK_pNGAL_SerineIntegrase_Cassette_sequenced.dna` [005].
The initial [002] GTS circuit design, featuring ΦC31 and Kan^R, and the cloning steps to arrive to plasmid [005] – the basic design for all subsequent GTS constructs.

sion. Further nomenclature will be introduced in Section 4.2.6 when referring to the inactive variants caused by various histidine-leucine recoding (summarised in Table 4.2). Although useful as an initial negative control, this meant that the in-frame stop codons needed to be reverted back to asparagines in order to ascertain a baseline activity level for the functional recombinase. After which, the wobble codons $Leu_{(CT\underline{T})}$ and $His_{(CA\underline{T})}$ could be introduced at the candidate histidine and leucine sites respectively to make it again a non-functional – but this time, reassignable – OCI reporter for testing the sense-to-sense codon reassignment system.

As for the composition of the Kan^R secondary reporter construct, aside from making synonymous leucine and histidine substitutions to remove wobble-pairing codons (so that the gene sequence becomes compatible for codon reassignment), *kanR* was also put under a synthetic RBS that was designed to moderate its translation initiation rate (TIR). By specifically lowering the expression rate of the heterologous reporter, it may be possible to reduce the burden of translating a potentially toxic (alternatively-coded), but essential (antibiotic resistant) product for host viability. Originally, the aim was to generate an RBS with a TIR output of 30,000 a.u.: 3% of the maximum translation rate that could be predicted using the Salis lab's RBS Calculator (`https://www.denovodna.com/software/design_rbs_calculator`). However, after including an `NdeI` cut site into the 3'-end of the synthetic RBS (as a contingency element for separating the RBS from the downstream *kanR* gene), the predicted TIR output was estimated at 6,900 a.u. (i.e. 0.69% of the maximum translation rate as predicted by the RBS Calculator). In the event that the proposed RBS was at an unsuitable level for use in a reassignment reporter,

an `SrfI` site was incorporated at the other end of the RBS sequence (where it joins to the *attB* end of the inversion switch sequence), such that the sequence could be easily replaced using `NdeI`/`SrfI` restriction. Finally, to complete the *kanR* operon, a *lambda t0* terminator was added to the downstream end of the CDS.

Having established the full arrangement of the GTS inversion prototype (the primary ΦC** reporter, the secondary Kan^R reporter, and *attB-pAmpR-attP* inversion switch; `Recombinase_circuit_VP.dna` [002], 3,269 bp), this linear construct was ordered for synthesis by Genewiz and arrived subcloned within their standard 2,665 bp pUC57-Kan [P002] vector. One additional feature/bug of note was that this synthesis resulted in a 5,801 bp `pUC57-KanSequence(2665bp)_est-seq_5801bp.dna` [003b] plasmid (Figure 4.10, top half), that contained two *kanR* variants: a codon-optimised version that was coupled to ΦC** as part of the GTS system circuit, and one from Genewiz's default [P002] vector. Whilst two copies of the *kanR* gene sequence would bias the host into retaining this plasmid (especially when exposed to the selection pressure of a kanamycin environment), the presence of a Kan^R that is already under constitutive expression would completely circumvent the GTS inversion switch designed: there would be no need to express and rescue the recoded ΦC** to invert the *pAmpR* and, in turn, activate the optimised Kan^R. Knowing that this issue would need to be addressed, the final design feature that was built into this basic GTS recombinase circuit [002], was the presence of two flanking restriction sites: an `XbaI` upstream of recombinase's RBS, and a `HindIII` site downstream of the full circuit, after the terminator beyond the *attP* recognition site (Figure 4.10). These two restriction sites would enable the majority of the [002] inversion construct to be subcloned into an appropriate vector containing a different selection marker. In this instance, upon arrival of the Genewiz-synthesised [003b] plasmid, the relevant GTS circuit was inserted into pNGAL97 [P005] (a 4,668 bp in-lab/house plasmid provided by post-doctoral researcher, Dr. Eszter Csibra), to create a 6,317 bp plasmid `20170511_AK_pNGAL_SerineIntegrase_Cassette_sequenced.dna` [005] – the GTS Inversion-Kan^R circuit (Figure 4.10, bottom half; Figure 4.11).

In the pNGAL97 [P005] vector, the ϕc^{**} of the Inversion-Kan^{R} circuit was placed under the tight transcriptional control of *pTetR/A* , and the selection marker associated with the plasmid was for ampicillin resistance (Amp^{R}, *bla*), thus should not be in conflict with the optimised Kan^{R} secondary reporter of the GTS construct. This [005] plasmid subsequently became the basic design for constructing the later inversion circuit derivations (Section 4.2.3, Figure 4.11; Section 4.2.4, Figure 4.13) and the GTS excision construct (Section 4.2.5, Figure 4.18).

4.2.3 The Inversion-KanR Circuit

With the basic GTS Inversion-Kan^{R} [005] construct created (Figure 4.10), the next step was to restore function to the recombinase to verify that the circuit prototype operates as designed. As described previously, this inactive ϕc^{**} gene sequence contained two amber stop ($*\text{Am}_{(\text{TAG})}$) codons at positions 200 and 204 that abolished recombination activity. Oligo YKH308 (refer to Table A.2a) was designed to revert both these stops back to the wild-type asparagine codons ($*\text{Am}_{(\text{TAG})}200\text{N}_{(\text{AAT})}$ and $*\text{Am}_{(\text{aTAG})}204\text{N}_{(\text{cAAC})}$) where, once restored, it would be possible to characterise the baseline recombinase activity, and begin to test the feasibility of the GTS Inversion-Kan^{R} switch created in Section 4.2.2.

Prior to reverting the ϕc^{**} in plasmid [005] back to its wild-type, one other control experiment was required: to confirm that the Inversion-Kan^{R} switch was truly inactive for both the recombinase (with its nonsense mutations), and the Kan^{R} secondary reporter (which was functional, but its *attB/attP* switch held the promoter in a *trans*-orientation to prevent transcription). The plasmid was transformed into the *E. coli* cloning strain DH10β, and plated onto solid LB growth media containing (i) 50 µg/mL or 100 µg/mL ampicillin (selection/positive control), and (ii) both 50 µg/mL ampicillin and 50 µg/mL kanamycin antibiotics (negative control). The expectation was that there would be no growth on the latter set of plates due to the recombinase being in its inactive state, which should prevent the promoter within the switch from being inverted and allowing transcription of *kanR*. This was not the case. Despite no apparent mechanism for expressing the relevant antibiotic resistance, the bacteria were able to grow in the presence of kanamycin. To allay

the possibility that the kanamycin was old or inactive, fresh stocks of the plates were made, but the same conclusion was reached: the strain bearing plasmid [005] was already kanamycin-resistant (data not shown).

Earlier Sanger sequencing results, performed during the construction of plasmid [005], had previously confirmed that the amber stops were in place within the recombinase and that there had been no indication that the *attB/attP* switch had been prematurely inverted (data not shown). Further sequencing of the Inversion-Kan^R [005] plasmid recovered from bacteria of both the ampicillin, and the double-antibiotic, plates subsequently showed that the constructs were still as designed: the recombinase was inactive, and the inversion switch had not yet been irreversibly flipped to allow *kanR* transcription.

To fully exclude the possibility that the inactive recombinase could have inadvertent effects on *kanR*, the $\phi c**$ gene was completely removed (using iPCR with primers YKH213/YKH289. In this way, there could be no possibility of the recombinase activating the switch and allowing transcription of the *kanR* sequence. This deletion had no effect on reducing or preventing the unexplained Kan^R expression (data not shown), and after subsequent attempts to troubleshoot the unexplained Kan^R signal (see Section 4.2.3.1 for discussions on alternative mitigation methods), this line of investigation was discontinued.

The remaining potential explanations for the continued Kan^R expression lay in the gene itself and the regions surrounding the 5'-end of the operon. For example, there could have been a spontaneous mutation or some underlying leaky expression, due to the selection pressure of kanamycin and/or a design flaw, that was exploited by the bacteria to allow expression of the resistance marker regardless of a viable promoter. Alternatively, despite having designed *kanR* to face the opposite direction to all other known operons within the [005] plasmid, transcriptional read-through could have occurred via an unknown, cryptic promoter at the start of the *kanR* sequence (around the synthetic RBS and the *attB-pAmpR-attP* switch) that was driving Kan^R expression without needing the recombinase to invert the switch.

In an attempt to identify the sequence that was initiating *kanR* transcription, the

operon was compared against other plasmids that contain the same resistance gene: i.e. pET29a(+) [P003], pUC57-Kan [P002], and pD881 [P008]. Although there were differences in the coding sequences, the protein sequence between the Kan^R variants were the same, with the exception of [P008] which lacked residues C10 and S11. The more pertinent differences, however, lay in the upstream region of the CDS: two had promoters (*pAmpR* for [P002] and *pCAT* for [P008]) directly before the transcription start site (TSS), suggesting that *kanR* does require a promoter for expression. In regards to the upstream portion to *kanR* in [P003], it was less clear: there was an RBS site and what could be a non-consensus promoter (consisting of a putative -10 Pribnow box and -35 region), but this was not clearly annotated. In spite of the uncertainty around the promoter region in the latter vector, it appears that *kanR* does not typically contain a promoter within its coding sequence. Hence, the mostly likely reason for the unexpected Kan^R expression was due to a cryptic promoter somewhere upstream of the CDS.

Given the introduction of the *attB/attP* inversion switch, a synthetic RBS, and an auxiliary `NdeI` restriction site directly upstream of *kanR*, it seemed prudent to first interrogate these sequences for a hidden promoter. It is possible that a promoter-like sequence, sufficient to transcribe *kanR*, could have been created when these disparate elements were combined. Whilst highly unlikely, it could also be possible that the well-utilised *pAmpR*, enclosed within the inversion switch, could in fact be a bidirectional promoter. To determine whether there may have been a cryptic transcription initiating element, bacterial promoter prediction tools, BPROM (`http://www.softberry.com/berry.phtml?topic=bprom&group=programs&subgroup=gfindb`; Solovyev and Salamov, 2011) and bTSSfinder (Shahmuradov et al., 2017) were used for analysing the 5'-sequence to the *kanR* CDS. Whilst there may have been possible promoter sequences with a *cis* orientation to Kan^R within this engineered region, there was no consensus between the prediction software on a particular site (data not shown).

As it was still unclear as to whether, and where, there was an upstream cryptic promoter to the *kanR* gene within the [005] construct, three parallel strategies were

developed to resolve this problem. First, to rule out the possibility of an upstream cryptic/bidirectional promoter to *kanR*, two alternative sequences were designed as replacements to the original switch (shown by [002] as simply *pAmpR* contained within *attB/attP*). The two subsequent designs for troubleshooting this problem were constructs: [007] TP, Terminator Plus a "decoy" RBS just upstream of the *pAmpR*, and [010] TP-Δ$pAmp^R$ as the former TP design minus *pAmpR*. These variants are shown in the far-right yellow box in Figure 4.11, and are further discussed in Section 4.2.3.1. In a similar vein, another solution would have been to subclone the *kanR* operon (including the synthetic RBS and downstream *lambda t0* terminator; 964 bp) into a vector with no known promoters to see if it still expresses. This approach was not pursued as it was not specifically furthering the inversion switch construct, but it could have indicated whether an isolated seemingly promoter-less *kanR* operon would have been expressed, and therefore may have provided further evidence to suggest that there was a hidden promoter around the switch sequence.

In the event where these alternate sequences still do not abolish the Kan^R activity (despite the non-functioning recombinase, ϕc**), and/or the marker is deemed too essential for the bacteria's viability, the second strategy was to replace the whole *kanR* operon with a less necessary reporter, such as sfGFP (shown as a generalised example in Figure 4.5, further detailed in Section 4.2.4). Given the nature of this proposed solution, the design and construction of this new secondary reporter for the inversion switch was implemented in parallel with troubleshooting the cryptic promoter strategy as described previously. Whichever approach that was first to give no expression of the proxy reporter would be continued.

Finally, should combinations of the aforementioned two strategies prove to be non-viable, the whole GTS inversion switch and secondary reporter system would be replaced by the GTS excision circuit. In this case, the initial design has the proxy reporter as the chromoprotein tsPurple, within an operon that contains a double-terminator excision switch in between the promoter and the RBS. The expectation is that the terminators would be sufficient in preventing tsPurple from being transcribed, but it is also possible that there may still be leaky expression, which could

make it difficult to differentiate the required signal from the background noise. The design of this approach is generalised in Figure 4.5, specified in Figure 4.17, and discussed in Section 4.2.5.

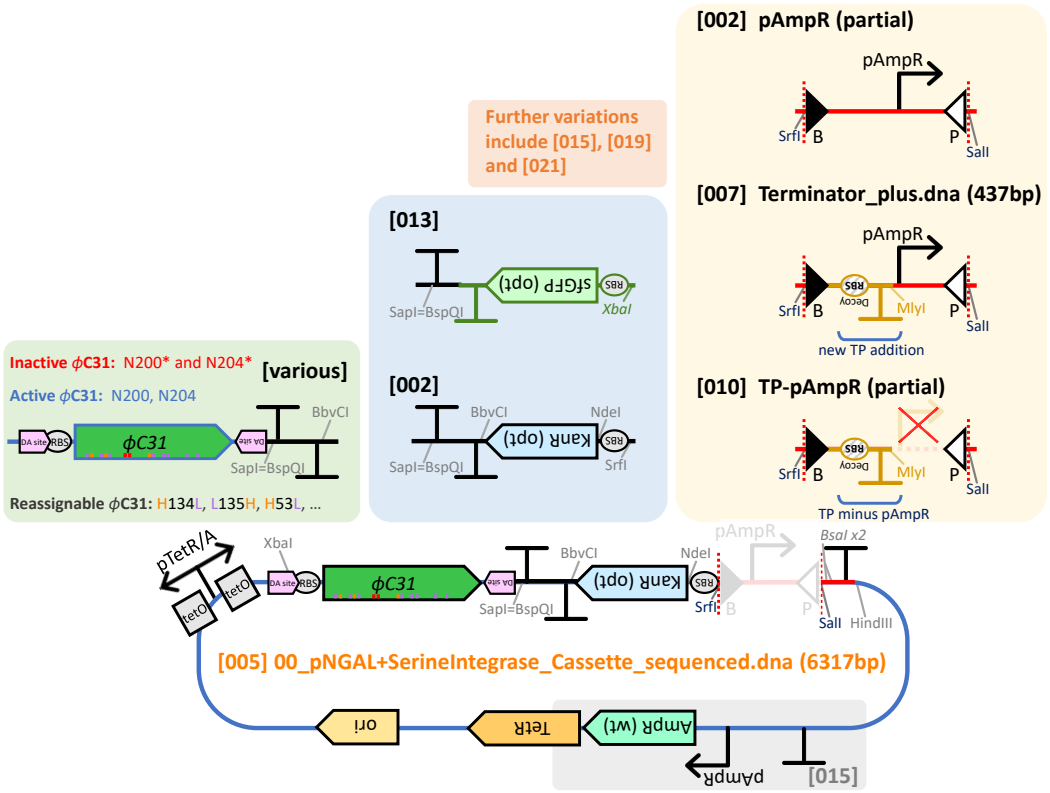


Figure 4.11: Variations upon the GTS Inversion constructs and its modular derivations

Plasmid [005] (construction previously shown in Figure 4.10) can be further modified as follows: in the green box, there are examples of alternative histidine-leucine recodings that are explored in Section 4.2.6; in the blue box, two secondary/proxy reporters for the Inversion circuits are shown ([002] is discussed in Section 4.2.3, whilst [013] is detailed in Section 4.2.4); in the yellow box, specific switch designs are shown (as explained in Section 4.2.4.1). Additional variations on this basic prototypical design were discussed: [019] and [021] are expanded further in Figure 4.13.

Although each of the designs could be systematically tested in the order as described, each could also be assessed independently from one another: with the path to integrating the GTS with the OCI constructs being determined by the strategy that most sufficiently achieves the desired stringent switch control. In reality, all three troubleshooting strategies were designed and implemented in parallel, but for ease of clarity, the investigation into each strategy follow as described.

4.2.3.1 Troubleshooting the *attB/attP* switch in the GTS Inversion-Kan^R construct – the GTS_v2 designs

A preliminary strategy for rectifying the unexplained Kan^R expression in the uninduced [005] GTS Inversion-Kan^R plasmid was to replace the internal sequence in between *attB* and *attP* with sequences that could theoretically terminate errant transcriptional signals directed towards *kanR*. As explained previously, two sequences were designed to replace the *pAmpR* switch: (i) [007] "Terminator Plus" (TP), which consisted of a "decoy" RBS, a terminator, and the original *pAmpR* sequence, and (ii) [010] "TP minus *pAmpR*" (TP-Δ*pAmp*^{R}) was built upon [007] by removing the promoter. These two alternative switches, [007] and [010], formed a part of the GTS_v2 designs discussed in this section (Figure 4.11, yellow box).

The [007] TP switch itself was composed of four features that, in theory, should be able to silence any read-through transcription from the hidden promoter. Reading from the direction of *attP* (towards the *kanR* sequence), the [007] design had the antisense sequence of *pAmpR*, followed by an engineered *rho*-independent *lambda t0* terminator (Scholtissek and Grosse, 1987) and a poly-T run. Whilst transcription terminators are not 100% efficient, the *lambda t0* variant is one of the highest efficiency prokaryotic terminators and, in combination with the poly-T tail, there should be sufficient destabilisation of the transcription complex and cleavage initiation to disrupt most of the unwanted read-through transcripts coming from the unknown promoter towards Kan^R. Following the poly-T run is the "decoy" RBS; this uses the consensus 5'-AGGAGG-3' sequence to initiate early translation of any rogue transcripts that bypass the terminator. The reason for allowing premature translation is due to its association with the final design element where stop codons were built into each of the six reading frames following the translational decoy. Theoretically, this should promptly truncate translation of all early, undesired transcripts before any can encounter the true (synthetic) RBS and *kanR* gene. Hence, translation of *kanR* should only come from the synthetic RBS, which in turn should only be expressing transcripts that result from the *pAmpR* that has been correctly inverted by a functioning ΦC31 recombinase. These new modifications resulted in an

extension of the internal *attB/attP* switch sequence from the 105 bp of the *pAmpR* promoter in [005] to a 272 bp sequence in [007] with redundant features that were expected to greatly reduce, or stop altogether, the Kan^R expression that was being produced even in the absence of the GTS Inversion-Kan^R switch being triggered.

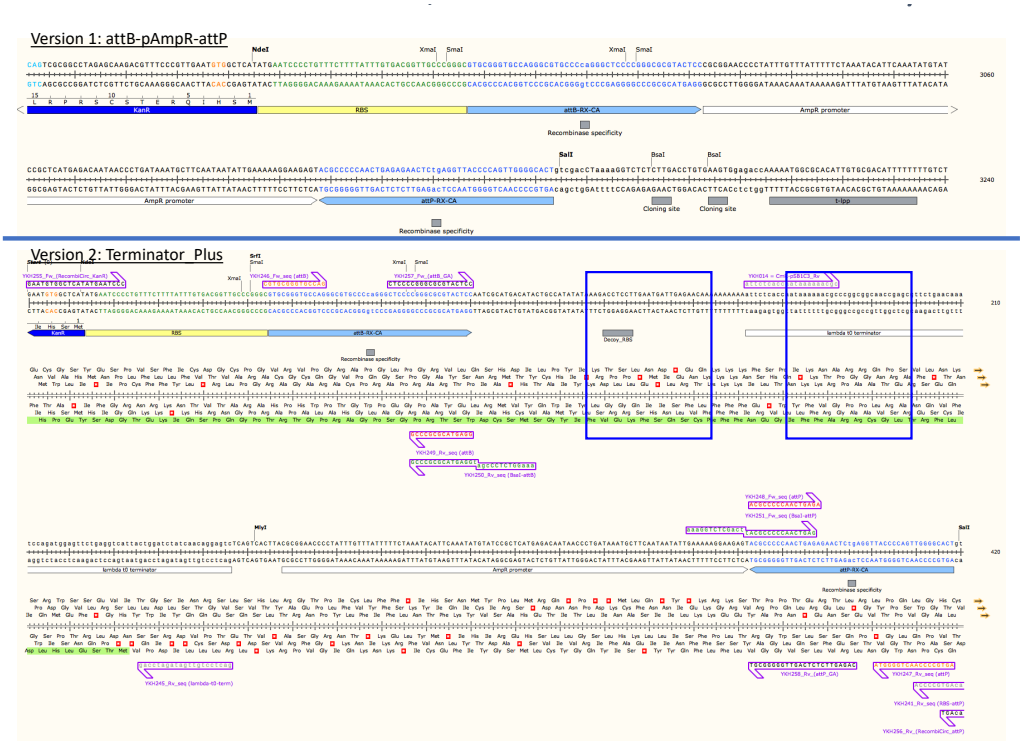


Figure 4.12: Comparison of [005] *pAmp^R* vs [007] TP sequence that was synthesised

Additional *in silico* analysis was performed as a means to identify any cryptic promoters. Results suggest that there may have been multiple promoter possibilities within the synthetic TP sequence but no specific consensus (data not shown).

In order to replace the switch in plasmid [005] with this new TP sequence, flanking ends that were homologous to where the switch would be inserted were placed around the new sequence to form the final 437 bp [007] fragment (Figure 4.12). This was subsequently synthesised as a double-stranded IDT gBlock®. The design was such that it could be inserted into [005] via Gibson Assembly (after iPCR to remove the *pAmpR* switch), or by classic restriction enzyme cloning using (`NdeI` or `SrfI`)/`SalI`. It was with the latter `SrfI`/`SalI` restriction cloning method that was taken to replace the *pAmpR* switch with that of TP, giving the resulting [008] `01_CnfKP+TP_6484bp.dna` (6,484 bp) plasmid.

Once sequence-verified, the [008] TP plasmid was used to make the construc-

tion carrying the third alternative switch: [010] TP-Δ$pAmp^R$. This latter plasmid was created by excising the *pAmpR* from the TP circuit using iPCR with primers YKH260/YKH248. The intention of this TP-Δ$pAmp^R$ switch variant was to be the negative control to TP; specifically, to assess whether the undesired transcriptional read-through was coming from the *pAmpR* region, which was hypothesised to hold an unknown bidirectional promoter.

As with the [005] *pAmpR* switch plasmid before, both [008] TP and [010] TP-Δ$pAmp^R$ were transformed into cloning strain DH10β, and plated again onto LB agar plates containing (i) 50 µg/mL or 100 µg/mL ampicillin (selection/positive control), and (ii) both 50 µg/mL ampicillin and 50 µg/mL kanamycin antibiotics (negative control). The expectation being that, in plates with kanamycin, the mitigating features introduced in [008] would result in minimal-to-no colony growth, and even less so in [010]. Yet despite such precautions, all strains containing the three GTS Inversion-Kan^R switch constructs (*pAmpR*, TP, and TP-Δ$pAmp^R$) generated thus far were still able to grow on media containing kanamycin (data not shown).

With hindsight, a similar test could have been achieved by simply removing the *pAmpR* from the original [005] construct. The inclusion of the extra TP switch elements (and subsequent selective removals in TP-Δ$pAmp^R$) arguably obfuscated the analysis unnecessarily in search of the cryptic promoter. That is, if the excision of *pAmpR* from the [005] plasmid still allowed Kan^R expression, then likely the transcription signal originated in the remaining switch and/or the *kanR* upstream elements, `NdeI` and synthetic RBS. The product of the potentially useful restriction site and an *in silico* prediction that had not previously been experimentally validated, could have very feasibly, although unintentionally, generated a cryptic promoter. Had this been this case, a new RBS and linker to the *attB* sequence would have been required.

At which point, combined with the expression problems already encountered with the multiple switches for the Kan^R proxy reporter, it seemed reasonable to select an alternative secondary reporter operon for the GTS system. Perhaps the

synthetic RBS and variant switch designs still harboured (or created anew) transcription initiation sequences that could not be easily eliminated, or that Kan^R was simply too essential given the selective environment – regardless, the reasons for its continued expression remain undiscovered. Given that this Inversion-Kan^R reporter system was not fit for purpose, I proceeded with the second troubleshooting strategy[4]: replacing the *kanR* operon with an alternate, and potentially less essential, operon. As such, we come to the GTS Inversion-sfGFP circuit (Section 4.2.4 – the GTS_v3 designs).

4.2.4 The Inversion-sfGFP Circuit

The fluorescent protein, sfGFP, was chosen as a suitable alternative reporter for troubleshooting the leaky Kan^R expression in the GTS Inversion-Kan^R circuit, for much the same reasons as when it was selected to first characterise the OCI strategy (Section 3.2.2.1): it is a stable, fast-expressing visual marker, and (particularly beneficial in this case) non-essential for host viability. Unlike Kan^R, which could be necessary for *E. coli* survival, expression of sfGFP has no such selective pressure (though equally no realistic need for the host to retain this added burden if the auxiliary antibiotic resistance marker, in this case Amp^R, is not required). Theoretically, as there would likely be less demand on the system to disproportionately express sfGFP, it should be easier to adhere to the designated circuit design: that is, to trigger the fluorescent protein only when the recombinase is active (rather than regardless of ΦC31).

In addition to replacing the Kan^R reporter, its associated synthetic RBS was also replaced. By doing so, it potentially bypasses an issue that was briefly raised in Section 4.2.3.1 wherein it was hypothesised that the cryptic promoter may have been located within the *in silico* designed RBS, not the *pAmp^R* within the *attB/attP* internal switch sequence. Whilst the RBS and surrounding region could have been the underlying explanation to the leaky Kan^R expression in the previous circuits, it still seemed prudent to test the three different switches ([005] *pAmp^R*, [007] TP,

[4]Work on this second strategy had begun alongside developing this first switch replacement strategy. Similarly, the necessary components for implementing the third strategy – from a GTS Inversion circuit to an Excision one – had been ordered at this point.

[010] TP-Δ$pAmp^R$; as designed in Section 4.2.3.1), with the new GTS_v3 Inversion-sfGFP design. It may be that with the change in RBS and reporter, the original promoter-only ($pAmp^R$) switch would in fact be sufficient control for the sfGFP secondary reporter. Equally, if the $pAmp^R$ switch did hide a cryptic/bi-directional promoter, the new Inversion-sfGFP context may provide another opportunity to assess whether the redundant terminating elements in the TP switch would be able to mitigate any undesirable basal expression, or have it eliminated altogether with the removal of the promoter using the diagnostic TP-Δ$pAmp^R$ construct.

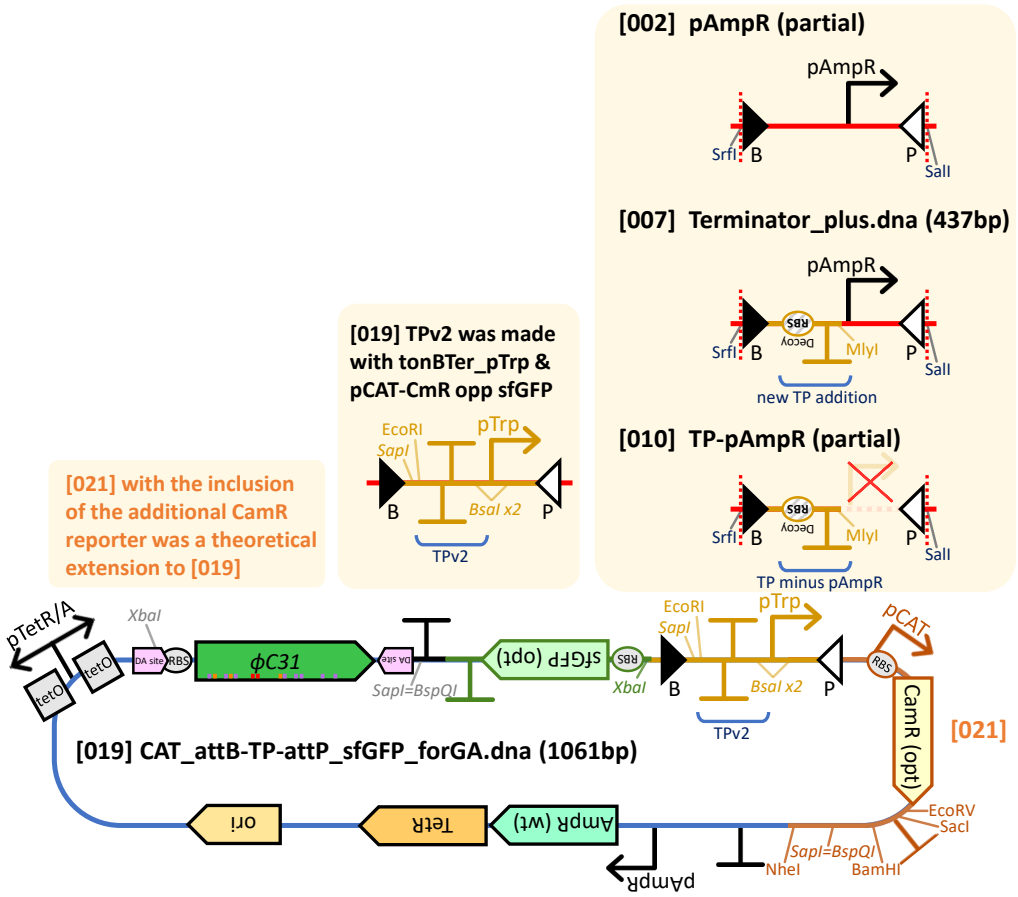


Figure 4.13: Additional variations on the basic GTS Inversion circuit design with constructs [019] and [021]

[019] TPv2 extended the design from Figure 4.11 further, and was used as an alternative switch design should it be needed.

Given that the three switches had previously failed to prevent Kan^R expression in the Inversion-Kan^R circuits, as a further precaution, I designed yet another switch – [019] TPv2 – that could be used in the Inversion-sfGFP construct as an

alternative contingency plan. [019] is a 1,061 bp linear fragment (synthesised by Twist Biosciences) that, when combined with the Inversion-sfGFP, would become a 7,053 bp plasmid [021] `Cnf_sfGFP_attB-TPv2-attP-CamR_7053bp.dna` (Figure 4.13). In the three former designs, transcriptional control was dependent on a (weak) constitutive promoter, TPv2 instead provides another layer of control: by incorporating an inducible promoter from the tryptophan operon, *pTrp* (Yanofsky et al., 1981; Yansura and Henner, 1990; Bass and Yansura, 2000), it may be possible to further regulate the secondary reporter. However, using *pTrp* would also mean adding and optimising for an additional inducer, 3-β-indoleacrylic acid (IAA) in this case, within the bacteria growth media. Learning from the Inversion-Kan^R switch designs, an additional *rho*-independent transcription terminating sequence, the bi-directional *tonB-P14* terminator (which has an estimated *in vivo* efficiency of 95% in the *tonB* direction, and 70% for *P14*; Postle and Good, 1983, 1985) was also included within [019] as an upstream element to the *pTrp* promoter. This was in anticipation of any potential cryptic (forward or reverse) read-through expression that might result from the TPv2 design[5]. In this instance, a variant of the decoy RBS was not included.

A final independent design element introduced alongside the construct was a tertiary reporter for the GTS Inversion-sfGFP circuit, which came in the form of a *cat* operon minus its *pCAT* promoter[6]. This operon was placed outside of the switch and downstream of the *attP* sequence (Figure 4.13, orange outlined features; similar also to the grey CDS portion within the general inversion circuit design from Figure 4.5). Once assembled together into plasmid [021], this triple-reporter construct (with its inclusion of the [019] TPv2 and CAT reporter variant of [210]) would become a possible first amalgamation of the GTS and OCI systems. Moreover, it

[5]The general concept of the TPv2 design was used again, and is further elaborated, in Section 4.2.5 with the design of the Excision-tsPurple constructs, albeit in a slightly different format (construct [022] and subsequent plasmid [033]).

[6]In the OCI [210] CAT plasmid, which was used to design the [019] TPv2 construct, the 104 bp *pCAT* promoter sequence included an RBS at its 3'-end, just before the TSS of the *cat* gene. In the [019] variant, 85 bp of the 5'-consensus sequence end of *pCAT* was removed as the intention was for transcription to be reliant on the *in cis pTrp* promoter within the *attB/attP* switch. The transcriptional terminator was also changed to that of *lambda t0*, instead of the *T7* terminator in the OCI version. The RBS and CDS remained unchanged.

would show the robustness and replicability of the OCI approach that was originally demonstrated in Section 3.2.4 in a slightly different context as the recoded reporter(s) and engineered orphan codon-invading anticodon-modified tRNAs would be on separate plasmids. Thus, in the event where any of the former alternative switch designs are able to demonstrate that the GTS Inversion-sfGFP circuit is robust and not leaky, this *cat* portion of the combined construct could potentially be used to further challenge the extent of the sense-to-sense codon reassignment platform. Supposing such a circuit can be implemented, this first OCI-GTS system (Figure 4.13) would proceed as follows: when induced, the *pTrp* should allow CAT expression under a normal untriggered switch state, whilst providing an opportunity to fine-tune an approximate range of inducer concentration required to transcribe its downstream gene (*cat* in the first instance, but *sfgfp* when the switch is inverted). If both CAT and ΦC31 are recoded to their inactive variants (i.e. under the rule of the alternative histidine-leucine code), and when paired with the rescuing codon reassignment tRNAs, it may be possible to see a change in signal from correctly translating CAT to producing the necessary ΦC31 for flipping the *pTrp* switch, and the expression of sfGFP instead[7].

4.2.4.1 Constructing the Inversion-sfGFP switch variants

To make the Inversion-sfGFP switch constructs, a 952 bp *sfgfp* operon fragment [012] was PCR amplified from [071] `00_sfGFP_pET29a_sequenced_5959bp.dna` (using primers YKH253/YKH223), and subcloned (via `BsaI` restriction) into the Inversion-Kan^R plasmids [005] ($pAmp^R$), [008] (TP), and [010] (TP-Δ$pAmp^R$): replacing the *kanR* operon in these three switch constructs with that of the *sfgfp* operon [012]. The three resulting Inversion-sfGFP switch circuits – [013a], [013b], [013c] respectively (refer to Figure 4.11 for their modular combinations) – were subsequently transformed into DH10β and grown on LB agar plates containing

[7]With hindsight, a better designed experiment here would actually be to not use an antibiotic resistance marker (given the troubles seen with Kan^R, and that it would be hard to implement a change in selection media with/without chloramphenicol), and instead use some other fluorescent or chromogenic protein, e.g. *mCherry* or *mScarlet*. In this manner, it would be easier to measure the activity of the engineered tRNAs by quantifying the ratio of red-to-green fluorescence over the time period of when the tRNAs are induced. For this to be implemented, a new reporter would simply need to be subcloned in place of the *cat* in the [021] design.

100 µg/mL ampicillin.

The TPv2 design was not built at the same time as the other Inversion-sfGFP circuits, partially as its construction was dependent on the former constructs, and partially as it was a contingency design in the event the [013] series also prove to be leaky. Had it been necessary, the steps to construct the [021] TPv2 plasmid, first requires *kanR* to be replaced with *sfgfp* (i.e. upon creation of any of the [013] plasmid series) as this would provide the requisite secondary reporter, as well as remove two of the three `SmaI` sites within the plasmid that would complicate the downstream step of changing the switch to the [019] TPv2 design. This would subsequently make it easier to linearise (and excise the existing switch sequence) with `SmaI/HindIII` and allow the synthesised [019] fragment to be subcloned in via Gibson Assembly. At the time, the creation of the TPv2 plasmid was put on hold, pending the assessment of the leakiness of sfGFP in the uninduced and inactive recombinase [013] plasmid series.

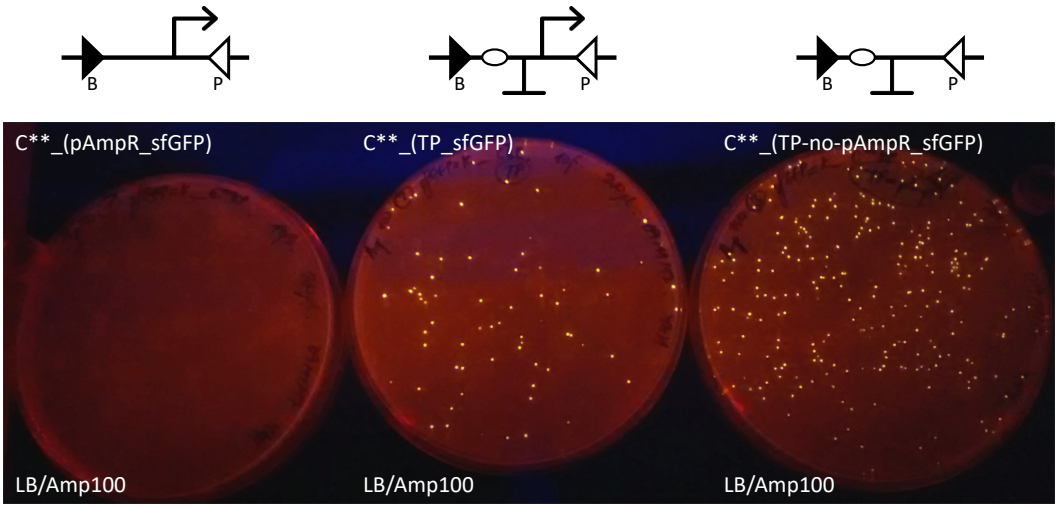


Figure 4.14: Comparison of the *pAmpR*, TP, and TP-Δ$pAmp^R$ switch constructs in the GTS Inversion-sfGFP circuits (plasmids [013a], [013b], [013c] respectively)

Contrary to expectations, it was the original *pAmpR* switch design (left plate) that should no fluorescence, whereas the subsequent TP (middle) and TP-Δ$pAmp^R$ (right) switches designed to reduce expected basal fluorescent expression showed increasingly leakiness.

As mentioned, the recombinase within the GTS constructs was still non-functional (ϕc**) at this stage. Thus, the new set of Inversion-sfGFP plasmids should theoretically be unable to allow transcription of the *sfgfp* gene, assuming

no cryptic promoters and/or other leaky expression within the three switch variants ([013a], [013b], [013c]). It was somewhat surprising, therefore, to discover that the switch that fared the best within these new constructs was the original one containing just $pAmp^R$ ([013a]): it did not appear to produce any background sfGFP expression (0/195 (0%) green CFUs; Figure 4.14, left). By contrast, the [013b] TP switch construct, which had been designed with both a terminator and decoy RBS to specifically prevent cryptic or reverse transcription from the $pAmp^R$ promoter, produced a fair amount of sfGFP expression (65/155 (~42%) green CFUs; Figure 4.14, middle). Moreover, the removal of $pAmp^R$ (within construct [013c] TP-$\Delta pAmp^R$) only served to generate an even greater degree of leakiness (236/384 (~61%) green CFUs; Figure 4.14, right). It could only be surmised that the designs intended to mitigate leaky transcription were not fit for purpose, and that the $pAmp^R$ promoter may even have helped curb some of the errant transcription signal stemming from the extra TP elements.

As it was, only the [013a] $pAmp^R$ construct was able to satisfy the requisite condition of producing minimal-to-no detectable level of sfGFP in the inactive (ϕc**) form of the Inversion-sfGFP circuit. At this point, it may have been possible to also create and test the potential leakiness of the [019] TPv2 circuit, but it seemed more time-efficient to proceed with the circuit that works as required. Plasmid [013a], therefore, became the design that was used in the next stage of the project: reverting ϕc** back to wild-type to examine the recombinase portion of the circuit (Section 4.2.4.2).

4.2.4.2 Characterising the GTS Inversion-sfGFP circuit with a restored ΦC31 recombinase

Having identified the GTS Inversion-sfGFP construct with the $pAmp^R$ switch [013a] as a suitable negative control baseline (i.e. with an inactive ΦC** that gave minimal downstream sfGFP expression), it was necessary to also establish the level of activity for the positive control (i.e. one with a functional ΦC31). Specifically, the primary aim of reverting the recombinase back to its functioning form was to confirm that ΦC31 was capable of inverting the *attB/attP* switch and allowing the

enclosed [013a] to initiate transcription of the *sfgfp* reporter construct (ideally to a quantifiable level of expression). A secondary objective was to assess how stringent the *pTetR/A* control was on the recombinase, and determine how much anhydrotetracycline (aTc) inducer would be required to activate ΦC31 and flip the [013a] switch.

To revert ϕc^{**} back to its functional $\phi c31$ form (a circuit later designated [013ai] ΦC31_$pAmp^R$), [013a] underwent iPCR with phosphorylated primers YKH215/YKH216 to correct the two nonsense stop codons back to asparagines. Although different concentrations of the aTc inducer (ranging from 0.0-0.2 µg/µL; Figure 4.15) was used, it seemed that even in the absence of the inducer, the *pTetR/A* control was leaky enough to allow some basal transcription of $\phi c31$, which in turn was sufficient enough to flip the $pAmp^R$ switch and express sfGFP; approximately 22/69 (32%) of the CFUs showed green fluorescence.

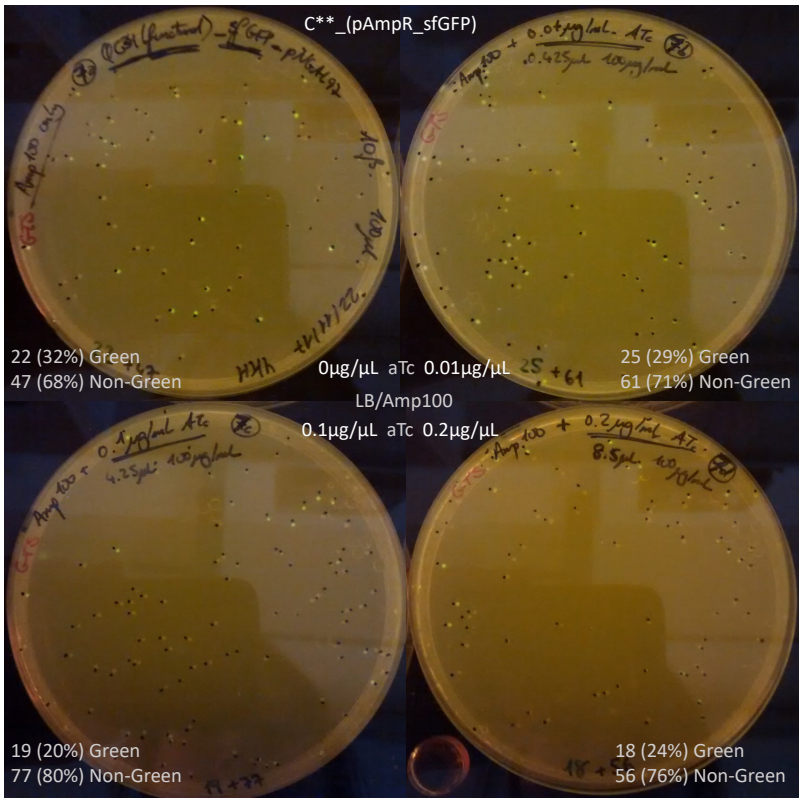


Figure 4.15: The active form of ΦC31 within the Inversion-sfGFP circuits appeared to show expression of the proxy sfGFP reporter even without aTc induction

The success of the recombinase reversion back to its functioning gene sequence

(ϕc** to *$\phi c31$*) was further confirmed through Sanger sequencing, wherein plasmids from the green fluorescent CFUs showed the expected changes both to the recombinase and the inverted switch. Unfortunately, the traces across constructs from a selection of non-fluorescing colonies also revealed that the iPCR approach taken to correct the recombinase had not been applied consistently. Although the two stop mutations had been corrected in most cases, some recovered [013ai] plasmid preparations had additional point mutations in the recombinase, whilst others possessed large deletions (including that of the *sfgfp* and *attB/attP* switch sequences). The high proportion of unforeseen errors within these constructs implied that: (i) most of the non-fluorescent colonies likely were incorrect re-assemblies; (ii) active recombinases were both sensitive and fast-acting, needing little to no induction for the putatively stringent *pTetR/A* control sequence to trigger *$\phi c31$* and the downstream *sfgfp* secondary reporter; (iii) fluorescent colonies likely had the corrected *$\phi c31$*. It was possible that the primers and/or DNA polymerase used for iPCR of the relatively long 6,290 bp plasmid were inefficient, and as such another approach would be required in future attempts to correct the recombinase[8].

Despite the lack of homogeneous corrections to the recombinase gene and the leakiness of the *$\phi c31$* operon, this set of experiments clearly demonstrated that a functioning recombinase would be both active and sensitive (even without induction). It was, therefore, a good indicator that an inactivated recombinase (such as one that has been recoded to follow an alternate histidine-leucine code) could have its function restored if the OCI anticodon-modified tRNAs are successful in implementing sense-to-sense codon reassignments. Equally, the fluorescence from colonies with the (re-)gain-of-function ΦC31 suggests that proxy sfGFP secondary reporter would be a sufficient visible signal that codon reassignment has been achieved.

In summary, the results from developing the GTS_v3 Inversion-sfGFP con-

[8]That said, in the next step of the project – the OCI-GTS stage – the Darwin Assembly method (further details in Section 4.2.6) would be used to recode the the recombinase (the ϕc** variant, to minimise premature switch inversion) for the sense-to-sense codon reassignment mutations. The different approach to incorporating mutations, along with the awareness of potential cloning issues, should mitigate future problems.

structs have demonstrated that: (i) the basic $pAmp^R$ inversion switch is a compatible component that does not present leaky expression of sfGFP in its "off" state (i.e. when the recombinase is inactive, ϕc^{**}); (ii) an active ΦC31 recombinase is sensitive enough to trigger the GTS circuit into a permanent "on" state (i.e. to invert the $pAmp^R$ and allow sfGFP expression); (iii) sfGFP is a suitable secondary/proxy reporter for sense-to-sense codon reassignment as it is not reliant on the alternate codes, and its fast and visual expression pattern shows that reassignment has occurred. Given these attributes, this GTS Inversion-sfGFP design should pose minimal toxicity to the host when its later recoded variant is used to convert a potentially more toxic signal (ad-mix translation population of naturally- and alternately-coded ΦC31) to a permanent, non-toxic response (naturally-coded sfGFP). Thus, this design construct (or more accurately the ϕc^{**} "off" state version [013a]) was taken forward to be combined with the histidine-leucine inactivating residues as identified in ΦC31 (from earlier Section 4.2.1) to make the circuit that could be used to detect sense-to-sense codon reassignment. The process for making and implementing this histidine-leucine inactivated ΦC31 is reviewed in Section 4.2.6. However, before doing so, it may be useful to complete the investigation into finding a non-leaky secondary reporter for the GTS system lest the third and final troubleshooting strategy – the GTS Excision-tsPurple circuit (Section 4.2.5) – proves to be an even more robust design suited for codon reassignment.

4.2.5 The Excision-tsPurple Circuit

Although Section 4.2.4 identified a GTS Inversion-sfGFP design [013a] that could be used to characterise sense-to-sense codon reassignment *in vivo* it was also worth considering whether an Excision circuit (Figure 4.16) could be a viable, and complementary, design for characterising reassignment as well. (notwithstanding that both troubleshooting strategies were developed in parallel). An Excision design could be used in instances where removal of a set of sequences within a circuit could be more favourable than retaining them via inversion. For example, an Excision circuit could be designed around the gene sequences of the OCI anticodon-modified tRNAs as a form of self-eliminating regulatory feedback loop to limit excessive ac-

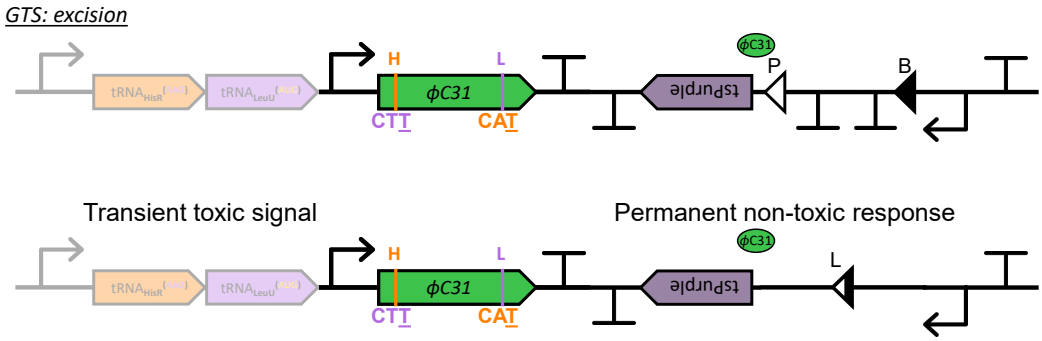


Figure 4.16: The prototypical GTS Excision circuit as seen in Figure 4.5

cumulation of mis-translated protein products within a multiple-coded host. That is, when expression of the engineered tRNAs reaches a sufficient level to activate the recoded recombinase, this would trigger the irreversible Excision switch (signifying that reassignment has occurred) to remove the gene sequences and prevent further production of the anticodon-modified tRNAs. Consequently, this could be a plausible model for demonstrating sense-to-sense codon reassignment *in vivo*, provide a self-monitored response to curtail the foreign anticodon-modified tRNA code, and allow the host to recover from having to endure multiple genetic codes – all without the need for direct external control. However, before contemplating further on yet another complex design[9] involving both the OCI and GTS strategies, it would be reasonable to first characterise an initial simplified model. This section, therefore, documents the creation and implementation of the GTS_v4 Excision-tsPurple design.

While the core recombinase component within the Excision circuit remains the same as in the [005] prototype plasmid, the switch and (optionally) the secondary reporter operon have a different architecture. Specifically, where the *attB* and *attP* switch sequences within the Inversion circuits oppose one another (*in trans*), these recombinase recognition sequences are oriented in the same direction (*in cis*) in the

[9] As with the [021] TPv2 OCI-GTS design discussed earlier (in Section 4.2.4 footnote 10), which also had potential in demonstrating a combination of the two systems, such an excision design would again require another proxy reporter as the *pTetR/A* control is insufficient in regulating a functional recombinase. That is, without an additional marker to signify reassignment has occurred, ΦC31 could feasibly be expressed and successfully excise the anticodon-modified tRNAs before its effects can be observed and acknowledged.

Excision design. This enables the portion within these sequences to be removed upon interaction with an active ΦC31 recombinase.

Such a design requires very stringent transcriptional regulation: if this secondary reporter is expressed in the absence of its recombinase trigger, it would be no better than the initial Inversion-Kan^{R} [005] construct to which this Excision circuit was designed to ameliorate. One theoretically plausible design that could provide the necessary tight regulatory control is the [019] TPv2 sequence that was developed during the earlier troubleshooting strategies (Section 4.2.4). The rationale being that its allegedly robust inducible promoter (*pTrp*) provides a degree of control to transcribe the downstream gene only when its inducer (IAA) is added, and the strong transcriptional terminator (the bi-directional *tonB-P14*) should truncate any leaky expression that may escape the control of the promoter[10]. It seemed reasonable, therefore, to co-opt these specific design features for the Excision-tsPurple construct.

Figure 4.17, displaying a linear 1,290 bp DNA segment [030], is the culmination of these considerations for the basic secondary reporter operon of the Excision construct. It consists of: a lead-in stretch of sequence that is homologous to its proposed insertion/replacement point in plamid [005], which also includes restriction sites `NheI`, `SapI`, `BamHI`, `XmnI` for alternate subcloning approaches; a strong, inducible promoter *pTrp* (Yanofsky et al., 1981; Yansura and Henner, 1990; Bass and Yansura, 2000); the Excision switch (i.e. the dropout fragment), beginning with *attB*, followed by the BioBrick BBa_B0015 *rrnBT1-T7TE* double-terminator composite part (`https://parts.igem.org/Part:BBa_B0015`), and closing with *attP*; the remaining portion of the reporter operon, which comprises of the consensus RBS sequence, a vibrant non-fluorescent chromoprotein TinselPurple (tsPurple, tsP; `http://parts.igem.org/Part:BBa_K1033905`; Liljeruhm et al., 2018), and the *lambda t0* transcription terminator. The construct is finished with a *t-lpp* terminating sequence, and one of the Darwin priming sites) to facilitate subcloning into its final destination plasmid [033].

[10]For further details on the TPv2 design justification and implementation, refer to the [021] plasmid designed (but not constructed) in Section 4.2.4.

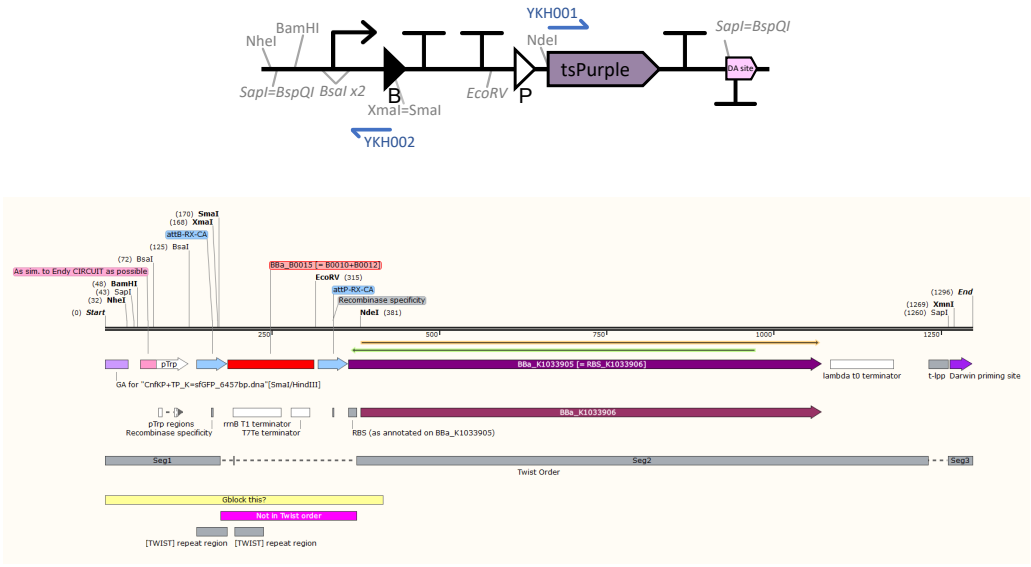


Figure 4.17: A schematic of the GTS Excision-tsPurple design

Refer to the generalised form of this construct in Figure 4.16, as well as the assembly into the full circuit in Figure 4.18.

As with the previous GTS circuits, [030] was designed to be modular in order to more easily facilitate any troubleshooting requirements. For example, two `BsaI` sites were planned around the *pTrp* promoter to enable the removal and exchange of promoters should *pTrp* prove to be too weak to induce transcription of its downstream gene or not stringent enough to prevent basal expression when not induced.

Similarly, the transcription terminators within the *attB/attP* switch could be replaced via restriction digest with the built-in `EcoRV` and `XmaI=SmaI` sites. This would permit the interchange of the proposed BBa_B0015 *rrnBT1-T7TE* double-terminator with the bi-directional *tonB-P14* terminator from [019] TPv2 (or other (double-)terminators from the relatively well-characterised iGEM Parts Repository (the Registry of Standard Biological Parts; `https://parts.igem.org/Main_Page`) should it be necessary. In the first instance, *rrnBT1-T7TE* was favoured over the bi-directional terminator as it was expected that the transcriptional signal would only be coming from one direction (the *pTrp* promoter side). Moreover, with a reported forward termination efficiency of at least 97% and a moderate reverse ter-

mination efficiency[11], it appears as stringent as *tonB-P14* and, therefore, should prove to be a suitable choice for the internal sequence within the *attB/attP* Excision switch.

TinselPurple (BioBrick BBa_K1033905; `http://parts.igem.org/Part:BBa_K1033905`) was chosen as the alternative secondary reporter, primarily as it exhibits a strong pink-purple colour when expressed and allows for instrument-free detection. This part originated from the Protein Paint Box collection of chromoproteins from DNA2.0 Inc. (now ATUM, CA, USA; `https://www.atum.bio/catalog/reagents/protein-paint-box`), and was part of a collection of eukaryotic chromoproteins that Forster and colleagues (2018) codon-optimised in order to produce a palette of *E. coli*-compatible BioBrick parts. In contrast to the fluorescent proteins used previously (where UV excitation, or fluorometry is needed for quantification), TinselPurple produces a pigment that can be viewed directly. As the intention of the GTS strategy was to provide a permanent, non-toxic proxy signal of sense-to-sense codon reassignment (assuming no prior leakiness in the system), it was not strictly necessary to quantify the strength of expression of the secondary reporter; merely the digital response (presence/absence) of TinselPurple was required. One minor note also is that the BBa_K1033905 part was technically a composite of TinselPurple and a non-specified RBS (but which encoded the consensus 5'-AGGAGG-3' Shine-Dalgarno sequence). Given that this composite part had previously been demonstrated to produce strong[12] and stable pigmentation (on solid agar and in liquid cultures; Liljeruhm et al., 2018), both parts were taken together for use as is. To finish, the *lambda t0* terminator was used, mirroring the secondary reporter operon in the Inversion-Kan^{R} [005] circuit.

[11] In the online documentation for BBa_B0015, two separate groups have assessed the efficiency of the double-terminator. One reports a forward efficiency of 98.4% and a reverse efficiency of 29.5%; the other measured it to be 97.0% and 62.0% respectively.

[12] It may be of relevance that Andreou and Nakayama (2018) reported the strength and speed of pigment development could be strain-dependant: with earlier/greater expression seen when chromoproteins were expressed in TOP10 (pigmentation seen after overnight culturing) rather than *E. coli* DH5α strain (NEB #) (DH5α) *E. coli* strains (after 24 hours). That said, GTS constructs (including the Inversion-tsPurple circuit) were typically transformed into DH10β, which have the same genotype as the TOP10 strain (as both are similar DH10β derivatives), hence should not display significantly different expression patterns.

Ultimately, the plan was for the Excision-tsPurple [030] fragment to be inserted (via Gibson Assembly) into the GTS prototype plasmid [005], and replace the original Inversion-Kan^R construct to make plasmid [033]. In this circuit, the recombinase should still be non-functional (ΦC**), thus would be treated as a negative control. If this construct expresses TinselPurple in the absence of a functional recombinase, then it (like the Inversion-Kan^R design) would be unsuitable for use in the GTS strategy, and the project would progress using the Inversion-sfGFP construct [013a] (*pAmp^R* switch; Section 4.2.4.2) as previously identified. However, if no expression is detected, it could also be taken to be a viable GTS circuit and the next step would be to revert the recombinase to its functional form and assess the extent to which triggering ΦC31 will have on activating the Excision-tsPurple circuit. This would rely on the use of aTc to activate ΦC31 expression, which should subsequently enable excision of the double-terminator that is interrupting the transcriptional signal driven by an IAA-induced *pTrp* promoter on the TinselPurple secondary reporter. If it can be demonstrated that ΦC31 can be used to switch on a downstream innocuous response in *E. coli*, it would then be possible to recode the recombinase in the manner (histidine-leucine replacements) as proposed in Section 4.2.1, for both Inversion-sfGFP and Excision-tsPurple designs, and thus implement sense-to-sense codon reassignment with minimal detriment (i.e. limited mis-translation from simultaneous decoding under the natural and synthetic codes) to the host organism.

4.2.5.1 Constructing the Excision-tsPurple switch variants

DNA synthesis of the 1,290 bp Excision-tsPurple [030] fragment, however, proved to be quite problematic, and necessitated minor adjustments to the design before construction could begin. The chief issues that made gene synthesis difficult were the presence of the three different transcription terminators (partial *t-lpp* on the 5'-end of the fragment, *rrnBT1-T7TE* within the switch, and *lambda t0* to terminate tsPurple), and two near-duplicate repeat regions (on both homologous flanking regions of the fragment that were to be used in subcloning).

To address these synthesis issues, the extra 29 bp *t-lpp* terminator downstream

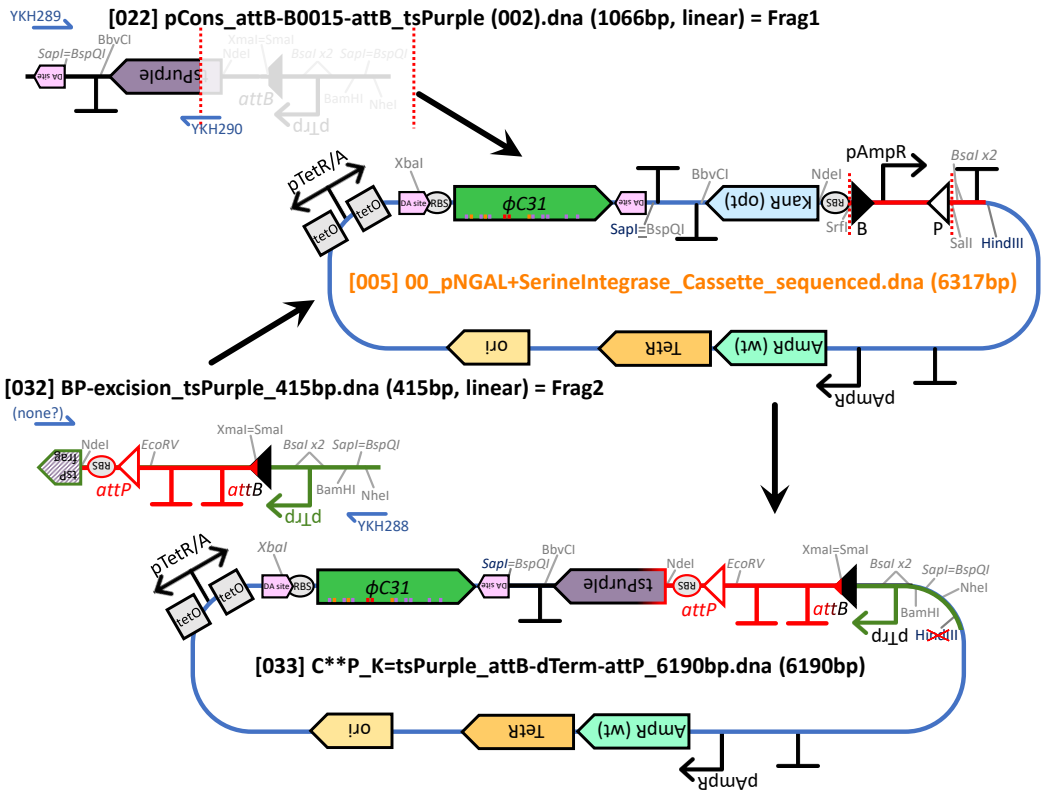


Figure 4.18: Construction of the GTS Excision-tsPurple circuit

The final [033] Excision-tsPurple construct resulted from an assembly of [022], [005], and [032].

of the TinselPurple operon was removed, and the Excision switch portion was separated from the rest of the fragment. The latter became a 415 bp switch fragment [032] that was ordered as a linear dsDNA gBlock® from IDT and the remainder of the former, a 1,066 bp segment [022], was sent to be synthesised by Twist Biosciences. Fragment [022] arrived in a "pTwist_Kan_MC" vector, becoming construct [023].

Given that the [030] design was separated into two fragments for synthesis, the final construction plan expanded into a three-part Gibson Assembly, involving the 415 bp gBlock® [032], an 891 bp PCR product of [023] (using primers YKH289/YKH290), and the original leaky Inversion-Kan^R [005] plasmid that had been digested with `HindIII/SapI`[13] (Figure 4.18). The richer 2xTY media was

[13] After restriction with `HindIII/SapI`, the digested [005] plasmid underwent gel extraction to isolate the required ~4,989 bp vector from the ~1,328 bp *kanR* operon and Inversion switch fragment. Doing so should have reduced the chances of re-ligation with the former fragments, and enabled greater success in producing the correct assembly of the new Excision-tsPurple [033] circuit.

used as *pTrp* is regulated by the availability of free tryptophan to the cell: high levels of L-tryptophan from the media was intended to help further repress transcription of the TinselPurple operon by activating the endogenous Trp repressors, and facilitate binding to the *trp* promoter/operator sequence, thereby blocking any transcription that may have bypassed the *rrnBT1-T7TE* double-terminator switch (Bass and Yansura, 2000). With no addition of the IAA inducer, increased repression of the *pTrp* promoter, and the presence of *rrnBT1-T7TE* interrupting the secondary reporter, the expectation was that this should be enough to prevent any basal *tsP* transcription. Moreover, as this Excision-tsPurple [033] plasmid still possessed an inactive recombinase gene sequence, ϕc^{**} (which itself was under control of another inducible promoter *pTetR/A*), no expression of TinselPurple should have been triggered from this part of the circuit.

Correspondingly, Figure 4.19 shows that no pink/purple pigmentation was detected in the CFUs on the agar plates post-transformation, nor in the colonies taken for culturing. Sanger sequencing of the subsequent plasmid preparations further showed that construct [033] had successfully incorporated the Excision-tsPurple fragments, and that the terminators within the *attB/attP* switch were still present within the recovered plasmid. This suggests that ϕc^{**} was still non-functional and unable to excise the switch, making plasmid [033] another suitable GTS circuit for taking onwards to test sense-to-sense codon reassignment.

4.2.5.2 Characterising the GTS Excision-tsPurple circuit with a restored ΦC31 recombinase

Having effectively made a suitable GTS Excision negative control circuit (i.e. containing the correct switch and secondary reporter, but coupled with a non-functional recombinase) in [033], the follow-up experiment was to create both the functional ΦC31 test-case and positive control plasmids. This was done in much the same process as when restoring recombinase function in the Inversion-sfGFP [013] constructs in Section 4.2.4.2. The ϕc^{**} recombinase was first reverted back to wild-type via iPCR with phosphorylated primers YKH215/YKH216, followed by self-ligation re-circularisation to create plasmid [033b], before being transformed into

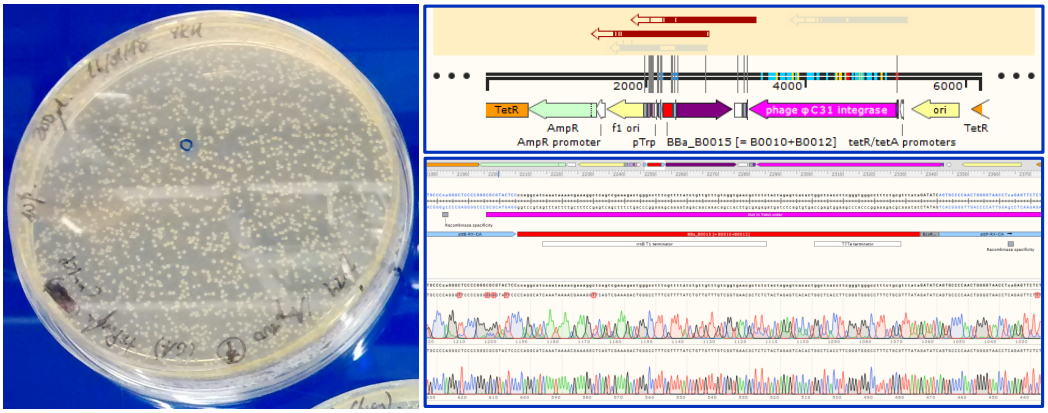


Figure 4.19: Plasmid [033] TinselPurple in DH10β and Sanger sequencing trace to show that when with ϕc** is inactive, there is no TinselPurple expression

Left: the circled colony (blue) was selected for Sanger sequencing and to revert the phage integrase PhiC31, inactive with double amber stop mutations (ΦC**) sequence back to the active wild-type ΦC31 form. Top-right: overview of the sequencing trace showing that there has been no spontaneous loss of the double-terminator switch, which corresponds to a lack of TinselPurple expression in the transformants. Bottom-right: sequencing trace at the DNA level showing the correct sequences around the switch.

DH10β.

Although the functionally-restored recombinase in the Inversion-sfGFP [013ai] construct had demonstrated that the *pTetR/A* control was not the most stringent (as the downstream sfGFP was expressed in the absence of aTc; Section 4.2.4.2, Figure 4.15), it was anticipated that the dual inducible promoter controls (*pTetR/A*, *pTrp*) together in the Excision-tsPurple approach could negate this. That is, even though the *pTetR/A*-controlled *$\phi c31$* operon was expected to also be leaky in the [033b] plasmid (sufficient to prematurely delete the double-terminator excision switch within the TinselPurple operon on its own), when coupled with *pTrp*, it was expected that the basal expression of the secondary reporter would be curtailed as this promoter would be without the required IAA inducer and in a tryptophan-rich (i.e. expression-repressed) environment. The theoretically unexcised and uninduced [033b] plasmid would, therefore, be the test-case variant (akin to a negative control), which when doubly-induced (i.e. to irreversibly excise the double-terminators and express the chromoprotein) should become the positive control construct [033c]. This would give a good indicator as to the level of TinselPurple expression expected when the GTS Excision-tsPurple is inactive and

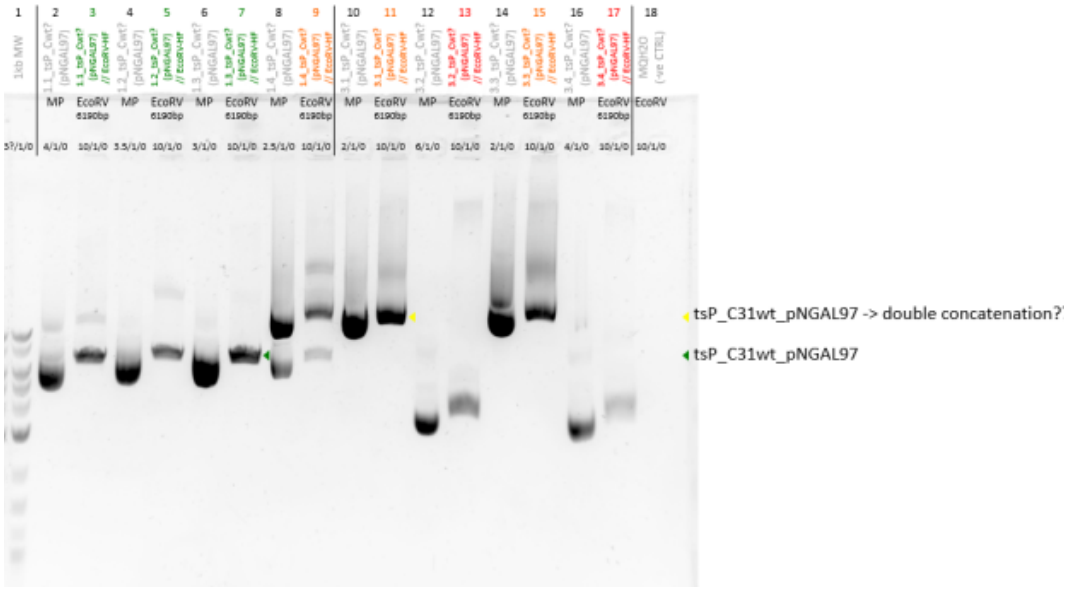
active respectively.

Despite appearing to display no detectable pigmentation when the ΦC31-restored [033b] transformants were plated on 2xTY agar plates (i.e. with no IAA induction), at least 3 of 8 subsequent overnight cultures of selected CFUs showed both pink colouration in the 2xTY liquid media, as well as in the cell pellets post-centrifugation (Figure 4.20a). Given that TinselPurple could be seen in samples #1.4, #3.1, and #3.3 at this point suggests that the dual-inducer promoter system is not as stringent as envisioned. The likelihood is that there was sufficient transcription bypassing the *pTetR/A* control on the functionally-restored ΦC31 recombinase – reaffirming the results shown in the Inversion-sfGFP circuits – but in this case resulting in the excision of the double-terminator switch. At the same time, there seemed to be insufficient repression of the *pTrp* promoter for *tsP*, coupled possibly with depletion of tryptophan in the 2xTY media in the cultures, to produce the premature chromoprotein expression.

As a first-pass screen to determine whether the 180 bp switch had been unintentionally excised, an `EcoRV` diagnostic digest was performed on the plasmid minipreps of the eight recovered [033b] construct samples. An `EcoRV` restriction site had been specifically designed within the Excision switch sequence as a fast diagnostic test (without the need for sequencing) to assess whether recovered plasmid constructs had lost the switch. Constructs that had retained the switch should be linearised, with an expected size of 6,190 bp. However, where excision has occurred, the plasmids should no longer have an `EcoRV` site, thus should remain in a circular configuration as the restriction digest would have no effect. The migration pattern for these was expected to run further than that of undigested control samples; both of which were expected to migrate further than the linearised fragment as they would likely present as coiled or super-coiled DNA structures.

Whilst, the agarose gel electrophoresis of the digested constructs were as expected (Figure 4.20b, lanes 2-7), there were also quite a number of unexpected results. The former suggests that samples #1.1, #1.2, and #1.3 were the expected and required construct [033b], having retained the Excision switch. This conclusion

(a) Chromoprotein expression from overnight liquid cultures of plasmid [033b] TinselPurple in DH10β



(b) EcoRV diagnostic digest of the plasmid preparations from the liquid cultures to screen for constructs that have retained the excision switch (expected size: 6,190 bp)

(c) Sanger sequencing traces of a subset of constructs from the EcoRV diagnostic digests showing the loss of the excision switches in three of the recovered plasmids

Figure 4.20: Assessing the stringency of the dual promoter control in a functionally-restored *ϕc31* recombinase on expression of the secondary reporter in the [033b] TinselPurple circuit from: (a) visual inspection, (b) restriction digests, and (c) Sanger sequencing

could be drawn for sample #1.4 (lanes 8-9) as well, however, this sample also suggested that there was a heterogeneous population of the recovered plasmid. Specifically, it seemed that the construct had doubled in size (estimated at 1.2 kb), implying that there had been a concatenation during the iPCR and re-ligation process to revert the $\phi c**$ to its functioning form. A similar pattern of concatenation was also visible for samples #3.1 and #3.3 (lanes 10-11, 14-15), and that the undigested samples ran slightly further than the digested ones indicated that at least one `EcoRV` site (and therefore one switch sequence) had been retained for the sample to be linearised. At the same time, there should have been a replicate within the concatenations that had lost the switch as these constructs showed TinselPurple expression; alternatively, the *pTrp* control was just leaky as well. It was safe to exclude samples #3.2 and #3.4 (lanes 12-13, 16-17) as these did not have the expected migration pattern of either the expected [033b] plasmid that had retained the Excision switch (lanes 2-9), nor that of a plasmid that had lost the switch prematurely to produce the pink-purple pigmentation (lanes 8-11, 14-15).

Given the results presented from the visual chromoprotein expression in the culture pellets, gel electrophoresis patterns from the diagnostics digests, and the lack of specific sequence-level information confirming the correct reversion of $\phi c**$ to *ϕc31*, the more promising plasmids recovered were taken forward for Sanger sequencing. As expected, the recombinase in samples #1.1 and #1.3 had correctly reverted to its functional *ϕc31* sequence and had retained the Excision switch within the *tsP* operon (Figure 4.20c). However, the sequencing also revealed that #1.3 had incurred a D107G point mutation in the *tsP* gene (data not shown), which may have rendered it inactive, even if the GTS system had been activated. As such, this latter sample was discarded in favour of #1.1 as the required [033b] construct. As for samples #1.4, #3.1, and #3.3, which had shown pink pigmentation in the cell pellets, the sequencing showed that they had indeed lost the 180 bp Excision switch. For #1.4, the heterogeneous pattern seen in the sequencing chromatogram also confirms the presence of an admix population. This could be attributed to the presence of the two bands for #1.4 in the agarose gel (Figure 4.20b, at ~6 kb and ~1.2 kb in lanes

8-9), which were assumed to be a double concatenation of the plasmid. However, if it is an exact duplication, the sequencing would not be able to differentiate this. The sequencing results for the presumably concatenated #3.1 and #3.3 samples (both with only bands at an estimated 1.2 kb, lanes 10-11 and 14-15) seemed to concur with this reasoning as neither displayed any heterogeneity in the chromatograms. Instead, upon closer inspection, sample #1.4 had an incomplete correction of the stop codon at residue 200 in the recombinase: it was deleted as opposed to being reverted to asparagine. It was likely this single codon frameshift that produced the admix sequencing results. Whilst interesting, these variations also precluded sample #1.4 from use in future experiments. With samples #3.1 and #3.3, given that these did have the expected functional sequences for the Excision-tsPurple plasmid (albeit perhaps as a concatenated construct), and had inadvertently provided the excised variants, these samples were kept on as potential [033c] positive TinselPurple controls.

In a final demonstration to confirm the functionality (and possible instability) of the recombinase portion of the [033b] Excision-tsPurple construct, this plasmid was cultured with and without 10 ng/mL aTc. Colony PCR (using primers that flanked the Excision switch) was performed on a dilution of each of the cultures, and the results were visualised on agarose gel (Figure 4.21). Construct [033] ΦC** was used as a negative control. A smaller amplicon (at 183 bp) was expected if the switch was excised (upon induction), whilst an unexcised switch would present as a 364 bp fragment. As can be seen, even without aTc induction, there was sufficient basal expression of the functionally-restored ΦC31 to cause the Excision switch to be dropped-out. With induction, the switch was completely removed from the circuit. Hence, this could also justify how the chromoprotein could be allowed to express, given the poorly regulated upstream recombinase.

To conclude, though the required [033b] Excision-tsPurple plasmid could be created – where the recombinase is functional but not yet been explicitly activated to excise the double-terminator switch stalling the secondary reporter operon – it was clear that the *pTrp* control for TinselPurple was not sufficiently stringent. Cul-

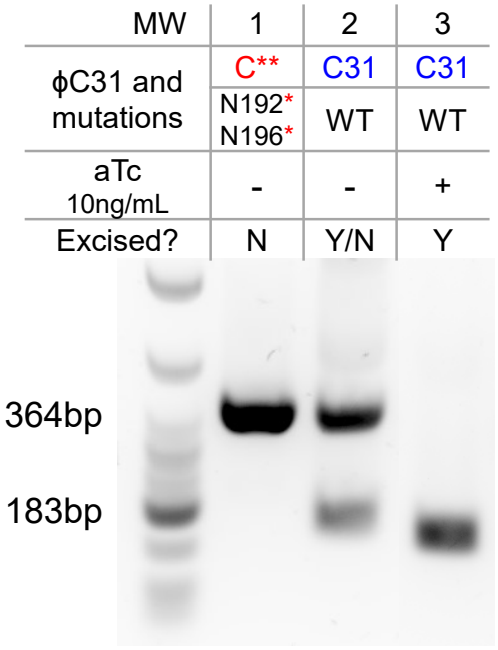


Figure 4.21: A colony PCR experiment to demonstrate both the functionality and instability of the functionally-restored ΦC31 recombinase within the [033b] Excision-tsPurple construct

If the switch is excised (whether spontaneously or through aTc induction), a fragment of 183 bp would be detected; if unexcised, the product would be 364 bp in size.

tures that had already lost the terminator switch (through the already established leaky *pTetR/A* control for the recombinase portion of the circuit) were able to express the chromoprotein, even in the absence of the IAA inducer. The basal expression from the uninduced, inducible *pTrp* promoter could, in effect, be treated as a weak constitutive promoter, much like the (true) constitutive $pAmp^R$ switch in the [013a] Inversion-sfGFP circuit. Essentially, the reason that TinselPurple was not expressing was due to the presence of the intervening double-terminators and a non-functional recombinase, rather than it being under the control of *pTrp*. However, regardless of the leakiness of the proxy reporter, it seems that in the negative control of both switch designs (the initially opposite-facing $pAmp^R$ in [013a] Inversion-sfGFP and the double-terminators within the [033] TinselPurple operon) the presence of a non-functional $\phi c**$ recombinase was sufficient in blocking transcription of their respective sfGFP and TinselPurple reporters. Thus, if the recombinase is non-functional – be it with stop codons ($\phi c**$), through histidine-leucine

recoding to render it incompatible with the natural code, or from tightening *pTetR/A* control even further – then there should be no measurable expression of the downstream proxy reporters, and hence this makes both circuits suitable for the next steps in the GTS approach.

4.2.5.3 Troubleshooting transcriptional control in the Excision-tsPurple circuit

At this point, we had two viable GTS circuits and two potential directions to further the project. As both Inversion-sfGFP and Excision-tsPurple were able to demonstrate sufficient off- (ϕc**) and on- ($\phi c31$) states, the next logical step would be to create the loss-of-function histidine-leucine recoded recombinase variants. For this, the histidine-leucine mutations that were identified in Section 4.2.1 would be used to create a library of alternatively-coded $\phi c31$ mutants in the GTS circuits. These would then be experimentally validated to identify which (combination) of point mutation candidates are able to disable ΦC31 function. The non-functional histidine-leucine recoded recombinases would become the off-state switches that would subsequently be combined with the OCI anticodon-modified tRNAs to determine whether the codon reassignment system can recover recombinase functionality with minimal toxicity to host viability (see Section 4.2.7). The other route the project could be taken in is to find alternative strategies to improve the stringency of transcriptional control within the GTS circuits. This could be in tightening the *pTetR/A* control over the recombinase operon, and/or the *pTrp* promoter for the Excision-tsPurple portion of the switch: the intention being that any basal expression bypassing the inducible *pTetR/A* promoter would have limited impact in expressing the downstream secondary reporter, at least until aTc is added.

Since development of these two approaches have benefits to both GTS circuits, it seemed judicious to use the Inversion-sfGFP [013a] plasmid to progress the sense-to-sense codon reassignment side of the project and produce histidine-leucine recoded recombinases (discussed in Section 4.2.6), whilst the Excision-tsPurple [033] circuit is to be used to optimise the transcriptional controls to reduce leaky reporter expression (this section). The solutions from each approach could then be applied to

both to make two different GTS circuits (Inversion and Excision) for demonstrating the robustness of implementing alternate genetic codes *in vivo*.

In hindsight, this latter approach for optimising transcriptional control was not strictly necessary: a fully non-functional recombinase would already be the definitive off-state required, and would pose far stronger control of the system over a more stringent (but not necessarily fully-controlled) promoter system for ΦC31. However, at the time, it seemed that being able to generate a less leaky recombinase expression by optimising at least the *pTetR/A* promoter would help provide a clearer signal-to-noise ratio for detecting codon reassignment.

With this in mind, promoter control optimisation began with increasing TetR repression of the *pTetR/A* promoter that was regulating the recombinase. If achieved, this would theoretically mean that (a higher concentration of) aTc would actually be required before the functional ΦC31 could be transcribed. Whilst increasing the degree of repression could be achieved by inserting an additional copy of *tetR* within the plasmids, doing so would have added another 627 bp gene duplication into a polycistronic portion of an already sizeable (6,190 bp) plasmid: not ideal, but a plausible supplementary method if required. Instead, the simpler course of action taken was to transform the Excision-tsPurple plasmid variants (the ΦC** negative control [033], the ΦC31 test-case [033b], and the ΦC31 positive control [033c] that had already excised the double-terminator) into another *E. coli* strain: DH5α_Z1. This DH5α derivative has two additional genomic copies of the *tetR* repressor[14] (driven by constitutive promoter *pN25*; Lutz, 1998a,b), which negated the need to insert an extra *tetR* copy into the plasmids. Despite having a relatively high transformation efficiency and being otherwise comparable to DH10β, a potential disadvantage of DH5α_Z1 was that its growth rate is moderately low (Lutz and Bujard, 1997). As such, it may take more time to reach an equivalent detectable level of both ΦC31 and TinselPurple expression in DH5α_Z1 than it would with DH10β.

[14] As well as the *tetR* gene copies, the DH5α_Z1 strains also carry at their *attB* locus two copies of Lac Repressor (*lacI* gene) under the constitutive *pLaci*^{q} promoter. A spectinomycin resistance (*spc*) gene also accompanies the LacI and TetR repressors.

Having re-transformed the three Excision-tsPurple plasmids into DH5α_Z1, the next step was to determine whether the additional *tetR* copies were able to repress ΦC31 sufficiently. Specifically, this was to assess whether – in the absence of aTc but presence of additional *tetR* – ΦC31 (the test-case [033b] plasmid) would still be spontaneously transcribed and excise the terminator switch within the *tsP* operon, or if the use of DH5α_Z1 is enough to curtail this activity when uninduced. The DH10β strains bearing the same plasmids from before were used as comparative controls, and the extent of TinselPurple colouration used as a proxy for the strength of TetR regulation. The expectation was that, in the absence of inducers, the extra *tetR* copies would limit transcription of the recombinase, thereby minimising downstream chromoprotein expression within DH5α_Z1 cultures carrying the ΦC31 test-case [033b] plasmid compared to DH10β strains carrying the same construct. It was also assumed that there would be little variation in TinselPurple expression between DH5α_Z1 and DH10β regarding the negative ([033] ΦC**) and positive ([033c] ΦC31) controls as additional TetR should do little to affect the absence and presence respectively of the downstream proxy chromoprotein signal in the two recombinase plasmid variants.

Not only were analyses made between the plasmids in the two different strains, a series of experiments using a combination of the aTc and IAA inducers, and growth in different media (LB, 2xTY, and M9; data not shown due to no discernable differences), were also conducted. Ultimately, this aimed to find optimal conditions for expressing the ΦC31 to lead to a strong proxy TinselPurple indicator for codon reassignment. An initial iteration of these experiments were explored first in solid media (Figure 4.22; additional optimisation data of growth in other media not shown), before continuing on to liquid cultures in order to determine more quantitative results.

The three Excision-tsPurple plasmids (i/1) [033b] ΦC31 (tsPurple switch not yet excised; test-case), (ii/2) [033c] ΦC31 (tsPurple switch excised; positive control), and (iii/3) [033] ΦC** (tsPurple switch not excise-able; negative control) that were transformed into *E. coli* strain DH10β (Figure 4.23a,c) were cultured along-

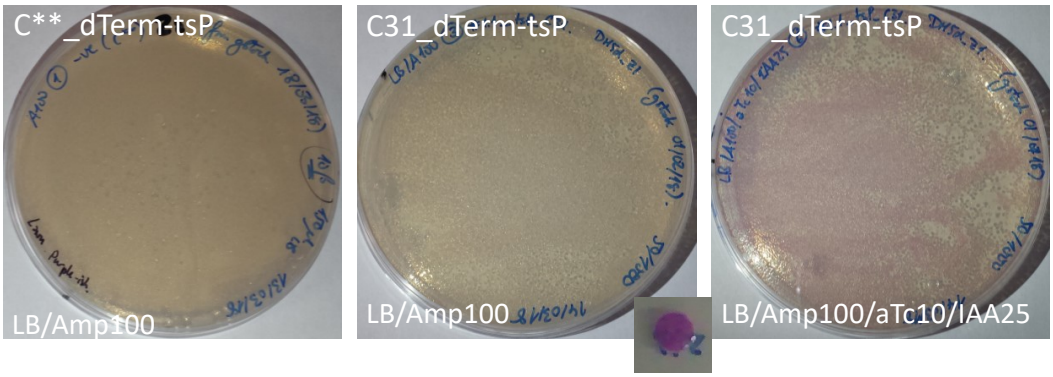
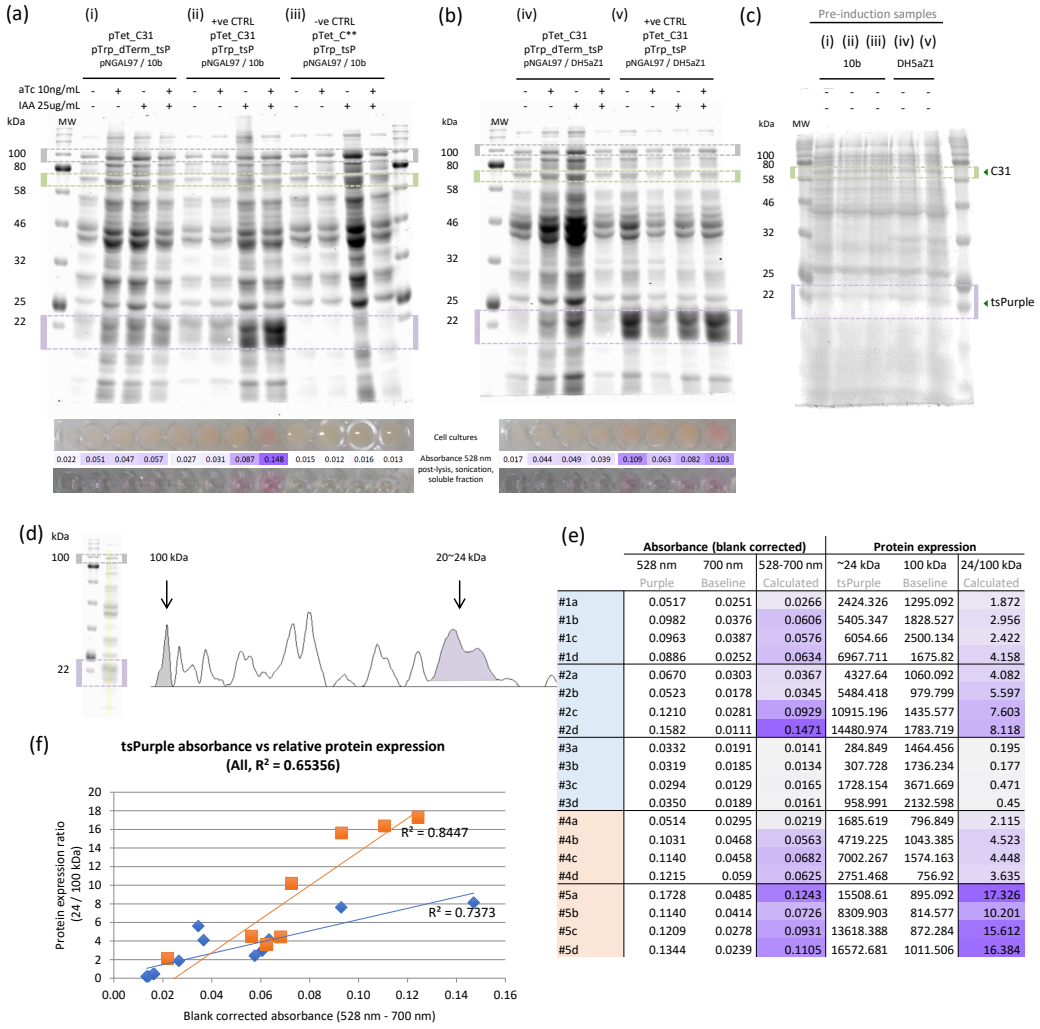


Figure 4.22: Comparison of the Excision-tsPurple in DH10β and DH5α_Z1 with and without the dual inducers

Left: the non-functional ΦC** [033] construct transformed into DH10β (negative control), and plated on 100 μg/mL ampicillin in LB media. Middle: the functionally-restored ΦC31 [033b] test-case construct transformed into DH5α_Z1, and plated on 100 μg/mL ampicillin in LB media. Right: [033b] transformed into DH5α_Z1, and plated on 100 μg/mL ampicillin in LB media supplemented with 10 ng/mL aTc and 25 µg/mL IAA to demonstrate expression of TinselPurple. Inset figure is of a colony picked from this plate and the accumulation of TinselPurple over time.

side the same test-case (iv/4) and positive control (v/5) plasmids that were transformed into DH5α_Z1 (Figure 4.23b,c). Each culture was used to seed four subsequent 3 mL cultures, which were supplemented with (a) no additional inducers; (b) only 10 ng/mL aTc (to activate the *pTetR/A* promoter driving the recombinase operon); (c) only 25 µg/mL IAA (to allow transcription of the *tsP* operon under *pTrp* control); (d) both 10 ng/mL aTc and 25 µg/mL IAA (for expression of both operons). The cultures were monitored over 4 hours to assess the level of TinselPurple expressed as a means to characterise the stringency of the dual promoter controls on the constructs – but specifically for the (i/1) and (iv/4) [033b] ΦC31 (test-case) – in DH10β and DH5α_Z1 respectively (Figure 4.23a,b). Endpoint absorbance measurements were taken from the cultures: one at 528 nm (absorption at green wavelengths) to quantify the pink-purple colour transmitted from the TinselPurple produced, and another at 700 nm as a baseline reading from which to standardise the measurements (Figure 4.23e). All samples, including pre-induction cultures (Figure 4.23c), were lysed by sonication and the soluble fraction was run on 10% SDS-PAGE (Figure 4.23a,b,c) in an attempt to further visualise and quantify the level of both recombinase (expected at ~68 kDa for ΦC31, ~21 kDa for the truncated ΦC** half) and TinselPurple (estimated at 25.5 kDa, but observed around



	Absorbance (blank corrected)			Protein expression		
	528 nm	700 nm	528-700 nm	~24 kDa	100 kDa	24/100 kDa
	Purple	Baseline	Calculated	tsPurple	Baseline	Calculated
#1a	0.0517	0.0251	0.0266	2424.326	1295.092	1.872
#1b	0.0982	0.0376	0.0606	5405.347	1828.527	2.956
#1c	0.0963	0.0387	0.0576	6054.66	2500.134	2.422
#1d	0.0886	0.0252	0.0634	6967.711	1675.82	4.158
#2a	0.0670	0.0303	0.0367	4327.64	1060.092	4.082
#2b	0.0523	0.0178	0.0345	5484.418	979.799	5.597
#2c	0.1210	0.0281	0.0929	10915.196	1435.577	7.603
#2d	0.1582	0.0111	0.1471	14480.974	1783.719	8.118
#3a	0.0332	0.0191	0.0141	284.849	1464.456	0.195
#3b	0.0319	0.0185	0.0134	307.728	1736.234	0.177
#3c	0.0294	0.0129	0.0165	1728.154	3671.669	0.471
#3d	0.0350	0.0189	0.0161	958.991	2132.598	0.45
#4a	0.0514	0.0295	0.0219	1685.619	796.849	2.115
#4b	0.1031	0.0468	0.0563	4719.225	1043.385	4.523
#4c	0.1140	0.0458	0.0682	7002.267	1574.163	4.448
#4d	0.1215	0.059	0.0625	2751.468	756.92	3.635
#5a	0.1728	0.0485	0.1243	15508.61	895.092	17.326
#5b	0.1140	0.0414	0.0726	8309.903	814.577	10.201
#5c	0.1209	0.0278	0.0931	13618.388	872.284	15.612
#5d	0.1344	0.0239	0.1105	16572.681	1011.506	16.384

Figure 4.23: Characterising the stringency of the dual promoter controls on TinselPurple expression from the Excision-tsPurple plasmids in DH10β and DH5α_Z1 strains

Sub-figures **(a-c)** are 10% SDS-PAGE gels of protein expression from Excision-tsPurple plasmids, where purple dotted boxes outline TinselPurple (~24 kDa), green dotted boxes for the recombinase protein (~68 kDa), and grey dotted boxes for background protein expression (~100 kDa). **(c)** shows constructs (i) [033b] ΦC31 (tsPurple switch not yet excised), (ii) [033c] ΦC31 (tsPurple switch excised), and (iii) [033] ΦC** (tsPurple switch not excise-able) that were transformed into *E. coli* strain DH10β; (iv) [033b] ΦC31 (tsPurple switch not yet excised), and (v) [033c] ΦC31 (tsPurple switch excised) that were transformed into the DH5α_Z1 strain. This was the protein expression pattern prior to aTc and/or IAA induction. **(a)** and **(b)** show the same constructs (i-v): without addition of any inducers; with only 10 ng/mL aTc (to transcribe the recombinase); with only 25 µg/mL IAA (to transcribe tsPurple); with both 10 ng/mL aTc and 25 µg/mL IAA (for both). **(d)** Using ImageJ (an image processing software), the total amount of protein expression (peaks) for TinselPurple (~24 kDa) and an arbitrary background protein (~100 kDa) were quantified for each sample/lane in **(a-c)**. **(e)** tabulates the values measured in **(d)** to calculate and compare the relative TinselPurple expression across each sample. The table also shows the absorbance measurements (528 nm for TinselPurple) corrected with background measurements for the cell cultures (700 nm). Note, #1 corresponds to (i), #2 to (ii), #3 to (iii), #4 to (iv), #5 to (v); similarly, a-d correspond respectively to the addition of inducers listed earlier.
(f) plots the standardised TinselPurple absorbance against its standardised protein expression levels in DH10β (blue markers and lines) against DH5α_Z1 (orange markers and lines).

22-24 kDa) expression across the different conditions considered.

Using ImageJ (`https://imagej.net/ij/`), the total amount of protein expression for TinselPurple and an arbitrary background protein at ~100 kDa were quantified for each sample/lane from the gel (Figure 4.23d,e). As can be seen in the gel images, there was quite a high degree of variation in the amount of each sample loaded onto the gels, despite attempts to normalise to equivalent optical density upon harvesting the cells. Hence, the post-imaging measurements at 100 kDa was a further bid to standardise the level of TinselPurple expression so that comparisons may be made across the different conditions (Figure 4.23e,f). Even so, the conclusions drawn were primarily taken as a general qualitative overview.

One clear result from the gel visualisations (which were equally confirmed through observations of the cultures and soluble fractions, and by the protein expression quantification) was the absence of TinselPurple expression from the [033] ΦC** negative controls (grown in DH10β), regardless of whether the samples were induced with aTc and/or with IAA (Figure 4.23aiii, purple box around 22 kDa; Figure 4.23e #3a-d). This suggests that, at least under the conditions tested, a fully non-functional recombinase leaves the Excision-tsPurple switch intact and prevents the secondary chromoprotein reporter from expressing (corroborating the results from Figure 4.21).

By comparison, the positive control samples in Figure 4.23aii (DH10β) and Figure 4.23bv (DH5α_Z1) show how strongly expressed TinselPurple was when induced. As the double-terminator blocking transcription of *tsP* had already been removed, addition of aTc should have limited/no effect on the circuit (other than perhaps produce $\phi c31$ transcripts that indirectly compete for translation); realistically, only IAA would be required to cause an increase of the chromoprotein. This was exemplified in the constructs within DH10β, which followed the expected pattern in expressing more TinselPurple when induced by IAA (samples ii/2c,d), but minimal when uninduced or with aTc (samples ii/2a,b). In DH5α_Z1 by contrast, there seemed to be an anomalous result whereby the uninduced positive control (sample v/5a) showed unexpected chromoprotein amounts (even accounting for

normalisation) that exceeded that of the IAA-only or double-induced conditions (samples v/5c,d). It is possible that there was an early mix-up when aliquoting samples from the 3 mL cultures to the 96-well plates for quantification that was propagated downstream, as sample iv/4d (the test-case [033b] in DH5α_Z1 that had been doubly-induced) also showed unexpected results with minimal TinselPurple, nearly equivalent to its uninduced condition (sample iv/4a), where it was predicted to have greater amounts than its other conditions. If this mix-up was indeed the error, then the expected results for the positive controls hold true. Equally, the earlier assumption that there would be little difference between the expression patterns for the positive controls between DH10β and DH5α_Z1 appear also to be accurate.

In terms of TinselPurple comparison between the [033b] test-case construct in DH10β and in DH5α_Z1: there did not appear to be any significant differences in the regulation of the circuits between the two strains, by visual observation or from protein expression quantification (Figure 4.23a,b,e). Both produced the expected pattern of expression upon aTc and/or IAA induction. For example, whilst the uninduced test-case samples (i/1a and iv/4a; with a functionally-restored ΦC31) in both strains were calculated to have approximately 10-fold higher TinselPurple over the normalised, equivalent condition in the negative control (sample iii/3a; with ΦC**, and where no TinselPurple was expected; 1.872/0.195 = 9.6 in DH10β; 2.115/0.195 = 10.8 in DH5α_Z1), neither test-case samples, as anticipated, reached a detectable level of chromoprotein expression. Again in both, the addition of aTc was sufficient in causing a slight increase in the proxy TinselPurple expression levels (samples i/1b and iv/4b), compared and normalised to that of their uninduced counterparts, of approximately 1.6 (= 2.956/1.872) and 2.1 (= 4.523/2.115) times in DH10β and DH5α_Z1 respectively. The conditions where only IAA was added were equivalent (or marginally, though not significantly, less after normalisation; 2.422/1.872 = 1.3 in DH10β, 4.448/2.115 = 2.1 in DH5α_Z1) to the ratio of aTc induced against uninduced conditions. Nevertheless, a tenuous argument could be made in justifying these slight differences in that whilst aTc could induce more constructs to drop the Excision switch, allowing additional expression of the down-

stream chromoprotein that was under control of an already slightly leaky *pTrp* promoter, addition of IAA only serves to increase transcription of existing excised constructs and not generate more functional *tsP* operons (hence the chromoprotein accumulated would not be as high as the previous condition).

Finally, with regards to the doubly-induced conditions: this is the only point where there was potentially a stronger distinction between the strains. As it stands, with the expression pattern in sample iv/4d (DH5α_Z1) – i.e. with TinselPurple at a level just greater than the uninduced condition – displayed as is, the reasonable conclusion was to rule this out as an error. Given all previous results, it is very unlikely for there to have been negligible levels of the chromoprotein if the recombinase had been activated to excise the switch and the appropriate IAA inducer provided to transcribe TinselPurple. If there had been a mix-up between samples iv/4d and v/5a (as discussed earlier), then an alternative conclusion – that falls in line with the expected hypothesis – can be drawn in that DH5α_Z1 provides a more stringent environment than DH10β. Using the v/5a sample in place of iv/4d would show the TinselPurple expression ratio between singly-induced with aTc and the doubly-induced condition to be 1.4 (= 4.158/2.956) times in DH10β and more than twice that at 3.8 (= 17.326/4.523) times in DH5α_Z1. The relative higher increases in TinselPurple produced in DH5α_Z1 over DH10β in the conditions from uninduced to induced with aTc, and also from this to the double inductions, could be explained by the added *tetR* copies in DH5α_Z1. In DH10β, the system could be so leaky[15] (and ΦC31 so efficient even at small amounts) that the addition of the inducers do not produce much more TinselPurple above the background basal expression. Whereas it is possible that in DH5α_Z1 the extra *tetR* was sufficient enough, such that the derepression with aTc (and IAA) was able to produce a significantly greater relative response in chromoprotein produced.

Although an attempt was made to quantify the differences in the strains and

[15]High basal transcription-translation of the recombinase could also explain why there does not seem to be a clear distinction for ΦC31 between uninduced and induced conditions. However, the corresponding logic would be that there would also be an equivalent level of the ΦC** transcribed in the negative control. As this gene would be truncated before residue 196, the protein product would be expected around 21 kDa and there is no sign of this in the negative control conditions.

understand if additional regulation on the recombinase portion of the Excision-tsPurple circuit would reduce leakiness in the system overall, no definitive conclusions could be drawn. The main issue is that despite efforts to control for differences in growth and protein expression rates, there were too many confounding factors that made it difficult to make true quantitative comparisons between the strains and conditions. For example, the experiment did not (and could not) factor in any variations in expression of the proxy reporter circuit (the IAA-induced *tsP* operon) that could skew the interpretations of the upstream *pTetR/A* (aTc-dependent) control on the recombinase. Moreover, there were not enough iterations (technical or biological repeats) of this experiment to derive any statistically significant results. The only anecdotal inference drawn is that expression in DH5α_Z1 strains could provide tighter control of the ΦC31 recombinase over expression in DH10β. However, there still remained basal expression of the recombinase, even in the DH5α_Z1 system.

A potential final, but unexplored (i.e. theoretical only), extension to tighten up regulation was to change the *pTrp* promoter to one of the two synthetic *pTac* variants (de Boer et al., 1983), or their subsequent derivatives, which are noted to have moderately high/er levels for expression control (Terpe, 2006). The hybrid promoters, *pTacI* and *pTacII*, were derived from the *pTrp* and *pLacUV5* promoters (as displayed in Figure 4.24). These promoters were reported to be approximately 11 and 7 times more efficient at directing transcription than the derepressed parental *pLacUV5*, and approximately 3 and 2 times more efficient than the *pTrp* promoter, respectively. The higher stringency of the *pTacI* promoter may be attributed to having both the consensus -35 (TTGACA) and -10 (TATAAT) Pribnow box sequences. A further extension to *pTacI* (also known as TAC16), where increasing the 16 bp spacer length between its -35 and -10 sequences by one or two bases produced promoter variants *pTrc* (TAC17) and *pTic* (TAC18) that were on average 90% and 65% as active *in vivo* as *pTacI*, respectively (Brosius et al., 1985). The efficiency of the latter two were described as somewhat surprising given that associated *in vitro* data suggests that *pTrc* is the stronger of the three (Mulligan et al., 1985), and

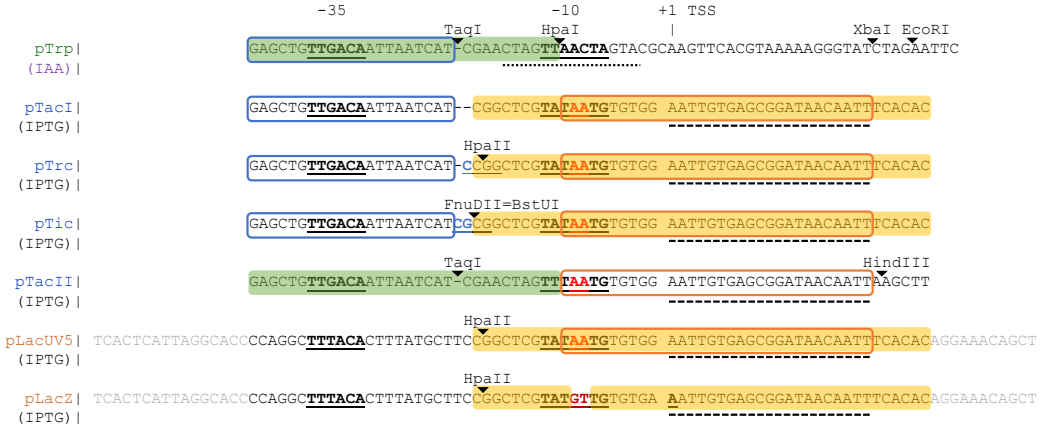


Figure 4.24: DNA sequences of the *pTrp*, *pLac*, and *pLacUV5* promoters, and their synthetic hybrid variants (*pTacI*, *pTacII*, *pTrc*, and *pTic*)

Promoters from *E. coli* are in green (*pTrp*) and orange (*pLac*, *pLacUV5*) font, with the synthetic variants in blue. *pLacUV5* differs from *pLac* by two bases in the -10 sequence (bases in red). *pTacI* and *pTacII* are hybrids derived from *pTrp* and *pLacUV5*; the highlights and coloured boxes over the DNA sequences show where the sequences match the parental and children promoters. *pTrc* and *pTic* are variants of the *pTacI* sequence, where one or two bases (in blue) respectively were inserted in between the -35 and -10 sequences (underlined). Each synthetic variant was described as being more efficient with controlling heterologous expression with respect to either of the *pTrp* and *pLacUV5* parental promoters (see main text for details). The inducers required to derepress the promoters are written in parentheses below the promoter names. The *trp* repressor and *lac* operator binding sites are shown by the dotted and broken lines respectively. The cut site for the restriction enzymes are shown by the black triangles below the enzyme names. The transcription start site (TSS), according to the *pLacUV5* promoter, is indicated by the +1; in the case of the hybrid promoters, it is assumed the TSS is in the same region.

that 17 bp spacers were usually more prevalent (and optimal) in *E. coli* promoter sequences (Hawley and McClure, 1983). Having said this, the authors also advised that the discrepancy between the relative activity of the TAC promoters measured *in vivo* and *in vitro* are not likely to be significant as the relative strengths differed by only a factor of two, which they believe is "within experimental error and the uncertainties underlying the calculations" (Mulligan et al., 1985).

Nonetheless, the reason why *pTacI* (or its derivatives) could assist with tightening control over the *tsP* operon in the DH5α_Z1 strain is because these hybrid promoters are controlled by the *lac* repressor (and can be derepressed with IPTG). This means that not only would the extra *tetR* genomic copies in DH5α_Z1 help with making the recombinase portion of the Excision-tsPurple more stringent, the additional *lacI* copies in this strain could also be used to control the *tsP* operon, if the promoter is switched from *pTrp* (IAA-controlled) to any of these *lac* repressor-

dependent hybrid promoters.

A parting caveat on optimising the Excision-tsPurple for greater stringency in DH5α_Z1 is that it is unlikely to be as applicable if the circuit is to be expressed in an *E. coli* B strain. The OCI circuits established in Section 3.2.1 put the LH-HL anticodon-modified tRNAs under control of the *pTetR/A* promoters, which when combined with the Excision-tsPurple system would be controlled in much the same way as the recombinase. However, if one of the OCI reporters, such as sfGFP that is under control of *pT7* (Section 3.2.2.2), is also linked into this GTS system either additional expression system components would be required for its use in DH5α_Z1 (e.g. inclusion of the *LysY* gene, the DE3 prophage/lysogen; knock-outs of *lon*, *ompT* protease genes), the promoter for the OCI reporter would need to be changed away from *pT7* (which has very high stringency, above that of *pLac*, *pTetR/A*, *pTacI*; Terpe, 2006), or the OCI-GTS circuits would need to be expressed in another *E. coli* strain (i.e. in a B strain that operates with T7 expression, like BL21(DE3) or T7 Express *lysY/I^q*). If the latter approach is taken, where there are no extra *tetR* (genomic) copies, it would negate the optimisation work implemented here, for better or worse.

Despite various attempts and proposals to improve promoter stringency there remained a level of basal expression of both ΦC31 and the proxy TinselPurple reporter that could not be removed. There were likely marginal gains in using the K-12 DH5α_Z1 strain with its additional repression (TetR and Lac repressors), and further features were also identified that could be beneficial in improving control of both GTS circuits in B strains. However, the conclusion remained that the most parsimonious approach for regulating expression is to completely abolish ΦC31 activity.

It was fortuitous, therefore, that it was at this point that the histidine-leucine mutations previously identified in Section 4.2.1 were found to produce a viable set of recombinase-inactivating mutants in the GTS Inversion-sfGFP plasmids (Section 4.2.6). As such, the focus reverted back to the Inversion-sfGFP circuit to verify that the new sets of recombinases were indeed non-functional. If so, the same muta-

tions could be applied to the Excision-tsPurple circuit for cross-validation, and then both would be used with the alternately-coding anticodon-modified tRNAs from the OCI approach to fully assess this improved sense-to-sense codon reassignment detection system (Section 4.2.7).

4.2.6 Creating a histidine-leucine inactive recombinase reporter for sense-to-sense codon reassignment

Now that at least one GTS circuit has been verified to be viable and non-leaky (when the recombinase is inactive ΦC** state), the next step was to generate an OCI-compatible off-state: one where the recombinase is (seemly) non-functional because key leucine codons have been recoded to histidine and vice versa, and not due to mis-sense stop codons replacing critical residues. By creating a recoded recombinase reporter within the GTS system, it would be possible to unite this approach with that of the anticodon-modified tRNAs of the OCI system, and fully characterise both halves of the histidine-leucine codon reassignment in a way that may be more amenable to host viability.

Earlier, in Section 4.2.1 (Table 4.1b, Figure 4.7), five candidate histidines and 14 leucines were identified as residues within the N-terminal catalytic domain of ΦC31 that may hamper activity if recoded to leucine and histidine respectively. From this selection, the *in silico* site-directed mutagenesis (SDM) simulations that displayed the most clashes with their nearby residues were: H61L (within a β-pleat); L83H; L117H, L118H, L120H (of the αD2 helix); L140H, L143H, L147L (forming part of the αE1 helix); L162H, L167H, L171H (in αE2 and a subsequent helical turn), and H234L (within a putative "recombinase" region in the C-terminal domain). Of interest also was H142(L), primarily due to its proximity to the potentially more crucial leucines within the αE1 helix rather than any particular disruption it may cause, and L468(H), a residue within a putative coiled-coil DNA binding domain that has been shown to have some, albeit not fully, detrimental impact on ΦC31 when mutated to L468P/A (Rowley et al., 2008; Liu et al., 2010). Note that it was not possible to model the L468H recoding as this region was missing from the crystal structure (PDB: 4BQQ). Similarly, an attempt was made at modelling the H234L mutation but this residue bordered a gap in the structured, which meant that it was not possible to visualise all its potential intra-molecular interactions.

With the exception of perhaps L143, however, it was unlikely that recoding these residues individually would be sufficient enough to produce an inactive re-

combinase. Hence, it was anticipated that a number of these residues would need to be mutated in tandem before an experimentally-validated, non-functional recombinase could be uncovered. If a combination of both leucine and histidine residues are recoded in a way that causes a loss-of-function to ΦC31, it would be possible to: (i) assess the histidine-to-leucine reassignment (that had not previously been achieved within the OCI CAT reporter), (ii) give further validation to the leucine-to-histidine reassignment (by replicating this reassignment in another recoded reporter), and (iii) be the first demonstration of bi-directional, non-synonymous sense-to-sense codon reassignment in a living organism.

Given the selection of candidates, and the opportunity to conceivably recode both histidine and leucine with one SDM experiment, the recombinase's αE1 region (encompassing residues L140, H142, L143, and L147) became one of the initial targets for creating a potentially inactive, recoded ΦC31. In the event that recoding both amino acids within the αE1 helix is insufficient in creating a loss-of-function ΦC31, a combination of the oligos listed in Table A.2 could also be used to generate a library of histidine-leucine recoded recombinases. To achieve this, Darwin Assembly (described in Section 2.2.18; Cozens et al., 2012) was used: this method was designed for simultaneous multi-site mutagenesis, and should be able to create all (or a combination of) the candidate recodings in a single experiment. The more sites that are recoded the greater the chance of making an inactive ΦC31, therefore one other initial proposal was to construct a variant that has all 12 recoding candidates in the N-terminal domain (i.e. residues 1–171) as an alternate experiment to the one mutating the four sites within the αE1 helix (Table 4.1). If all the above recodings truly fail to generate a single loss-of-function mutant, there is still some structural data up until residue 379 from which to make *in silico* mutagenesis predictions for further leucine sites (e.g. L255H, L359H, L372H), however, H234 is the last potentially crucial histidine where the impact of its recoding could be modelled; recoding residues beyond this point would be entirely reliant on experimental (guess)work. Regardless, it was anticipated that recoding both amino acids in the pre-selected sites would be sufficient in producing non-functional recombi-

nases ready for testing sense-to-sense codon reassignment.

A note regarding the construction of histidine-leucine recoded ΦC31 variants concerns the choice of the source plasmids. As the GTS circuits contain irreversible unidirectional switches, it was necessary to ensure that the plasmids used to construct the reassignment-compatible recombinases are all in their negative control (off-states) prior to recoding ΦC31: a circuit that has already been permanently triggered to invert or excise its genetic switch, and allow expression of its downstream reporter, would no longer be useful for testing sense-to-sense codon reassignment. To fully ensure that the secondary reporters (sfGFP in the case of the Inversion circuits, and TinselPurple for Excision) would be in their off-states, their corresponding ΦC**-containing circuits (plasmids [013a] and [033] respectively) were used as the starting material for recording. Recall, however, that these recombinases are only non-functional due to two Amber stop codons in place of asparagines at positions 200 and 204[16]. Hence, when creating a newly recoded loss-of-function recombinase circuit that is compatible for histidine-leucine reassignment, it is necessary to revert the stop codons to their wild-type residues, alongside recoding the pre-selected histidine and leucine sites to be compatible with the alternate code.

It should also be acknowledged that there is a fine balance between recoding enough residues to sufficiently inactivate the recombinase but not too many so as to make it difficult to assess if sense-to-sense codon reassignment has occurred. As all previous synthetic systems for codon reassignment (Döring and Marlière, 1998, and the OCI system from this work) have only recoded and demonstrated reassignment of just one residue – albeit a crucial one – it seemed prudent to also create a selection recombinases with a minimal number of recodings.

Consequently, alongside implementing the proposed many multi-site histidine-leucine recodings across the recombinase's N-terminal domain (discussed next in Section 4.2.6.1), three minimally/singly-recoded sites were also planned for residues considered most likely to cause a loss-of-function in the recombinase

[16] Although it shall not be explicitly noted hereinafter, the SDM oligo (YKH308) that is to restore the two Amber stop codons to asparagine also included a synonymous recoding of $I_{ata}203I_{atC}$ to help with the oligo design. The only other references to this are in Table A.2 and Table 4.2, where residues that are to undergo synonymous codon change are denoted by grey font or a yellow background cell.

(detailed in Section 4.2.6.2). This two-pronged approach also serves as a risk-mitigation strategy: ensuring that a recoded recombinase variant could be created and tested for loss-of-function before being used to assess sense-to-sense codon reassignment. As described previously, the higher-risk approach uses the novel Darwin Assembly method to mutate, simultaneously, up to 14 sites within the recombinase and create a library of histidine-leucine recorded mutants, thereby providing more opportunities for a loss-of-function variant to be uncovered. In the best-case scenario, this would produce all the expected histidine-leucine recodings with many suitably inactive variants for testing reassignment. In less optimal conditions, there may be need for multiple rounds of screening and sequencing before suitable variants can be selected. In the worst case, the assembly method may create recoded recombinases with additional off-target mutations that require several rounds of reassemblies to correct.

By comparison, the low-risk strategy would use the well-established iPCR method to implement single residue recodings for the subset of the critically key histidine and leucine candidates within ΦC31. Specifically, the key target residues are: (i) H61L (one of the few potentially crucial histidine recoding that was modelled); (ii) H142L_L143H (a putatively essential leucine, with a likely less key adjacent histidine), and (iii) L468H (an unmodelled recoding that purportedly impacts recombinase function but may not provide a sufficient loss-of-function on its own). Once sequencing has confirmed the correct recoding for these three recombinase variants, each would require only one further round of iPCR to revert the two Amber stop codons within the recombinases to wild-type, before screening for loss-of-function activity and for sequence verification. The flaw of this approach is that if none of the three recoded targets, after the missense stop codons are corrected, are critical enough to create an inactive recombinase, then additional rounds of iPCR would be required to find a minimally recoded function-less variant. In this instance, it would be prudent to combinatorially recode these three target codons together first, before expanding the search further. But note also that the more codons that need to be iteratively recoded means that it may be more time-efficient to proceed

with the newer Darwin Assembly method instead.

As, at the time, there were two viable GTS constructs (the Inversion-sfGFP circuit and the Excision-tsPurple design), it was possible to tackle both high- and low-risk recoding strategies in tandem. The former circuit (with plasmid [013a]) could be used to implement the multi-site histidine-leucine recodings using Darwin Assembly, whilst the latter (with plasmid [033]) would take the iPCR approach to create the three minimally-recoded recombinases. The strategy that is able to give more practical results (i.e. viable inactive recombinases) for testing sense-to-sense codon reassignment in the next stage of the project would be pursued going forward.

4.2.6.1 Many multi-site histidine-leucine recodings for an inactive ΦC31 within the Inversion-sfGFP switch

Using the Darwin Assembly method (as described in Section 2.2.18) and the oligos from Table A.2a,c , three multi-site recoding experiments were carried out on the ΦC31 recombinase within the Inversion-sfGFP [013a] plasmid. Mutagenesis experiment: [1A] used inner oligos YKH308, YKH309, YKH310, YKH311, YKH314, and YKH315[17] to revert the two stop codons to asparagine and recode all 12 of the candidate histidine and leucine sites (Table 4.1a,b); [1B] used only oligos YKH308 and YKH314 to correct the stop codons and recode the four sites within the αE1 helix (L140H, H142L, L143H, L147H); [1C] was the same as [1B] but with the addition of oligo YKH315 to also recode the three leucines in the nearby αE2 helix-turn (L162H, L159H, L163H). It should be noted that although Darwin Assembly should be able to achieve complete mutagenesis of all sites as encoded by the inner oligos, there is a chance that not all oligos would be bound fully, leading to variation in mutagenesis products and an amplicon pool that is a heterogeneous mix of partially recoded recombinases.

The other necessary oligos used in these Darwin Assembly recoding experiments were the boundary oligos YKH304 and YKH305, and outnest oligos

[17] Alongside recoding L162H, L167H, and L171H, SDM oligo YKH315 also included a synonymous recoding of $N_{aaC}166N_{aaT}$ to help with the oligo design. The only other references to this are in Table A.2 and Table 4.2, where residues that are to undergo synonymous codon change are denoted by grey font or a yellow background cell.

YKH276 and YKH277. The 3'-end of boundary oligo YKH304 and the 5'-end of YKH305 bind at specific Darwin priming sites that sit beyond the binding regions of the inner oligos, designating the bounds of the *ϕc31* region that are to be recoded. Once the polymerase has filled in the junctions between boundary oligos and inner oligos, and the nicks have been joined by the ligase, the outnest oligos would be able to amplify the final pool of recoded recombinase products for use in downstream assemblies. Specific features of note in boundary oligos are that the 5'-end of YKH304 is biotinylated to enable purification post-amplification, whilst YKH305 has a 3'-inverted dT modification to inhibit both degradation by 3' exonucleases and extension by DNA polymerases; both bind on the same strand as each other and the inner oligos. Internal to both these modifications are two 20 bp regions that are specific binding sites for the outnest oligos only (Table A.2d, sequences in purple). `SapI` restriction sites are placed further along the boundary oligos (sequences in pink, with cut sites in blue and yellow) to allow the eventual recoded recombinase products to be inserted back into its original plasmid through TIIS assembly.

As for generating the backbone portion of the constructions, the Inversion-sfGFP [013a] plasmid initially underwent iPCR using primers YKH306 and YKH307, both of which also contained the `SapI` restriction site. After being cut with `SapI`, both the circuit vector and the recombinase product from Darwin Assembly, were ligated together to reform, once again, the Inversion-sfGFP plasmid – the exception being that the recombinases in these variants (known as plasmid library [049]) were recoded per the mutations introduced using the inner oligos.

Although seemingly straightforward, the multi-site mutagenesis of the recombinase constructs using Darwin Assembly was fraught with implementation issues. Figure 4.25 displays a streak/stock plate from the initial mutagenesis experiments [1A], [1B], and [1C] (bottom three rows of the plate, in reverse order), with eight colonies from each. Visual inspection of the colony streaks show little difference between the fluorescent levels for the majority of these samples compared to the two ΦC** inactive variants (top row of the plate), which suggests (by proxy) that

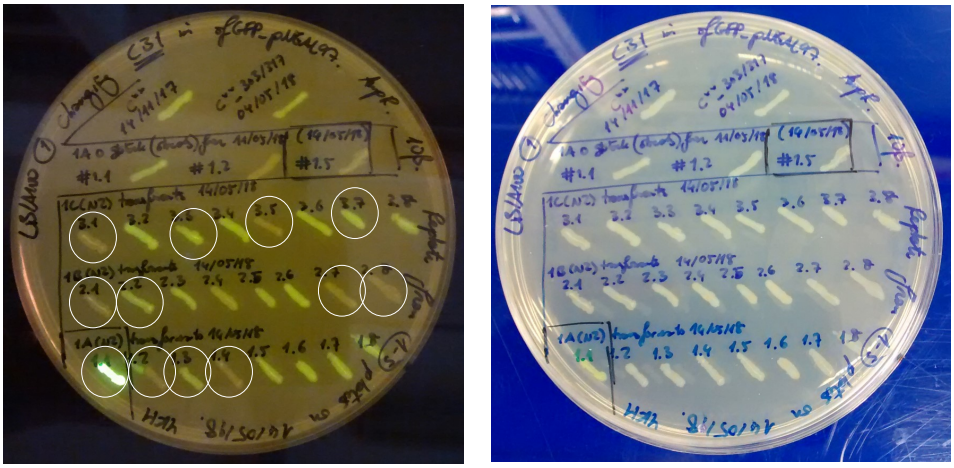


Figure 4.25: Early Darwin Assembly attempts to generate the sfGFP library of recoded mutants in the Inversion-sfGFP constructs

Each colony sequenced (ringed) all contained errors, be it in the switch portion of the circuit, or the proxy sfGFP reporter. Some constructs failed to be sequenced altogether.

the histidine-leucine mutagenesis experiments were able to generate non-functional, recoded recombinases. Previously, the rationale for selecting the Inversion-sfGFP construct [013a] for the recombinase recoding process was due to the lack of any noticeable background fluorescence from the secondary reporter when the recombinase was inactive (ΦC**; Figure 4.14), at least compared to the other construct switch variants (i.e. [013b], [013c]). However, from the low-level fluorescence of a re-streak of this sequence-verified [013a] plasmid, alongside the newly-made (presumably inactive) recoded recombinase samples, it was clear that there had been some leaky expression of sfGFP in the circuits afterall, even with the $pAmp^R$ promoter Inversion switch facing away from the fluorescent protein (i.e. switch still in its off-state).

In order to assess the extent of recoding in the mutated recombinase samples, and to confirm that the background fluorescence was due to poor transcriptional-translational control rather than spontaneous inversion of the switch allowing sfGFP expression, a selection of these weakly fluorescing samples were cultured in preparation for Sanger sequencing (with primers YKH150 and YKH317). As an initial appraisal, four samples from each mutagenesis experiment were selected: for each, two samples were chosen that were less fluorescent, and two that were more so (the

samples ringed in white in Figure 4.25). Given that the proxy signal of fluorescence was clearly visible by eye in the previous Inversion-sfGFP positive control construct [013ai] (i.e. where the ΦC** stop codons were reverted to wild-type to allow regain-of-function; Figure 4.15), it was assumed that inactive recoded recombinases, when coupled with the OCI engineered tRNAs for histidine-leucine codon reassignment, would also demonstrate fluorescence that is significantly higher than the current levels seen in Figure 4.25. If, upon testing reassignment, the sfGFP is active but indistinguishable from this background fluorescence, then the GTS system would require yet another redesign to boost the signal-to-noise ratio for detection.

By contrast, sample 1A.1 (from experiment [1A]; bottom-left of Figure 4.25) was evidently expressing sfGFP at a very high level. This was unexpected as these constructs were meant to be inactive through recoding: experiment [1A] in particular was expected to have the highest chances of producing loss-of-function ΦC31 mutants as it was to simultaneously recode 12 different histidine and leucine sites across the recombinase. Given the unusual pattern of expression, this sample was also taken forward for sequencing. If, for some reason, the mutations in sample 1A.1 still makes it viable as a GTS construct, it could serve as a possible positive control in differentiating against the weak fluorescence as seen in the other inactive Inversion-sfGFP constructs.

In reality, sequencing of these initial recoded recombinase constructions yielded a library of circuits with unfavourable results. Despite displaying the expected phenotype (i.e. minimal fluorescence), none of the plasmids recovered from these transformants matched the original circuit design: half of those sequenced possessed only a short 5'- or 3'-terminal segment of the recombinase (with ends covering less than 300 bp and 200 bp, respectively, of the full 1,839 bp gene); half had lost both the recombinase and the *sfgfp* sequence altogether (but retained the *attB-pAmp*^{R}*-attP* switch). For example, the least unsuccessful of the mutagensis and re-assembly experiments was sample 1B.7: the recombinase was accurately constructed up until residue D109, whereupon the sequence skips straight to L140 to correctly recode this residue to histidine, as well as mutate the following H142L,

L143H, and L147H as designed. However, the rest of *ϕc31* then failed to be assembled and instead proceeded to the Darwin priming site (that is placed directly after where the recombinase sequence would have ended). Whilst all four of the planned residues were recoded, it was clearly still non-viable given that the majority of the recombinase sequence was missing. The rest of the sequenced constructs had lost even more segments (discussed in greater depth later alongside additional data), hence it was no surprise that, in all instances, the Inversion switch remained un-inverted.

Ultimately, none of these initial (sequenced) assemblies contained the full recombinase sequence, let alone one that has been recoded for histidine-leucine reassignment suitable for assessing sense-to-sense codon reassignment. However, analysis of the range of truncated, partially recoded inserts from these first iterations (Figure 4.25) suggests that there were some fundamental issues in the implementation of the multi-mutagenesis method. For instance, possible problems could have been that: the inner oligos were not binding sufficiently to allow full extension in between the sites marked for mutagenesis, the molar ratios of oligos to ssDNA template could have been sub-optimal, the number of different mutagenesis oligos mixed together could have formed secondary structures/heterodimers that hindered binding by the DNA polymerase or caused it to drop-off prematurely, there may not have been sufficient time for the polymerase or ligase to extend and join up the recoded fragments leading to partial assemblies, and/or the mutagenesis assembly mixture was incorrectly prepared or contained inactive enzymes. Given that the method was, at the time, still in development, assessing the former possibilities would have required significant optimisations. Instead, it was more feasible to first eliminate the latter two possibilities by making a fresh formulation of the reaction mixture and carrying out the mutagenesis experiments over a more prolonged period of time.

A first step to troubleshooting the assembly issues was to understand if, and to what extent, there was obvious heterogeneity in the Darwin Assembly products (e.g. incomplete amplification, truncated amplicons, etc.). The initial mutagenesis

failures highlighted the need to ensure that future attempts produce excess material that could be verified (e.g. visualisation using gel electrophoresis), prior to taking any constructs forward to assembly. Whilst the first attempt had been limited by the amount of available input material, by ensuring that future attempts produce a surfeit of DNA, it would be possible to screen for correctly-sized products before continuing on to downstream assemblies. Moreover, if the screening shows that there is a heterogenous mix of differently-sized products, yet the required one is of adequately high yield, the relevant fragment could be gel extracted and assembled with the vector directly.

Assembly optimisation was not limited to the process of producing the recoded recombinase sequences: it appears that the conditions for amplifying the vector also needed improving. The original primer pair YKH306/YKH307 used should have produced a complementary, linearised vector much like that of [013a] but without the recombinase. This should have included the downstream secondary reporter sequence, *sfgfp*: yet some of the sequenced constructs had lost this operon and more (~1,044 bp in total). Notably, having missed out *sfgfp*, the constructs (samples 1A.2, 1C.1, 1C.5; data not shown) were found to all begin at the *attB* site, and also to all incorporate an extra `SapI` after the Inversion switch and `BsaI` cloning site immediately following, before continuing on as per the design sequence. The consistent absence of the *sfgfp* portion and downstream insertion of the `SapI` sequence (presumably from the unbound 5'-end of YKH307) across all affected constructs suggests that this primer was poorly designed with multiple binding sites, both at the designated site and at a secondary site downstream. Likely, the lack of binding specificity of primer YKH307 to the [013a] plasmid template was exacerbated by the fact that the vector itself retained too many vestige sequences from previous assemblies. Further inspection of the vector design reveals that there were three regions with partial *t-lpp* terminator sequences to which YKH307 could potentially bind: of which all three share 40 bp direct repeats, whilst two have up to 70 bp sequence identity. Hence, to prevent this issue again in the later assemblies, the original Inversion-sfGFP [013a] underwent iPCR amplification with primers YKH303/YKH317, fol-

lowed by self-/re-ligation, to remove a 102 bp fragment that contained a double *t-lpp* repeat, formerly situated downstream of the *attP* Inversion switch site and upstream of the f1 ori[18]. The resulting 6,188 bp plasmid [045] did still retain a partial 12 bp *t-lpp* fragment in this region and another full *t-lpp* sequence approximately ~1,200 bp away (as the terminating sequence downstream of the recombinase). As a further precaution to prevent non-specific binding from YKH307 (from the original vector amplifying primer pair YKH306/YKH307), a new YKH318 primer was designed as a replacement for YKH307: this was slightly offset and placed more towards the 3'-end of the recombinase so as to reduce further multi-site binding issues during amplification. As with the recombinase recoding products, later attempts in amplifying the vector (using primers YKH306/YKH318) factored in the production of additional material for visualisation by gel electrophoresis and for gel extraction-purification so as to prevent further incorrect recombinase inserts and vector backbone assemblies.

Figure 4.26 shows one such later iteration of amplifying both the vector (PCR of plasmid [045] using new primers YKH306/YKH318), and the recoded recombinase products of Darwin Assembly (PCR with original outnest primers YKH276/YKH277). Although there were other products generated alongside the vector (lane 2; 4,301 bp), these were at such low amounts relative to the required amplicon that it seemed a more optimal use of time to simply gel extract and purify the product for assembly. As for the recombinase products of experiment [1A], [1B], and [1C] (lanes 3-5, respectively; 1,977 bp), it seems that [1B] could also be gel extracted, but the products of the other two did not appear to produce as high a yield as to be efficient for assembly[19]. Nonetheless, an attempt was made

[18]These repeat *t-lpp* regions had previously been flagged as a potential issue during an earlier stage when troubleshooting the Excisions-tsPurple construct (which had a similar vector sequence; Section 4.2.5.3), and iPCR primers YKH302/YKH303 had been designed to remove 18 bp of one of the *t-lpp* duplicates (making plasmid [039]). In anticipation of the same problem here, the iPCR of the Inversion-sfGFP [013a] vector with YKH303/YKH317 to remove the 102 bp of the repeat sequences (making plasmid [045]) was performed alongside this first Darwin Assembly attempt on the recombinase. Plasmid [045] can be seen as the other ΦC** negative control construct (Figure 4.25 top row, right sample); had the first Darwin Assembly attempt on ϕc** in the Inversion-sfGFP [013a] been successful, the precautionary creation of [045] would have been unnecessary.

[19]The amplicons recovered using streptavidin beads (lanes 7-9), though demonstrated a greater recovery of products from [1B], highlighted that this method was no more efficient than direct recovery

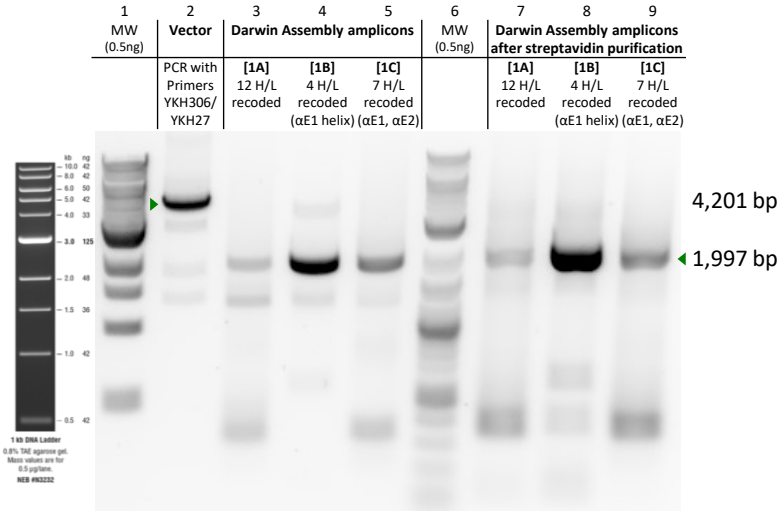


Figure 4.26: A subsequent Darwin Assembly attempt to for multi-site recoding of the recombinase within the Inversion-sfGFP circuit

Lane 2 is the vector backbone (amplified with primers YKH306/YKH318; expected size: 4,301 bp). Lanes 3-5 are amplicons from new Darwin Assemblies (made with fresh master mixes) of [1A] recoding of all 12 histidine and leucine sites, [1B] recoding the four sites in the αE1 helix (L140H, H142L, L143H, L147H), [1C] as the recoding for the αE1 helix as well as three leucines in the nearby αE2 helix-turn (L162H, L159H, L163H), respectively (expected size: 1,977 bp). Lanes 7-9 are as lanes 3-5 but were purified using streptavidin beads. Lanes 1 and 6 are DNA ladders from NEB (1 kb and 1 kb Plus, respectively).

to recover the correctly sized products of [1A] and [1C], whilst repeating the mutagenesis process once again for these two experiments. This subsequently proved somewhat successful for [1A] (shown later), but not so for [1C] (discontinued; data not shown).

Although isolating the correctly-sized insert and vector goes some way to increasing the chances of assembling the correct constructs, there were additional (post-transformation) screening steps that could be taken before sending samples for sequence verification. For instance, from examining the colony phenotypes of the first assembly attempt (Figure 4.25) with their corresponding sequencing results, it became apparent that even when the $pAmp^R$ within the Inversion switch faced away from the *sfgfp* operon, colonies that had retained a functional reporter sequence (samples 1A.1 - brightly fluorescent; samples 1A.3, 1B.2, 1C.3, 1C.7 -

post-amplification. The higher recovery of truncated products smaller than 500 bp, but absence of the products around 1.5 kb, also showed that there had likely been an assembly problem with ligating and extending from the biotinylated 5'-boundary oligo to the recoding inner oligos around the αE1 helix, with the downstream oligos to the boundary oligo on the 3'-end.

weakly fluorescent) displayed more green background fluorescence compared to those that had lost *sfgfp* (no/negligible fluorescence: samples 1A.2, 1C.1, 1C.5 - confirmed loss; samples 1A.4, 1B.1, 1B.8 - presumed loss[20]; sample 1B.7 - a T63S point mutation likely led to loss-of-function). It was equally clear from the comparison of the weakly fluorescing samples (1A.3, 1B.2, 1C.3, 1C.7) to the strongly expressing sample 1A.1 that there is a significant difference when sfGFP is fully active. Unfortunately, as expected, sample 1A.1 also had substantial changes relative to its assembly design: whilst the *sfgfp* sequence remained unchanged, it had lost the full recombinase sequence, and in its place was a 30 bp sequence that appeared to be a partial match to the reverse complement of mutagenesis oligo YKH315, followed by a 207 bp duplication of the *attB-pAmp^R-attP* switch. Both were unexpected insertions (likely products of non-specific priming) into what seemed to be the correct vector backbone. Moreover, both the *pAmp^R* promoter from the original and additional Inversion switches faced away from the *sfgfp* sequence (albeit upstream and downstream of the reporter respectively), and with transcriptional terminators in between the latter promoter and *sfgfp*. As such, there was no clear indicator as to what transcriptional mechanisms allowed the resulting strong fluorescence. A further puzzling result stems from the sequencing of sample 1C.7, which was near identical to that of sample 1A.1, and yet its level of fluorescence appeared similar to

[20]The sequencing primers used were YKH150 (which bound on the 3-'end of the vector and would amplify into the recombinase insert to confirm correct insertion and orientation) and YKH317 (pairing to a site on the 5'-end of the vector going back towards Inversion switch to confirm that it had not flipped, and which would continue further towards the insert direction to check that at least the majority of *sfgfp* was intact). As explained previously, PCR primer YKH307 (for originally amplifying the vector) had its designated binding site between the recombinase and *sfgfp*, but also (as sequencing results suggested) an unexpected binding site between *sfgfp* and the *attB* end of the Inversion switch, and a known alternate binding site after the *attP* end of the switch. However, sequencing primer YKH317 would bind in between *attP* and the alternate YKH307 binding site, which means that if YKH307 generated a vector sequencing when bound at this alternate site, it would lose not only *sfgfp*, but also the Inversion switch, and this YKH317 binding region for verifying sequencing. For samples 1A.4, 1B.1, and 1B.8, the sequencing returned clean when using YKH150, but completely failed to give any results with YKH317, suggesting that the binding site had been lost due to the YKH307 alternate binding when producing the vector. The complete lack of fluorescence from these colonies, coupled with the same phenotype seen in those that had been sequence-verified to have lost *sfgfp*, indicated that these semi-sequenced samples likely lost the gene during assembly. At the time, there was no other primer that could be used to verify this theory, nor did it seem productive to design another sequencing primer to check this as these constructs were already deemed to be useless for the project.

all other variants that retained *sfgfp*. However, given that the sequencing quality for 1C.7 dropped significantly (as as to be indeterminate) between residues 170-225 of the *sfgfp*, the reason for the differences in fluorescence likely lay in this region.

In short, it is theoretically possible to use low-level fluorescence as an initial screen for recoded colonies that have retained a functioning *sfgfp* sequence, but it should be coupled with a more rigorous assessment to determine the sequence identity of the plasmid construct. If necessary, the brightly fluorescent sample 1A.1 could be treated as a crude positive control for indicating when sfGFP is fully expressing, but it lacks as a true comparative control for the full Inversion-sfGFP circuit as it does not have the key recombinase element within the plasmid.

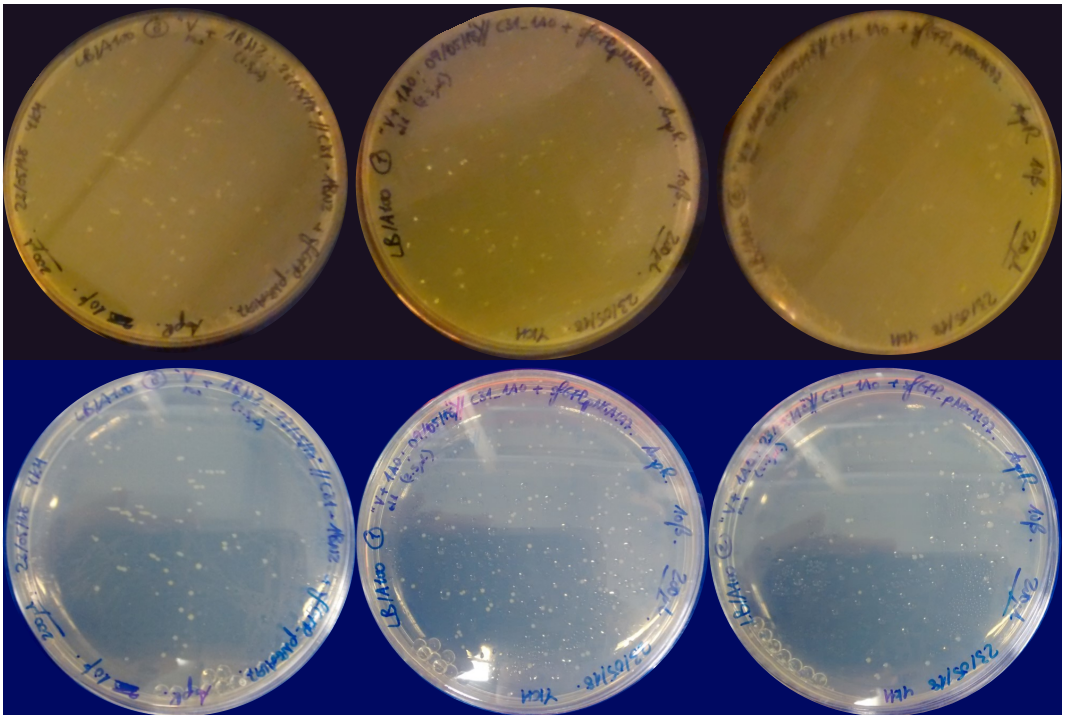


Figure 4.27: A second Darwin Assembly attempt to correctly create the Inversion-sfGFP circuit

All colonies were picked for screening via an over-day micro-culturing experiment for monitoring green fluorescence expression (Figure 4.28). When this failed to produce any clear indication of whether any had viable sfGFP constructs, a subset was selected for colony PCR screening (Figure 4.30).

In practice, the visual screening for basal fluorescence of the recoded recombinase Inversion-sfGFP transformants using subsequent Darwin Assembly attempts (two iterations of experiment [1A] and one of experiment [1B], Figure 4.27;

[1C] again failed to produce sufficient material for assembly and was subsequently discarded, data not shown) was not a definitive method for identifying successful assemblies, especially in the absence of a clearly defined negative and positive control for side-by-side comparison. This was particularly evident when visually inspecting individual agar plates of recoded constructs, given that a single plate of transformants would be unlikely to have both non-fluorescent and brightly fluorescent controls within it for direct comparison.

In an attempt to establish a better measure of the experimental transformants against the controls, nearly all colonies from the three plates shown in Figure 4.27 were picked and: (i) grown in individual liquid micro-cultures within two 96-well plates over the course of a day to see if any samples produce weakly detectable fluorescence to signify that they contain a functioning sfGFP reporter that is in its off-switch orientation (Figure 4.28), (ii) re-streaked onto solid agar plates to further assess fluorescence after overnight incubation (Figure 4.29). Alongside the 94 colonies of experiment [1A] (attempt from 20180509), all 33 colonies of [1A] and 39 colonies of [1B] (attempts from 20180523), it was possible to include a negative (LB only) and a positive (the brightly fluorescent sample 1A.1[21] from the first recoding experiments) control within both these experimental formats for direct comparison. Moreover, from these samples, it was also possible to perform colony PCR (Figure 4.30) to more accurately determine whether a recoded recombinase has been assembled as expected into the vector, and to scale up culturing to prepare purified plasmids for Sanger sequencing verification. As these later assessments would require more limiting resources than the aforementioned screening methods, these final confirmation measures were reserved until after some preliminary fluorescence data could be acquired from the over-day liquid cultures to help reduce the search space for valid Inversion-sfGFP circuits with fully recoded recombinases.

Unfortunately, the over-day liquid micro-culturing did not produce definitive

[21] Although also not ideal as it was not a correct construction in terms of the recombinase side of the Inversion-sfGFP plasmid, it was the only available control with a functioning sfGFP at the time. An ideal positive control would also have the recombinase, and where the $pAmp^R$ in its Inversion switch would have been flipped to allow expression of sfGFP, be it via a functional ΦC31 (e.g. [013ai]), or with a non-functional recombinase (e.g. [013a], [045]) that has had its switch artificially inverted via PCR (not yet created at this point).

Plate A	Date: 24/05/2018 | Time: 13:51:40 | 14:00:20												
Test ID: 1626	'[170524_Cmuts-sfG-pNG.xlsx]Sheet2'!												
Test ID: 5618	'[180524_C31-sfGFP-pNGAL97_C31muts-DA_1AO_1AO-1BN2.xlsx]Sheet2'!												
		1	2	3	4	5	6	7	8	9	10	11	12
A	Blank Corrected Raw Data (600nm)	0.001	0.266	-0.009	-0.098	-0.096	0.022	-0.169	-0.064	-0.023	**0.419**	0.118	-0.045
	Raw Data (Ex470nm/Em520nm)	11597	10517	9804	9650	9504	8751	8980	9808	10406	10422	10183	10980
B	Blank Corrected Raw Data (600nm)	-0.099	-0.061	0.209	-0.045	0.086	-0.019	0.041	0.096	-0.157	-0.023	-0.150	-0.045
	Raw Data (Ex470nm/Em520nm)	9582	9465	9491	9505	8994	9337	9365	9297	9470	9200	9225	10105
C	Blank Corrected Raw Data (600nm)	-0.032	-0.041	0.010	0.291	-0.012	-0.045	**0.484**	0.061	-0.032	-0.014	0.153	0.122
	Raw Data (Ex470nm/Em520nm)	10013	10429	10147	9709	9544	9223	9478	9284	9482	9773	9005	9093
D	Blank Corrected Raw Data (600nm)	0.085	0.190	-0.034	0.226	0.109	-0.001	-0.019	-0.059	0.122	0.046	0.013	0.066
	Raw Data (Ex470nm/Em520nm)	9887	9603	10334	8822	9010	9424	9248	9556	9356	9465	9317	10325
E	Blank Corrected Raw Data (600nm)	**0.408**	0.261	**0.468**	0.019	0.216	**0.363**	0.206	**0.334**	0.005	0.173	0.070	0.160
	Raw Data (Ex470nm/Em520nm)	10556	9842	10070	10486	9600	9422	9518	9340	10003	9386	9748	10197
F	Blank Corrected Raw Data (600nm)	**0.462**	**0.426**	0.199	0.166	0.265	0.070	0.205	**0.370**	0.240	**0.328**	0.261	0.282
	Raw Data (Ex470nm/Em520nm)	10630	10325	9742	9482	9648	10322	9244	9586	9289	9377	9734	10216
G	Blank Corrected Raw Data (600nm)	0.265	0.102	0.169	0.156	0.105	**0.474**	0.298	0.153	**0.332**	0.047	0.051	**0.511**
	Raw Data (Ex470nm/Em520nm)	10482	11453	10783	10622	10194	9511	9460	9268	9256	10153	9927	10549
H	Blank Corrected Raw Data (600nm)	0	0.269	0.270	0.048	**0.322**	0.145	0.077	0.283	**0.359**	0.155	**0.453**	**0.516**
	Raw Data (Ex470nm/Em520nm)	13236	10889	9583	11315	9976	10067	10170	10110	10101	9623	9586	**22626**

Plate B	Date: 24/05/2018 | Time: 13:53:45 | 14:06:52												
Test ID: 1627	'[170524_Cmuts-sfG-pNG.xlsx]Sheet5'!												
Test ID: 5619	'[180524_C31-sfGFP-pNGAL97_C31muts-DA_1AO_1AO-1BN2.xlsx]Sheet5'!												
		1	2	3	4	5	6	7	8	9	10	11	12
A	Blank Corrected Raw Data (600nm)	0.043	0.097	-0.026	**0.156**	0.066	-0.073	0.045	-0.091	-0.169	**0.169**	-0.144	-0.043
	Raw Data (Ex470nm/Em520nm)	19643	17334	17868	17493	17251	16677	16753	17149	18713	17292	18355	20286
B	Blank Corrected Raw Data (600nm)	-0.010	-0.029	-0.076	-0.018	-0.123	0.016	-0.052	0.027	-0.157	-0.021	0.008	-0.079
	Raw Data (Ex470nm/Em520nm)	18336	17634	19021	18070	19025	16894	16745	16846	18301	17620	17304	17804
C	Blank Corrected Raw Data (600nm)	-0.079	-0.081	-0.090	-0.103	-0.075	0.024	-0.049	-0.107	**0.177**	<0.085	<0.088	<0.084
	Raw Data (Ex470nm/Em520nm)	19193	18350	19257	19166	19492	17142	17072	19633	17214	[illegible]	[illegible]	[illegible]
D	Blank Corrected Raw Data (600nm)	0.031	-0.010	0.086	0.015	-0.004	0.072	0.055	-0.033	-0.078	-0.067	-0.051	0.020
	Raw Data (Ex470nm/Em520nm)	18863	19523	17659	18802	18219	17924	17113	17615	18942	18557	18014	19298
E	Blank Corrected Raw Data (600nm)	0.053	0.066	0.052	0.077	0.077	0.041	0.056	0.030	**0.400**	0.068	0.003	**0.187**
	Raw Data (Ex470nm/Em520nm)	19427	17783	19583	18137	18934	17991	17807	17709	17765	19198	18558	17858
F	Blank Corrected Raw Data (600nm)	**0.213**	0.098	0.080	0.055	0.098	**0.202**	0.035	-0.013	**0.124**	0.002	-0.022	**0.321**
	Raw Data (Ex470nm/Em520nm)	19004	19497	18571	18954	19748	17976	17964	17345	18133	18990	19639	17962
G	Blank Corrected Raw Data (600nm)	**0.107**	**0.154**	0.018	<0.003	<0.003	<0.007	<0.006	<0.007	<0.002	<0.004	<0.006	<0.01
	Raw Data (Ex470nm/Em520nm)	18795	17966	20719	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
H	Blank Corrected Raw Data (600nm)	0.000	0.002	0.000	<0.003	<0.002	<0.001	<0.001	<0.005	<0.002	<0.008	<0.004	**0.264**
	Raw Data (Ex470nm/Em520nm)	21947	20384	19759	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	**22916**

Figure 4.28: Fluorescence and OD600 results from an over-day micro-culturing experiment of all colonies from the Darwin Assembly attempt for creating a library of recoded Inversion-sfGFP constructs

Characterisation results following the colonies picked from Figure 4.27. Cells highlighted in a thick black border were taken forward for colony PCR screening (Figure 4.30).

results with regards to basal fluorescence that could materially inform the selection of samples for colony PCR. In the first 96-well plate (Plate A from Figure 4.28, which held the 94 colonies from the [1A] experiment on 20180509), the green fluorescent signal produced in the majority of the samples did not differ significantly from that of the negative control in well H01, from both visual inspection and from measurements taken from a platereader. This could be (incorrectly) interpreted as the samples having all lost the *sfgfp* sequence in yet another unsuccessful bid at assembling Inversion-sfGFP plasmids with recoded recombinases. However, the fundamental flaw in this deduction is in using an only LB sample as the negative control for correcting background fluorescence. As both the media and bacteria itself each display varying low-level fluorescence, the LB only control would not accurately correct for the intrinsic auto-fluorescence from *E. coli*-containing samples, let

alone be suitable for interpreting any additionally fluorescence from basal expression emitted by sfGFP-bearing recoded Inversion-sfGFP plasmid transformants. In short, the LB only negative control would not correct for background fluorescence from the experimental samples. In hindsight, a better control that should have also been included was one of the confirmed inactive Inversion-sfGFP circuits, e.g. sample 1B.7 with the loss-of-function *sfgfp*. Even though it still lacked a good portion of the inactively-recoded recombinase, at least its associated fluorescent reporter had already been demonstrated to be inactive.

A moderately better control (with regards to *sfgfp*; see footnote 26 also) was that of the strongly fluorescent sample 1A.1 (in well H12), which at least demonstrated that none of the 94 [1A] transformants produced fluorescence to the same degree as this positive control. Thus, if these transformants did contain the Inversion-sfGFP plasmid, at least the switch had not yet flipped to allow fluorescent expression.

Plate B used the same controls, in the same positions, as in Plate A, but also held the 33 [1A] and 39 [1B] transformants of the later recoding attempt (from 20180523) in wells A01-C09 and D01-G03 respectively. The remaining wells were filled with LB media. Unlike the fluorescence measurements of Plate A, Plate B appeared to indicate that all wells were comparable to the positive control, despite an obviously visible difference. This was later determined to be an error in measurement (i.e. poor calibration and mixing prior to detection) as follow-on recombinase induction assays[22] demonstrated the positive control to have 1.5- to 4.5-fold higher fluorescence expression relative to the transformant samples, and with an otherwise similar expression pattern to Plate A (data not shown).

[22]For the induction assays, the transformant samples were split into two further plates, one of which was left as is, whilst samples in the other were each induced with 10 ng/mL aTc. If the samples contained correctly assembled Inversion-sfGFP plasmids, and if any of the recoded recombinases were still functional, the expectation was that there would be a gradual increase in fluorescence signal (due to the recombinase flipping the Inversion switch to allow sfGFP expression). As it was, there were no detectable fluorescence above background levels in either induced Plate A or B samples. This could be interpreted as the constructs all successfully bearing inactively-recoded recombinases, that the plasmids were either lost or contain incorrect assemblies that are unable to express sfGFP, or not enough inducer was used. No further interpretation could be made without results from the later colony PCR experiments or from sequencing the plasmids directly.

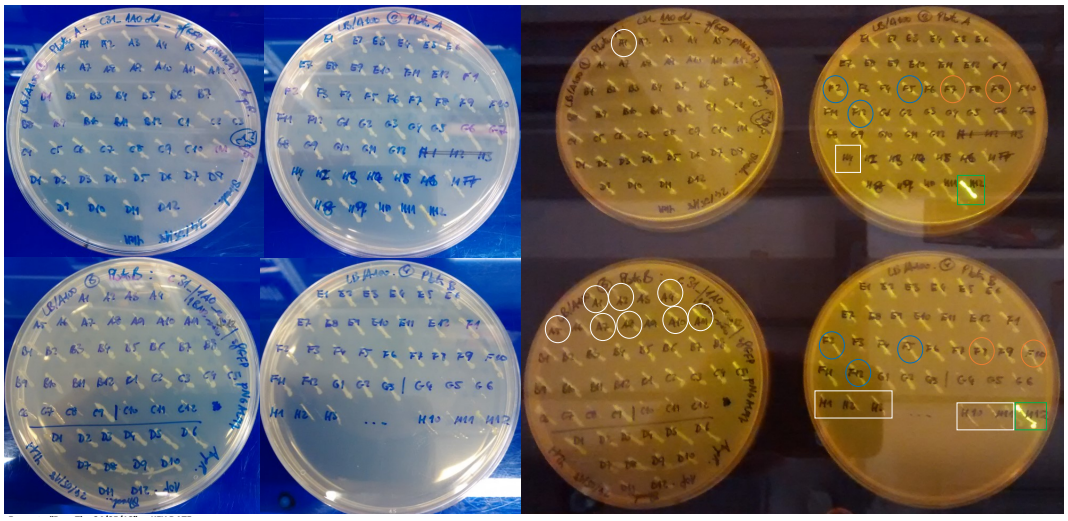


Figure 4.29: Stock plates of colonies recovered from the second Darwin Assembly attempt

The CFU from Figure 4.27 that were picked and re-streaked onto fresh agar plates for storage and visualisation for any unexpected fluorescent expression. Those circled were taken forward for colony PCR and Sanger sequencing. The green box indicates the sfGFP positive control, whilst white boxes indicate the negative controls.

At this point in time, the colony re-streaks/stock agar plates had not yet been incubated long enough to see any re-growth, let alone observe basal fluorescence, and with the over-day growth-and-induction/fluorescence culturing experiments having provided inconclusive evidence for identifying correctly assembly constructs, it was necessary to use a more definitive screening method: colony PCR. For transformants whose colony PCR results that subsequently indicate any potentially successful assemblies – that is, with full-length $\phi c31$ recombinase and *sfgfp* sequences, both in the expected orientation – overnight liquid cultures could be prepared in advance. Doing so would help accelerate the process for plasmid preparations and subsequent dispatch for sequence verification. Colony PCR screening, however, is only sufficient in informing the presence and position of the gene sequences. Determining whether the recombinases have been recoded can only be confirmed though sequencing of a subset of these confirmed constructs.

Should these screenings (and sequencing) all fail to produce reasonably inactive recoded recombinase variants, the stock agar plates could be used to screen additional constructs and select a new batch of samples to investigate. As it was, the agar plates subsequently showed that most of the colonies had some basal flu-

orescent expression, but none as bright as the positive H12 control (Figure 4.29). This suggests that most should have taken up an Inversion-sfGFP plasmid, and each with a functioning sfGFP and hopefully a recoded recombinase sequence.

As discussed, with no prior direction of which samples to take for colony PCR, it seemed prudent to simply screen for a slightly larger selection of transformants than previously planned in order to increase the chances of selecting an inactively-recoded variant. Thus, 12 (of the 94) [1A] samples across Plate A and eight (of 33) from Plate B were taken for colony PCR, whilst for the [1B] experiments, 12 (of 39) were taken forward for screening. The initial samples selected were based on the rows that contained the most samples with 0D600nm measurements above an arbitrary cut-off value of 0.1 (selected wells are denoted as cells with thick black borders in Figure 4.28). For [1A], this was row F on Plate A[23], and select samples of row A in Plate B. Likewise, row F on Plate B was picked for assessing the [1B] transformants.

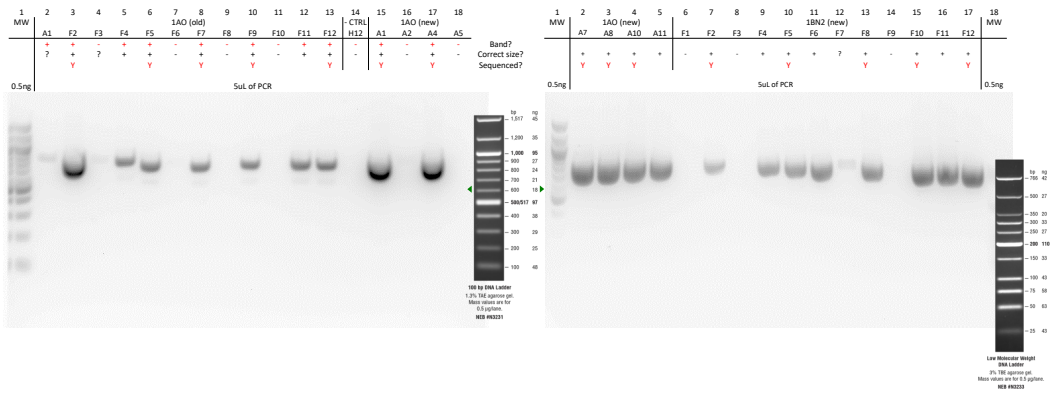


Figure 4.30: Colony PCR results of a subset of transformants from the Darwin Assembly construction of the Inversion-sfGFP library

A subset of transformants from the later Darwin Assembly construction of the Inversion-sfGFP library (Figure 4.27) that were taken forward for colony PCR screening. Correct assemblies were expected to have a length of 754 bp (amplified with primers YKH316/YKH153). Of these, a further subset were sent for Sanger sequencing.

Primers YKH316 and YKH153 were chosen to ensure that the colony PCR screen would provide sufficient information on both the presence of the *sfgfp* in

[23] Sample A01 was accidentally seeded into one of the colony PCR reactions, as opposed to F01. Instead of making a new mix, the reaction continued with A01 on the premise that it could theoretically be a negative control/wild-card sample.

the vector sequence, as well as the correct insertion and orientation of the recoded recombinase. A positive amplification would replicate a 754 bp fragment from residue R450 on ΦC31 through to the 6xHis tag at the end of the *sfgfp* sequence. In contrast, no product would be produced if either the recoded recombinase was not inserted correctly and/or if alternate priming of the vector results in the loss of the *sfgfp* operon. Knowing that the sfGFP positive control 1A.1 lacked the recombinase, this served as a suitable negative control for the colony PCRs (as evident in the absence of product in Figure 4.30, lane 14). Whilst possible, a no-template negative and positive control (using [045]) were not included. However, as can be seen in the figure, at least 22 (of 32) of the samples from both [1A] and [1B] assemblies produced amplicons around the expected size: finally providing some evidence of correct reassembly of the Inversion-sfGFP plasmid. It should be noted that some of the samples (lanes 2, 4, 5, 30) were a little larger than expected, whereas a few (lanes 3, 15, 17) were slightly smaller. Likely the former samples gained some inadvertent insertions during assembly; the latter instances may be an artifact of being overloaded as the samples appeared to have quite high yields.

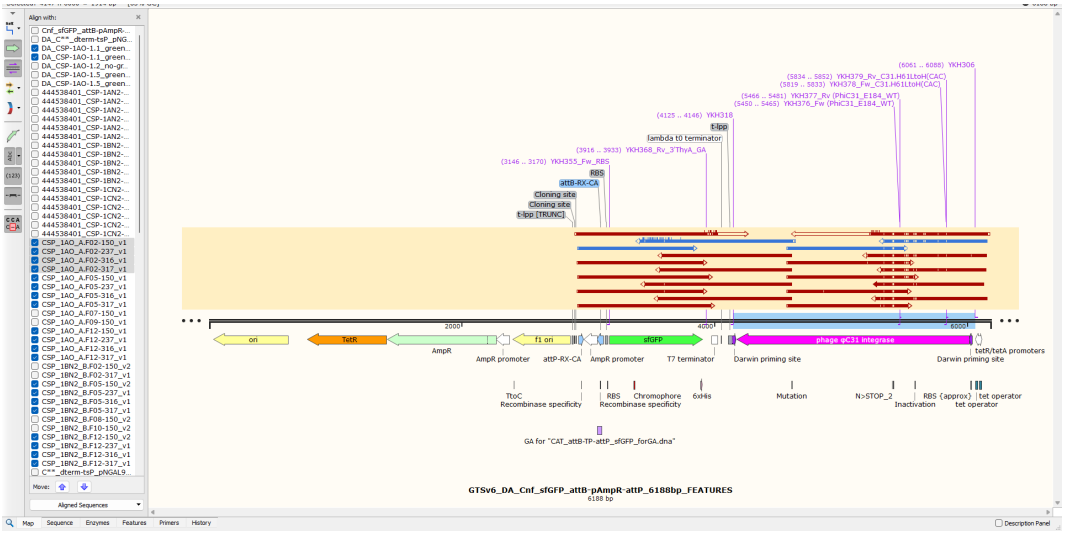


Figure 4.31: An overview of the sequencing results from a selection of the Inversion-sfGFP library constructs

Although some of these discrepancies could be investigated further during sequencing, there needed to be a pragmatic approach to the choice of samples taken

forward as the area of interest spans ~3,300b bp, and would require four primers (YKH150, YKH237, YKH316, and YKH317) to ensure good coverage. Together, these overlapping sequencing results should give clear evidence for whether the full recombinase sequence has been inserted and in the right orientation, if the target histidines and leucines within it had been recoded as designed, that the *sfgfp* is still present and functional, and that the Inversion switch has remained in its off-state (i.e. $pAmp^R$ has not yet flipped). Consequently, a set of five from each mutagenesis assemblies were sent for sequencing: from [1A], these were constructs F02, F05, F07, F09, and F12; from [1B], F02, F05, F08, F10, and F12 were selected. Note, in later analyses, these samples are simply referred to as A.F02, B.F12, etc. An overview of the sequencing results for these samples can be seen in Figure 4.31: for example, the blue arrows displays the sequence coverage of sample A.F02, where a solid fill denotes sequence identity match, whereas unfilled portions show mismatches. The expectation is that the mismatches represent where Darwin Assembly has been successful in recoding the designated codon sites, and that doing so would inactivate the recombinase such that the Inversion switch would not be flipped.

On the whole, the multi-site mutagenesis of ϕc^{**} into histidine-leucine recoded variants and subsequent re-assemblies were reasonably successful. To begin with, it appears that each sample retained a functioning *sfgfp* and, given that all switches remained in their off-states, it is likely that the recoded recombinases were all inactive variants. Table 4.2 shows in more detail the relevant recombinase sequencing results for three of the samples from each of the [1A] and [1B] mutagenesis experiments: cells in green show where the recoding has been implemented as expected, whilst those in pink highlight deviations from the design.

Recall that the [1A] mutagenesis experiments were to recode 10 potentially critical leucines to histidine, two histidines to leucine, and revert two Amber stop codons back to asparagine (Table 4.1a,b), whereas [1B] was to edit only the four candidate residues in the αE1 helix (L140H, H142L, L143H, L147H) and correct the two stop codons. As can be seen in Table 4.2, A.F02 was successful in generating all the required recodings, as was A.F09 (but with an additional synonymous

	[1A]O			**[1B]N2**		
	A.F02(/9)	**A.F05**	**A.F12**	B.F02(/8)	**B.F05**	**B.F12**
M1		RBS SDM: G>A		Δ1nt: post-*pTetR/A*, pre-RBS		
G14				$G14S_{(Agt)}$		
H61	$H61L_{(cTT)}$	$H61L_{(cTT)}$	$H61L_{(cTT)}$			
L83	$L83H_{(cAT)}$	$L83H_{(cAT)}$	$L83H_{(cAT)}$			
L117	$L117H_{(cAT)}$					
L118	$L118H_{(cAT)}$					
L120	$L120H_{(cAT)}$					
L140	$L140H_{(cAT)}$	$L140H_{(cAT)}$	$L140Q_{(cAG)}$	$L140H_{(cAT)}$	$L140H_{(cAT)}$	$L140H_{(cAT)}$
H142	$H142L_{(cTT)}$	$H142L_{(cTT)}$		$H142L_{(cTT)}$		$H142L_{(cTT)}$
L143	$L143H_{(cAT)}$	$L143H_{(cAT)}$		$L143H_{(cAT)}$		$L143H_{(cAT)}$
L147	$L147H_{(cAT)}$	$L147H_{(cAT)}$	$L147H_{(cAT)}$	$L147H_{(cAT)}$	$L147H_{(cAT)}$	$L147H_{(cAT)}$
H151						
L162	$L162H_{(cAT)}$	$L162H_{(cAT)}$	$L162H_{(cAT)}$			
N166	$N166N_{(aaT)}$	$N166N_{(aaT)}$	$N166N_{(aaT)}$			
L167	$L167H_{(cAT)}$	$L167H_{(cAT)}$	$L167H_{(cAT)}$			
L171	$L171H_{(cAT)}$	$L171H_{(cAT)}$	$L171H_{(cAT)}$			
*200	$*200N_{(AaT)}$	$*200N_{(AaT)}$	$*200N_{(AaT)}$	$*200N_{(AaT)}$	$*200N_{(AaT)}$	$*200N_{(AaT)}$
I203	$I203I_{(atC)}$	$I203I_{(atC)}$	$I203I_{(atC)}$	$I203I_{(atC)}$	$I203I_{(atC)}$	$I203I_{(atC)}$
*204	$*204N_{(AaT)}$	$*204N_{(AaT)}$	$*204N_{(AaT)}$	$*204N_{(AaT)}$	$*204N_{(AaT)}$	$*204N_{(AaT)}$
L206						
H208						
L213						
R225	$R225Q_{(cAg)}$	$R225Q_{(cAg)}$	$R225Q_{(cAg)}$	$R225Q_{(cAg)}$	$R225Q_{(cAg)}$	$R225Q_{(cAg)}$
H234						$H234N_{(Aac)}$
E269		$E269V_{(gTg)}$				
W349						
L359						
S360						
E386					E386Δ-aa	
L468					Δ-aa	
Correction	YKH323/324	YKH323/324	YKH326/324	–	YKH323/324	YKH323/324
PCR I:V	YKH321/322	YKH321/322	YKH321/325	–	YKH321/322	YKH321/322

Table 4.2: A summary of the ΦC31 residues in the Inversion-sfGFP circuit that were re-coded using Darwin Assembly, as confirmed by Sanger sequencing

Cells in green show where recoding has been implemented as expected; cells in pink highlight deviations from the designs. Subsequent PCRs using the primer pairs at the bottom of the table were able to correct all the mis-assemblies (pink). The five constructs in bold were selected as suitable for testing with the OCI system. Table 4.3 shows an equivalent summary for histidine-leucine recodings in the Excision-tsPurple constructs.

recoding of residue $V5V_{(gtC)}$; data not shown), and nearly so in A.F07 (with the exception of L162, which was not recoded; data not shown). However, it was expected that there would be further assemblies with more incomplete recodings, such as in A.F05 and A.F12. In both instances, oligo YKH311 may not have bound efficiently, leading to no recoding of L117H, L118H, L120H in the αD2 helix. With A.F12, it appears oligo YKH314 also had problems in binding, leading to only L147H in the αE1 helix being recoded as designed; there was additionally a mis-coding of $L140Q_{(cAG)}$, and H142 and L143 were not recoded. A.F05 too suffered a mis-coding of $E269V_{(gTg)}$, and a G-to-A transition mutation in the putative RBS sequence. All these minor recoding mistakes suggest that the implementation of the Darwin Assembly mutagenesis method was not fully optimal, but sufficient enough to generate a range of recoded recombinases that, after slight corrections, would be fit to test sense-to-sense codon reassignment.

On the other hand, more amendments were needed with all the [1B] assemblies in order to make them suitable for assessing reassignment. Although four of the five constructs sent for sequencing incorporated all four recodings in the αE1 helix and reverted the two stop codons to asparagine, each also gained various other point mutations or construction error. B.F02 and B.F08 were identical, but both had acquired a $G14S_{(Agt)}$ mutation within the recombinase and a single nucleotide deletion in the segment directly upstream of the RBS. Likewise, B.F12 had accrued an addition $H234N_{(Aac)}$ mutation. With B.F10, this construct probably had all the required elements (with no addition recodings), but the quality of the sequencing dropped shortly after the stop codon corrections and, given that there were other constructs that contained the core recodings (as well as an understanding of how to correct their additional errors), it was not deemed essential to further sequence-verify the low quality regions of this construct (data not shown). By comparison, B.F05, the only construct to not incorporate all the expected αE1 helix recodings, was worth examining in more depth. It had only recoded L140H and L147H, and was not able to modify the critical L143, nor the potentially interesting but likely non-essential H142, so was a useful case to see if these adjacent mutations alone were sufficient

to render ΦC31 inactive. In its current form, however, this assessment could not yet be made: a single point deletion of nucleotide G, at where residue E386 was expected to be, caused a frameshift throughout the remainder of the construct.

One additional error that was consistent with all the recombinases within the [1A] and [1B] assemblies was mutation R225Q$_{(cAg)}$: though it is unlikely that the loss of activity in the recombinases were solely due to this mutation as all the assemblies had multiple recodings (and other build errors). It was, however, unclear how this mutation came too be. Most likely it was introduced during the creation of the final Darwin Assembly template (plasmid [045], where most of the multiple *t-lpp* replicates within the sequence were removed), yet there was no evidence of this when [045] was sequenced-verified. Regardless, to confirm that the loss-of-function was a result of the proposed recombinase residue recodings, the construction errors were corrected with PCR using the primers as designated at the bottom of Table 4.2. Note, the top primer pair (either YKH323/YKH324 or YKH326/YKH324) amplifies the relevant portion within the recombinase where the key residues were recoded from each of the respective constructions (A.F02, A.F05, A.F12, B.F05, and B.F12 as templates; insert, I, 536/279 bp), whilst the bottom pair (either YKH321/YKH322 or YKH321/YKH325) amplifies and corrects this mutation (using construct A.F02 as the template) to form the complementary vector sequence (V, 5,702/5,959 bp). B.F02 was not taken forward as the rectified version would be the same as B.F12. The five corrected inserts and vectors were re-ligated again to reconstruct the recoded plasmids and, in all cases, subsequent Sanger sequencing shows that the errors (pink cells in Table 4.2) were corrected to wild-type. The exception being A.F12, where the L140Q$_{(cAG)}$ was recoded to L140H$_{(cAT)}$. In all cases, the background sfGFP expression – while still present – was at a minimal level.

At this point, these five recoded Inversion-sfGFP constructs were ready to be combined with the OCI anticodon-modified tRNAs to characterise sense-to-sense codon reassignments via its proxy sfGFP signal. The primary concern was whether histidine-leucine reassignment could be demonstrated sufficiently using these vari-

ants, considering that all previous instances of reassignment was based on a single residue. For example, like the OCI CAT.H193L construct, B.F05 would only be unidirectional, albeit with the opposite, not previously been demonstrated, histidine-to-leucine reassignment. B.F05 would also extend reassignment to two targets, L140H$_{(cAT)}$ and L147H$_{(cAT)}$). On the other end, A.F02 would pose a far bigger challenge, with the need to reassign 12 residues: both recoded histidines (of which there are two), and leucines (ten). Nevertheless, the intention was to use these as later constructs for assessing how far we can take histidine-leucine reassignment in an otherwise naturally-coded organism.

4.2.6.2 Minimal histidine-leucine recodings for an inactive ΦC31 within the Excision-tsPurple switch

In Section 4.2.6.1, five Inversion-sfGFP constructs were established where the recombinase was rendered non-functional by recoding two to 12 different histidine and leucine residues, all in preparation for testing sense-to-sense codon reassignment using the OCI approach. At the same time, the Excision-tsPurple [033] plasmid was being modified using the established method of iPCR to make targeted single/minimal residue recodings at sites purported to impact ΦC31 functionality: specifically, residues (i) H61L, (ii) H142L_L143H, and (iii) L468H to independently test leucine-to-histidine and histidine-to-leucine codon reassignment.

Despite being conceivably simplistic constructions for each of these ΦC31 variants, theoretically requiring only two iPCRs – one to revert the amber stop codons at residues 200 and 204 to asparagines, and a second to make the single/dual histidine-leucine recodings – this approach developed a number of problems. For starters, it appeared that there was a contamination in the original [033] Excision-tsPurple stock as this ΦC** negative control started expressing TinselPurple after growth overnight (data not shown). A pre-emptive removal of a repeat *t-lpp* sequence in the [033] Excision-tsPurple circuit (making it plasmid [039] as described in footnote 23; theoretically to facilitate easier re-assemblies) seemed to produce more issues with the introduction of an L359P mutation in the recombinase sequence that was not detected prior to the recodings, resulting in its propagation in

the subsequent recoded recombinase constructs. However, the biggest problems lay in the assembly of the recoded recombinases and the resulting Inversion-tsPurple phenotypes. Ignoring that the ΦC** negative control constructs (all #1.1-8 samples in Figure 4.32) inexplicably displayed TinselPurple expression, some of the biological replicates for the recoded recombinase constructs for *ϕc31*.L468H (at least #3.2, #3.4, #3.5, #3.7, #3.8 to a minor extent), *ϕc31*.H142L_L143H (at least #4.2 and

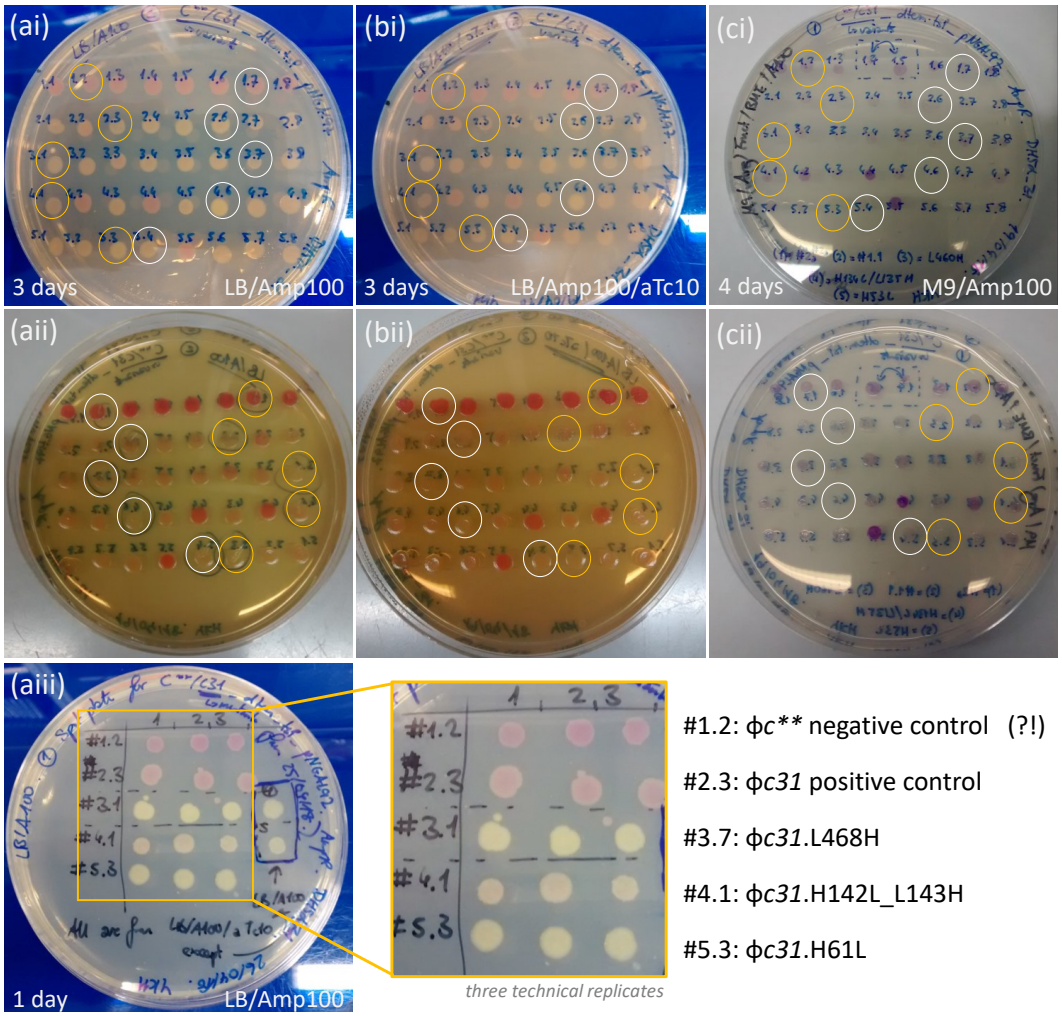


Figure 4.32: Plate screening for ΦC31 recombinases with the H61L, H142L_L143H, and L468H recoded variants via the absence of TinselPurple expression

Samples with the same numbering are the same across all three plates (a), (b), (c) series. #1.x refers to *ϕc*** negative controls, #2.x are for the *ϕc31* positive controls, #3.x are *ϕc31*.L468H, #4.x are *ϕc31*.H142L_L143H, and #5.x are *ϕc31*.H61L; x is the biological replicate number. The top row (i) shows the colony spots as viewed from the bottom, the middle row (ii) shows the spots from the top (agar plate open), and the bottom row (iii) are technical replicates of a selected colony from each variant that was selected to continue with subsequent experiments. Colonies circled were sequenced and discontinued (white) or continued (yellow).

#4.4 significantly; #4.8 minor extent), and $\phi c31$.H61L (at least #5.1 significantly so) also accumulated pink-purple colouration. If recoding these residues was insufficient at knocking out recombinase function, all replicates would be expected to show TinselPurple expression. The isolated incidences, and the fact that the expression occurred in the same samples with (Figure 4.32b) and without (Figure 4.32a,c) aTc induction, suggests that it may have been a problem in the recoding/re-assembly process that gave rise to the variation. As such, two colonies from each variant that displayed the least TinselPurple expression were initially selected for further characterisation and in search of the recoded recombinase circuit(s) that would be suitable for assessing codon reassignment.

Selecting against colonies with chromoprotein expression meant that the ones with functional recombinases or that had otherwise irreversibly lost the double-terminator in the *tsP* operon would already be screened out. Those remaining (with none-to-minimal colour) would likely be variants that possess a non-functional recoded recombinase; else, are constructs that have been assembled incorrectly in some other manner (e.g. with a non-functional tsPurple circuit). Given that it had already been demonstrated that the Excision-tsPurple circuit is leaky (even if mildly ameliorated in DH5α_Z1; Section 4.2.5.3) and that very little induction is needed to activate a functional recombinase, another method for screening out unsuitable variants would be to culture the candidates with aTc. The initial attempt at inducing the variants on solid media revealed that TinselPurple can be more easily observed in LB if expression is high or allowed to accumulate for longer, whilst early expression (with light pink colonies) can be hard to distinguish against the amber-coloured agar background. Hence, it was non-trivial to visually differentiate expression between the uninduced and uninduced sets (Figure 4.32a,b). The signal was cleaner using M9 minimal media (purple colonies on pale yellow agar; Figure 4.32c), but the time before chromoprotein expression can be observed took significantly longer: approximately 1 day in LB vs 3-4 days in M9 media. A faster, intermediary screen would be to use colony PCR (with primers YKH288/YKH248) across the double-terminator disrupting the *tsP* operon. This could be performed

both on colonies that have been grown with or without aTc, and on agar plates (i.e. from Figure 4.32) or from liquid cultures, even before TinselPurple can be detected. Screening with colony PCR in advance would also serve to minimise the number of samples needing to be sequenced, as at least three primers would be required to fully cover the ~ 3,200 Excision-tsPurple circuit for each variant. The expected colony PCR results if the recoded recombinases are non-functional and unable to excise the double-terminator would be the production of a 364 bp amplicon; those that have excised the sequence would instead produce a 183 bp product. For those that show a lack of excision, uninduced liquid cultures would be prepared to extract the DNA for sequencing and verify the construction at the sequence level.

Having determined the characterisation parameters, a subset of samples circled in Figure 4.32 underwent colony PCR: results of the variants that were amplified from samples grown without aTc are seen in lanes 6-11 of Figure 4.33 (directly from the agar plate of Figure 4.32a), or with the inducer (lanes 12-14; from a later experiment while selected samples were being sequenced). Samples selected for sequencing had a preliminary scan across the relevant regions in the recombinase (with primer YKH150) and over the 5'-end of the *tsP* operon (primer YKH288) to verify the presence/absence of the double-terminator, whilst those circled in yellow (from Figure 4.32) were sequenced further (primers YKH237, YKH316) and were taken forward to the next stage. The latter were also re-spotted as technical replicates on LB Amp 100 μg/mL agar plates to check again for TinselPurple expression and, in this case (bar that of the #1.2 negative control), the pattern was as expected: TinselPurple present in the #2.3 test-case/"positive"[24] control, and negligible to no expression in the recombinase-recoded variants.

For the most part, the sequencing results of the selected colonies (summarised in Table 4.3) also matched the phenotypes displayed (Figure 4.32). For example,

[24]Sample #2.3 is considered "positive" in the sense that as the recombinase is now functionally-restored to *ϕc31* and as the system is known to be leaky, it would only be a matter of time for the double-terminators to be excised from all instances of the *tsP* operon, allowing significant TinselPurple expression and making it a "true positive". This plasmid is equivalent to [033b] as created in Section 4.2.5.2, minus an 18 bp *t-lpp* sequence. Its provenance being from the 6,190 bp [033] ΦC** plasmid, to the 6,172 bp [039] ΦC** construct (following the *t-lpp* removal), and to the 6,172 bp sample #2.3 (with ΦC31 restored).

	As recoded from the Excision-tsPurple plasmid [039]				
	#1.2(/7)	#2.3	**#3.7(/1) H** L468H	**#4.1 LH** H142L.L143H	**#5.3 L** H61L
M1					
H61					H61L$_{(cTT)}$
L83					
L117					
L118					
L120					
L140					
H142				H142L$_{(cTT)}$	
L143				L143H$_{(cAT)}$	
L147					
H151					
L162					
N166					
L167					
L171					
*200		*200N$_{(AaT)}$	*200N$_{(AaT)}$	*200N$_{(AaT)}$	*200N$_{(AaT)}$
I203		I203I$_{(atC)}$	I203I$_{(atC)}$	I203I$_{(atC)}$	I203I$_{(atC)}$
*204		*204N$_{(AaT)}$	*204N$_{(AaT)}$	*204N$_{(AaT)}$	*204N$_{(AaT)}$
L206					
H208					
L213					
R225					
H234					
E269					
W349				W349R$_{(Cgg)}$	
L359		L359P$_{(cCg)}$	L359P$_{(cCg)}$	L359P$_{(cCg)}$	L359P$_{(cCg)}$
S360					S360P$_{(Cca)}$
E386					
L468			L468H$_{(cAT)}$		
Correction PCR I:V	YKH323/324 YKH321/322	YKH323/324 YKH321/322	YKH323/330* YKH327/328 then YKH329/322	YKH323/324 YKH321/322	YKH323/324 YKH321/322

Table 4.3: A summary of the ΦC31 residues in the Excision-tsPurple circuit that were recoded using iPCR, as confirmed by Sanger sequencing

Cells in green show where recoding has been implemented as expected; cells in pink highlight deviations from the designs. Subsequent PCRs using the primer pairs at the bottom of the table were able to correct all the mis-assemblies (pink). The three constructs in bold were selected as suitable for testing with the OCI system. #1.2 and #2.3 were used as negative and positive controls, respectively. Table 4.2 shows an equivalent summary for histidine-leucine recodings in the Inversion-sfGFP constructs.

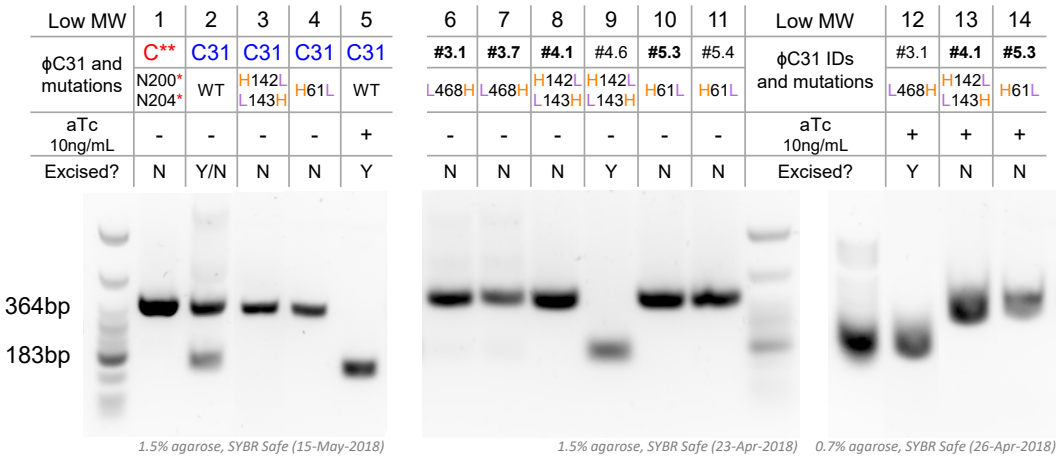


Figure 4.33: A colony PCR screen of recoded ΦC31 variants with loss-of-function activity

Lanes 6-11 reflect colony PCR of the initial transformants of recoded recombinases in the ExcisionTsPurple variants from Figure 4.32. Lanes 12-14 are a subset of these variants that were induced with 10 ng/mL aTc to assess whether the recoded recombinases remain non-functional with induction. Lanes 1-5 are a summary of the ΦC** control, the ΦC31 test-case (with and without aTc), and the most promising recoded recombinases (*ϕc31*.H142L_L143H, *ϕc31*.H61L). Non-functional recombinases that are unable to excise the double-terminator are expected to produce 364 bp products; those with excision activity should produce amplicons at 183 bp.

in #2.3 the chromatogram across the *tsP* part of the construct shows an admix population, with some having lost the double-terminator within the operon (data not shown). This corresponds with the pink tinge seen in Figure 4.32; particularly so in (aiii), which was re-plated a week after the plasmid was transformed into (ai/ii), having had more time for ΦC31 to act. Interestingly, the heterogeneous sequencing pattern of the chromoprotein part of the circuit in #2.3 was comparable to that of the allegedly negative control samples of #1.2 and #1.7, indicating that the double-terminator was also (being) lost in these samples, thereby explaining the presence of TinselPurple expression in these controls. However, the reason as to why excision was occurring in #1.2 and #1.7 was unclear as the recombinase still retained its non-functional *ϕc*** sequence (data not shown). Possible hypotheses could be that the strain bearing the original [033] ΦC** negative control that was selected for storage was not taken from a single colony (though previous sequencing does not indicate that this was the case; Figure 4.19), or that there was a spontaneous drop-out of the double-terminator sequence (not yet seen before). By comparison, in #2.3, sequencing of the recombinase reveals that the stop codons were successfully

corrected back to asparagine – reflecting the re-gain-of-function. Also of interest is that the ΦC31 recombinase still had activity despite an unexpected L359P mutation, suggesting that had this leucine been selected as a recoding candidate, it would not have drastically hindered enzyme activity. As it was, this leucine was not previously considered as it was not a part of any particular domain of interest, and annotations on the PDB structure indicated no secondary structure in this region (Figure 4.6).

As for the recombinase-recoded variants: in both #3.1 and #3.7, the sequencing traces for the recombinase and *tsP* operons showed a homogeneous population: bar the accidental (but seemingly insignificant) L359P mutation, the L468H recoding was correctly incorporated and the stop codons reverted, all with no indication of the recombinase causing excision of the double-terminator, nor any expression of TinselPurple (Figure 4.32aiii). This seemed contradictory to previous research, which indicated that L468A/P was only able to partially abolish recombinase activity when present as a single point mutation (Table 4.1; Liu et al., 2010; Rowley et al., 2008). It was only when this *ϕc31*.L468H was cultured again in 10 ng/mL aTc, however, that it became clear the earlier findings held true: colony PCR of #3.1 post-induction shows that the recombinase was able to excise the double-terminator from the *tsP* operon (Figure 4.33, lane 12). That the recombinase was still functional despite the L359P and L468H mutations was further confirmation that neither on their own are able to fully suppress ΦC31 activity. As an aside: it was later found that there was a G-to-A transition within *tsP*, $R92H_{(CaT)}119$, that may have caused a loss-of-function in the chromoprotein, hence the lack of pigmentation in #3.1 from the proxy reporter (Figure 4.32aiii), despite the positive confirmation of the recombinase activity. As it was, these results meant that L468H (and likely L359, if recoded to histidine) would be insufficient – on their own, and in this case, as a double mutation – to inhibit the ΦC31 recombinase and, therefore, would not be suitable for demonstrating the histidine-to-leucine reassignment capabilities of the OCI approach.

In terms of the *ϕc31*.H142L_L143H mutations, the early colony PCR screen found that the recombinase in #4.6 was able to excise the double-terminator, thus

had still retained functional activity (Figure 4.33, lane 9). As such, this sample was not taken forward for sequence verification. Sample #4.1, on the other hand, whilst arguably displayed very faint chromoprotein expression in its proxy reporter upon being transformed into DH5α_Z1 (Figure 4.32aiii), did not display any excision activity from its upstream recombinase component (lane 8 and 3), nor was the enzyme active when induced with aTc (lane 13). Sequencing confirmed that both the reversion of the internal stop codons and the recoding of the double H142L and L143H residues were correctly incorporated into the recombinase. However, besides the likely innocuous L359P substitution carried through from an earlier assembly, a new W349R mutation was also introduced (Table 4.3); other attempts failed to attain a fully correct sequence (with #4.3, #4.7; data not shown). It was, therefore, possible that the lack of recombinase activity could be attributed to either this mutation, or a combination of W349R, H142L, and L143H. Besides noting its placement in an alpha helix, it was difficult to ascertain the extent of the effect of W349R on ΦC31 function; it would be necessary to correct this point mutation before checking again for recombinase (and downstream TinselPurple) activity.

The candidates for the *ϕc31*.H61L recoding was similarly correctly recoded, but also additionally confounded by off-target mutations. In the case of sample #5.3, candidate residue H61L was properly recoded, the nonsense stop codons corrected, but in addition to the L359P that was carried through, there was a further S360P mutation. Regardless, no drop-out of the double-terminator was observed in the sequencing trace, nor presence of the chromoprotein in the agar plate. That said, given that this latter mutation had not been characterised before, it was unclear whether it also contributed to the lack of functionality seen in the recombinase. As a counterpoint, S360P is adjacent to L359P in an area of no known secondary structure, and as the latter had limited effect on recombinase function, the consequence of the S360P mutation may not be as significant either. It should be noted, however, that sequencing revealed a nucleotide point deletion at $V_{(gTG)}119$ in *tsP*, which would likely cause a frameshift and non-expression of the chromoprotein. Consequently, it is likely there were two issues with the #5.3 *ϕc31*.H61L sample: some combi-

nation of mutations in ΦC31 rendered the recombinase non-functional, resulting in retention of the double-terminator; a frameshift deletion in *tsP* prevented any further indication of leaky expression from the chromoprotein proxy reporter. Unfortunately, the other *ϕc31*.H61L sample #5.4 proved to be a mis-assembly, lacking the majority of the recombinase (data not shown). A screening of additional samples from the stock plate also failed to find a correct assembly for both recombinase and *tsP* genes: #5.2 and #5.8 had different point mutations in the chromoprotein; #5.5 was confirmed to have fully lost the double-terminator but an otherwise intact Excision circuit, hence the strong pink-purple expression (data not shown).

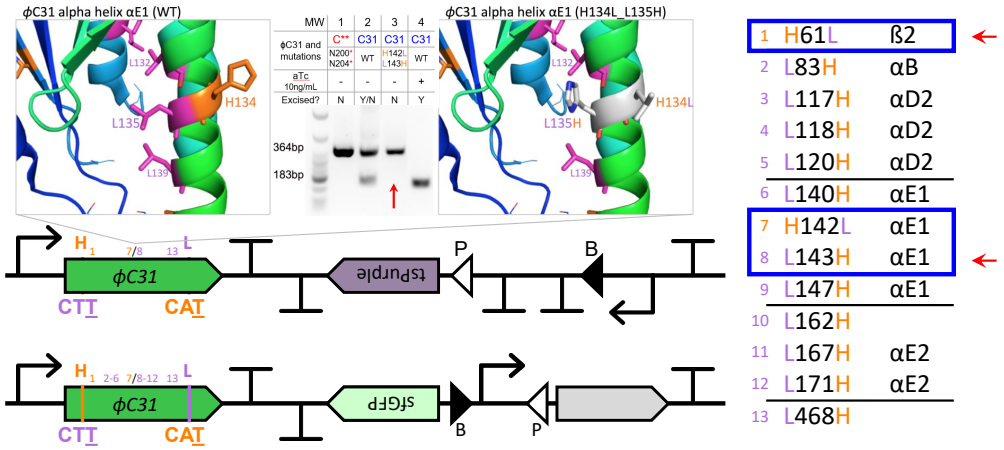


Figure 4.34: ΦC31 variants with candidate inactivating point mutation(s)

The final OCI-GTS experiments, combining both the engineered OCI tRNAs and the GTS Inversion-sfGFP, were subsequently focused to the recoded H61L and H142L_L143H sites (in the blue boxes, indicated by the red arrows) within the ΦC31 recombinase. The intention was that should there be success in reassigning these minimal recoding targets, it would be possible to next characterise reassignment in the many multi-site histidine-leucine recoded variants that were initially created (see Table 4.2).

The conclusion from the analysis of the recoded ΦC31 variants from the Excision-tsPurple constructs, therefore, was that whilst the expected targets for minimal ΦC31 recodings were achieved individually (H61L, H142L_L143H, L468H), the Excision-tsPurple was not reliable enough in its current state to be used as a clear indicator for sense-to-sense codon reassignment.

Be that as it may, it was still possible to make use of the recodings by subcloning the relevant portions into the earlier Inversion-sfGFP constructs. The final row in Table 4.3 indicated the primers used to successfully implement these

transfers, with the resulting key minimal recodings for consideration summarised in Figure 4.34.

Having finally established a viable GTS circuit, as well as finding sites within the new OCI ΦC31 recombinase reporter for assessing histidine-leucine reassignment, it was now possible to co-transform these plasmids with the anticodon-modified tRNAs to test the codon reassignments (Figure 4.35).

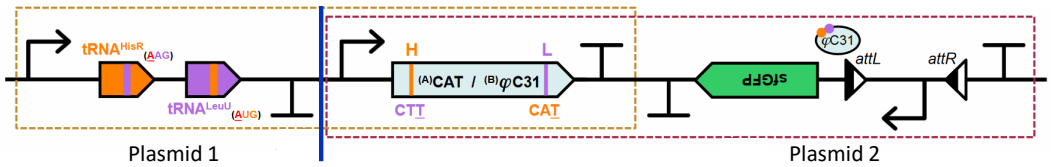


Figure 4.35: A schematic of the combined OCI-GTS system for characterising sense-to-sense codon reassignment

The GTS plasmid (to the right of the blue vertical line) was co-transformed with a minimal anticodon-modified tRNA plasmid (to the left of the blue vertical line) to assess the feasibility of implementing histidine-leucine reassignment. The ΦC31 recombinase forms an overlap of the OCI (dotted golden box) and GTS (dotted maroon box) systems, as it is the recoded reporter target of the alternatively-coding tRNAs from the former system, but also forms the integral part to the Inversion-sfGFP circuit; it is the crux to flip from the transient signal of the engineered tRNAs to the permanent response of the naturally-coded sfGFP proxy secondary reporter.

4.2.7 Combining the OCI and GTS systems together

With the anticodon-modified tRNAs from Section 3.2.1 and the set of inactive recombinase reporter variants from Section 4.2.6, I now possessed all the key elements for testing the full OCI-GTS system together.

Yet there still remained a few issues to resolve before the full circuit could be characterised. As the original designs for both OCI and GTS circuits were based on the same ColE1 family plasmids, it would not be possible to co-transform the two without encountering incompatibility/stability issues. The original intention had been to subclone the 400 bp engineered tRNA transcription unit from the OCI plasmid into the GTS Inversion-sfGFP plasmid, thereby removing the need to co-transform altogether. However, having encountered the number of cloning issues with establishing the GTS circuit, it seemed more prudent to modify the plasmid bearing the engineered tRNAs to make the two plasmids compatible instead.

With the Inversion-sfGFP plasmid, the key features of note here were the use of

the ColE1 origin of replication and the ampicillin resistance marker. The same features existed in the engineered tRNA plasmid, alongside the CAT gene that was the original associated OCI reporter. Hence, the key element that needed to be changed was the origin of replication for the tRNA plasmid. As it was, there had been early research within the Pinheiro group by Eleftheria Stamou and Ana Riesco to mutate the diversifying loop regions within the ColE1 element to create new viable origins that were compatible with the wild-type (research subsequently published in Chaillou et al., 2022). From the initial mutational studies of the RNAI region of ColEI, a sequence termed the "alpha" ori had been identified that appeared to be orthogonal to the wild-type. As such, primers YKH339/YKH340 were used to PCR amplify out the original ColE1 ori, and the alpha ori was inserted (via SapI restriction cloning) in place of it.

In terms of the AmpR selection marker within both the Inversion-sfGFP and the tRNA plasmids: whilst there are benefits of both plasmids maintaining the same marker (i.e. only needing to select with ampicillin for both plasmids), if removed from the tRNA plasmid (leaving only the CamR marker), this could be a strategy for verifying that both plasmids had been retained within the host. That is, when the two plasmids are co-transformed, if there the host becomes sensitive to chloramphenicol, then it would suggest that the tRNA plasmid had been lost (such as if alpha truly was incompatible with the wild-type ColE1 from the GTS plasmid). Likewise, a loss of ampicillin resistance would suggest that the Inversion-sfGFP plasmid had been lost. Thus, subsequent to the replacement of the origin of replication, primers YKH231/YKH284 were used to amplify out the *bla* gene and the resulting fragment was blunt ligated together.

Having made these changes to the tRNA construct, the resulting 3,436 bp plasmid was co-transformed with the Inversion-sfGFP plasmid into DH10β *E. coli* strains, with no indication that there was a loss of either plasmids on carbenicillin/ampicillin and chloramphenicol selection media (data not shown).

4.2.8 Analysis of the combined OCI and GTS system for detecting the signal of histidine-leucine reassignment using flow cytometry

Whilst it was possible to observe a slight increase in fluorescence between experimental and control conditions of the strains containing both portions of the OCI and GTS Inversion-sfGFP construct, it was difficult to quantify these differences. As such, flow cytometry was used to statistically quantify the level of fluorescence emitted from the cells across a range of conditions.

To set up the baseline measurements, a positive control condition was established with a DH10β strain that contains only the Inversion-sfGFP construct, with a functional and active ΦC31 recombinase, that had already been irreversibly switched to its "on" state (i.e. constructively expressing the proxy sfGFP; Figure 4.36, "POS" on the top row, far right). A negative control was established using an inactive variant of the GTS construct ($\phi c31$.H53L) that was coupled not to sfGFP but to a non-fluorescent ThyA reporter (Figure 4.36, "NEG" on the top row, second from the right). As the two populations are distinctly different, a gate was set-up between the two, at 10^3 fluorescence intensity of GFP (BL1-H): a point which was sufficient in grouping 99.01% of the cells as non-fluorescing, and 97.35% of the cells as being fluorescent. Note for each run, two technical replicates of each condition was performed, each of which processed 10,000 cell samples. The results shown are an average of the replicates.

Having set up a threshold for delineating between fluorescent and non-fluorescent cells, the cell populations for the different codon reassignment assay conditions could be measured. The three conditions characterised were: D6, the DH10β strain bearing only the GTS $\phi c31$.H53L construct with the sfGFP proxy reporter; D7, the DH10β strain bearing both the GTS $\phi c31$.H53L construct with the sfGFP proxy reporter, and the histidine-leucine reassigning tRNAs plasmid, in the absence of any inducers; D8, the DH10β strain bearing both the GTS $\phi c31$.H53L construct with the sfGFP proxy reporter, and the histidine-leucine reassigning tRNAs plasmid, but which had been cultured with 20 ng/mL aTc (Figure 4.36, bot-

tom row, left to right). For sample D6, 81.05% were within the fluorescent gate (vs 18.95% in non-fluorescent), but median intensity measurement was at 1,945.0; sample D7, 34.88% were within the fluorescence range (vs 65.12% non-fluorescent), with median intensity at 769.0; sample D8, showed 84.92% as fluorescent (vs 15.08% non-fluorescent), with median intensity at 2,482.5. For comparison, the median intensity for the negative control was at 103.0, whilst the positive control was 18,931.5.

The differences between the different conditions are more clearly depicted in the left two sub-figures on the top row of Figure 4.36. The sub-figure on the far left shows the following conditions with increasing fluorescent intensity: negative control (grey, median at 103.0), sample D6 (baseline test condition; light blue, median at 1,945.0), and the positive control (green, median at 18,931.5). This suggests that the GTS construct, in the absence of the engineered tRNAs for implementing sense-to-sense codon reassignment has a background level of fluorescence, likely owing to the presence of the sfGFP construct. The sub-figure that is second from the left shows the relevant conditions with the addition of the histidine-leucine reassigning tRNAs, and without (dark blue) and with (red) the aTc inducer. From this figure, it appears that the addition of the tRNA plasmid reduces the overall fluorescent intensity of the cell population, shifting the peak lower, with a median intensity of 769.0 – 40% lower than in the condition without the engineered tRNAs. The addition of both synthetic tRNAs and the inducer, however, causes an overall increase in fluorescent intensity of the population. The distribution of the population shifts higher, albeit only at a point only slightly greater than that of the baseline D7 population, but the median level of fluorescent intensity is over three times higher (328%, median at 2,482.5) than that of the uninduced condition. The increase in fluorescence suggests that the histidine-leucine tRNAs were able to be transcribed at a sufficient level to correctly reassign the $\phi c31$.H53L construct, which in turn was able to irreversibly invert the Genetic Toggle Switch to allow expression of the proxy sfGFP reporter. Whilst on the whole, the population of the reassignment test condition (D7) did not appear to have significantly greater fluorescent intensity over the

baseline condition (D6), it was clear that the addition of the co-transformed tRNA plasmid and the inducer did make a tangible difference that could be considered confirmation that codon reassignment – albeit likely just leucine-to-histidine reassignment – is possible and quantifiable with the OCI-GTS system developed. As such, it could be a fesaible platform with which to develop and characterise other sense-to-sense codon reassignment rule sets within an otherwise naturally-coded organism.

4.3 Conclusion to GTS

The two aims of this chapter was to develop a system that would allow the transient expression of the toxic anticodon-modified tRNAs within a naturally-coded organism that could output a permanent non-toxic response, and to find an alternative OCI recoded reporter that would be able to demonstrate the reverse histidine-to-leucine reassignment. Of these, I was able to create a process – GTS system – with a library of alternative recoded ΦC31 recombinase reporters that could be adapted for use in characterising a range of sense-to-sense codon reassignments. With ΦC31, I demonstrated again the leucine-to-histidine reassignment, and whilst a range of leucine residues were identified that could be recoded for use in implementing the reverse reassignment, I was not able to create a viable, recoded recombinase for this characterisation in the timeframe of the project.

In terms of the GTS system, there were two switch circuits that were designed and which showed potential for establishing a downstream naturally-coded proxy reporter of reassignment: the Inversion and the Excision circuit. Two variants of the Inversion circuit were assessed, one that used Kan^R as the secondary reporter, and the other that used sfGFP. Whilst Kan^R was initially selected as a complement of sorts to the OCI CAT reporter (and also as it potential to be another OCI recoded reporter too with its two adjacent residues L131_H132), the operon itself proved too leaky to be a feasibly proxy reporter. In hindsight, it may be better in general not to use antibiotic resistance markers as primary recoded or as secondary proxy reporters altogether. The reason being that (even if construction of the recoded

reporter proved facile and stringent) it would be difficult to implement a change in selection media during transient expression of the anticodon-modified tRNAs. Instead, a fluorescent or chromogenic protein (such as sfGFP, *mCherry*, tsPurple per the subsequent GTS designs) that is not dependent on a specific substrate would make it easier to detect when reassignment has occured.

Whilst the final GTS system followed the route of the Inversion-sfGFP design, it is likely that the Excision-tsPurple would have also been a feasible method for demonstrating reassignment. The primary limitation in its current form (other than issues with plasmid construction) is again in the leaky basal expression. That said, as it was, the Inversion-sfGFP also turned out to be less stringent than previously observed. Regardless, had there been more time, it is likely that the various approaches applied to increase its stringency would have been able to achieve a level that had a sufficient signal-to-noise ratio. Likewise, if a quick, visual output was not directly required, carrying out a colony PCR on the culture after the reassignment had been induced would likely have served as a suitable strategy for detecting reassignment; especially as preliminary tests of this approach was able to demonstrate a clear output of excision activity.

Finally, in terms of using flow cytometry for characterising the proxy sfGFP signal for reassignment: this appeared to be a relatively quick and quantitative method for measuring the efficiency of the full OCI-GTS system. Whilst the time constraints of the project meant that the approach was not optimised further, nor was I able to test other ΦC31 recoded constructs. Nevertheless, the preliminary results showed promise that the Inversion-sfGFP circuit was a viable approach for characterising sense-to-sense codon reassignment not only single/minimal recoding variants but also the many multi-site recoded recombinases created. It is with hope that the key elements for characterising these histidine-leucine recoded reporters are already established, and would require only a final few co-transformations and flow cytometry analyses to demonstrate definitively that the complete histidine-leucine reassignment can be attained.

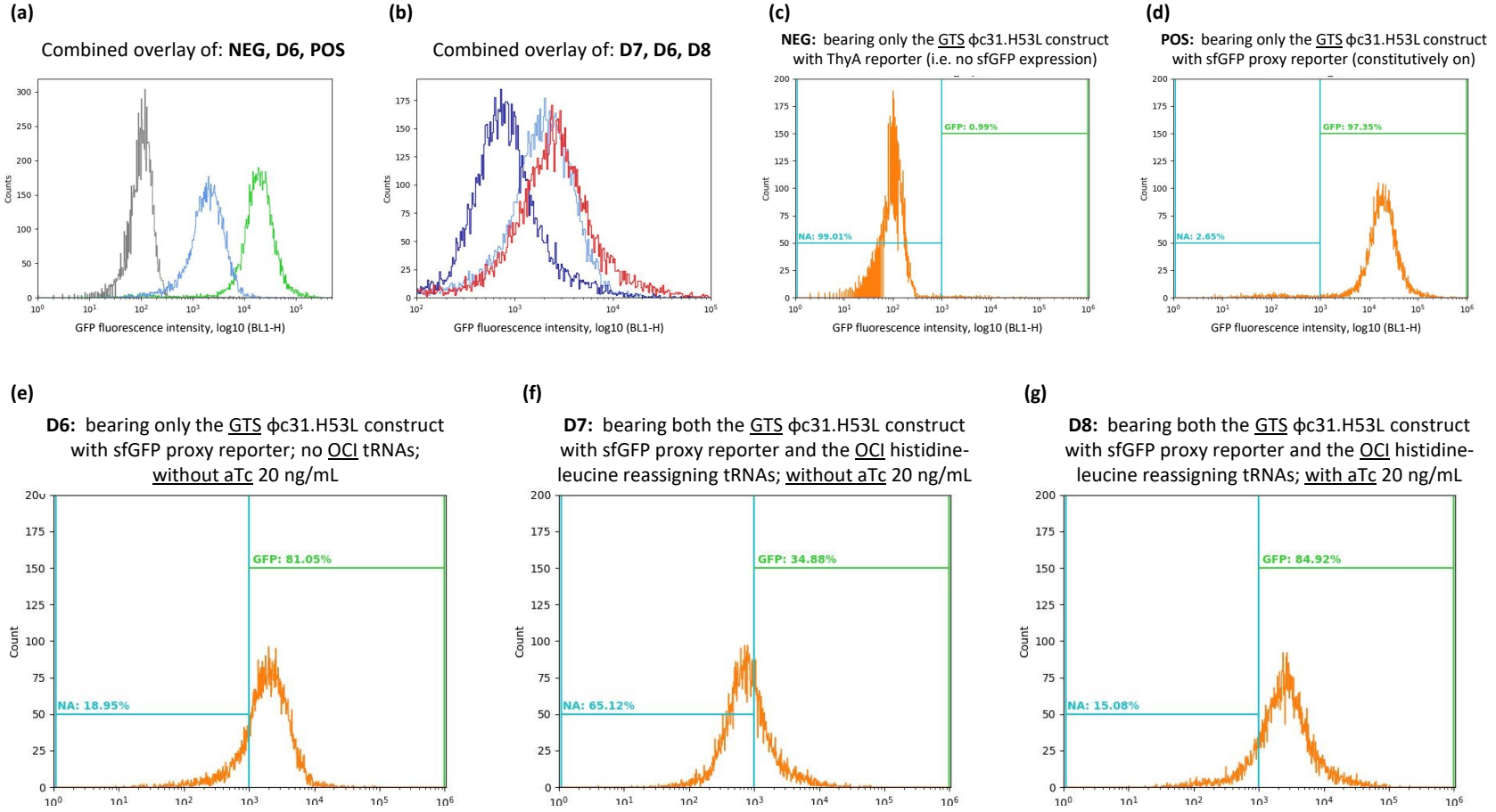


Figure 4.36: Flow cytometry of the combined OCI-GTS system, characterised using the Inversion-sfGFP circuit with recoded recombinase ΦC31.H53L

Top row, left to right: (a) histogram overlay of the control conditions, with grey showing the negative control (no fluorescence), red showing the positive control (constitutively expressing sfGFP), and light blue showing the baseline test condition (D6) with *E. coli* DH10β strain bearing only the GTS $\phi c31$.H53L construct with the sfGFP proxy reporter; (b) the same D6 sample condition is shown again in light blue, whilst the D7 sample condition shows the DH10β strain bearing both the GTS $\phi c31$.H53L construct with the sfGFP proxy reporter, and the histidine-leucine reassigning tRNAs plasmid, but in the absence of any inducers, and the D8 sample is as D7 but with the addition of 20 ng/mL aTc; (c) the negative control; (d) the positive control. Bottom row, left to right shows the test conditions: (e) D6, (f) D7, (g) D8, and how the population falls across the non-fluorescence/fluorescence threshold.

Chapter 5

Final Discussions and Conclusions

5.1 General Discussions on the Research Outputs

The aim of this thesis was to investigate genetic code plasticity – specifically the feasibility of implementing sense-to-sense codon reassignment *in vivo* – through the development of genetic circuits that can function under both synthetic and natural coding systems. Demonstrating that it is possible to change these fundamental rules of the quasi-universal genetic code provides unequivocal evidence that the basic molecular biology translational machinery is remarkably mutable, and yet remains a field of research that is under-explored. Regardless, the implications of this research has potential for use in designing alternative coding systems, with direct benefits as an additional strategy for semantic biocontainment when engineering biology, such as in the development of orthogonal genetically modified or recoded organisms and production of any downstream synthetic protein products (Torres et al., 2016).

In terms of the achieved outcomes of this project, I was able to create a functional platform for validating sense-to-sense codon reassignment in the *E. coli* system and establish a set of heterologously-expressed recoded reporters to verify that the uni-directional leucine-to-histidine reassignment could be realised without severe adverse effects on an otherwise naturally-coded host organism. The establishment of the reassignment system was a two-part process: one, in creating engineered anticodon-modified tRNAs ($tRNA^{His}_{\underline{A}AG}$ and $tRNA^{Leu}_{\underline{A}UG}$) that could decode orphan sense codons ($Leu_{(CU\underline{U})}$ and $His_{(CA\underline{U})}$ respectively) yet still be aminoacylated

by native aminoacyl-tRNA synthetases within *E. coli*; two, in identifying visual or measurable reporter constructs that could be recoded (synonymously and non-synonymously) to be non-functional under the natural code, but could be rescued upon application of the synthetic, codon-reassigning tRNAs. This process was detailed in Chapter 3 when establishing what is known as the Orphan Codon Invasion (OCI) system.

Whilst a number of recoded reporters were created, successful codon reassignment was detected in only the CAT antibiotic resistance marker and ΦC31 recombinase library variants (as shown in Chapter 3 and Chapter 4 respectively). Notably, only leucine-to-histidine reassignment could be determined with these constructs: i.e. rescuing heterologously-expressed reporter sequences that were rendered non-functional by having key histidine codon sites initially recoded to the orphan leucine codon, $\text{Leu}_{(\text{CU}\underline{\textbf{U}})}$. Moreover, I was able to demonstrate that the engineered $\text{tRNA}^{\text{His}}_{\underline{\textbf{A}}\text{AG}}$ was both necessary and sufficient for reassigning (leucine-to-histidine) and recovering the activity of the recoded *cat*.H193L reporter.

By contrast, I was not able to show the reverse histidine-to-leucine reassignment. In spite of having identified various candidate leucine residues that were potentially critical to activity in the recombinase reporter (and to the signal sequences of two periplasmic systems that were selected for recoding but proved difficult to construct), and which when recoded to histidine orphan $\text{His}_{(\text{CA}\underline{\textbf{U}})}$ should have given the necessary reporters for assessing the reverse reassignment. I hypothesised that the reason this sense-to-sense codon reassignment could not be achieved (at least not in the timeframe of the research) was due to issues with my synthetic tRNA designs. This was perhaps in part due to an incomplete understanding of both the tRNA sequence-structure of the native *Eco*$\text{tRNA}^{\text{Leu}}_{\underline{\textbf{G}}\text{AG}}$ (*leuU*) and *Eco*$\text{tRNA}^{\text{His}}_{\underline{\textbf{G}}\text{UG}}$ (*hisR*) from which the anticodon-modified tRNA was based upon, and on the unknown interactions between the engineered $\text{tRNA}^{\text{Leu}}_{\underline{\textbf{A}}\text{UG}}$ isoacceptor with the native *Eco*LeuRS, rather than not finding the crucial residues for inactivating the recoded reporters prior to reassignment.

However, it was only through a deeper review of the literature on tRNA biol-

ogy, when troubleshooting the histidine-to-leucine reassignment issues, that it became apparent the problem likely resided in codon-anticodon interactions. More specifically, the absence of leucine incorporation into orphan codon $His_{(CA\underline{U})}$ could actually be due to the effects – or lack thereof – of tRNA post-translational modifications on the engineered $tRNA^{Leu}_{\underline{A}UG}$. The initial hypothesis was that modifying the tRNA anticodons to have full Watson-Crick complementarity to the orphan codons would be sufficient to compete with the native tRNAs for codon invasion, particularly as the key determinants for the selected tRNAs did not rely (heavily) on the anticodon: for *Eco*$tRNA^{His}_{\underline{G}UG}$, for example, the major recognition element was at the opposite end of the tRNA structure, at position G-1 (Arnez et al., 1995). As such, I overlooked the potential effects of post-translational modifications (see Figure 3.7). In fact, it appears that queuosine (Q), a hyper-modified guanosine nucleoside – found specifically at the anticodon wobble (position 34) – in tRNAs with GUN anticodons (i.e. Tyr, His, Asn, Asp; Harada and Nishimura, 1972) is conserved in near uniformity across all species (with the exceptions of yeast, *T. thermophilus*, and the archaeal bacteria). As such, the introduction of an adenosine at the wobble base in the engineered $tRNA^{Leu}_{\underline{A}UG}$, to try make it complementary to the orphan codon $His_{(CA\underline{U})}$, would never have been converted to the near-universally conserved queuosine. Whilst it is not clear the role of queuosine on tRNA interactions, it has been associated with stabilisation of the anticodon loop structure. More importantly, it appears that when the wobble position is unmodified, the tRNAs have a stronger preference for the Watson-Crick NA<u>C</u> codon, whereas a queuosine-modified tRNA may bias more towards the wobble NA<u>U</u> codon (Morris and Elliott, 2001; Meier et al., 1985). Given the lack of leucine incorporation at the orphan histidine codons designed into the recoded reporters, it would imply that the absence of the queuosine modification negated the potential benefit of the anticodon-modified tRNAs having full Watson-Crick anticodon complementarity to the target codons.

In eubacteria, like *E. coli*, the queuosine modification for *Eco*$tRNA^{His}_{\underline{G}UG}$ is produced *de novo* via a multi-step enzymatic synthesis pathway. Starting with GTP, the queuosine precursor modification is produced, before being transferred to the

wobble base of the tRNA by tRNA-guanine transglycosylase (TGT), where it is converted to queuosine-tRNA by QueA and QueG (Winther et al., 2021; Morris and Elliott, 2001; Iwata-Reuyl, 2003). As post-translational modifications can be produced *de novo* in *E. coli*, a revised hypothesis for future work in implementing the histidine-to-leucine reassignment could be that the anticodon-modified tRNA should have Watson-Crick complementarity at positions 35 and 36 of the anticodon, but the base at position 34 should remain as guanine such that it has the potential to be decorated with a queuosine modification. Whilst leaving the wobble position as guanosine provides the opportunity for it to be modified to queuosine, it also leaves the anticodon as a wobble base-pair. In fact, doing so means that the anticodon of the revised engineered tRNA (i.e. of $\text{tRNA}^{\text{Leu}}_{\underline{\textbf{G}}\text{UG}}$ in this case, rather than the original $\text{tRNA}^{\text{Leu}}_{\underline{\textbf{A}}\text{UG}}$ synthetic design) is exactly the same as the native *Eco*$\text{tRNA}^{\text{His}}_{\underline{\textbf{G}}\text{UG}}$ isoacceptor. If all else being equal, and we account for only codon-anticodon sequence complementarity, the revised anticodon-modified tRNA should theoretically be no better or worse than the native tRNA^{His} for invading the orphan codon target.

Assuming the revised tRNA design for histidine-to-leucine reassignment can be established, the presumably equal level of native and synthetic tRNA on reassigning the heterologously-expressed recoded reporters may not be at a sufficient yield as to be easily detected, especially not in that of a naturally-coded host organism. For essential proteins, like ThyA (as per Döring and Marlière, 1998) or CAT expression in the presence of chloramphenicol (this work, Chapter 3), the selective environment could necessitate the retention of both the indispensable exogenously-expressed genes-phenotypes and the alternatively-coding tRNAs. There may even be cause to overexpress the anticodon-modified tRNAs to match the required need, despite the penalties it may incur to the host, e.g. cellular toxicity from excessive mis-translation of genomic/endogenous genes. However, the signal of reassignment may not be suitably sufficient. The overexpression of both synthetic tRNAs and recoded reporters could reduce host viability to a level where reassignment of the reporters is indistinguishable from a low yield expression and poorly-growing cell. Spontaneous mutations in the host may revert or lose the alternatively-coded trans-

lation components. There may be need to use a reporter protein that does not, or can not, demand an excess of the host's resources to adequately display reassignment. The resulting Genetic Toggle Switch (GTS) system (Chapter 4) was created to address these issues.

The GTS circuits permitted a method with which to transmit a transient signal of the alternative genetic code (as set up by the anticodon-modified tRNAs and the recoded reporters), whilst permitting expression and amplification of a permanent non-toxic response (via a secondary, naturally-coded proxy reporter). Using the serine recombinase ΦC31 as the switch to go from the dual synthetic and natural genetic codes, to just continuing with a proxy naturally-coded reporter, proved an effective system to demonstrate reassignment in a sustainable manner. After an extended period of troubleshooting, two visual and viable genetic switch systems were developed: an Inversion-sfGFP circuit and an Excision-tsPurple construct, with ΦC31 acting as the irreversible OCI recoded reporter upon detection of sense-to-sense codon reassignment.

While a number of key leucine and histidine residues were identified in the N-terminal catalytic domain of ΦC31 as potentially inactivating sites of interest for testing histidine-leucine reassignment, a significant amount of time was given over to creating and verifying the creation of this library of histidine-leucine recoded variants. As a result of this manually-intensive process, only three of the Inversion-sfGFP recombinase switch constructs were ready for testing with the full OCI-GTS system (i.e. with the anticodon-modified tRNAs): (i) *ϕc31*.H61L, (ii) *ϕc31*.H142L_L143H, and (iii) *ϕc31*.L468H. However, from literature reviews and earlier characterisation assays, it was known that *ϕc31*.L468H on its own was not sufficient to inactivate the recombinase, thus would not be a valid test for histidine-to-leucine.

Similarly, whilst it was the promising variant to demonstrate the bi-directional histidine-leucine reassignment, *ϕc31*.H142L_L143H was also flawed: like *ϕc31*.L468H, the H142L mutation was reported to only have a weakly inactivating effect on recombinase activity, which meant that this variant would effectively be

demonstrating histidine-to-leucine reassignment. This, in and of itself, was still of some value as there had not yet been a recoded reporter that had demonstrated this reverse reassignment. Although a couple attempts were made in expressing this construct alongside the anticodon-modified tRNAs, each effort failed to yield positive results. Given the earlier discussions on the likely need for the queuosine modification in the native *Eco*tRNA *hisR* with native anticodon 5'-GUG-3' ($tRNA^{His}_{\underline{G}UG}$), it is possible that this was also the crux of the problem with not being able to show the histidine-to-leucine reassignment. In a final attempt to show multiple directional assignments for sense-to-sense codon reassignment, I also endeavoured to replicate the isoleucine-to-cysteine ThyA experiments by Döring and Marlière (1998), but was unable to complete this within the timeframe of the thesis.

In the end, $\phi c31$.H61L was the only GTS construct that could be paired with the OCI anticodon-modified tRNAs in the full OCI-GTS system for characterising sense-to-sense codon reassignment and its effect on host viability. The benefits of being able to test this recombinase reporter in the combined system was that it was possible to use flow cytometry to replicate and quantify the number of cells that had: (i) only the $\phi c31$.H61L construct (with none of the alternatively-coding tRNAs); (ii) $\phi c31$.H61L and the engineered tRNAs (although strictly-speaking only the $tRNA^{His}_{\underline{A}AG}$ variant was required), and with no aTc induction; (iii) as per the second condition but with the addition of aTc with which to transcribe the tRNAs and allow expression of the recombinase. A limitation, as noted, was that this experiment required only the leucine-to-histidine reassigning tRNA isoaccepter, which meant that it was again only showing a uni-directional mode of sense-to-sense codon reassignment. It was, however, demonstrating the robustness of the leucine-to-histidine reassigment strategy as it was using the same engineered reassignment tRNAs but with a different recoded reporter (CAT antibiotic resistance vs ΦC31 recombinase), and a different analysis method (plate-based assay vs flow cytometry).

Recall that in the *cat*.H193L reassignment assay (Figure 3.23), the presence of the leucine-to-histidine $tRNA^{His}_{\underline{A}AG}$ isoacceptor alone was both necessary and sufficient to rescue the *E. coli* bearing both the recoded reporter and the engineered

tRNA. Note that the growth of this strain on the agar plate containing chloramphenicol was much reduced compared to the equivalent strains containing a fully-active *cat*.H193H variant, and that the negative control *cat*.H193L in the presence of the histidine-to-leucine isoacceptors ($tRNA^{Leu}_{\underline{A}UG}$) correctly failed to grow at all. This assay was very much a qualitative characterisation, and relied on an extended period of time (i.e. days) to confirm the lack of growth in the negative control. In the meantime, each condition was subject to the constant selective pressure to retain a functional copy of both the *cat* gene and the engineered tRNAs.

By comparison, the *ϕc31*.H61L flow cytometry assays were expressed over an afternoon and quantified within an hour. Given the design of the OCI-GTS system, there was minimal selective pressure to continue transcribing the synthetic tRNAs once the recombinase switch had been triggered. A limitation, however, was that the proxy sfGFP signal for reassignment may not have been given sufficient time to mature prior to analysis on the flow cytometer, and as such may not be as comparable to assess the effect on cell viability over a longer time period. This interpretation could arguably be observed in the flow cytometry results (n = 2 technical replicates; Figure 4.36): with an average of 84.92% of the test condition (*ϕc31*.H61L, with $tRNA^{His}_{\underline{A}AG}$, and 20 ng/mL aTc) displaying the proxy sfGFP expression, compared to 81.05% of the cells showing background sfGFP fluorescence in the condition that contained only the Inversion-sfGFP *ϕc31*.H61L GTS circuit. It was noticeable, however, that the test condition displayed significantly more sfGFP expression over the condition that had both parts of the OCI-GTS system but lacked the inducer (only 34.88% sfGFP fluorescence detected). Note also that the negative control displayed 0.99% fluorescence, whilst the positive control was measured at 97.35% sfGFP expression. Had there been more time to repeat this experiment, I would have induced expression of the circuits for a range of time periods to see whether is an an optimal condition to detect a greater difference in the proxy fluorescence signal.

Nevertheless, even though it was not possible to characterise histidine-to-leucine reassignment during the course of this research, and the assays for testing

leucine-to-histidine reassignment whilst positive were only preliminary results, the output of having created the OCI-GTS system means that there is a functional workflow and platform with which to create and characterise transient sense-to-sense codon reassignment experiments.

5.2 Limitations of the Research

As with every project, there are bounds as to what is within scope and what is feasible in the timeframe and resources available at the time. In terms of the GTS half of my research, one of the initial limitations was in the lack of available data on the full ΦC31 crystal structure. This meant that the search space for identifying critical histidine and leucine residues was restricted to the N-terminal catalytic domain of the recombinase. As this region is a key component for protein activity, the lack of data for the rest of the structure was not a severe limitation, but it would have been useful to be able to explore all histidine and leucine residues across the entire protein, particularly as there were less histidine residues available for modelling site-directed mutagenesis compared to leucines. Recent advancements with AlphaFold could have made it easier to predict the structure of from the sequence of the missing domains. Moreover, with regards to using PyMOL to estimate intermolecular clashes: whilst the mutagenesis modelling gave a general overview of the potential impact of a given mutation, the model is only able to show a limited view of the "true" impact of the mutations on the overall stability of the protein. This meant that it would have still been necessary to experimentally validate the mutations, which could easily have been an entire project in itself. That said, had it been possible to cheaply and rapidly automate the synthesis of the different mutagenesis combinations, a larger combination of recoded histidines and leucines could have been generated for testing the extent of codon reassignment and re-gain-of-function of the recombinase. Had I been able to generate a larger library of multi-site recoded reporters, it may have been possible to find a variant that could be used to demonstrate the full bi-directional histidine-leucine reassignment.

With hindsight, narrowing down the possible amino acids and their associated

tRNAs to the ones that did not use the anticodon as a major identity element was beneficial in reducing the search space of which codons to reassign, however, given that Döring and Marlière (1998) was able to demonstrate isoleucine-to-cysteine and methionine-to-cysteine (amino acids that have translational machinery that are more particular about anticodon recognition), there may have been scope to focus on other tRNA isoacceptors with more obscure orphan codons. Whether doing so would have required using more essential genes for the recoded reporters is less clear. Regardless, one clear lesson learned from my research is that more care must be taken to understand the tRNAs that are used to model the anticodon modifications, as it is not as simple as just changing the anticodons. Hence, more time should be spent confirming that synthetic tRNAs can be aminoacylated (be it with a native synthetase, or co-evolved with an engineered variant), can sufficiently bind with (orphan) codons, and is compatible in general with all other translation machinery.

5.3 Overall Impact and Potential for Future Work

Ultimately, the GTS and OCI-GTS system was in and of itself a workaround as whole genome recoding to remove the impact of cross-talk between natural and synthetic genetic codes was not feasible at the time this research was carried out. As of yet, large-scale genome recoding is still limited to select research groups (such as the Chin and Church groups). Until such a time as to this technology becoming more readily accessible, feasible, and replicable, the OCI-GTS platform developed in this research should be a sufficient tool for researchers who are interested in exploring possible other sense-to-sense codon reassignment combinations. With a bit more refinement of the optimal transcription times of the anticodon-modified tRNAs and the window of expression time needed to see a significant shift in the proxy Inversion-sfGFP reporter signal from the GTS recombinase construct, the system could conceivably be used to test any number of amino acid reassignment pairs.

Beyond looking at just tRNA pairings, the GTS developed could also be used to better understand basic tRNA/aaRS (tRNA/aaRS) biology – much in the same

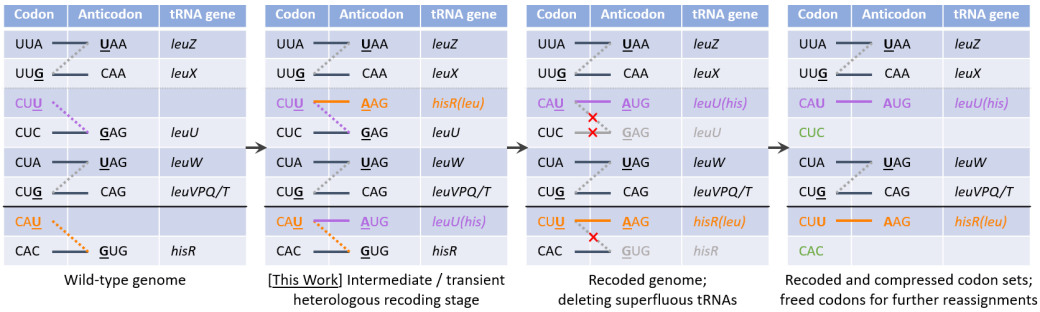


Figure 5.1: Proposed wholesale recoding of the *E. coli* to accept an alternative histidine-leucine genetic code

From the natural genetic code (far left box), to the alternative genetic code with the proposed histidine-leucine admix natural-synthetic code (second from left box; this work), to the proposed steps in the future for a for fully histidine-leucine synthetic genetic code (right two boxes). Reproduction of Figure 4.1.

way as my original knock-out/rescue strategy. In a similar vein, it could be possible to use the switch to test orthogonal tRNA/aaRS pairings also, be it with translation components within a species or between different species. The natural extension of gaining a better understanding of how different tRNAs and synthetases interact between species could open up further opportunities to develop novel semantic biocontainment systems that are specific to a given species. At this point in time, biocontainment strategies are still in relatively early stages of development, and as such, the safest way to limit genetic escape is to use a number of overlapping strategies (genetic, semantic, auxotrophic, kill switches, etc.). However, there is hope that the research presented in this thesis will be of value as advances are made in the development of novel and orthogonal genetically recoded organisms of the future.

Appendix A

Oligonucleotides, Sequences, and Plasmid Maps

A.1 Key genetic constructs

Background information, rationale for the methods, and description of specific key genetic constructs created and used within this body of work is as follows. A summary of the key constructs have previously been listed in Section 2.1.3.

A.1.1 *E. coli* and *T. thermophilus* leucyl-tRNA synthetase

DNA and amino acid sequences for *Eco*LeuRS (*leuS*; 2,580 bp, 860 AA, 97 kDa) were obtained from NCBI (*E. coli* strain K-12 sub-strain MG1655 NC_000913.3, locus tag b0642 [FASTA]). Genomic DNA extraction was performed on an in-house *E. coli* K-12 strain using the Monarch Genomic DNA Purification Kit (NEB #T3010), and primers YKH007/YKH008 were used to PCR amplify the *leuS* sequence. These primers also added an `NdeI` and `HindIII` respectively to each end of the sequence. The amplified product was purified and subcloned into [P003] `pET29a(+)` (4,852 bp) via `NdeI`/`HindIII` restriction cloning into the MCS, whilst simultaneously removing the upstream thrombin and S-tag sequences.

The gene sequence for *Tth*LeuRS (2,634 bp, 878 AA, 101 kDa) was also obtained from NCBI (*T. thermophilus (Tth)* NC_006461.1, locus tag TTH_RS00845 [FASTA]) was similarly obtained by post-doctoral researcher Dr. Antje Krüger through genomic DNA extraction and subcloning into a modified `pET29a(+)` vec-

tor.

Both *Eco*LeuRS and *Tth*LeuRS were used in the earlier strategy to implement a genomic knock-out and subsequent rescue with heterologous expression of the synthetases and tRNAs from plasmids.

A.1.2 *E. coli* histidyl-tRNA synthetase

DNA and amino acid sequences for *Eco*HisRS (*hisS*) were obtained from NCBI (*E. coli* strain K-12 sub-strain MG1655 NC_000913.3, locus tag b2514 [FASTA]). Designs were made using this sequence in the earlier knock-out/rescue strategy for sense-to-sense codon reassignment but this approach was subsequently discontinued in favour of the OCI-GTS system.

A.1.3 *E. coli* leucyl-tRNA and histidyl-tRNA

The wild-type/native *Eco*tRNA$^{\mathrm{Leu}}$ has five isoacceptors, encoded by eight gene sequences, to base-pair with six codons. *Eco*tRNA$^{\mathrm{His}}$ has a single isoacceptor, encoded by one gene sequence, to base-pair with two codons. The engineered tRNA$^{\mathrm{Leu}}$ and tRNA$^{\mathrm{His}}$ for constructing the anticodon-modified tRNAs of the Orphan Codon Invasion (OCI) system were based on the gene sequences *leuU* (tRNA$^{\mathrm{Leu}}_{\underline{\mathbf{G}}\mathrm{AG}}$) and *hisR* (tRNA$^{\mathrm{His}}_{\underline{\mathbf{G}}\mathrm{UG}}$) respectively. In order to achieve histidine-leucine reassignment, the approach taken was to alter the anticodon of these tRNAs such that the former becomes tRNA$^{\mathrm{Leu}}_{\underline{\mathbf{A}}\mathrm{UG}}$ (yet recognising histidine codon His$_{(\mathrm{CA}\underline{\mathbf{T}})}$) and the latter as tRNA$^{\mathrm{His}}_{\underline{\mathbf{A}}\mathrm{AG}}$ (yet recognising leucine codon Leu$_{(\mathrm{CT}\underline{\mathbf{T}})}$). In this manner, it should be possible to produce histidine-to-leucine and leucine-to-histidine reassignments respectively.

As this forms the fundamental premise of my research, the design (Section 3.1.3), creation (Section 3.2.1), and implementation (Section 3.2.5) of these engineered tRNAs are described in more depth in the Chapter 3 sections indicated.

A.1.4 Cre and lambda Red Recombinase Constructs

DNA and amino acid sequences for the Cre recombinase (X03453.1 [EMBL] [FASTA]; P06956) were obtained from the European Nucleotide Archive and UniProt respectively. Of the 1,553 bp sequence (which included also a 34 bp *loxP*

site), only the 1,032 bp *cre* gene was taken forward for design modifications prior to its initial incorporation into the [P007] pBAD30 (4,919 bp) vector. In order to do so, a 42 bp sequence homologous to the 3'-end of the araBAD promoter (including an NheI restriction site) in vector pBAD30 and an EcoRI site was added the 5'-end of the gene, whilst a HindIII site and another 42 bp sequence matching part of the vector's MCS was placed downstream of the CDS. The final design of the 1,128 bp *cre* construct (YKH036) was commercially synthesised (IDT gBlocks®, Belgium). Vector pBAD30 was cut with restriction enzymes EcoRI/HindII, and the dsDNA gBlocks® fragment was directly incorporated via Gibson Assembly (downstream of the araBAD promoter), creating a final plasmid of 5,906 bp. Had it been necessary, primers YKH026/YKH027 could have been used to amplify the YKH036 *cre* construct, prior to subcloning into the pBAD30 vector via EcoRI/HindII restriction cloning.

Attempts to create an equivalent *cre* construct in vector [P008] pD881 (2,289 bp) proved unsuccessful. As such, primers YKH059/YKH060 were used to amplify the rhaB promoter (whilst adding also flanking SapI restriction sites; totalling 152 bp) from the pD881 vector for use with the *cre* construct. Similarly, iPCR using primers YKH057/YKH058 was performed on the *cre*_pBAD30 plasmid to remove both the *araC* regulatory gene and the araBAD promoter, producing a linear 4,536 bp construct of the *cre* recombinase and the rest of pBAD30 flanked on either side by SapI restriction sites. Following Golden Gate Assembly, the two fragments were reconstituted together (with the rhaB regulatory system replacing araBAD) to form a 4,650 bp plasmid where *cre* was under control of the rhaB promoter. A subsequent iPCR amplification with YKH076/YKH077 on this plasmid was used to insert a 9 bp (5'-TGAAATTCT-3') sequence upstream of the RBS as a means to control leaky expression of the Cre recombinase. This was followed by SapI restriction and self/re-ligation to create a 4,657 bp plasmid that had a stronger, more regulated control of the recombinase.

Finally, as the intention was to use the Cre (a tyrosine recombinase)[1] in

[1] Alternative recombinases that were considered for use in this system included Bxb1 and the TP901 integrase. These were kindly provided by Tom Folliard as part of a 6,613 bp "Dual controller

conjunction with the lambda Red homologous recombination system, primers YKH093/YKH094 were used to amplify the relevant segment – containing the araC regulator, the araBAD promoter, and its three downstream lambda Red components (Gam, Beta, Exo) – from the [P010] `pKD46` plasmid, and combine this 3,420 bp construct via Gibson Assembly into an `NheI`-linearised plasmid containing the *cre* recombinase that was under control of the rhaB promoter. The sizeable (8,023 bp) resulting Cre/lambda Red plasmid was taken forth as the means to create the initial genomic integration/knock-out genetic switch for codon reassignment – the precursor of the final OCI-GTS system.

A.1.5 ΦC31 Recombinase Construct Variants

The DNA sequence for the *ϕc31* serine recombinase (1,866 bp, 621 AA, 69 kDa) was kindly provided in a [P003] `pET29a(+)` vector (totalling 7,080 bp) by then-PhD candidate, Hugo Villanueva[2]. Whilst the physical construct provided was not eventually used, its gene sequence was taken as the base design for the *ϕc31* recombinase that was subsequently used in the prototypical Genetic Toggle Switch (GTS) circuit in Chapter 4. The reason being that the recombinase needed to be codon-optimised for use as a potential sense-to-sense codon reassignment reporter, which required synonymous recoding of all histidine and leucine residues that were not to be reassigned, identifying key histidines that need to be recoded to the reassignable leucine codon (and vice versa), as well as introducing two Amber stop codons ($*\text{Am}_{(\text{TAG})}$) to render the protein non-functional in its initial state in order to create the basic GTS circuit. The final 3,269 bp [002] design construct (Figure A.1) was commercially synthesised (Twist Biosciences, USA).

Like the engineered tRNAs are to the OCI system, the ΦC31 recombinase is the central element to the GTS system, thus Chapter 4 is dedicated to detailing its design through to implementation. In brief, Section 4.2.1 begins with a description

plasmid", where Bxb1 was under control of a pTetO promoter, TP901 was regulated by the araBAD promoter, and the plasmid had a ColE1 ori and Cam resistance marker.

[2]Hugo also provided Bxb1 (1,566 bp, 521 AA, 58.8 kDa) in `pET29a(+)` (totalling 6,780 bp). Whilst he was able to purify Bxb1, he did not have success in using it *in vitro*; *in vivo* expression was not tested. A collaborator, Dr. Andy Osborne, who was provided with the same constructs, was able to express and use the ΦC31 construct, hence, the GTS system was based on this recombinase.

of its design within the prototypical GTS circuit (Figure A.1), alongside a preliminary *in silico* modelling of the recombinase's active and likely inactive states if various key histidines and leucines were recoded/substituted with the other. This is followed by the construction of a stop-inactivated *ϕc31* (*ϕc***) into the basic GTS circuit in Section 4.2.2. After various iterations to develop a viable secondary reporter that works in conjunction with the recombinase, Section 4.2.6 concludes the chapter by reporting on the process for making the recombinase into a potentially reassign-able reporter by recoding multiple key histidine-leucine residues within *ϕc31* (as identified earlier through the mutagenesis models). These inactive recoded recombinases, therefore, act as another OCI reporter system through which the engineered anticodon-modified tRNAs can be used to characterise the alternative histidine-leucine genetic code proposed for sense-to-sense codon reassignment. It should be noted that there were two approaches for recoding ΦC31 into an inactive variant. For making multi-site histidine-leucine recombinase recodings within the Inversion-sfGFP switch, the new Darwin Assembly method (Section 2.2.18) was deployed. As such, Section 4.2.6.1 was dedicated to describing this approach, including verification and characterisation of the library of created constructs. The second approach uses a more typical SDM approach with iPCR amplification to create either one or two point mutations at single sites within the recombinase in the Excision-tsPurple switch. The design rationale and construction of these ΦC31 recodings were detailed in Section 4.2.6.2.

A.1.6 Fluorescent and Chromoprotein Reporter Constructs

Fluorescent Proteins – the sfGFP and EBFP variants:

Fluorescent proteins, particularly superfolder green fluorescent protein (sfGFP), were used in the development of both OCI recoded reporters for demonstrating sense-to-sense codon reassignments (Chapter 3), as well as in the creation of a GTS circuit with a naturally-coded secondary/proxy reporter for amplifying the signal of histidine-leucine reassignment without adversely affecting the viability of the host organism (Chapter 4). The gene construct was created and kindly provided as plasmid [203] (within a modified [P003] `pET29a(+)` vector) by post-doctoral

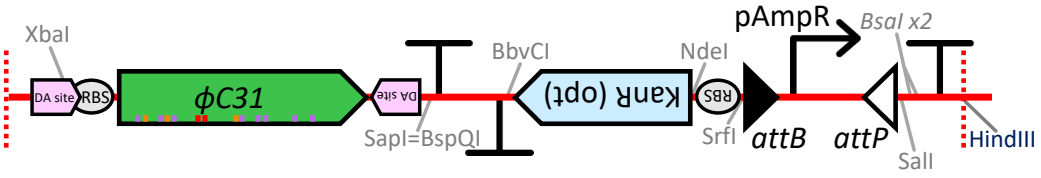


Figure A.1: The design of the prototypical GTS ΦC31 recombinase circuit [002]

Chapter 4 discusses the role of the ΦC31 recombinase (in green) within the wider GTS circuit [002]. ΦC31 forms one half of this 3,269 bp GTS Inversion-Kan^R circuit that was commercially synthesised (Twist Biosciences, USA). The purple notches within the green *ϕc31* represents leucine codon sites that could be recoded, whilst orange notches show histidine sites. The two red notches identify two asparagines (at positions 200 and 204) that were recoded to $*Am_{(TAG)}$ stop codons. Notches are not positioned to scale. The other half of the circuit comprises of a Kan^R (in blue) operon. The only difference this presents from a typical operon is that the promoter is enclosed within *attB/attP* site-specific recombination sequences and that the promoter in this initial state is positioned facing the opposite direction to the rest of the Kan^R operon. Refer to Figure 4.9 for the full [002] design.

researcher, Dr. Chris Cozens.

In Chapter 3, alternative fluorescent variants of sfGFP – potential OCI codon reassignment reporters – were created through site-directed mutagenesis with the introduction of either a Y66H or Y66L substitution into plasmid [203]. For the two blue fluorescent variants, termed EBFP1.1, a reassign-able $Y66H_{(CAT)}$ variant was made with iPCR using primers YKH195/YKH201, whilst the synonymous but not reassign-able $Y66H_{(CAC)}$ was made with YKH165/YKH199. Similarly, non-fluorescent versions, termed noFP1.1, were produced using the full set of leucine codons, of which the key two were a reassign-able $Y66L_{(CTT)}$ variant (with YKH176/YKH201), and the synonymous but not reassign-able $Y66L_{(CTG)}$ (with YKH190_v2/YKH199). The primers used to make the remaining synonymous leucine codon substitutions at residue 66 are listed in Table 2.3. After amplification, each product was treated with `DpnI/ExoI` (to remove the template plasmid and the ssDNA primers), phosphorylated with T4 DNA PNK (to enable ligation), and self-/re-ligated with T4 DNA ligase, before being transformed into both *E. coli* DH10β cloning and T7 Express *lysY/I^q* expression strains. A more detailed description of the generation of these sfGFP and EBFP1.1 variants, including the characterisation process is in Section 3.2.2.2.

Another set of blue fluorescent protein variants were also created from a known and established EBFP2 as obtained from FPbase. Not only was the sequence design

codon-optimised for use in both *E. coli* and for reassignment, additional flanking ends (including pT7 promoter, LacO operator, and RBS on the 5'-end, and T7 terminator on the 3'-end of the gene sequence) were added to make it suitable for expression and compatible for direct assembly into `pET29a(+)`. This 1,021 bp [025] construct was synthesised commercially (Twist Biosciences, USA). A further blue version, EBFP2.1 was made from EBFP2 by introducing the three mutations F99S, M153T, V163A, to bring it closer to that of sfGFP. Likewise, a non-fluorescent noFP2.1 followed on from EBFP2.1 with a simple H66L substitution.

A full summary of the green, blue, and non-fluorescent proteins that were synthesised and modified for use as OCI reassignment reporters are listed in Table 3.3 and are discussed in more detail in Section 3.2.2.2.

As for the use of sfGFP in the GTS system in Chapter 4: the *sfgfp* gene sequence from plasmid [071] (a replica of [203]) was simply subcloned into the basic Inversion switch circuit, as described in Section 4.2.4.1. Unlike in the OCI system, the purpose of sfGFP in the GTS Inversion-sfGFP circuits was just as a naturally-coded secondary/proxy reporter with which to verify the correct reassignment of the primary ΦC31 recombinase reporter (i.e. the true target of histidine-leucine reassignment using the OCI anticodon-modified tRNAs; Figure A.2).

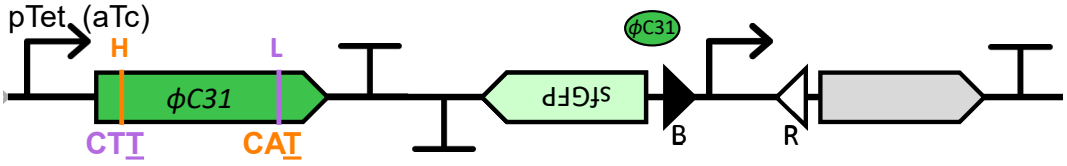


Figure A.2: A schematic of the GTS Inversion-sfGFP circuit design

This design is one of the GTS circuits designed in Chapter 4. The primary reporter is the histidine-leucine recoded ΦC31 recombinase (in dark green), with the naturally-coded sfGFP (in light green) as the secondary/proxy reporter for augmenting the signal of ΦC31. The ΦC31 and sfGFP operons were designed in opposing directions to prevent accidental read-through. Refer to Figure 4.5 for the full implementation of this circuit.

Chromoproteins – TinselPurple:

TinselPurple is a vibrant non-fluorescent chromoprotein (Liljeruhm et al., 2018), sourced from the iGEM Parts Registry (BBa_K1033905, a composite part including the consensus RBS). Like the sfGFP used in the GTS circuit, TinselPurple was another naturally-coded secondary/proxy reporter of the Excision-tsPurple circuit that

is used in Chapter 4 to augment the signal of reassignment in the ΦC31 histidine-leucine recoded OCI reporter. The design of the tsPurple portion of the circuit was broadly that of a typical operon, except that it had a double-terminator sequence enclosed within two site-specific recombination sites (*attB/attP*) that interrupted the operon between the promoter and the RBS-tsPurple sequence (Figure A.3).

Whilst the design of the tsPurple portion of the circuit was relatively straightforward (as detailed in Section 4.2.5), the construction was more of an endeavour. In brief, the double-terminators proved to be a problematic sequence for synthesis, thus the construct needed to be split into two fragments for commercial synthesis: a 415 bp switch fragment (including a 5'-segment of tsPurple) [032] was synthesised as a linear dsDNA gBlock® (IDT, Belgium), whilst the remaining 1,066 bp segment [022] was synthesised by Twist Biosciences (USA) and arrived in a pTwist_Kan_MC vector. The full construction is shown in Figure 4.18, and fully detailed in Section 4.2.5.1.

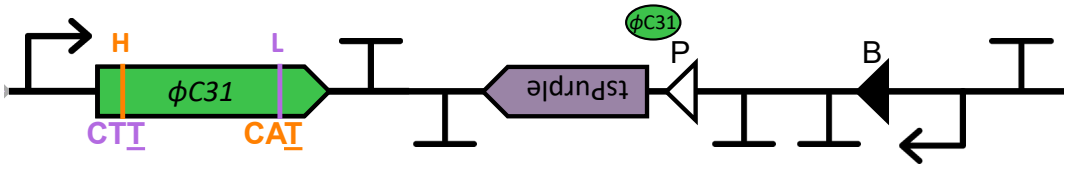


Figure A.3: A schematic of the GTS Excision-tsPurple circuit design

This design is one of the GTS circuits designed in Chapter 4. The primary reporter is the histidine-leucine recoded ΦC31 recombinase (in green), with the naturally-coded TinselPurple (in purple) as the secondary/proxy reporter for augmenting the signal of ΦC31. The ΦC31 and TinselPurple operons were designed in opposing directions to prevent accidental read-through. Refer to Figure 4.5 for the full implementation of this circuit.

A.1.7 Periplasmic Secretor Constructs

Like the EBFP variants, two periplasmic secretor constructs were designed as recoded reporters for use in the OCI system (Chapter 3). This was based on the two major pathways used by bacteria to secrete proteins across the cytoplasmic membrane: the general *Sec*retion route (the Sec-pathway) that transports unstructured proteins (i.e. proteins in their unfolded conformation) across the transmembrane, and the *T*win-*a*rginine *t*ranslocation pathway (the Tat-pathway) that catalyses the translocation of folded proteins. One of the recoded reporters was formed of the

signal sequences from OmpA (the initial 66 bp from NCBI *E. coli* NC_000913.3, locus tag b0957 [FASTA]) of the Sec-pathway, whilst the other was from SufI (the initial 81 bp from NCBI *E. coli* NC_000913.3, locus tag b3017 [FASTA]) of the Tat-pathway; both were coupled to the *lacZα* sequence (225 bp, 75 AA; from 4-228 bp, residues 2-76 AA from NCBI *E. coli* NC_000913.3, locus tag b0344 [FASTA]).

Section 3.2.3.2 describes the design and construction of both the OmpA-LacZα construct [027] (438 bp) and the SufI-LacZα segment YKH388 (205 bp) in more depth, including how the latter could be made from the former once OmpA-LacZα had replaced the recoded reporter sfGFP (via Gibson Assembly) within the OCI construct. Construct [027] was commercially synthesised and subcloned into [P011] pTwist-Kan vector by Twist Biosciences (USA), whilst segment YKH388 (205 bp) was synthesised as a dsDNA gBlock® (IDT, Belgium).

A.2 Oligos, primers, and sequences

Full list of all primers/oligos used within this body of work are as follows.

Oligo/Seq ID	**Item Name**	**Date, dd/mm/yyyy**	T_m, °C	**Sequence 5'->3'**	**Size, nt**
YKH001	tRNA_AAC_IVT	14/04/2015	73.6	TGG TGC CGG GGG CGG GAC TTG AAC CCG CAC GCC CTT GCG GGC ACA TGA CCA ACA ATC ATG CGC GTC TAC CAA TTC CGC CAC CCC GGC TAT AGT GAG TCG TAT TAC TTT TTG TAA TAC GAC	120
YKH002	tRNA_AAC_IVT2	09/07/2015	72.8	TGG TGC CGG GGG CGG GAC TTG AAC CCG CAC GCC CTT GCG GGC ACA TGA CCA ACA ATC ATG CGC GTC TAC CAA TTC CGC CAC CCC GGC TAT AGT GAG TCG TAT TAC TTT TTG TAA TAC GAC tca cta tag g	130
YKH003	Seq-pVPO1TthLRS_F2	28/07/2015	56.2	CAA GGG CCG CGT TAC ATA TC	20
YKH004	Seq-pVPO1TthLRS_R1	28/07/2015	56.6	GTA GAA CTC AGG GTG GGC TT	20
YKH005	TthLeuRS-iPCR-Fw_YK	28/07/2015	49.3	/5Phos/ATG TAT ATC TCC TTC TTA AAG TTA AAC	27
YKH006	TthLeuRS-iPCR-Rv_YK	28/07/2015	51.1	/5Phos/ATG GAA AAA TAT AAC CCG CAC	21

YKH007	NdeI_EcoLeuRS_Fw	05/08/2015	59.3	aaCTCATATGCAAGAGCAATACCGCC	26
YKH008	HindIII_EcoLeuRS_Rv	05/08/2015	62.9	AACAAGCTTGCCAACGACCAGATTGAGG	28
YKH009	Seq-EcoLRS_Fw	14/08/2015	55.5	TCTTGACCATGAAGCGGC	18
YKH010	Seq-EcoLRS_Rv	14/08/2015	53.7	CTGCTTTAATCGGGCTGG	18
YKH011	ori-pSB1C3_Fw	29/09/2015	52.2	ctgtcagaccaagtttactcat	22
YKH012	ori-pSB1C3_Rv	29/09/2015	52.7	ggtatcagctcactcaaagg	20
YKH013	CmR-pSB1C3_Fw	29/09/2015	53.2	tgatcgggcacgtaaga	17
YKH014	CmR-pSB1C3_Rv	29/09/2015	49.6	attctcaccaataaaaaacgc	21
YKH015	seq_CmR-homR-KO_Fw	29/09/2015	51.3	ccgtttgtgatggcttc	17
YKH016	seq_ori-homL-KO_Fw	29/09/2015	51.4	tctgacttgagcgtcg	16
YKH017	seq_CmR-homL-ori_Rv	29/09/2015	48.9	gtctggttataggtacattga	21
YKH018	homL_EcoLRS-KO_Fw	29/09/2015	51.5	CACCTCAACGCTACATTTG	19
YKH019	homR_EcoLRS-KO_Rv	29/09/2015	52.8	CCGCCAGAGATAACAACAAT	20

YKH020	homL_EcotRNA-VPQ-KO_Fw	29/09/2015	58.3	CCACTAGAATGCGCCTCCG	19
YKH021	homL_G-SDM-A_EcotRNA-VPQ-KO_Fw	29/09/2015	54.7	CCACTAGAATaCGCCTCCG	19
YKH022	homR_EcotRNA-VPQ-KO_Rv	29/09/2015	52.1	GAATCAGGGATGAAAAATCAAGG	23
YKH023	homL_EcotRNA-T-KO_Fw	29/09/2015	57.6	GTCGTGGGTTCGAATCCCATTAG	23
YKH024	homL_GC-SDM-TA_EcotRNA-T-KO_Fw	29/09/2015	53.2	GTCGTGGtTTaGAATCCCATTAG	23
YKH025	homR_EcotRNA-T-KO_Rv	29/09/2015	52	TACTCGCCGAAATACTGC	18
YKH026	EcoRI_Cre_Fw	29/09/2015	52.6	CAACTCTCTACTGTTTCTCCATAC	24
YKH027	HindIII_Cre_Rv	29/09/2015	52.6	CTGTATCAGGCTGAAAATCTTCT	23
YKH028	pBAD-Cre-FP-seq	29/09/2015	50.1	AAGACATCTCACGTACTGA	19

YKH029	pBAD-Cre-RP-seq	29/09/2015	52.3	CACTTCTGAGTTCGGCAT	18
YKH033	EcoLRS_Koadaptor _spacer	01/10/2015	N/A	tgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagCCGCCAGAGATAACAACAATGTTGCCAGATATCGCACGCTTCCTCCCGCGCtattatATAACTTCGTATAATGTATGCTATACGAAGTTATattctcaccaataaaaaacgcccggcggcaaccgagcgttctgaacaaatccagatggagttctGAGCGGtagtgatcttatttcattatggtgaaagttggaacctcttacgtgcccgatcaATAACTTCGTATAATGTATGCTATACGAAGTTATGGCAGCCAGTGGTCCTGTTTTCAATACGGCTACAAATGTAGCGTTGAGGTGggtatcagctcactcaaaggcggtaatacggttatccacagaatcagggg	399

YKH034	EcotRNA-LeuVPQ_KOadaptor	01/10/2015	N/A	tgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagGAATCAGGGATGAAAAATCAAGGAAGAAACAAGAAAGGAAGTAAAGATAATATAACTTCGTATAAGGTATACTATACGAAGTTATattctcaccaataaaaaacgcccggcggcaaccgagcgttctgaacaaatccagatggagttctGAGCGGtagtgatcttatttcattatggtgaaagttggaacctcttacgtgcccgatcaATAACTTCGTATAAGGTATACTATACGAAGTTATATACCATCAATTCTTAAAAAGAATTGCTACCACGGAGGCGCATTCTAGTGGgtatcagctcactcaaaggcggtaatacggttatccacagaatcagggg	393

YKH035	EcotRNA-LeuT_KOadaptor	01/10/2015	N/A	tgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagTACTCGCCGAAATACTGCTTTTTGAATTTTTAGTTCAATTCTTTAAAGTCGATAACTTCGTATAAAGTATCCTATACGAAGTTATattctcaccaataaaaaacgcccggcggcaaccgagcgttctgaacaaatccagatggagttctGAGCGGtagtgatcttatttcattatggtgaaagttggaacctcttacgtgcccgatcaATAACTTCGTATAAAGTATCCTATACGAAGTTATATTGTCACAACTTCTAATAATGGGGTGGCTAATGGGATTCGAACCCACGACggtatcagctcactcaaaggcggtaatacggttatccacagaatcagggg	393

YKH036	EcoRI-Cre-bricked-HindIII	01/10/2015	N/A	CAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCTA GCGAATTCATGTCCAATTTACTGACCGTACACCAAAATTT GCCTGCATTACCGGTCGATGCAACGAGTGATGAGGTTCGC AAGAACCTGATGGACATGTTCAGGGATCGCCAGGCGTTTT CTGAGCATACCTGGAAAATGCTTCTGTCCGTTTGCCGGTC GTGGGCGGCATGGTGCAAGTTGAATAACCGGAAATGGTTT CCCGCAGAACCTGAGGATGTTCGCGATTATCTTCTATATC TTCAGGCGCGCGGTCTGGCAGTAAAAACTATCCAGCAACA TTTGGGCCAGCTAAACATGCTTCATCGTCGGTCCGGGCTG CCACGACCAAGTGACAGCAATGCTGTTTCACTGGTTATGC GGCGGATTCGAAAAGAAAACGTTGATGCCGGTGAACGTGC AAAACAGGCTCTAGCGTTCGAACGCACTGATTTCGACCAG GTTCGTTCACTCATGGAAAATAGCGACCGCTGCCAGGATA TACGTAATCTGGCATTTCTGGGGATTGCTTATAACACCCT GTTACGTATAGCCGAAATTGCCAGGATCAGGGTTAAAGAC ATCTCACGTACTGACGGTGGGAGAATGTTAATCCATATTG	1128

YKH036 cont.				GCAGAACGAAAACGCTGGTTAGCACCGCAGGCGTAGAGAA GGCACTTAGCCTGGGGGTAACTAAACTGGTCGAGCGATGG ATTTCCGTCTCTGGTGTAGCTGATGATCCGAATAACTACC TGTTTTGCCGGGTCAGAAAAAATGGTGTTGCCGCGCCATC TGCCACCAGCCAGCTATCAACTCGCGCCCTGGAAGGGATT TTTGAAGCAACTCATCGATTGATTTACGGCGCTAAGGATG ATTCTGGTCAGAGATACCTGGCCTGGTCTGGACACAGTGC CCGTGTCGGAGCCGCGCGAGATATGGCCCGCGCTGGAGTT TCAATACCGGAGATCATGCAAGCTGGTGGCTGGACCAATG TAAATATTGTCATGAACTATATCCGTAACCTGGATAGTGA AACAGGGGCAATGGTGCGCCTGCTGGAAGATGGCGATTAG AAGCTTGGCTGTTTTGGCGGATGAGAGAAGATTTTCAGCC TGATACAG	
YKH030	homL_EcotRNA-W-KO_Fw	29/09/2015	54.6	ATAATGATGGGGTCACAGGTTC	22
YKH031	homR_EcotRNA-W-KO_Rv	29/09/2015	56.8	CGGTGCCTTACCGCTTG	17

YKH032	homR_A-SDM-T_EcotRNA-W-KO_Rv	29/09/2015	57.4	CGGTGCCTTtCCGCTTG	17
YKH079	EcotRNA-LeuW_KOadaptor	03/02/2016	N/A	tgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagCGGTGCCTTACCGCTTGGCGATACCCCATCCGTACAACGCTTTCTGGTGAAATAACTTCGTATAATGTATACTATACGAAGTTATattctcaccaataaaaaacgcccggcggcaaccgagcgttctgaacaaatccagatggagttctGAGCGGtagtgatcttatttcattatggtgaaagttggaacctcttacgtgcccgatcaATAACTTCGTATAATGTATACTATACGAAGTTATAAAAAAAGATGGTGGCTACGACGGGATTCGAACCTGTGACCCCATCATTATggtatcagctcactcaaaggcggtaatacggttatccacagaatcagggg	393
YKH037	Seq-pASK75_Fw	07/10/2015	51.3	atgacccgacaccatc	16
	Seq-pASK75_Rv			acacccgccgc	11
YKH039	BsaI-pASK75_Fw	07/10/2015	63.3	attGGTCTCTgctaCTGCGTCACGGA	26
YKH040	BsaI-pASK75_Rv	07/10/2015	60.4	atgGGTCTCAcaaaCTATTCATTTCACTTTTCTCTATCA	39
YKH041	BsaI-EcoLRS_Fw	07/10/2015	58.8	tgtGGTCTCAtttgTTTAACTTTAAGAAGGAGATATACAT A	41

YKH042	BsaI-EcoLRS_Rv	07/10/2015	58.1	ttaGGTCTCTtagcAGCCAACTCAG	25
	Alternative_Seq-pASK75_Rv			gggcgctggcaagtgtag	18
	Seq-pASK75_Rv2			cgtaaccaccacaccc	16
YKH038	Seq-EcoLRS_Fw2	07/10/2015	51.2	TGTTTATGATGTTTGCTTCTCC	22
YKH043	iPCR_BsaI_pKD46_Fw	05/11/2015	61.7	GGTCTCCtaacCTGTCAGACCAAGTTTACTCAT	33
YKH044	iPCR_BsaI_pKD46_Rv	05/11/2015	63	GGTCTCCatttCCCCGAAAAGTGCCA	26
YKH045	BsaI_SmR_Fw	05/11/2015	62.3	GGTCTCAaaatAGTCTCACGCCCGGAG	27
YKH046	BsaI_SmR_Rv	05/11/2015	59.9	GGTCTCTgttaGACATTATTTGCCGACTACC	31
YKH047	Seq_pKD46_Fw	05/11/2015	50.7	GGTTTTGACGATCAACTCTATT	22
YKH048	Seq_pKD46_Rv	05/11/2015	52.5	CGAAAACTCACGTTAAGGGA	20
YKH049	homL_pD881_Cre_1	11/11/2015	66.3	tttaagaaggagatatacatGCTAGCGAATTCATGTCCAA TTTACTGACCGTACACCAAA	60
YKH050	homL_pD881_Cre_2	11/11/2015	65.3	TTGGACATGAATTCGCTAGCatgtatatctccttcttaaa aagaatttcattacgaccag	60
YKH051	homR_Cre_pD881_3	11/11/2015	72.3	AAGATGGCGATTAGAAGCTTgcggccgccaccgctgagca ataactagcataaccccttg	60

YKH052	homR_Cre_pD881_4	11/11/2015	76.4	tgctcagcggtggcggccgcAAGCTTCTAATCGCCATCTT CCAGCAGGCGCACCATTGCC	60
YKH053	seq_pD881_Rv	11/11/2015	53.5	tctgaggctcgtcctg	16
YKH054	seq_pD881_Fw	11/11/2015	52.5	cgttcatctttccctggtt	19
YKH055	BsaI_CmR_Fw	17/11/2015	60.5	ggtctcaAAATgatcgggcacgtaaga	27
YKH056	BsaI_CmR_Rv	17/11/2015	58.2	ggtctctGTTAattctcaccaataaaaaacgc	32
YKH057	SapI-pBAD-Cre-iPCR_Fw	15/12/2015	64.4	aaaGCTCTTCgtttGGGCTAGCGAATTCATGTCC	34
YKH058	pBAD-Cre-SapI-iPCR_Rv	15/12/2015	64	aaaGCTCTTCcgccGACATCACCGAT	26
YKH059	SapI-pRhm-iPCR_Fw	15/12/2015	64.9	aaaGCTCTTCtggcCCACCACAATTCAGCAAAT	33
YKH060	pRhm-SapI-iPCR_Rv	15/12/2015	60.4	aaaGCTCTTCaaaaAGAATTTCATTACGACCAGTCTAAAA AG	42
YKH061	pCDF-BsaI-iPCR_Fw	19/01/2016	60.9	AATggtctcgaattcAATAACTAGCATAACCCCTTG	36
YKH062	pCDF-BsaI-iPCR_Rv	19/01/2016	59.6	TTAggtctcactagtAAAATTATTTCTACAGGGGAATTGT TA	42
YKH063	A_Fw_tRNALeuPQW	19/01/2016	68.6	aaaGGTCTCACTAGTAAAACCACGTTGATATTGCTCGCAC TGGGCGAAGG	50

YKH064	B_Rv_tRNALeuPQW	19/01/2016	73.7	GGACTTGAACCCCCACGTCCGTAAGAACACTAACACCTGA AGCTAGCGCGTCTACCAATTCCGCCACCTTCGCCCAGTGC	80
YKH065	C_Fw_tRNALeuPQW	19/01/2016	72.6	GTGGGGGTTCAAGTCCCCCCCCCTCGCACCAAACGAGGCGA TATCAAAAAAAAGTAAGATGACTGTGCGAAGGTG	73
YKH066	D_Rv_tRNALeuPQW	19/01/2016	73.2	GACTTGAACCCCCACGTCCGTAAGGACACTAACACCTGAA GCTAGCGCGTCTACCAATTCCGCCACCTTCGCACAGTCAT	80
YKH067	E_Fw_tRNALeuPQW	19/01/2016	71.5	CGTGGGGGTTCAAGTCCCCCCCCCTCGCACCAATTATCTTT AACTACTTTATGTAGTCTCCGCCGTGTAGCAAGAAATTGA	80
YKH068	F_Rv_tRNALeuPQW	19/01/2016	71.3	CGCCAGAACCTAAATCTGGTGCGTCTACCAATTTCGCCAC TCCCGCAAAAAAAGATCTTCTCAATTTCTTGCTACACGGC	80
YKH069	G_Fw_tRNALeuPQW	19/01/2016	74.1	CCAGATTTAGGTTCTGGCGCCGCAAGGTGTGCGAGTTCAA GTCTCGCCTCCCGCACCATTCACCAGAAAACGAGGC	76
YKH070	H_Rv_tRNALeuPQW	19/01/2016	66.9	aaaGGTCTCGAATTCGCTCTTCTTTTGATATCGCCTCGTT TTCTGGTGAAT	51
YKH071	tRNALeu-adpt-1	19/01/2016	64.2	GGTAGACGCGCTAGCTTCAGGTGTTAGTGT	30
YKH072	tRNALeu-adpt-2	19/01/2016	62.4	ATCTTTTTTTGCGGGAGTGGCGAAATTGGTA	31
YKH073	pBAD-Cre-RP2-seq	19/01/2016	51.3	CCAAATGTTGCTGGATAGTTT	21

YKH074	Cre-SDM-gAT_FP	29/01/2016	57.6	GATGGCgATTAGAAGCTTGGCT	22
YKH075	Cre-SDM_RP	29/01/2016	56.2	TTCCAGCAGGCGCAC	15
YKH076	RBS-Cre_FP	29/01/2016	59.9	aagaaggagGAATTCATGTCCAATTTACTGACC	33
YKH077	Cre-pRham-SDM_RP	29/01/2016	54.1	CCAAAAGAATTTCAgTACGACCAG	24
YKH078	Cre-pBAD_RP	29/01/2016	53.5	CCAAAAAAACGGGTATGGAGAA	22
YKH080	D_Rv_tRNALeuPQW _17trunc	03/03/2016	72.1	CCGTAAGGACACTAACACCTGAAGCTAGCGCGTCTACCAA TTCCGCCACCTTCGCACAGTCAT	63
YKH081	E_Fw_tRNALeuPQW _20add	03/03/2016	71.9	GGTGTTAGTGTCCTTACGGACGTGGGGGTTCAAGTCCCCC CCCTCGCACCAATTATCTTTAACTACTTTATGTAGTCTCC GCCGTGTAGCAAGAAATTGA	100
YKH082	seq-ELRS-KO_Fw	08/03/2016	52.4	TGTATGAATAATCCGGGCATC	21
YKH083	seq-ELRS-KO_Rv	08/03/2016	53.9	GGATCGCCTGAGTCCA	16
YKH084	seq-EctRNA-VPQ-KO_Fw	08/03/2016	53.2	GATGATGCTGTATCACAAACAGT	23
YKH085	seq-EctRNA-VPQ-KO_Rv	08/03/2016	53.9	TACCAAACCTGCACGCTA	18
YKH086	seq-EctRNA-T-KO_Fw	08/03/2016	53.9	GAGCTGCGGTGGTAGTA	17

YKH087	seq-EctRNA-T-KO_Rv	08/03/2016	52.8	CGAAGCGGGGTTCAAA	16
YKH088	seq-EctRNA-W-KO_Fw	08/03/2016	52.9	GGCGATACCCCAATAACC	18
YKH089	seq-EctRNA-W-KO_Rv	08/03/2016	53.1	CGCTGCAAGGCTCTATAC	18
YKH090	pBAD-Cre-RP3-seq	17/03/2016	49.9	ATA GTT TTT ACT GCC AGA CC	20
YKH091	Cre-RBS-SapI_FP	04/04/2016	60.2	aatGCTCTTCAaagaaggagGAATTCATGTCC	32
YKH092	Cre-RBS-SapI_RP	04/04/2016	61.9	aatGCTCTTCActtCCAtAAGAATTTCAgTACGACCAG	38
YKH093	pKD46-LambdaRed-GA_Fw	04/04/2016	66.7	gtcgccttgcgtataatatttgccgTTTATTATGACAACT TGACGGCTACATCA	54
YKH094	pKD46-LambdaRed-GA_Rv	04/04/2016	65.2	acatagtaagccagtatacactccgTTTCTACTGGTATTG GCACAA	46
YKH095	Seq_p15a_Rv	04/04/2016	51	AGCGTCAGATTTCGTGA	17
YKH096	Seq-pBAD_Fw	04/04/2016	49.1	GCCATTCATCCGCTTA	16
YKH097	Seq_pKD46-Beta_Fw	04/04/2016	52	GGTCCCGCATCATCAA	16
YKH098_Fw	Cre-3'SDM_Fw	12/04/2016	55	caaACCTGCgagtGATGGCgATTAGAAGCTTGGCTGTT	38
YKH099_Rv	Cre-3'SDM_Rv	12/04/2016	56	acaACCTGCgataCATCTTCCAGCAGGCG	29

YKH100_Fw _seq	Seq_pCDF-pLacI_Fw	12/04/2016	51	GCCATTCGATGGTGTC	16
YKH101_Rv _seq	Seq_pCDF-T7term_Rv	12/04/2016	50	CCTCAAGACCCGTTTAG	17
YKH102_Fw	pCDF-BsaI-iPCR_Fw2	18/04/2016	54	AATggtctcgaattCCACCGCTGAGCAA	28
YKH103_Fw	A_PQW_Fw	18/04/2016	53	aaaGGTCTCACTAGTAAAACCAC	23
YKH104_Rv	H_PQW_Rv	18/04/2016	54	aaaGGTCTCGAATTCGCTC	19
YKH105_Rv _seq	Seq_pKD46-Gam_Rv	26/04/2016	51	TCCTGAATGTAATAAGCGTTC	21
YKH106_Fw				tttACCTGCtcggATCTTCCCCATCGGTGAT	31
YKH107_Rv				tttACCTGCcgtcTTCCGCTCGCCGC	26
YKH108_Fw				tttACCTGCcggcggaaATAACTTCGTATAATGTgTaCTA TACGAAGTTATggaaatggcttacgaacg	69
YKH109_Rv				tttACCTGCgtggagatATAACTTCGTATAGtAcACATTA TACGAAGTTATtggggaagatcgggctcgcc	71
YKH110_Fw _seq				TTGTGCCAATACCAGTAG	18

YKH111_Rv _seq				TACGACCAGTCTAAAAAGc	19
YKH112_Fw		20/09/2016	56	ggtGGTCTCAACTTTAAGAAGGAGATATACATATGCAAGA	40
YKH113_Rv		21/09/2016	55	ataGGTCTCTCGATCCTCTACGCCG	25
YKH114_Fw		22/09/2016	62	atcgCATTATTGCAATTAATAAACAACTAACGGACAATTC TACCTAACAATTTTGTTTA	59
YKH115_Rv		23/09/2016	62	aagtTAAACAAAATTGTTAGGTAGAATTGTCCGTTAGTTG TTTATTAATTGCAATAATG	59
YKH116_Fw		24/09/2016	64	atcgTCATGAACAATAAAACTGTCTGCTTACATAAACAGT AATACAAGGGGTGTTATTTTGTTTA	65
YKH117_Rv		25/09/2016	64	aagtTAAACAAAATAACACCCCTTGTATTACTGTTTATGT AAGCAGACAGTTTTATTGTTCATGA	65
YKH118_Fw		26/09/2016	64	atcgGGCACGTAAGAGGTTCCAACTTTCACCATAATGAAA TAATTTTGTTTA	52
YKH119_Rv		27/09/2016	64	aagtTAAACAAAATTATTTCATTATGGTGAAAGTTGGAAC CTCTTACGTGCC	52

YKH120_Fw		28/09/2016	62	atcgACATTCAAATATGTATCCGCTCATGAGACAATAACA TTTTGTTTA	49
YKH121_Rv		29/09/2016	62	aagtTAAACAAAATGTTATTGTCTCATGAGCGGATACATA TTTGAATGT	49
YKH122_Fw		30/09/2016	65	atcgTTGACAGCTAGCTCAGTCCTAGGTATAATGCTAGCA TTTTGTTTA	49
YKH123_Rv		01/10/2016	65	aagtTAAACAAAATGCTAGCATTATACCTAGGACTGAGCT AGCTGTCAA	49
YKH124_Fw		02/10/2016	68	atcgTTGACgGCTAGCTCAGTCCTAGGTAcAgTGCTAGCA TTTTGTTTA	49
YKH125_Rv		03/10/2016	68	aagtTAAACAAAATGCTAGCAcTgTACCTAGGACTGAGCT AGCcGTCAA	49
YKH126_Fw		04/10/2016	67	atcgTTGACAGCTAGCTCAGTCCTAGGTActgTGCTAGCA TTTTGTTTA	49
YKH127_Rv		05/10/2016	67	aagtTAAACAAAATGCTAGCAcagTACCTAGGACTGAGCT AGCTGTCAA	49

YKH128_Fw		06/10/2016	66	atcgTTGACAGCTAGCTCAGTCCTAGGTATtgTGCTAGCA TTTTGTTTA	49
YKH129_Rv		07/10/2016	66	aagtTAAACAAAATGCTAGCAcaATACCTAGGACTGAGCT AGCTGTCAA	49
YKH130_Fw		08/10/2016	67	atcgTTGACgGCTAGCTCAGTCCTAGGTATAgTGCTAGCA TTTTGTTTA	49
YKH131_Rv		09/10/2016	67	aagtTAAACAAAATGCTAGCAcTATACCTAGGACTGAGCT AGCcGTCAA	49
YKH132_Fw		10/10/2016	66	atcgTTtACgGCTAGCTCAGTCCTAGGTATAgTGCTAGCA TTTTGTTTA	49
YKH133_Rv		11/10/2016	66	aagtTAAACAAAATGCTAGCAcTATACCTAGGACTGAGCT AGCcGTaAA	49
YKH134_Fw		12/10/2016	66	atcgTTtACgGCTAGCTCAGTCCTAGGTAcAATGCTAGCA TTTTGTTTA	49
YKH135_Rv		13/10/2016	66	aagtTAAACAAAATGCTAGCATTgTACCTAGGACTGAGCT AGCcGTaAA	49

YKH136_Fw		14/10/2016	67	atcgTTGACAGCTAGCTCAGTCCTAGGgActATGCTAGCA TTTTGTTTA	49
YKH137_Rv		15/10/2016	67	aagtTAAACAAAATGCTAGCATagTcCCTAGGACTGAGCT AGCTGTCAA	49
YKH138_Fw		16/10/2016	67	atcgTTGACAGCTAGCTCAGTCCTAGGgATtgTGCTAGCA TTTTGTTTA	49
YKH139_Rv		17/10/2016	67	aagtTAAACAAAATGCTAGCAcaATcCCTAGGACTGAGCT AGCTGTCAA	49
YKH140_Fw		18/10/2016	65	atcgcTGAtAGCTAGCTCAGTCCTAGGgATtATGCTAGCA TTTTGTTTA	49
YKH141_Rv		19/10/2016	65	aagtTAAACAAAATGCTAGCATaATcCCTAGGACTGAGCT AGCTaTCAg	49
YKH142_Rv _seq		20/10/2016	52	CGGTGAAGAATTTCTGTTCC	20
YKH143_Fw	YKH143_Fw_LH-HL	06/11/2016	55	aaaGGTCTCACTAGTccgc	19
YKH144_Rv	YKH144_Rv_LH-HL	06/11/2016	56	aaaGGTCTCGAATTCgggAAG	21

YKH145_Fw	YKH145_Fw_pNGAL-vec	06/11/2016	54-55	cacGGTCTCaagcttGACCTGTGAAGTG	28
YKH146_Rv	YKH146_Rv_pNGAL-vec	06/11/2016	56	ggcGGTCTCtctagaTTTTTGTCGAACTATTCATTTCACT	40
YKH147_Rv	YKH147_Rv_pNGAL-vec	06/11/2016	56	ctcGGTCTCgtcATTTTTTGCCCTCGTTATCTAGATTTTT GT	42
YKH148_Fw	YKH148_Fw_sfGFP	06/11/2016	57	cgtGGTCTCaagcttATCTCGATCCCGCGAAAT	33
YKH149_Rv	YKH149_Rv_sfGFP	06/11/2016	57	ggcGGTCTCgaattcCCGGATATAGTTCCTCCTTTCAG	38
YKH150_Fw	YKH150_Fw_seq	06/11/2016	52	GCCTTTTTACGGTTCCTG	18
YKH151_Fw	YKH151_Fw_seq	06/11/2016	54	GGTAGAGCCCTGGATTAAGAT	21
YKH152_Rv	YKH152_Rv_seq	06/11/2016	53	CGGCGGAGACTACATAAAG	19
YKH153_Rv	YKH153_Rv_seq	06/11/2016	53	ATCCGGCTGCTAACAAAG	18
YKH154_Fw	YKH154_Fw_seq	06/11/2016	53	AACAGTTCTTCACCTTTGCTA	21
YKH155_Rv	YKH155_Rv_seq	06/11/2016	53	GAGATCCGTGACGCAG	16
YKH156_Fw	YKH156_Fw_seq	06/11/2016	52	CTCGATCCCGCGAAAT	16
YKH157_Rv	YKH157_Rv_seq	06/11/2016	53	CGGATATAGTTCCTCCTTTCAG	22

YKH158	YKH158_EcotRNAs (LH-HL)	06/11/2016	N/A	aaaGGTCTCACTAGTccgcGGATCCAAAACCACGTTGATA TTGCTCGCACTGGGATTTGTAGTGGTGGCTATAGCTCAGT TGGTAGAGCCCTGGATTAAGATTCCAGTTGTCGTGGGTTC GAATCCCATTAGCCACCCCATTATTAGAAGAACGAGGCGA TATCAAAAAAAGTAAGATGACTGTTAGTATCCGTGCCGAG GTGGTGGAATTGGTAGACACGCTACCTTATGGTGGTAGTG CCCAATAGGGCTTACGGGTTCAAGTCCCGTCCTCGGTACC AAATTCCAGAAACTACTTTATGTAGTCTCCGCCGTGTAGC AAGAAATTGAGAAGAGAAGAGCAAAATGGCGCACATTGTG CGACATTTTTTAAGCTTcccGAATTCGAGACCttt	395
YKH159	YKH159_ (pET_assembly_vec_F)	06/11/2016	56	gagtcaggtctcAGGAAGCTGAGTTGGCT	29
YKH160	YKH160_ (pET_assembly_vec_R)	06/11/2016	57	gagtcaggtctcATCCGCTCACAATTCCCCTATA	34
YKH161	YKH161_ (pET29_DA_F)_S-nicked	06/11/2016	56	CCCCTTATTAGCGTTTGCCAggtctcGCGGATAACAATTC CCCTCTAG	48

YKH162	YKH162_ (pET29_DA_R)_S-nicked	06/11/2016	56	CTAACAAAGCCCGAAAGGAAGgagaccCATCTGACCTCTG TGCTGCT/3invdT/	47
YKH163	YKH163_ (Out-nest_Fw=S/Rv=AS)	06/11/2016	56	CCCCTTATTAGCGTTTGCCA	20
YKH164	YKH164_ (Out-nest_Rv=S/Fw=AS)	06/11/2016	58	AGCAGCACAGAGGTCAGATG	20
YKH165	YKH165_Y67H (wt*CAC)_CrP	06/11/2016	55	ccaccctgaccCACggcgttcagtg	25
YKH166	YKH166_H26H (wt*CAC)	06/11/2016	55	atgtgaatggccaCaaatttagcgt	25
YKH167	YKH167_H78H (wt*CAC)	06/11/2016	55	ctatccggatcaCatgaaacgcca	24
YKH168	YKH168_H82H (wt*CAC)	06/11/2016	55	aCatgaaacgccaCgatttctttaaaagc	29
YKH169	YKH169_H78,82H (wt*CAC)	06/11/2016	56	tatccggatcaCatgaaacgccaCgatttctttaaaag	38

YKH170	YKH170_H140H (wt*CAC)	06/11/2016	57	gcaacattctgggtcaCaaactggaatataat	32
YKH171	YKH171_H149H (wt*CAC)	06/11/2016	55	atttcaacagccaCaatgtgtatattaccg	30
YKH172	YKH172_H182H (wt*CAC)	06/11/2016	55	ctggcggatcaCtatcagcagaat	24
YKH173	YKH173_H200H (wt*CAC)	06/11/2016	56	gccggataatcaCtatctgagcacc	25
YKH174	YKH174_H218H (wt*CAC)	06/11/2016	58	gaaaaacgtgatcaCatggtgctGct	26
YKH175	YKH175_H243-245,247H (wt*CAC)	06/11/2016	55	aaggcagccaccaCcaCcaCcaccaCtaatgagatcc	37
YKH176	YKH176_Y/Hwt*67L (scrCTT)_CrP	06/11/2016	55-57	caccctgaccCTTggcgttcagtg	24
YKH177	YKH177_L65H (scrCAT)_Hwt*67_S4	06/11/2016	58	tgaccacccATaccCACggcgt	22

YKH178	YKH178_L65H (scrCAT)_L (scrCTT)67_S4+CrP	06/11/2016	57	gtgaccacccATaccCTTggcgt	23
YKH179	YKH179_L65H (scrCAT)_Y/Hwt*67L (scrCTT)_S4+CrP	06/11/2016	55-56	tggtgaccacccATaccCTTggcgttcagtg	31
YKH180	YKH180_L45H (scrCAT)_S2	06/11/2016	56	acggtaaactgacccATaaatttatttgcacc	32
YKH181	YKH181_L43H (scrCAT)_(S1)	06/11/2016	56	gaccaacggtaaacATacccTGaaatttatttgc	34
YKH182	YKH182_L43,45H (scrCAT)_S2(S1)	06/11/2016	56	gcgaccaacggtaaacATacccATaaatttatttgcacc	39
YKH183	YKH183_L61H (scrCAT)_(S3)	06/11/2016	57	tggccgacccATgtgaccaccc	22
YKH184	YKH184_Hwt*149L (scrCTT)_S5	06/11/2016	55	aatttcaacagccTTaatgtgtatattaccgcc	33

YKH185	YKH185_L221H (scrCAT)_S6	06/11/2016	55	tgatcaCatggtgcATctggaatttgttacc	31
YKH186	YKH186_L222H (scrCAT)_(S7)	06/11/2016	56	tcaCatggtgctgcATgaatttgttaccg	29
YKH187	YKH187_L221-222H (scrCAT)_S6(S7)	06/11/2016	55	gtgatcaCatggtgcATcATgaatttgttaccgcc	35
YKH188	YKH188_Y67L(CTC)	06/11/2016	55	ccaccctgaccCTCggcgttcagtg	25
YKH189	YKH189_Y67L(CTA)	06/11/2016	55	ccaccctgaccCTAggcgttcagtg	25
YKH190	YKH190_Y67L(CTG)	06/11/2016	55	ccaccctgaccCTAggcgttcagtg	25
YKH191	YKH191_ (pET29_DA_F)_AS-nicked	06/11/2016	56	AGCAGCACAGAGGTCAGATGggtctcCTTCCTTTCGGGCT TTGTTAG	47
YKH192	YKH192_ (pET29_DA_R)_AS-nicked	06/11/2016	56	CTAGAGGGGAATTGTTATCCGCgagaccTGGCAAACGCTA ATAAGGGG/3invdT/	48
YKH193	YKH193_Y67H (wt*CAC)_RC	06/11/2016	55	ctgaacgccGTGggtcagggtg	22

YKH194	YKH194_Y67L (CTT)_RC	06/11/2016	55	cactgaacgccAAGggtcagggtg	24
YKH195	YKH195_Y67H (scrCAT)	07/04/2017	54	caccctgaccCATggcgttca	21
YKH190_v2	YKH190_Y67L (CTG)_v2	07/04/2017	53	ccaccctgaccCTGggcgttcagt	24
YKH196	YKH196_Y67L(TTA)	07/04/2017	54	caccctgaccTTAggcgttcagt	23
YKH197	YKH197_Y67L(TTG)	07/04/2017	54	caccctgaccTTGggcgttcagt	23
YKH198	YKH198_Rv	07/04/2017		TCACCAGGGTCGGC	14
YKH199	YKH199_Rv	07/04/2017		TCACCAGGGTCGGCC	15
YKH200	YKH200_Y67	07/04/2017		caccctgaccTATggcg	17
YKH201	YKH201_Rv2			gTCACCAGGGTCGGC	15
YKH202	YKH202_Rv2			gTCACCAGGGTCGGCc	16
	YKH###_ pET29a_F1_seq			ACAGTCGTTGCTGATTGG	18
	YKH###_ pET29a_R1_seq			GCCGCTTTACAGGCTTC	17

	YKH###_ pET29a_F2_seq			CCGATGCCCTTGAGAG	16
	YKH###_ pET29a_R2_seq			GCCGAAAATGACCCAGAG	18
	YKH###_ pET29a_F3_seq			GTCAGAGGTGGCGAAAC	17
	YKH###_ pET29a_R3_seq			AGGGAGAAAGGCGGAC	16
	YKH###_ pET29a_F4_seq			ATCCGTACTCCTGATGATG	19
	YKH###_ pET29a_R4_seq			TCGTGATTGCGCCTG	15
	YKH###_ pET29a_dRop_seq			CGCCCTGCACCATTATGTTC	
YKH203		28/04/2017		TCCGCggatccAAAAC	16
YKH204		28/04/2017		cgacaaaaatCTAGTCCGCggatccAA	27
YKH205		28/04/2017		GgaattcGGGaagcttAAAAAATG	24

YKH206		28/04/2017		GGAACTATATCCGGgaattCgggAAGCTTaaaaaatg	37
YKH207		28/04/2017		CGGATATAGTTCCTCCTTTCAG	22
YKH208		28/04/2017		cttCCCgaattcCCGGATATAGTTCCTCCTT	29
YKH209		28/04/2017		ctcgatcccgcgaaat	16
YKH210		28/04/2017		ACAGGTCaagcttatctcgatcccgcga	28
YKH211		28/04/2017		ataagcttGACCTGTGAAGTG	21
YKH212		28/04/2017		gggatcgagataagcttGACCTGTGAAGTG	30
YKH213		28/04/2017		CTAGatttttgtcgaactattcatttc	27
YKH214		28/04/2017		gatccGCGGACTAGatttttgtcgaactattcatttc	37
YKH215	YKH215_Fw (C31-*toN)	15/05/2017	56	AATgtcgtcatcAACaaactggcgcactcgacc	33
YKH216	YKH216_Rw C31	15/05/2017	54	gaccattcggccgttg	16
YKH217	YKH217_Fw (BsaI[/SalI]-RBS-ELRS)	15/05/2017	57	cggGGTCTCgtcgaATTTTGTTTAACTTTAAGAAGGAGATATACATATGC	50

YKH218	YKH218_Rv (ELRS-TAA-TGA-HindIII/BsaI)	15/05/2017	56	cggGGTCTCAAGCTTTCATTAGCCAACGACCAGATTGAGG	40
YKH219	YKH219_Fw (RBStrunc-attB)	15/05/2017	59	aaaGGTCTCtaGGCGTGCGGGTGC	24
YKH220	YKH220_Rv lambda-t0-term	15/05/2017	58	cggGGTCTCagaatAATTCATGAGCGGATGGAG	33
YKH221	YKH221_Fw (RBS-pET29a)	15/05/2017	53	ggcGGTCTCaataaTTTTGTTTAACTTTAAGAAGGAGATA TACA	44
YKH222	YKH222_Rv (lacO-pET29a)	15/05/2017	55	cttGGTCTCgAGAGGGGAATTGTTATCCG	29
YKH223	YKH223_Fw (BsaI-XbaI-RBS-pET29a)	15/05/2017	55	cagGGTCTCgcTCTAGAAATAATTTTGTTTAACTTTAAGA AGGAG	45
YKH224	YKH224_Rv (ELRS-TTA-TGA-BsaI)	15/05/2017	56	ccaGGTCTCgttatTCATTAGCCAACGACCAGATTGAG	38
YKH225	YKH225_Rv (TLRS-TTA-TGA-BsaI)	15/05/2017	56	cgcGGTCTCgttatTCATTAACCGCGCACCAC	32

YKH226	YKH226_Fw (lacOtrunc _RBS_XLRS_sfGFP)	15/05/2017	57	cctGGTCTCcgcctACAATTCCCCTCTAGAAATAATTTTG TTTAAC	46
YKH227	YKH227_Rv T7term	15/05/2017	58	atgGGTCTCgattcCGGATATAGTTCCTCCTTTCAGC	37
YKH228	YKH228_Rv (ELRStrunc-BsaI)	16/05/2017	56	ttaGGTCTCagcagGGTCAAGCTTTCATTAGCCAAC	36
YKH229	YKH229_Fw (RBS?-TetR)	16/05/2017	54	ggcGGTCTCccagGAATTAATGATGTCTCGTTTAGATAA	39
YKH230	YKH230_Fw (AmpRtrunc-TetR)	16/05/2017	56	gtcGGTCTCgcaggCACTGATTAAGCATTGGTAGGAATTA	40
YKH231	YKH231_Fw (BsaI-BBa_B0030-TetR)	16/05/2017	55	ggaGGTCTCgcaggGATTAAAGAGGAGAAAgGAATTAATG ATGTCTCGTTTAGATAA	57
YKH232	YKH232_Fw (terms-CamR)	16/05/2017	56	caaGGTCTCcctgcataacgcgaagtaatc	30
YKH233	YKH233_Rv (SacI-EcoRV-CamR)	16/05/2017	55	tctGGTCTCgcctggagctcgatatcaaattacgcc	36

YKH234	YKH234_Rv (BsaI-CamR)	16/05/2017	56	tctGGTCTCgcctgatcaaattacgccccgc	31
YKH235	YKH235_Fw (pSB1C3_full_linear_R)	16/05/2017	58	cTCTAGAagcggccgcGAATTCcagaaatcatccttagcg	40
YKH236	YKH236_Rv (pSB1C3_full_linear_F)	16/05/2017	58	tACTAGTagcggccgCTGCAGtccg	25
YKH237	YKH237_Rv_seq (C31)	16/05/2017	54	tcctgccaacgggtc	15
YKH238	YKH238_Fw_seq (C31)	16/05/2017	52	gttgtcacgaagtccacta	19
YKH239	YKH239_Rv_seq (ELRS_RBS-sfGFP)	16/05/2017	52	GGTCGTTGGCTAATGAataaT	21
YKH240	YKH240_Fw_seq (XbaI-lacO-attB)	16/05/2017	52	TTATTTCTAGAGGGGAATTGTaGG	24
YKH241	YKH241_Rv_seq (RBS-attP)	16/05/2017	54	aTtcgacAGTGCCCCA	16

YKH242	YKH242_Fw_seq (ELRS)	16/05/2017	55	GTTGGCTAATGAAAGCTTGACC	22
YKH243	YKH243_Fw_seq (TetR)	16/05/2017	52	AGCGGGCTTTGCTC	14
YKH244	YKH244_Rv_seq (TetR)	16/05/2017	52	ATCTTCCAATACGCAACCT	19
YKH245	YKH245_Rv_seq (lambda-t0-term)	16/05/2017	52	gactcctgttgatagatccag	21
YKH246	YKH246_Fw_seq (attB)	17/05/2017	57	CGTGCGGGTGCCAG	14
YKH247	YKH247_Rv_seq (attP)	17/05/2017	58	AGTGCCCCAACTGGGGTA	18
YKH248	YKH248_Fw_seq (attP)	17/05/2017	56	ACGCCCCCAACTGAGA	16
YKH249	YKH249_Rv_seq (attB)	17/05/2017	56	GGAGTACGCGCCCG	14
YKH250	YKH250_Rv_seq (BsaI-attB)	17/05/2017	55	aaaGGTCTCCcgatGGAGTACGCGCCCG	28

YKH251	YKH251_Fw_seq (BsaI-attP)	17/05/2017	54	aaaGGTCTCGacttACGCCCCCAACTGAG	29
YKH252	YKH252_Rv_seq (bact-term)	17/05/2017	52	cacgaggcagaatttcag	18
YKH253	YKH253_Rv (BsaI-HindIII-T7term)	17/05/2017	57	cggGGTCTCAAGCTTCGGATATAGTTCCTCCTTTCAGC	38
YKH254	YKH254_Fw (BsaI[/SalI]-RBS-TLRS)	17/05/2017	57	cggGGTCTCgtcgaGAAGGAGATATACACTTCATGGC	37
YKH255	YKH255_Fw_ (RecombiCirc_KanR)	28/07/2017	55	GAATGTGGCTCATATGAATCCC	22
YKH256	YKH256_Rv_ (RecombiCirc_attP)	28/07/2017	56	AGAGACCttttAGgtcgacAGT	22
YKH257	YKH257_Fw_(attB_GA)	28/07/2017	65	CTCCCCGGGCGCGTACTCC	19
YKH258	YKH258_Rv_(attP_GA)	28/07/2017	63	CAGAGTTCTCTCAGTTGGGGGCGT	24
YKH259	YKH259_Rv (t-lpp-BsaI)	18/08/2017	54	ttaGGTCTCagcagGGTCAAGCTtCGGCAGA	31

YKH260	YKH260_Rv_seq (lambda-t0-term)	05/09/2017	55	CGTAAGTGACTGAgactcctg	21
YKH261	YKH261_LacZa	26/09/2017	52	accatgattacggattcact	20
YKH262	YKH262_M31fwd	26/09/2017	53	gtaaaacgacggccagt	17
YKH263	YKH263_OmpA _LacZa_Fw	26/09/2017	55	ACTTCACAGGTCaagcttatAGG	23
YKH264	YKH264_OmpA _LacZa_Rv	26/09/2017	55	GGCTTTGTTAGCAGCCg	17
YKH265	YKH265_Boundary-Fw_5Biotin	27/09/2017	64	/5BiotinTEG/CCCCTTATTAGCGTTTGCCAGGTCTCaa tggtttcttagacgtcaggtggcactc	55
YKH266	YKH266_Boundary-Rv_3invdT	27/09/2017	63	gccgggcgtttttttattggtgagaatcc'AAGC,CGAGAC CCATCTGACCTCTGTGCTGCT/3invdT/	59
YKH267	YKH267_pSB1C3A2-Fw	27/09/2017	65	gagtcaGGTCTCcaagcctcgagctgtcagaccaagt	37
YKH268	YKH268_pSB1C3A2-Rv	27/09/2017	65	gagtcaGGTCTCaccattattatcatgacattaacctata aaaataggcgtatcacgagg	60

YKH269	YKH269_H17caC _H21caC	27/09/2017	60	atcccaatggcaCcgtaaagaacaCtttgaggcatt	36
YKH270	YKH270_H70caC	27/09/2017	55	gatgaatgctcaCccggaatttcg	24
YKH271	YKH271_H96caC	27/09/2017	56	acaccgttttccaCgagcaaactg	24
YKH272	YKH272_H192caC _H193cat(WT)	27/09/2017	57	cgattcaggttcaCcatgccgtttg	25
YKH273	YKH273_H200caC	27/09/2017	57	gatggcttccaCgtcggcagaat	23
YKH274	YKH274_Fw_seq	27/09/2017	54	GACAGATCGCTGAGATAGGT	20
YKH275	YKH275_Rv_seq	27/09/2017	54	actcacgttaagggattttgg	21
YKH276	YKH276_outnest1	28/09/2017	56	CCCCTTATTAGCGTTTGCCA	20
YKH277	YKH277_outnest2	28/09/2017	58	AGCAGCACAGAGGTCAGATG	20
YKH278	YKH278_H192caC _H193L(CTT)	28/09/2017	56	gcgattcaggttcaCcTtgccgtttgT	27
YKH279	YKH279_H200caC_v2	28/09/2017	57	atggcttccaCgtcggcagaatg	23
YKH280	YKH280_CAT_OCI-GA_Fw	28/09/2017	57	ACTTCACAGGTCaagcttatAGGTCTCAatcgGTCGACtg atcgggcacgtaagagg	57

YKH281	YKH281_CAT_OCI-GA_Rv	28/09/2017	57	GGCTTTGTTAGCAGCCggatAGGTCTCAaagtGCGGCCGC ttattacgccccgccc	56
YKH282	YKH282_Fw (BsaI-KanR)	28/09/2017	53	agcGGTCTCggagtATGAGCCATATTCAACGGGAA	35
YKH283	YKH283_Rv (BsaI-KanR)	28/09/2017	53	tctGGTCTCtcctgTTAGAAAAACTCATCGAGCATCAAAT	40
YKH284	YKH284_Rv (BsaI-pAmpR)	28/09/2017	54	ttaGGTCTCtactctTCCTTTTTCAATATTATTGAAGCA	39
YKH285	YKH285_Fw_(BsaI-t0term)	19/10/2017	53	ggcGGTCTCaagctAgaatAATTCATGAGCGGATG	35
YKH286	YKH286_Rv_ (BsaI_attB)	19/10/2017	53	aaaGGTCTCtagagGTGCGGGTGCCAG	27

YKH287	BP-excision_tsPurple	08/01/2018		AATGTGCGCCATTTTTCACTTCACAGGTCAagctAGCTCT TCAggatGGATCCTCAGTCACTTAAGGTCTCAatcggctg ttgacaattaatcatcgaactagttaactagtacgcaagt tcacgacttTGAGACCTGTGCGGGTGCCAGGGCGTGCCCc aGGGCTCCCCGGGCGCGTACTCCccaggcatcaaataaaa cgaaaggctcagtcgaaagactgggcctttcgttttatct gttgtttgtcggtgaacgctctctactagagtcacactgg ctcaccttcgggtgggcctttctgcgtttataGATATCAG TGCCCCAACTGGGGTAACCTcaGAGTTCTCTCAGTTGGGG GCGTaccttaggaggtaaacatatggcgagcttggttaag aaagatatgtgtgtt	415
YKH288	YKH288_Fw_(t-lpp)	08/01/2018	54	AATGTGCGCCATTTTTCAC	19
YKH289	YKH289_Rv_ (DA_priming_site)	08/01/2018	56	GCAATGAACGGACTTGACCT	20
YKH290	YKH290_Fw_ (tsPurple_NdeI)	08/01/2018	54	AAAcatatggcgagcttgg	19

YKH291	YKH291_Fw_ (tsPurple)	09/01/2018	53	aagagtcctttccagaggg	19
YKH292	YKH292_Rv_(tsPurple)	09/01/2018	58	gacgcacggctccca	15
YKH293	YKH293_Fw_(attB)	23/01/2018	63	GGGCACGCCCTGGCAC	16
YKH294	YKH294			ccgttcgccggactt	15
YKH295	YKH295_L468H(cat)			gcgaacCATgttgcggagcg	20
YKH296	YKH296_L143H(cat)			ATGgtgaatcaggtccatgac	21
YKH297	YKH297_H142L(ctt)			cagAAGaatcaggtccatgacgt	23
YKH298	YKH298_H142L(ctt) _L143H(cat)			ATGAAGaatcaggtccatgacgt	23
YKH299	YKH299			attatgcggCTGgacg	16
YKH300	YKH300			gaacctgaaccggcc	15
YKH301	YKH301_H61L(ctt)			gtcgggCTTttcagcgaagcg	21
YKH302	YKH302_Fw_iPCR-t-lpp	16/04/2018	53	GCAGACAAAAAAAATGTCGC	20
YKH303	YKH303_Rv_iPCR-f1ori	16/04/2018	53	CACGGATCTCCACGC	15

YKH304	YKH304_Boundary-Fw_5Biotin_SapI	16/04/2018	64	/5BiotinTEG/CCCCTTATTAGCGTTTGCCAGCTCTTCt'aag,aGCTTCACAGGTCAAGTCCGTTCATTG	58
YKH305	YKH305_Boundary-Rv_3invdT_SapI	16/04/2018	62(-63)	GAAGTGTATATCTCCTTTCTAGATTTTTGTCGAACTATTCATTTc'act,tGAAGAGCCATCTGACCTCTGTGCTGCT/3invdT/	76
YKH306	YKH306_pCtsP_pNGAL97-Fw_SapI	16/04/2018	59	gagtcaGCTCTTCc'act,tTTCTCTATCACTGATAGGGAGTG	41
YKH307	YKH307_pCtsP_pNGAL97-Rv_SapI	16/04/2018	60	gagtcaGCTCTTCA'ctt,cGTGATTGTCTGCCGTTTAC	37
YKH308	YKH308_*192(200)N(AAT)_I*196(204)N(AaC)	17/04/2018	56	agtgcgccagtttGTTGatgacgacATTgaccattcgg	38
YKH309	YKH309_H53(61)L(cTT)	17/04/2018	58	cttcgctgaaAAGcccgacgaacct	25
YKH310	YKH310_L75(83)H(cAT)	17/04/2018	55	cggcattcgttATGgatgcgttcg	24

YKH311	YKH311_L109(117) |L110(118) |L112(120)H(cAT)	17/04/2018	55	caatcgtcacgccATGggcATGATGttccgagacaatc	38
YKH312	YKH312_L135(143) |L139(147)H(cAT)	17/04/2018	58	tgcgacgcgtcATGccgcataatATGgtgaatcag	35
YKH313	YKH313_L132(140) |L135(143)H(cAT) _H134(142)L(cTT)	17/04/2018	52	ATGAAGaatATGgtccatgacgtttcc	27
YKH314	YKH314_L132 |L135 |L139H(cAT) _H134L(cTT)	17/04/2018	57	tgcgacgcgtcATGccgcataatATGAAGaatATGgtcca tgacgtttc	49
YKH315	YKH315_L154(162) |NL159(167) |L163(171)H(cAT)	17/04/2018	57	gtacccgccATGttcgcgctgATGAttcttcgtgtcATGa atcttcgc	48
YKH316	YKH316_Fw_seq (C31)	24/04/2018	54	cgacgcttcggcaag	15

YKH317	YKH317_Fw_iPCR-t-lpp	24/04/2018	53	GTGCGCCATTTTTggtc	17
YKH318	YKH318_pC-sfGFP-flip_pNGAL97-Rv_SapI	24/04/2018	61	gagtcaGCTCTTCA'ctt,cGTGAAAAATGGCGCACAT	36
YKH319	YKH319_C31-R225_Rv	29/05/2018	56	Cggattacgtcgggctc	17
YKH320	YKH320_C31-R225_Fw	29/05/2018	55	gtggtggtggcgtga	15
YKH321	YKH321_C31-R225_Fw_SapI	31/05/2018	54	aatGCTCTTCtccGgTGGTGGTGGCGT	27
YKH322	YKH322_C31-G54_Rv_SapI	31/05/2018	55	aatGCTCTTCcggccCCCGTC	21
YKH323	YKH323_C31-G54_Fw_SapI	31/05/2018	60	attGCTCTTCTgccgGTTCAGGTTCGTCG	29
YKH324	YKH324_C31-R225_Rv_SapI	31/05/2018	60	aatGCTCTTCaCggaTTACGTCGGGCTCGAAC	32

YKH325	YKH325_C31-G54_Rv_SapI	31/05/2018	55	aatGCTCTTCtAtggTCCATGACGTTTCCC	30
YKH326	YKH326_C31-G54_Fw_SapI	31/05/2018	60	ttaGCTCTTCccaTaTTCACCTGATTATGCGGCatGACG	33
YKH327	YKH327_C31-L359_Fw	01/06/2018	62	gggCTGtcacgggggcaag	19
YKH328	YKH328_C31-L359_Rv	01/06/2018	62	cttgccgcgcccccct	15
YKH329	YKH329_L468H_Fw_SapI	01/06/2018	56	tgtGCTCTTCtcgaacCATgttgcggag	28
YKH330	YKH330_L468H_Rv_SapI	01/06/2018	59	aatGCTCTTCttcgcccgttcgccg	25
YKH329_v2	YKH329_L468H_Fw_SapI_[v2:sfGFP_pNGAL97]	03/06/2018	57	tgtGCTCTTCtcgaacCATgttgcggagcgcg	32

YKH331	YKH331_Fw_BamHI_ ?LtoH_ ?EcotRNAHis(CUU)	04/06/2018	52	gggGGATCCAAAAAAAGTAAGATGACTGTTAGTATCc	29
YKH332	YKH332_Rv_BamHI_ ?LtoH_ ?EcotRNAHis(CUU)	04/06/2018	54	aaaggatccGCGGACTAG	18
YKH333	YKH333_Fw_SapI_ ?HtoL_ ?EcotRNALeu(CAU)	04/06/2018	53	cacAAGAGAAGAGCAAAATGGC	22
YKH334	YKH334_Rv_SapI_ ?HtoL_ ?EcotRNALeu(CAU)	04/06/2018	53	cggGCTCTTCaCTTATCGCCTCGTTCTTCTAATAATG	37
YKH335	YKH335_?CAT_ linkerTop_BsaI-ready	04/06/2018	54	aagtGCttCCGCttaGTCGACata	24
YKH336	YKH336_?CAT_ linkerBottom_BsaI-ready	04/06/2018	54	atcgtatGTCGACtaaGCGGaaGC	24

YKH337	YKH337_Fw_Alpha-ori_SapI	20/07/2018	55	tttGCTCTTCTCGTTCCACTGAGCGTC	27
YKH338	YKH338_Rv_Alpha-ori_SapI	20/07/2018	56	attGCTCTTCGGCGTTTTTCCATAGGCT	28
YKH339	YKH339_Fw_pUC-ori-3'_SapI	20/07/2018	56	accGCTCTTCACGCCAGCAACGC	23
YKH340	YKH340_Rv_pUC-ori-5'_SapI	20/07/2018	55	taagGCTCTTCTACGAAAACTCACGTTAAGGG	32
YKH341	YKH341_AF_OXB20-KanR_Fw	29/08/2018	59	CAGGAGGCTTTCGCATGATTGA	22
YKH342	YKH342_AF_OXB20-KanR_Rv	29/08/2018	58	TCAGAGCAGCCGATTGTCTG	20
YKH343	YKH343_AF_Nissle1917-terB-Tus_Fw	29/08/2018	57	GGTCAAGTCACCACCACTGT	20

YKH344	YKH344_AF_ Nissle1917-terB-Tus_Rv	29/08/2018	58	GCTCTTGTTCCATCTGGCGA	20
YKH345	YKH345_Rv_SapI_ ?tRNAs	04/12/2018	54	aaacagGCTCTTCcCCAGTGCGAGCA	26
YKH346	YKH346_Fw_tRNAHL	04/12/2018	54	ggtGCTCTTCctggTAGTATCCGTGCCGAG	30
YKH347	YKH347_Rv_tRNAHL +Spacer54-53	04/12/2018	53	tttGCTCTTCccttGTGCTTATCGCCTCGTTTTCTGGAAT TTGGTACCGAG	51
YKH348	YKH348_Fw_tRNAHL +(8bp-5')	04/12/2018	55	cgtGCTCTTCctggCTAAAAAGTAGTATCCGTGCCGAGGT GG	42
YKH349	YKH349_Rv_tRNAHL +(9bp-3')+Spacer54-53	04/12/2018	53	tttGCTCTTCccttGTGCTTATCGCCTCGcAGCGTCTCTT TTCTGGAATTTGGTACCGAG	60
YKH350	YKH350_Fw_5'-CysT	13/12/2018	67	tggCCTGAAGAATTTGGCGCGTTAACAAAGCGGTTATGTA GCGGATTtatAATCCGTCTAGTCCGGTTCGACTCC	75
YKH351	YKH351_Rv_CysT-3'	13/12/2018	67	CTTGTGCTTATCGCCTCGTTTCGGGAAGAAAGTGGAGGCG CGTTCCGGAGTCGAACCGGACTAGACGGATTATaAATCCG CTACATAACC	90

YKH352	YKH352_Fw_CysT (VAT)_blunt-iPCR	13/12/2018	50	VATAATCCGTCTAGTCCG	18
YKH353	YKH353_Rv_CysT_ blunt-iPCR	13/12/2018	50	AATCCGCTACATAACCG	17

YKH354	YKH354_ThyA (gblock)	13/12/2018		GGGCTTTGTTAGCAGCCGGATCTCaTTAGATAGCCACCGG CGCTTTgATGCCCGGATGCGGATCGTAGCCTTCgATCTCA AAGTCTTCGAAACGGTAGTCGAAGATGGATTCGGGTTTAC GTTTGATgATCAgCTTCGGCAGCGGACGCGGTTCGCGGCT cAgTTGCAGATGAGTTTGATCCATATGGTTGCTGTACAGA TGCGTGTCGCCACCGGTCCAGACAAAATCACCCACTTCCA GATCGCACTGCTGCGCCATCATATGCACCAgcAgCGCGTA GCTGGCgATGTTGAACGGCAGGCCcAGGAAGACGTCACAG GAGCGCTGATAcAGCTGGCAAGAcAGTTTGCCGTCTGCCA CATAGAACTGGAAGAATGCATGGCACGGTGCCAGCGCCAT TTTATCCAGTTCGCCTACGTTCCACGCTGAAACgATgATG CGGCGCGAATCCGGGTCGTTTTTCAGCTGGTTCAGTACCG TAGTGATCTGGTCgATATGACGACCATCTGGCGTTGGCCA GGCGCGCCACTGTTTACCATACACTGGCCCcAGGTCGCCG TTTTCATCGGCCCATTCGTCCCAGATGGTGACATTGTTTT CGTGcAGATAAGCgATGTTAGTGTCGCCCTGCAGAAACCA	867

YKH354 cont.				CAGCAGTTCATGGATGATGGAACGCAGGTGGCAACGTTTA GTTGTCACCAGCGGGAATCCATCTTGCAGGTTAAAACGCA TCTGATGACCAAAgATGGAcAGCGTTCCGGTTCCGGTACG GTCGTTTTTCTGTGTGCCTTCGTCcAGCACTTTTTGCATC AGTTCcAgATACTGTTTCATtGAAGTGTATATCTCCTTCT TAAAGTTAAACAAAATTATTTCTAGAg	
YKH355	YKH355_Fw_RBS	13/12/2018	51	CATTGAAGTGTATATCTCCTTCTTA	25
YKH356	YKH356_Fw (allele Cys(TGC)16Ile(AUW) blunt-iPCR	13/12/2018	56	CTGGCACCGATwCAtGCATTCTTCCAGTTC	30
YKH357	YKH357_Rv (allele Cys(TGC)16Ile(AUW) blunt-iPCR	13/12/2018	56	CGCCATTTTATCCAGTTCGC	20
YKH358	YKH358_RM01_TS1	13/12/2018	68	GTGCTCGACGAAGGCACACAG	21
YKH359	YKH359_RM02_TS2	13/12/2018	68	CGTAATTAGATAGCCACCGGCG	22
YKH360	YKH360_RM03_EM1	13/12/2018	80	GCGCGCGTGTTGATAGTGCAGTATC	
YKH361	YKH361_RM04_EM2	13/12/2018	84	GCGCGCCCCCGTAGGCGCTAGGG	

YKH362	YKH362_Fw (allele Cys(TGC)) blunt-iPCR	25/01/2019	56	CTGGCACCGTGCCAT	15
YKH363	YKH363_Fw (ThyA_C192)	25/01/2019	52	TGCGATCTGGAAGTGG	16
YKH364	YKH364_Rv	25/01/2019	53	CTGCTGCGCCATCATAT	17
YKH365	YKH365_Fw_tRNAIC_ SapI	25/01/2019	50	aaacagGCTCTTCCtggCCTGAAGAATTTG	30
YKH366	YKH366_Rv_tRNAIC_ SapI	25/01/2019	51	TTTGCTCTTCTCTTGTGC	18
YKH367	YKH367_Fw (allele Cys(TGC)16Ile(AUA) blunt-iPCR	05/02/2019	56	CTGGCACCGATaCAtGCAT	19
YKH368	YKH368_Rv_3' ThyA_GA	05/02/2019	53	CTTTGTTAGCAGCCGGAT	18
YKH369	YKH369_Fw_RBS +ThyA_GA	05/02/2019	53	TAAGAAGGAGATATACACTTCAATGAA	27

YKH370	YKH370_Rv_SapI_ ?HtoL _?EcotRNALeu(CAU)	13/02/2019	53	cggGCTCTTCaCCAATCGCCTCGTTCTTCTAATA	34
YKH371	YKH371_Fw _ErmR_L73	14/02/2019	52	CTGAACATTCGTGTCACTTTA	21
YKH372	YKH372_Rv _Erm_L145	14/02/2019	52	GCAAGAGCAACCCTAGT	17
YKH373	YKH373_LH-HL_37-hisR(GTG>AAG)_38-leuT(CAG>ATG)	08/03/2019		gaaatgaatagttcgacaaaaatCTAGTTTAGTCCCGGCG CTTGAGCTGCGGTGGTAGTAATACCGCGTAACAAGATTTG TAGTGGTGGCTATAGCTCAGTTGGTAGAGCCCTGGATTAA GATTCCAGTTGTCGTGGGTTCGAATCCCATTAGCCACCCC ATTATTAGAAGTTGTGACAATGCGAAGGTGGCGGAATTGG TAGACGCGCTAGCTTATGGTGTTAGTGTCCTTACGGACGT GGGGGTTCAAGTCCCCCCCCTCGCACCACGACTTTAAAGA ATTGAACTAAAAATTCAAAAAGCAGTATTTCGGATATAGT TCCTCCTTTCAG	332

YKH374	YKH374_IC-HL_75-cysT(GCA>TAT)_76-leuZ(TAA>ATG)	08/03/2019		gaaatgaatagttcgacaaaaatCTAGGTTTAAAAGACAT CGGCGTCAAGCGGATGTCTGGCTGAAAGGCCTGAAGAATT TGGCGCGTTAACAAAGCGGTTATGTAGCGGATTtATAATC CGTCTAGTCCGGTTCGACTCCGGAACGCGCCTCCACTTTC TTCCCGAGCCCGGATGGTGGAATCGGTAGACACAAGGGAT TATGAATCCCTCGGCGTTCGCGCTGTGCGGGTTCAAGTCC CGCTCCGGGTACCATGGGAAAGATAAGAATAAAATCAAAG CAATAAGCAGTGTCGTGAAACCACCTTCGGGTGGTTTTTT TGTGCGGATATAGTTCCTCCTTTCAG	346
YKH375	YKH375_LH-IC_37-hisR(GTG>AAG)_75-cysT(GCA>TAT)	08/03/2019		gaaatgaatagttcgacaaaaatCTAGTTTAGTCCCGGCG CTTGAGCTGCGGTGGTAGTAATACCGCGTAACAAGATTTG TAGTGGTGGCTATAGCTCAGTTGGTAGAGCCCTGGATTAA GATTCCAGTTGTCGTGGGTTCGAATCCCATTAGCCACCCC ATTATTAGAAGTTGTGACAATATGTCTGGCTGAAAGGCCT GAAGAATTTGGCGCGTTAACAAAGCGGTTATGTAGCGGAT TtATAATCCGTCTAGTCCGGTTCGACTCCGGAACGCGCCT CCACTTTCTTCCCGACGGATATAGTTCCTCCTTTCAG	317

YKH376	YKH376_Fw (PhiC31_E184_WT)	18/03/2019	53	gctggtttcggagacg	16
YKH377	YKH377_Rv (PhiC31_E184_WT)	18/03/2019	54	Tcgaagccgtaaggcg	16
YKH378	YKH378_Fw_C31.H61 LtoH(CAC)	21/03/2019	57	ttcagcgaagcgccg	15
YKH379	YKH379_Rv_C31.H61 LtoH(CAC)	21/03/2019	58	GTGcccgacgaacctgaac	19
YKH380	YKH380_Fw_37hisR-LH(AAG)	27/03/2019	51	GTAACAAGATTTGTAGTGGTGGct	24
YKH381	YKH381_Rv_ 37hisR-LH(AAG)	27/03/2019	51	GTCACAACTTCTAATAATGGGGTgg	25
YKH382	YKH382_Fw_37hisR-LH(AAG)+HL/IC	27/03/2019	53	aCCCCATTATTAGAAGTTGTGAC	23
YKH383	YKH383_Rv_37hisR-LH(AAG)+HL/IC	27/03/2019	52	CCACCACTACAAATCTTGTTAC	22

YKH384	YKH384_Fw_Pre-tRNAs	27/03/2019	52	gaaatgaatagttcgacaaaaatCTAG	27
YKH385	YKH385_Rv_Post-tRNAs	27/03/2019	53	CTGAAAGGAGGAACTATATCCG	22
YKH386	YKH386_Fw _CAT.A125	10/04/2019	53	gcgtgttacggtgaaaac	18
YKH387	YKH387_Rv _CAT.G217	10/04/2019	53	gccctgccactcatc	15
YKH388	SufI(Tat) signal seq	26/09/2017		TCACTTCACAGGTCaagcttatAGGTCTCAatcgGTCGAC tttacggctagctcagccctaggtattatgctagcCTTAT CATCAAAGAGGGAGCTAAAAatgtcaCTCagtcggcgtca gttcattcaggcatcggggattgcaCNTtgtgcaggcgct gttcccCTGaaggccagcgcaTCCaccatgattacggatt cactg	205

...

Table A.1: Full list of all primers/oligos used within this body of work.

A.3 Oligos used in Darwin Assembly for recoding ΦC31

Listed in Table A.2 are the oligos required to generate a library of leucine-histidine recoded ΦC31 variants as used for designing the basic GTS circuit (Section 4.2.2) and the subsequent library construction (Section 4.2.6). The results of which are shown in Table 4.2 and Table 4.3.

	Oligo ID	**Oligo DNA Sequence (5' to 3')**
(a)	YKH308_*192(200)N(AAT) _I195(203)I(atC) _*196(204)N(AaC)	agtgcgccagtttGTTGatgacgacATTgac cattcgg
(b)	YKH297_H134(142)L(cTT)	cagAAGaatcaggtccatgacgt
	YKH296_L135(143)H(cAT)	ATGgtgaatcaggtccatgac
	YKH298_H134(142)L(cTT) _L135(143)H(cAT)	ATGAAGaatcaggtccatgacgt
(c)	YKH309_H53(61)L(cTT)	cttcgctgaaAAGcccgacgaacct
	YKH310_L75(83)H(cAT)	cggcattcgttATGgatgcgttcg
	YKH311_L109(117)H(cAT) _L110(118)H(cAT) _L112(120)H(cAT)	caatcgtcacgccATGggcATGATGttccga gacaatc
	YKH312_L135(143)H(cAT) _L139(147)H(cAT)	tgcgacgcgtcATGccgcataatATGgtgaat cag
	YKH313_L132(140)H(cAT) _H134(142)L(cTT) _L135(143)H(cAT)	ATGAAGaatATGgtccatgacgtttcc
	YKH314_L132(140)H(cAT) _H134(142)L(cTT) _L135(143)H(cAT) _L139(147)H(cAT)	tgcgacgcgtcATGccgcataatATGAAGaat ATGgtccatgacgtttc
	YKH315_L154(162)H(cAT) _N158(166)N(aaT) _L159(167)H(cAT) _L163(171)H(cAT)	gtacccgccATGttcgcgctgATGAttctt cgtgtcATGaatcttcgc
	YKH295_L460(468)H(cAT)	gcgaacCATgttgcggagcg

Table A.2: Oligo sequence designs that were used for site-directed mutagenesis using the Darwin Assembly method to create the inactive recoded ΦC31 sequence variants

	Oligo ID	**Oligo DNA Sequence (5' to 3')**
(d)	YKH304_Boundary-Fw _5Biotin_SapI	/5BiotinTEG/CCCCTTATTAGCGTTTGCCA GCTCTTCt'aag,aGCTTCACAGGTCAAGTCC GTTCATTG
	YKH305_Boundary-Rv _3invdT_SapI	GAAGTGTATATCTCCTTTCTAGATTTTTGTCG AACTATTCATTTc'act,tGAAGAGCCATCTG ACCTCTGTGCTGCT/3invdT/
	YKH276 = YKH163_Outnest1	CCCCTTATTAGCGTTTGCCA
	YKH277 = YKH164_Outnest2	AGCAGCACAGAGGTCAGATG
	YKH306_pCtsP_pNGAL97 -Fw_SapI	gagtcaGCTCTTCc'act,tTTCTCTATCACT GATAGGGAGTG
	YKH307_pCtsP_pNGAL97 -Rv_SapI	gagtcaGCTCTTCA'ctt,cGTGATTGTCTGC CGTTTAC
	YKH318_pC-sfGFP-flip _pNGAL97-Rv_SapI	gagtcaGCTCTTCA'ctt,cGTGAAAAATGGC GCACAT
(e)	YKH319_C31-R225_Rv	Cggattacgtcgggctc
	YKH320_C31-R225_Fw	gtggtggtggcgtga
	YKH321_C31-R225_Fw_SapI	aatGCTCTTCtccGgTGGTGGTGGCGT
	YKH324_C31-R225_Rv_SapI	aatGCTCTTCaCggaTTACGTCGGGCTCGAAC
	YKH325_C31-G54_Rv_SapI	aatGCTCTTCtAtggTCCATGACGTTTCCC
	YKH326_C31-G54_Fw_SapI	ttaGCTCTTCccaTaTTCACCTGATTATGCGG CatGACG
	YKH327_C31-L359_Fw	gggCTGtcacgggggcaag
	YKH328_C31-L359_Rv	cttgccgcgcccccct
	YKH329_L468H_Fw_SapI	tgtGCTCTTCtcgaacCATgttgcggag
	YKH329_v2_L468H_Fw_SapI _[v2:sfGFP_pNGAL97]	tgtGCTCTTCtcgaacCATgttgcggagcgcg
	YKH330_L468H_Rv_SapI	aatGCTCTTCttcgcccgttcgccg
(f)	YKH289_Rv_(DA_primingsite)	GCAATGAACGGACTTGACCT

Table A.2: (Continued) Oligo sequence designs that were used for site-directed mutagenesis using the Darwin Assembly method to create the inactive recoded ΦC31 sequence variants

Appendix B

Flow Cytometry Scripts

These Python scripts, using the FlowCal library package, were developed with the invaluable support of Dr. Charmian Dawson, and were used for analysing the flow cytometry data from Section 4.2.8. Charmian wrote the Python scripts in Jupyter Notebooks, executed in Google Colab, and provided advice on the analyses; I adapted the scripts for my data.

```
# (a) Importing FlowCal package and sourcing data:
# using the FlowCal library to analyse the .fcs files
!pip install FlowCal===1.3.0

import FlowCal
import matplotlib.pyplot as plt

from google.colab import drive
drive.mount('/content/drive')


# (b) Importing data and generating histograms and gated plots:
savepath = "/content/drive/MyDrive/Colab Notebooks/yan-kay_scripts/"
rootpath = "/content/drive/MyDrive/yan-kay/AttuneNxT/190523_Flow1/"
folders = ['Flow1_TRep1/190523_sfGFP_',
           'Flow1_TRep2/Results_Export/190523_sfGFP_TRep2_']
```

```
files = ['190523_sfGFP_PBS_d1000_D2_PBS',
        '190523_sfGFP_PBS_d1000_D3_NEG_C31H53L_ThyA_ATA_10b',
        '190523_sfGFP_PBS_d1000_D4_NEG_C31H53L_ThyA_WT_10b',
        '190523_sfGFP_PBS_d1000_D5_POS_C31WT_sfGFPwt_10b',
        '190523_sfGFP_PBS_d1000_D6_C31H53L_sfGFPwt_10b',
        '190523_sfGFP_PBS_d1000_D7_asD6_LHHLold_noATC_10b',
        '190523_sfGFP_PBS_d1000_D8_asD6_LHHLold_ATC20_10b',
        '190523_sfGFP_PBS_d1000_D9_asD6_LHHLold_ATC200_10b']
fc = ".fcs"
ss = []
#a list of all the titles
titls = ['','',"NEG: C31.H53l_ThyA.WT","POS: C31.WT_sfGFP.WT","D6:
    C31.H53L_sfGFP.WT","D7: C31.H53L_sfGFP.WT + LH_HL // -aTc 20
    ng/mL", "D8: C31.H53L_sfGFP.WT + LH_HL // +aTc 20 ng/mL"]
for i in folders:#goes through both folders and processes the data
    and puts it in the ss list above
  for j in files:
    filename = rootpath+i+j+fc
    #print(j)
    s = FlowCal.transform.to_rfi(FlowCal.io.FCSData(filename))
    ss.append(s)
    #print(FlowCal.trans)
    #FlowCal.plot.hist1d(s,6,'log',histtype='step',xlim=[1,1e5])
    #plt.show()
#print(s.channels)#('Time', 'FSC-A', 'SSC-A', 'BL1-A', 'FSC-H',
    'SSC-H', 'BL1-H', 'FSC-W', 'SSC-W', 'BL1-W')
neg = 2#just helping myself know which numbers in the ss list refer
    to which datasets
pos = 3
d6 = 4
d7 = 5
d8 = 6#does not look like the graph in the poster, but the poster
```

```
    histogram has clearly been gated, while the others haven't which
    is a bit weird...
d9 = 7
xlabl = "GFP fluorescence intensity, log10 (BL1-H)"
xlabl2 = "BL1-H"
#these two graphs are one-dimensional histograms of the two graphs
    with multiple datasets on each
FlowCal.plot.hist1d([ss[neg],ss[pos],ss[d6]],6,'log',500,
    histtype='step', xlim=[1,500000],
    edgecolor=['gray','limegreen','cornflowerblue'], xlabel=xlabl)
##above - the datasets to plot, 6=BL1-H, log is the x axis, 500 = the
    bin size, step means lines but not filled, xlim is the limits of
    the x axis, then the colours for each dataset, and xlabl is the
    label for the x axis (a few lines up)
#plt.savefig(savepath+"graph1.jpg")#savepath is defined above, and
    the name of this graph will be graph1.jpg
plt.show()
print('medians - gray: ', FlowCal.stats.median(ss[neg],6),', green:
    ', FlowCal.stats.median(ss[pos],6), ', blue: ',
    FlowCal.stats.median(ss[d6],6))
FlowCal.plot.hist1d([ss[d6],ss[d7],ss[d8]],6,'log',500,
    histtype='step', xlim=[1e2,1e5],
    edgecolor=['cornflowerblue','darkblue','red'], xlabel=xlabl)
#plt.savefig(savepath+"graph2.jpg")
plt.show()
print('medians - blue: ', FlowCal.stats.median(ss[d6],6),', darkblue:
    ', FlowCal.stats.median(ss[d7],6), ', red: ',
    FlowCal.stats.median(ss[d8],6))


ylabl = 'Count'#the default for the y axis is counts
#the info below is for positioning the blue and green lines on the
    smaller graphs
```

```
ypoints = [0,200]
xpoint = 1.08
xpoint2 = 1e3
xpoint3 = 1e6
xpoints = [xpoint,xpoint]
xpoints2 = [xpoint2,xpoint2]
xpoints3 = [xpoint3,xpoint3]

#this function is used to plot graphs all in the same way - ds is a
    number that is the position in the ss list
def makeSecondGraphType(ds):
  FlowCal.plot.hist1d(ss[ds],6,'log',1000, histtype='step',
      ylim=[0,200], edgecolor='tab:orange', xlabel=xlabl2, ylabel =
      ylabl)
  lows = FlowCal.gate.high_low(ss[ds],6,high=1e3)#this line counts
      how many events are below the 1e3 threshold.
  blueperc = round((lows.shape[0]/ss[ds].shape[0])*100,3)#calculates
      the percentage of events below the threshold.
  bluestr = str(blueperc) + "%"
  greenstr = str(round(100-blueperc,3)) + "%" #works out the rest of
      the events as a percentage
  #these lines plot the green and blue lines
  plt.plot(xpoints,ypoints,color="tab:cyan")
  plt.plot([xpoint,xpoint2],[50,50],color="tab:cyan")
  plt.annotate(' NA: '+bluestr,xy=(1,53), weight='bold', color =
      "tab:cyan")
  plt.plot(xpoints2,ypoints,color="tab:cyan")
  plt.plot(xpoints3,ypoints,color="limegreen")
  plt.annotate(' GFP: '+greenstr,xy=(1e3,153), weight='bold', color =
      "limegreen")
  plt.plot([xpoint2,xpoint3],[150,150],color="limegreen")
```

```
    plt.title(titls[ds])#selects the appropriate title from the list
    #plt.savefig(savepath+"graph_small_"+str(ds)+".jpg")#saves the
        figure
    plt.show()
    print('medians: ',FlowCal.stats.median(ss[ds],6))

#this plots the graphs from the ss list above, from 2 to 6 inclusive
for i in range(2,7):
    makeSecondGraphType(i)


# (c) Comparing data between the technical replicates:
reps = ['rep1','rep2']
for i in range(len(files)):
    filename1 =
        "/content/drive/MyDrive/yan-kay/AttuneNxT/190523_Flow1/"
        +folders[0]+files[i]+".fcs"
    filename2 =
        "/content/drive/MyDrive/yan-kay/AttuneNxT/190523_Flow1/"
        +folders[1]+files[i]+".fcs"
    print(filename1)
    s1 = FlowCal.transform.to_rfi(FlowCal.io.FCSData(filename1))
    s2 = FlowCal.transform.to_rfi(FlowCal.io.FCSData(filename2))
    FlowCal.plot.hist1d([s1,s2],6,'log',histtype='step')
    plt.legend(reps)
    plt.show()
    FlowCal.plot.density2d(s1,[1,2],mode='scatter',
        xscale='log',yscale='log')
    plt.show()
    FlowCal.plot.density2d(s2,[1,2],mode='scatter',
        xscale='log',yscale='log')
    plt.show()
```

Bibliography

Hui-Wang Ai, Nathan C Shaner, Zihao Cheng, Roger Y Tsien, and Robert E Campbell. Exploration of New Chromophore Structures Leads to the Identification of Improved Blue Fluorescent Proteins. *Biochemistry*, 46:5904–5910, 2007. doi: 10.1021/BI700199G.

O. Allner and L. Nilsson. Nucleotide modifications and tRNA anticodon-mRNA codon interactions on the ribosome. *RNA*, 17(12):2177–2188, 2011. ISSN 1355-8382. doi: 10.1261/rna.029231.111.

J Christopher Anderson, Thomas J Magliery, and Peter G Schultz. Exploring the Limits of Codon and Anticodon Size. *Chemistry & Biology*, 9(2):237––244, 2001. doi: 10.1016/S1074-5521(02)00094-7. URL `https://www.sciencedirect.com/science/article/pii/S1074552102000947`.

J Christopher Anderson, Ning Wu, Stephen W Santoro, Vishva Lakshman, S King, Peter G Schultz, and David S King. An expanded codon quadruplet code with a functional. *Proceedings of the National Academy of Sciences*, 101(20):7566–7571, 2004. ISSN 0027-8424. doi: 10.1073/pnas.0401517101.

Andreas I Andreou and Naomi Nakayama. Mobius Assembly: A versatile Golden-Gate framework towards universal DNA assembly. *PLoS ONE*, 13(1):e0189892, 2018. doi: https://doi.org/10.1371/journal.pone.0189892. URL `https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0189892`.

J G Arnez, D C Harris, A Mitschler, B Rees, C S Francklyn, and D Moras. Crys-

tal structure of histidyl-tRNA synthetase from *Escherichia coli* complexed with histidyl-adenylate. *The EMBO journal*, 14(17):4143–4155, 1995. ISSN 0261-4189. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC394497/?page=6.

Marcello Barbieri. Evolution of the genetic code: The ambiguity-reduction theory. *BioSystems*, 185(104024):1–7, 2019. doi: 10.1016/j.biosystems.2019.104024. URL https://pubmed.ncbi.nlm.nih.gov/31499094/.

Steven H Bass and Daniel G Yansura. Application of the *E. coli trp* promoter. *Molecular Biotechnology*, 16:253–260, 2000. doi: https://doi.org/10.1385/MB:16:3:253. URL https://link.springer.com/article/10.1385/MB:16:3:253.

Matthew D Berg and Christopher J Brandl. Transfer RNAs: diversity in form and function. *RNA Biology*, 18(3):316–339, 2021. doi: 10.1080/15476286.2020.1809197. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7954030/.

Matthew J. Betts and Robert B. Russell. Amino-Acid Properties and Consequences of Substitutions. *Bioinformatics for Geneticists: A Bioinformatics Primer for the Analysis of Genetic Data: Second Edition*, 4:311–342, 2003. ISSN 0009-9147. doi: 10.1002/9780470059180.ch13. URL https://onlinelibrary.wiley.com/doi/10.1002/9780470059180.ch13.

Ana R Bezerra, João Simões, Wanseon Lee, Johan Rung, Tobias Weil, Ivo G Gut, Marta Gut, Mónica Bayés, Lisa Rizzetto, Duccio Cavalieri, Gloria Giovannini, Silvia Bozza, Luigina Romani, Misha Kapushesky, Gabriela R Moura, and Manuel a S Santos. Reversion of a fungal genetic code alteration links proteome instability with genomic and phenotypic diversification. *Proceedings of the National Academy of Sciences of the United States of America*, 110(27):11079–11084, 2013. ISSN 1091-6490. doi: 10.1073/pnas.1302094110. URL https://www.pnas.org/content/early/2013/06/12/1302094110.

Ana R Bezerra, Ana R Guimarães, and Manuel A S Santos. Non-Standard Genetic Codes Define New Concepts for Protein Engineering. *Life (Basel, Switzerland)*, 5(4):1610–1628, 2015. ISSN 2075-1729. doi: 10.3390/life5041610. URL http://www.mdpi.com/2075-1729/5/4/1610/htm.

Wil Biddle, Margaret A Schmitt, and John D Fisk. Modification of orthogonal tRNAs: unexpected consequences for sense codon reassignment. *Nucleic Acids Research*, pages 1–9, 2016. ISSN 1759-6653. doi: 10.1093/gbe/evw245. URL http://fdslive.oup.com/www.oup.com/pdf/production_in_progress.pdf.

G R Björk, J U Ericson, C E Gustafsson, T G Hagervall, Y H Jönsson, and P M Wikström. Transfer RNA modification. *Annual review of biochemistry*, 56:263–287, 1987. ISSN 00664154. doi: 10.1146/annurev.biochem.56.1.263.

Frederick R Blattner, Guy Plunkett III, Craig A Bloch, Nicole T Perna, Valerie Burland, Monica Riley, Julio Collado-Vides, Jeremy D Glasner, Christopher K Rode, George F Mayhew, Jason Gregor, Nelson Wayne Davis, Heather A Kirkpatrick, Michael A Goeden, Debra J Rose, Bob Mau, and Ying Shao. The Complete Genome Sequence of Escherichia coli K-12. *Science*, 277(5331):1453–1462, 1997. doi: 10.1126/science.277.5331.1453. URL https://pubmed.ncbi.nlm.nih.gov/9278503/.

Michal T Boniecki and Susan A Martinis. Coordination of tRNA synthetase active sites for chemical fidelity. *Journal of Biological Chemistry*, 287(14):11285–11289, 2012. ISSN 00219258. doi: 10.1074/jbc.C111.325795.

Jerome Bonnet, Peter Yin, Monica E Ortiz, Pakpoom Subsoontorn, and Drew Endy. Amplifying genetic logic gates. *Science*, 340(6132):599–603, 2013. ISSN 1095-9203. doi: 10.1126/science.1232758. URL http://www.ncbi.nlm.nih.gov/pubmed/23539178.

Kristin Breitschopf, Tilmann Achsel, Kristina Busch, and Hans J. Gross. Identity elements of human tRNALeu: Structural requirements for converting hu-

man tRNASer into a leucine acceptor in vitro. *Nucleic Acids Research*, 23 (18):3633–3637, 1995. ISSN 03051048. doi: 10.1093/nar/23.18.3633. URL https://pubmed.ncbi.nlm.nih.gov/7478989/.

Jürgen Brosius, Mary Erfle, and John Storella. Spacing of the -10 and -35 regions in the tac promoter. Effect on its in vivo activity. *The Journal of Biological Chemistry*, 260(6):3539–3541, 1985. URL https://pubmed.ncbi.nlm.nih.gov/2579077/.

Geraldine Butler, Matthew D Rasmussen, Michael F Lin, Manuel A Santos, Sharadha Sakthikumar, Carol A Munro, Esther Rheinbay, Manfred Grabherr, Anja Forche, Jennifer L Reedy, Ino Agrafioti, Martha B Arnaud, Steven Bates, Alistair J P Brown, Sascha Brunke, Maria C Costanzo, David A Fitzpatrick, Piet W de Groot, David Harris, Lois L Hoyer, Bernhard Hube, Frans M Klis, Chinnappa Kodira, Nicola Lennard, Mary E Logue, Ronny Martin, Aaron M Neiman, Elissavet Nikolaou, Michael A Quail, Janet Quinn, Maria C Santos, Florian F Schmitzberger, Gavin Sherlock, Prachi Shah, Kevin A Silverstein, Marek S Skrzypek, David Soll, Rodney Staggs, Ian Stansfield, Michael P Stumpf, Peter E Sudbery, Thyagarajan Srikantha, Qiandong Zeng, Judith Berman, Matthew Berriman, Joseph Heitman, Neil A R Gow, Michael C Lorenz, Bruce W Birren, Manolis Kellis, and Christina A Cuomo. Evolution of pathogenicity and sexual reproduction in eight Candida genomes. *Nature*, 459(7247):657–662, 2009. ISSN 0028-0836. doi: 10.1038/nature08064. URL http://www.ncbi.nlm.nih.gov/pubmed/19465905.

R E Campbell and M W Davidson. Fluorescent Reporter Proteins. *Molecular Imaging with Reporter Genes*, page 1, 2010. ISSN 0521882338. doi: 10.1128/AEM.00701-10. URL http://books.google.com/books?hl=en&lr=&id=oVqH-OENT44C&oi=fnd&pg=PA3&dq=Fluorescent+Repoter+Proteins&ots=aNPE_N_Nop&sig=3P4Arivnz3OAl7dNqcDG1KuWmOI%5Cnpapers3://publication/uuid/07C0E190-00D5-4DBA-95AA-F27062906F5E.

Charles W. Carter and William L. Duax. Did tRNA synthetase classes arise on

opposite strands of the same gene? *Molecular Cell*, 10(4):705–708, 2002. ISSN 10972765. doi: 10.1016/S1097-2765(02)00688-3.

Charles W. Carter and Richard Wolfenden. tRNA acceptor stem and anticodon bases form independent codes related to protein folding. *Proceedings of the National Academy of Sciences*, 2015(19):201507569, 2015. ISSN 0027-8424. doi: 10.1073/pnas.1507569112. URL http://www.pnas.org/lookup/doi/10.1073/pnas.1507569112.

Francesca Ceroni, Rhys Algar, Guy-Bart Stan, and Tom Ellis. Quantifying cellular capacity identifies gene expression designs with reduced burden. *Nature Methods*, 12(5):415–418, 2015. doi: 10.1038/nmeth.3339. URL https://www.nature.com/articles/nmeth.3339.

Santiago Chaillou, Pinelopi-Eleftheria Stamou, Leticia L Torres, Ana B Riesco, Warren Hazelton, and Vitor B Pinheiro. Directed evolution of colE1 plasmid replication compatibility: a fast tractable tunable model for investigating biological orthogonality. *Nucleic Acids Research*, 50(16):9568–9579, 2022. doi: 10.1093/nar/gkac682. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9458437/.

Anargyros Chaliotis, Panayotis Vlastaridis, Dimitris Mossialos, Michael Ibba, Hubert D Becker, Constantinos Stathopoulos, and Grigorios D Amoutzias. The complex evolutionary history of aminoacyl-tRNA synthetases. *Nucleic Acids Research*, 43(3):1059–1068, 2017. doi: 10.1093/nar/gkw1182. URL https://academic.oup.com/nar/article/45/3/1059/2605964.

Robert W Chambers. On the Recognition of tRNA by Its Aminoacyl-tRNA Liqase. In *Progress in Nucleic Acid Research and Molecular Biology*, volume 11, pages 489–525. Academic Press, 1971. doi: 10.1016/s0079-6603(08)60336-0. URL https://pubmed.ncbi.nlm.nih.gov/4934251/.

Patricia P Chan and Todd M Lowe. GtRNAdb: A database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Research*, 37:D93–D97, 2009.

ISSN 03051048. doi: 10.1093/nar/gkn787. URL https://academic.oup.com/nar/article/37/suppl_1/D93/1010599.

Patricia P Chan and Todd M Lowe. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Research*, 44 (D1):D184–D189, 2016. doi: 10.1093/nar/gkv1309. URL https://academic.oup.com/nar/article/44/D1/D184/2503100.

Abhishek Chatterjee, Sophie B. Sun, Jennifer L. Furman, Han Xiao, and Peter G. Schultz. A versatile platform for single- and multiple-unnatural amino acid mutagenesis in escherichia coli. *Biochemistry*, 52(10), 2013. ISSN 00062960. doi: 10.1021/bi4000244. URL https://pubs.acs.org/doi/10.1021/bi4000244.

Abhishek Chatterjee, Marc J Lajoie, Han Xiao, George M Church, and Peter G Schultz. A bacterial strain with a unique quadruplet codon specifying non-native amino acids. *ChemBioChem*, 15(12):1782–1786, 2014. ISSN 14397633. doi: 10.1002/cbic.201402104.

Jason W Chin, T Ashton Cropp, Stephanie Chu, Eric Meggers, and Peter G Schultz. Progress Toward an Expanded Eukaryotic Genetic Code. *Chemistry & Biology*, 10:511–519, 2003. ISSN 0960-9822. doi: 10.1016/S.

Shaileja Chopra and John Reader. tRNAs as antibiotic targets. *International Journal of Molecular Sciences*, 16(1):321–349, 2015. ISSN 14220067. doi: 10.3390/ijms16010321.

Dmitriy M Chudakov, Mikhail V Matz, Sergey Lukyanov, and Konstantin A Lukyanov. Fluorescent Proteins and Their Applications in Imaging Living Cells and Tissues. *Physiological Reviews*, 90(3):1103–1163, 2010. ISSN 00319333. doi: 10.1152/physrev.00038.2009. URL http://physrev.physiology.org/content/90/3/1103.short.

Brendan P Cormack, Raphael H Valdivia, and Stanley Falkow. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene*, 173(1):33–38, 1996.

doi: 10.1016/0378-1119(95)00685-0. URL https://www.sciencedirect.com/science/article/pii/0378111995006850.

Christopher Cozens and Vitor B Pinheiro. Darwin Assembly: fast, efficient, multi-site bespoke mutagenesis . *Nucleic Acids Research*, 46(8):e51, 2018. doi: 10.1093/nar/gky067. URL https://pubmed.ncbi.nlm.nih.gov/29409059/.

Christopher Cozens, Vitor B Pinheiro, Alexandra Vaisman, Roger Woodgate, and Philipp Holliger. A short adaptive path from DNA to RNA polymerases. *Proceedings of the National Academy of Sciences of the United States of America*, 109(21):8067–72, may 2012. ISSN 1091-6490. doi: 10.1073/pnas.1120964109. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3361454&tool=pmcentrez&rendertype=abstract.

Andreas Crameri, Erik A Whitehorn, Emily Tate, and Willem P C Stemmer. Improved Green Fluorescent Protein by Molecular Evolution Using DNA Shuffling. *Nature Biotechnology*, 14:315–319, 1996. doi: 10.1038/nbt0396-315. URL https://www.nature.com/articles/nbt0396-315.

Paula J Cranfill, Brittney R Sell, Michelle A Baird, John R Allen, Zeno Lavagnino, H Martijn de Gruiter, Gert-Jan Kremers, Michael W Davidson, Alessandro Ustione, and David W Piston. Quantitative assessment of fluorescent proteins. *Nature Methods*, 13:557–562, 1996. doi: 10.1038/nmeth.3891. URL https://www.nature.com/articles/nmeth.3891.

Donald M Crothers, Takahiro Seno, and Dieter Söll. Is There a Discriminator Site in Transfer RNA? *Proceedings of the National Academy of Sciences*, 69(10):3063–3067, 1972. doi: 10.1073/pnas.69.10.3063. URL https://www.pnas.org/doi/abs/10.1073/pnas.69.10.3063.

Andrew B Cubitt, Leslie A Woollenweber, and Roger Heim. Chapter 2: Understanding Structure—Function Relationships in the Aequorea victoria Green Fluorescent Protein. In Kevin F. Sullivan and Steve A. Kay, editors, *Green Fluorescent Proteins*, volume 58 of *Methods in Cell Biology*, pages 19–30. Academic Press,

1998. doi: https://doi.org/10.1016/S0091-679X(08)61946-9. URL https://www.sciencedirect.com/science/article/pii/S0091679X08619469.

James F Curran. *Modified Nucleosides in Translation*, chapter 27, pages 493–516. John Wiley & Sons, Ltd, 1998. ISBN 9781683672685. doi: https://doi.org/10.1128/9781555818296.ch27. URL https://onlinelibrary.wiley.com/doi/abs/10.1128/9781555818296.ch27.

Richard N Day and Michael W Davidson. The fluorescent protein palette: tools for cellular imaging. *Chemical Society reviews*, 38(10):2887–921, 2009. ISSN 1460-4744. doi: 10.1039/b901966a. URL https://pubs.rsc.org/en/content/articlelanding/2009/CS/b901966a.

Herman A de Boer, Lisa J Comstock, and Mark Vasser. The *tac* promoter: A functional hybrid derived from the *trp* and *lac* promoters. *Proceedings of the National Academy of Sciences*, 80(1):21–25, 1983. doi: https://doi.org/10.1073/pnas.80.1.21. URL https://www.pnas.org/doi/abs/10.1073/pnas.80.1.21.

Daniel de la Torre. *Rewriting the genome of Escherichia coli*. PhD thesis, University of Cambridge, Cambridge, UK, March 2020.

A C Dock-Bregeon, R Sankaranarayanan, P Romby, J Caillet, M Springer, B Rees, C S Francklyn, C Ehresmann, and D Moras. Transfer RNA-mediated editing in threonyl-tRNA synthetase: The class II solution to the double discrimination problem. *Cell*, 103(6):877–884, 2000. ISSN 00928674. doi: Doi10.1016/S0092-8674(00)00191-4.

Hengjing Dong, Lars Nilsson, and Charles G Kurland. Co-variation of tRNA Abundance and Codon Usage in *Escherichia coli* at Different Growth Rates. *Journal of molecular biology*, 260(5):649–663, 1996. ISSN 0022-2836. doi: 10.1006/jmbi.1996.0428. URL http://www.sciencedirect.com/science/article/pii/S0022283696904283.

Volker Döring and Philippe Marlière. Reassigning cysteine in the genetic code of Escherichia coli. *Genetics*, 150(2):543–551, 1998. ISSN 00166731. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1460352/.

Morana Dulic, Nevena Cvetesic, John J. Perona, and Ita Gruic-Sovulj. Partitioning of tRNA-dependent editing between pre- and post-transfer pathways in class I aminoacyl-tRNA synthetases. *Journal of Biological Chemistry*, 285(31):23799–23809, 2010. ISSN 00219258. doi: 10.1074/jbc.M110.133553.

Morana Dulic, John J. Perona, and Ita Gruic-Sovulj. Determinants for tRNA-dependent pretransfer editing in the synthetic site of isoleucyl-tRNA synthetase. *Biochemistry*, 53(39):6189–6198, 2014. ISSN 15204995. doi: 10.1021/bi5007699.

Daniel L Dunkelmann, Sebastian B Oehm, Adam T Beattie, and Jason W Chin. A 68-codon genetic code to incorporate four distinct non-canonical amino acids enabled by automated orthogonal mRNA design. *Nature Chemistry*, 13:1110––1117, 2021. doi: 10.1038/s41557-021-00764-5. URL https://www.nature.com/articles/s41557-021-00764-5.

Özlem Doğan Ekici, Mark Paetzel, and Ross E Dalbey. Unconventional serine proteases: Variations on the catalytic Ser/His/Asp triad configuration. *Protein Science*, 17(12):2023–2037, 2008. doi: 10.1110/ps.035436.108. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2590910/.

A R Fersht and M M Kaethner. Enzyme hyperspecificity. Rejection of threonine by the valyl-tRNA synthetase by misacylation and hydrolytic editing. *Biochemistry*, 15(15):3342–3346, 1976. ISSN 0006-2960. doi: DOI10.1021/bi00660a026. URL http://www.ncbi.nlm.nih.gov/pubmed/182209.

C Francklyn and P Schimmel. Enzymatic aminoacylation of an eight-base-pair microhelix with histidine. *Proceedings of the National Academy of Sciences of the United States of America*, 87(21):8655–9, 1990. ISSN 0027-8424. doi: 10.1073/pnas.87.21.8655. URL

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=55016&tool=pmcentrez&rendertype=abstract.

Julius Fredens, Kaihang Wang, Daniel de la Torre, Louise F H Funke, Wesley E Robertson, Yonka Christova, Tiongsun Chia, Wolfgang H Schmied, Daniel L Dunkelmann, Václav Beránek, Chayasith Uttamapinant, Andres Gonzalez Llamazares, Thomas S Elliott, and Jason W Chin. Total synthesis of *Escherichia coli* with a recoded genome. *Nature*, 569:514–518, 2019. doi: 10.1038/s41586-019-1192-5. URL https://doi.org/10.1038/s41586-019-1192-5.

Angela Wai Shan Fung, Charles Chung Yun Leung, and Richard Peter Fahlman. The determination of tRNA Leu recognition nucleotides for Escherichia coli L/F transferase. *RNA*, 20(8):1210–1222, 2014. doi: 10.1261/rna.044529.114. URL https://rnajournal.cshlp.org/content/20/8/1210.

Richard Giegé, Marie Sissler, and Catherine Florentz. Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Research*, 26(22):5017–35, 1998. ISSN 0305-1048. doi: 10.1093/nar/26.22.5017. URL http://www.ncbi.nlm.nih.gov/pubmed/9801296.

Richard Giegé, Frank Jühling, Joern Pütz, Peter Stadler, Claude Sauter, and Catherine Florentz. Structure of transfer RNAs: similarity and variability. *Wiley Interdisciplinary Reviews: RNA*, 3(1):37–61, 2012. ISSN 17577004. doi: 10.1002/wrna.103.

Richard Giegé, Gilbert Eriani, Richard Giegé, and Gilbert Eriani. Transfer RNA Recognition and Aminoacylation by Synthetases. *eLS*, pages 1–18, 2014. doi: 10.1002/9780470015902.a0000531.pub3. URL http://doi.wiley.com/10.1002/9780470015902.a0000531.pub3.

Ana C Gomes, Isabel Miranda, Raquel M Silva, Gabriela R Moura, Benjamin Thomas, Alexandre Akoulitchev, and Manuel A S Santos. A genetic code alteration generates a proteome of high diversity in the human pathogen Candida albicans. *Genome Biology*, 8(10):1–15, 2007. doi: 10.1186/gb-2007-8-10-r206.

Henri Grosjean, Claude Houssier, Pascale Romby, and Roland Marquet. *Modulatory Role of Modified Nucleotides in RNA Loop-Loop Interaction*, chapter 7, pages 113–133. John Wiley & Sons, Ltd, 1998. ISBN 9781683672685. doi: https://doi.org/10.1128/9781555818296.ch7. URL https://onlinelibrary.wiley.com/doi/abs/10.1128/9781555818296.ch7.

Amy C Groth and Michele P Calos. Phage Integrases: Biology and Applications. *Journal of Molecular Biology*, 335(3):667–678, 2004. doi: 10.1016/j.jmb.2003.09.082. URL https://linkinghub.elsevier.com/retrieve/pii/S0022283603013561.

Ita Gruic-Sovulj, Jasmina Rokov-Plavec, and Ivana Weygand-Durasevic. Hydrolysis of non-cognate aminoacyl-adenylates by a class II aminoacyl-tRNA synthetase lacking an editing domain. *FEBS Letters*, 581(26):5110–5114, 2007. ISSN 00145793. doi: 10.1016/j.febslet.2007.09.058.

Li-Tao Guo, Yane-Shih Wang, Akiyoshi Nakamura, Daniel Eiler, Jennifer M Kavran, Margaret Wong, Laura L Kiessling, Thomas A Steitz, Patrick O'Donoghue, and Dieter Söll. Polyspecific pyrrolysyl-tRNA synthetases from directed evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 111(47):16724–9, 2014. ISSN 1091-6490. doi: 10.1073/pnas.1419737111. URL http://www.ncbi.nlm.nih.gov/pubmed/25385624.

Min Guo, Paul Schimmel, and Xiang Lei Yang. Functional expansion of human tRNA synthetases achieved by structural inventions. *FEBS Letters*, 584(2):434–442, 2010. ISSN 00145793. doi: 10.1016/j.febslet.2009.11.064. URL http://dx.doi.org/10.1016/j.febslet.2009.11.064.

Andrew Hadd and John J Perona. Coevolution of specificity determinants in eukaryotic glutamyl- and glutaminyl-tRNA synthetases. *Journal of Molecular Biology*, 2014. ISSN 1089-8638. doi: 10.1016/j.jmb.2014.08.006. URL http://www.ncbi.nlm.nih.gov/pubmed/25149203.

Fumio Harada and Susumu Nishimura. Possible anticodon sequences of tRNAHis,

tRNAAsn, and tRNAAsp from Escherichia coli. Universal presence of nucleoside Q in the first position of the anticodons of these transfer ribonucleic acid. *Biochemistry*, 11(2):301–308, 1972. doi: 10.1021/bi00752a024. URL https://pubs.acs.org/doi/abs/10.1021/bi00752a024.

Roland K Hartmann, Markus Gößringer, Bettina Späth, Susan Fischer, and Anita Marchfelder. Chapter 8 The Making of tRNAs and More - RNase P and tRNase Z. *Progress in Molecular Biology and Translational Science*, 85(C):319–368, 2009. ISSN 18771173. doi: 10.1016/S0079-6603(08)00808-8.

Sanchita Hati, Brigitte Ziervogel, Julius Sternjohn, Fai Chu Wong, Maria C. Nagan, Abbey E. Rosen, Paul G. Siliciano, Joseph W. Chihade, and Karin Musier-Forsyth. Pre-transfer editing by class II prolyl-tRNA synthetase: Role of aminoacylation active site in "selective release" of noncognate amino acids. *Journal of Biological Chemistry*, 281(38):27862–27872, 2006. ISSN 00219258. doi: 10.1074/jbc.M605856200.

Scott Hauenstein, Chun-Mei Zhang, Ya-Ming Hou, and J John Perona. Shape-selective RNA recognition by cysteinyl-tRNA synthetase. *Nature Structural & Molecular Biology*, 11:1134–1143, 2004. doi: 10.1038/nsmb849. URL https://www.nature.com/articles/nsmb849.

Diane K Hawley and William R McClure. Compilation and analysis of Escherichia coli promoter DNA sequences. *Nucleic Acids Research*, 11(8):2237–2255, 1983. doi: 10.1093/nar/11.8.2237. URL https://pubmed.ncbi.nlm.nih.gov/6344016/.

Roger Heim and Roger Y Tsien. Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Current Biology*, 6(2):178–182, 1996. doi: 10.1016/s0960-9822(02)00450-5. URL https://www.cell.com/current-biology/fulltext/S0960-9822(02)00450-5.

Roger Heim, Douglas C Prasher, and Roger Y Tsien. Wavelength mutations and

posttranslational autoxidation of green fluorescent protein. *Proceedings of the National Academy of Sciences*, 91(26):12501–12504, 1994. ISSN 0027-8424. doi: 10.1073/pnas.91.26.12501. URL http://apps.webofknowledge.com.proxy1.lib.uwo.ca/full_record.do?product=WOS&search_mode=GeneralSearch&qid=18&SID=2BZ4h2AnJwmLzAAoHCu&page=1&doc=6.

S Henikoff and J G Henikoff. Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences*, 89:10915–10919, 1992. doi: 10.1073/pnas.89.22.10915. URL https://www.pnas.org/content/89/22/10915.

Stephanie Herring, Alexandre Ambrogelly, Sarath Gundllapalli, Patrick O'Donoghue, Carla R Polycarpo, and Dieter Söll. The amino-terminal domain of pyrrolysyl-tRNA synthetase is dispensable in vitro but required for in vivo activity. *FEBS Letters*, 581(17):3197–3203, 2007. ISSN 00145793. doi: 10.1016/j.febslet.2007.06.004.

Hyouta Himeno, Shukuko Yoshida, Akiko Soma, and Kazuya Nishikawa. Only one nucleotide insertion to the long variable arm confers an efficient serine acceptor activity upon Saccharomyces cerevisiae tRNA(Leu) in vitro. *Journal of Molecular Biology*, 268(4):704–711, 1997. ISSN 0022-2836. doi: 10.1006/jmbi.1997.0991.

Joanne M Ho, Noah M Reynolds, Keith Rivera, Morgan Connolly, Li Tao Guo, Jiqiang Ling, Darryl J Pappin, George M Church, and Dieter Söll. Efficient Reassignment of a Frequent Serine Codon in Wild-Type Escherichia coli. *ACS Synthetic Biology*, 5(2):163–171, 2016. ISSN 21615063. doi: 10.1021/acssynbio.5b00197. URL https://pubs.acs.org/doi/10.1021/acssynbio.5b00197.

Anita K Hopper and Eric M Phizicky. tRNA transfers to the limelight. *Genes & Development*, 17:162–180, 2003. doi: 10.1101/gad.1049103.162. URL https://genesdev.cshlp.org/content/17/2/162.long.

Qian Huang, Peng Yao, Gilbert Eriani, and En Duo Wang. In vivo identification

of essential nucleotides in tRNALeu to its functions by using a constructed yeast tRNALeu knockout strain. *Nucleic Acids Research*, 40(20):10463–10477, 2012. ISSN 03051048. doi: 10.1093/nar/gks783. URL https://academic.oup.com/nar/article/40/20/10463/2414722.

T Hussain, V Kamarthapu, S P Kruparani, M V Deshmukh, and R Sankaranarayanan. Mechanistic insights into cognate substrate discrimination during proofreading in translation. *Proc Natl Acad Sci U S A*, 107(51): 22117–22121, 2010. ISSN 1091-6490. doi: 1014299107[pii]\r10.1073/pnas.1014299107. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3009766/pdf/pnas.1014299107.pdf.

Michael Ibba and Dieter Soll. Aminoacyl-tRNA synthesis. *Annual Review of Biochemistry*, 69:617–650, 2000. doi: 10.1146/annurev.biochem.69.1.617. URL https://pubmed.ncbi.nlm.nih.gov/10966471/.

Michael Ibba and Dieter Söll. Aminoacyl tRNA: setting the limits of the genetic code. *Gene & Development*, 18:731–738, 2004. ISSN 08909369. doi: 10.1101/gad.1187404.

Michael Ibba, Christopher Francklyn, and Stephen Cusack. *The Aminoacyl-tRNA Synthetases*. Landes Bioscience, Georgetown, Texas, USA, 2005. ISBN 1587061899. doi: 10.1002/047120918X.emb0061.

James S Italia, Partha Sarathi Addy, Chester J J Wrobel, Lisa A Crawford, Marc J Lajoie, Yunan Zheng, and Abhishek Chatterjee. An orthogonalized platform for genetic code expansion in both bacteria and eukaryotes. *Nature Chemical Biology*, pages 446–450, Feb 2017. doi: 10.1038/nchembio.2312.

Yuzuru Itoh, Markus J Bröcker, Shun-ichi Sekine, Gifty Hammond, Shiro Suetsugu, and Dieter Söll. Decameric SelA•tRNASec Ring Structure Reveals Mechanism of Bacterial Selenocysteine Formation. *Science*, 340(6128):75–78, 2013. ISSN 1. doi: 10.1126/science.1229521. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3976565/.

Dirk Iwata-Reuyl. Biosynthesis of the 7-deazaguanosine hypermodified nucleosides of transfer RNA. *Bioorganic Chemistry*, 31(1):24–43, 2003. doi: 10.1016/S0045-2068(02)00513-8. URL https://www.sciencedirect.com/science/article/pii/S0045206802005138?via%3Dihub.

Jane E Jackman and Eric M Phizicky. tRNAHis guanylyltransferase adds G-1 to the 5' end of tRNAHis by recognition of the anticodon, one of several features unexpectedly shared with tRNA synthetases. *RNA (New York, N.Y.)*, 12(6):1007–14, 2006. ISSN 1355-8382. doi: 10.1261/rna.54706. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1464847&tool=pmcentrez&rendertype=abstract.

H Jakubowski and E Goldman. Editing of errors in selection of amino acids for protein synthesis. *Microbiological reviews*, 56(3):412–429, 1992. ISSN 0146-0749. doi: 10.1007/s12104-010-9253-6.

Haeyoung Jeong, Valérie Barbe, Choong Hoon Lee, David Vallenet, Dong Su Yu, Sang-Haeng Choi, Arnaud Couloux, Seung-Won Lee, Sung Ho Yoon, Laurence Cattolico, Cheol-Goo Hur, Hong-Seog Park, Béatrice Ségurens, Sun Chang Kim, Tae Kwang Oh, Richard E Lenski, F William Studier, Patrick Daegelen, and Jihyun F Kim. Genome Sequences of Escherichia coli B strains REL606 and BL21(DE3). *Journal of Molecular Biology*, 394(4):644–652, 2009. doi: 10.1016/j.jmb.2009.09.052. URL https://www.sciencedirect.com/science/article/pii/S0022283609011917?via%3Dihub.

Quan-Quan Ji, Zhi-Peng Fang, Qing Ye, Ruan Zhi-Rong, Xiao-Long Zhou, and En-Duo Wang. C-terminal Domain of Leucyl-tRNA Synthetase from Pathogenic Candida albicans Recognizes both tRNASer and tRNALeu. *Journal of Biological Chemistry*, 291(7):3613–3625, 2016. doi: 10.1074/jbc.M115.699777. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4751399/.

Linda Johansson, Guro Gafvelin, and Elias S J Arnér. Selenocysteine in proteins—properties and biotechnological use. *Biochimica et Biophysica*

Acta, 1726(1):1–13, 2005. ISSN 1. doi: 10.1016/j.bbagen.2005.05.010. URL https://www.sciencedirect.com/science/article/abs/pii/S0304416505001522.

Frank Jühling, Mario Mörl, Roland K Hartmann, Mathias Sprinzl, Peter F. Stadler, and Joern Pütz. tRNAdb 2009: Compilation of tRNA sequences and tRNA genes. *Nucleic Acids Research*, 37(SUPPL. 1):D159–D162, 2009. ISSN 03051048. doi: 10.1093/nar/gkn772. URL https://academic.oup.com/nar/article/37/suppl_1/D159/1009888.

Sebastian Kirchner and Zoya Ignatova. Emerging roles of tRNA in adaptive translation, signalling dynamics and disease. *Nature Reviews Genetics*, 16(2):98–112, 2015. ISSN 1471-0056. doi: 10.1038/nrg3861. URL http://www.ncbi.nlm.nih.gov/pubmed/25534324.

Lev L Kisselev. The Role of the Anticodon in Recognition of tRNA by Aminoacyl-tRNA Synthetases. *Progress in Nucleic Acid Research and Molecular Biology*, 32:237–266, 1985. doi: 10.1016/S0079-6603(08)60350-5. URL https://www.sciencedirect.com/science/article/abs/pii/S0079660308603505?

Colin Kleanthous, Paul M Cullis, and William V Shaw. 3-(Bromoacetyl)chloramphenicol, an active site-directed inhibitor for chloramphenicol acetyltransferase. *Biochemistry*, 24(20):5307–5313, 1985. doi: 10.1021/bi00341a006. URL https://pubs.acs.org/doi/pdf/10.1021/bi00341a006.

Andreas Knapp, Myriam Ripphahn, Kristina Volkenborn, Pia Skoczinski, and Karl-Erich Jaeger. Activity-independent screening of secreted proteins using split GFP . *Journal of Biotechnology*, 258:110–116, 2017. doi: 10.1016/j.jbiotec.2017.05.024. URL https://pubmed.ncbi.nlm.nih.gov/28619616/.

Radha Krishnakumar and Jiqiang Ling. Experimental challenges of sense codon reassignment: An innovative approach to genetic code expansion. *FEBS Letters*,

588(3):383–388, 2014. ISSN 00145793. doi: 10.1016/j.febslet.2013.11.039. URL http://dx.doi.org/10.1016/j.febslet.2013.11.039.

Radha Krishnakumar, Laure Prat, Hans-Rudolf Aerni, Jiqiang Ling, Chuck Merryman, John I Glass, Jesse Rinehard, and Dieter Söll. Transfer RNA misidentification scrambles sense codon recoding . *Chembiochem*, 14(15):1967–1972, 2013. doi: 10.1002/cbic.201300444. URL https://pubmed.ncbi.nlm.nih.gov/24000185/.

Antje Krüger, Leticia L Torres, and Vitor B Pinheiro. Rewriting the genetic code through aminoacyl-tRNA synthetase engineering, 2015.

Joseph A Krzycki. The direct genetic encoding of pyrrolysine. *Current Opinion in Microbiology*, 8(6):706–712, 2005. ISSN 13695274. doi: 10.1016/j.mib.2005.10.009. URL https://pubmed.ncbi.nlm.nih.gov/16256420/.

Inchan Kwon, Kent Kirshenbaum, and David A. Tirrell. Breaking the degeneracy of the genetic code. *Journal of the American Chemical Society*, 125(25):7512–7513, 2003. ISSN 00027863. doi: 10.1021/ja0350076.

Marc J Lajoie, S Kosuri, Joshua A Mosberg, C J Gregg, D Zhang, and George M Church. Probing the limits of genetic recoding in essential genes. *Science*, 342 (6156):361–3, 2013. ISSN 1095-9203. doi: 10.1126/science.1241460. URL https://www.science.org/doi/10.1126/science.1241460.

Talley J Lambert. FPbase: a community-editable fluorescent protein database. *Nature Methods*, 16:277–278, 2019. doi: 10.1038/s41592-019-0352-8. URL https://www.nature.com/articles/s41592-019-0352-8.

Keith E Langley and Irving Zabin. beta-Galactosidase alpha complementation: properties of the complemented enzyme and mechanism of the complementation reaction . *Biochemistry*, 15(22):4866–4875, 1976. doi: 10.1021/bi00667a018. URL https://pubmed.ncbi.nlm.nih.gov/791361/.

Keith E Langley, Merna R Villarejo, Audree V Fowler, Patrice J Zamenhof, and Irving Zabin. Molecular basis of beta-galactosidase alpha-complementation . *Proceedings of the National Academy of Sciences USA*, 72(4):1254–1257, 1975. doi: 10.1073/pnas.72.4.1254. URL https://pubmed.ncbi.nlm.nih.gov/1093175/.

Andrew G W Leslie, Peter C E Moody, and William V Shaw. Structure of chloramphenicol acetyltransferase at 1.75-A resolution. *Proceedings of the National Academy of Sciences USA*, 85(12):4133–4137, 1988. doi: 10.1073/pnas.85.12.4133. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC280380/.

Ann Lewendon, Iain A Murray, Colin Kleanthous, Paul M Cullis, and William V Shaw. Substitutions in the active site of chloramphenicol acetyltransferase: role of a conserved aspartate. *Biochemistry*, 27(19):7385–7390, 1988. doi: 10.1021/BI00419A032. URL https://pubs.acs.org/doi/abs/10.1021/bi00419a032.

Ann Lewendon, Iain A Murray, William V Shaw, Michael R Gibbs, and Andrew G W Leslie. Replacement of catalytic histidine-195 of chloramphenicol acetyltransferase: Evidence for a general base role for glutamate. *Biochemistry*, 33(7):1944–1950, 1994. doi: 10.1021/BI00173A043. URL https://pubs.acs.org/doi/abs/10.1021/bi00173a043.

Josefine Liljeruhm, Saskia K Funk, Sandra Tietscher, Anders D Edlund, Sabri Jamal, Pikkei Wistrand-Yuen, Karl Dyrhage, Arvid Gynnå, Katarina Ivermark, Jessica Lövgren, Viktor Törnblom, Anders Virtanen, Erik R Lundin, Erik Wistrand-Yuen, and Anthony C Forster. Engineering a palette of eukaryotic chromoproteins for bacterial synthetic biology. *Journal of Biological Engineering*, 12 (8):–, 2018. doi: https://doi.org/10.1186/s13036-018-0100-0. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5946454/.

Jiqiang Ling, Noah Reynolds, and Michael Ibba. Aminoacyl-tRNA synthesis and

translational quality control. *Annual review of microbiology*, 63:61–78, 2009. ISSN 0066-4227. doi: 10.1146/annurev.micro.091208.073210.

Jiqiang Ling, Kaitlyn M. Peterson, Ivana Simonovic, Dieter Söll, and Miljan Simonovic. The mechanism of pre-transfer editing in yeast mitochondrial threonyl-tRNA synthetase. *Journal of Biological Chemistry*, 287(34):28518–28525, 2012. ISSN 00219258. doi: 10.1074/jbc.M112.372920.

Chang C Liu and Peter G Schultz. Adding new chemistries to the genetic code. *Annual review of biochemistry*, 79:413, 2010. ISSN 1545-4509. doi: 10.1146/annurev.biochem.052308.105824. URL `http://www.annualreviews.org/doi/abs/10.1146/annurev.biochem.052308.105824`.

Cuiping Liu, M Jeffrey Sanders, M John Pascal, and Ya-Ming Hou. Adaptation to tRNA acceptor stem structure by flexible adjustment in the catalytic domain of class I tRNA synthetases. *RNA*, 18(2):213–221, 2012. doi: 10.1261/rna.029983.111. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3264908/`.

Min Liu, Bin Wang, Fei Wang, Zhi Yang, Dan Gao, Chenyao Zhang, Lixin Ma, and Xiaolan Yu. Soluble expression of single-chain variable fragment (scFv) in Escherichia coli using superfolder green fluorescent protein as fusion partner. *Applied Microbiology & Biotechnology*, 103:6071–6079, 2019. doi: 10.1007/s00253-019-09925-6. URL `https://link.springer.com/article/10.1007/s00253-019-09925-6`.

Shaohui Liu, Jinfang Ma, Wei Wang, Maoxiang Zhang, Xin Qingting, Siman Peng, Rongxiu Li, and Huanzhang Zhu. Mutational Analysis of Highly Conserved Residues in the Phage PhiC31 Integrase Reveals Key Amino Acids Necessary for the DNA Recombination. *PLoS ONE*, 5(1):e8863–, 2010. doi: https:/doi.org/10.1371/journal.pone.0008863. URL `https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0008863`.

N P Lukashenko. Expanding genetic code: Amino acids 21 and 22, selenocysteine and pyrrolysine. *Russian Journal of Genetics*, 46(8):899–916, 2010. ISSN

8. doi: 10.1134/S1022795410080016. URL https://link.springer.com/article/10.1134/S1022795410080016.

Rolf Lutz. pZ Host Strains. http://www.expressys.com/main_strains.html, 1998a. Accessed: 24-August-2022.

Rolf Lutz. Applications: 2.1 Using Tetracycline (Tc), Anhydrotetracycline (aTc) or Doxycycline (Dox) as inducer. http://www.expressys.com/main_applications.html, 1998b. Accessed: 24-August-2022.

Rolf Lutz and Hermann Bujard. Independent and Tight Regulation of Transcriptional Units in Escherichia Coli Via the LacR/O, the TetR/O and AraC/I1-I2 Regulatory Elements. *Nucleic Acids Research*, 25(6):1203–1210, 1997. doi: 10.1093/nar/25.6.1203. URL https://academic.oup.com/nar/article/25/6/1203/1197243.

Magdalena A Machnicka, Anna Olchowik, Henri Grosjean, and Janusz M Bujnicki. Distribution and frequencies of post- transcriptional modifications in tRNAs. *Integrative Biology of the Cell (I2BC)*, CNRS(December):1619–1629, 2014. doi: 10.4161/15476286.2014.992273.

Thomas J Magliery, J Christopher Anderson, and Peter G Schultz. Expanding the genetic code: selection of efficient suppressors of four-base codons and identification of "shifty" four-base codons with a library approach in Escherichia coli. *Journal of Molecular Biology*, 307(3):755—769, 2001. doi: 10.1006/jmbi.2001.4518. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7125544/.

Susan A Martinis and Michal T Boniecki. The balance between pre- and post-transfer editing in tRNA synthetases. *FEBS Letters*, 584(2):455–459, 2010. doi: 10.1016/j.febslet.2009.11.071.

Denisa D Mateus, João A Paredes, Yaiza Español, Lluís Ribas de Pouplana, Gabriela R Moura, and Manuel A S Santos. Molecular reconstruction of a fungal genetic code alteration. *RNA Biology*, 10(6):968–980, 2013. doi:

10.4161/rna.24683. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4111736/.

Felix Meier, Beat Suter, Henri Grosjean, Keith Gerard, and Eric Kubli. Queuosine modification of the wobble base in tRNAHis influences 'in vivo' decoding properties. *The EMBO Journal*, 4(3):823–827, 1985. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC554263/.

Kexin Meng, Christina Z Chung, Dieter Söll, and Natalie Krahn. Unconventional genetic code systems in archaea. *Frontiers in Microbiology*, 13:1664–302X, 2022. ISSN 1664-302X. doi: 10.3389/fmicb.2022.1007832. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9499178/.

May Meroueh and Christine S Chow. Thermodynamics of RNA hairpins containing single internal mismatches. *Nucleic Acids Research*, 27(4):1118–1125, 1999. ISSN 03051048. doi: 10.1093/nar/27.4.1118.

A Merrick, Christine, Jia Zhao, and Susan J Rosser. Serine Integrases: Advancing Synthetic Biology. *ACS Synthetic Biology*, 7(2):299–310, 2018. doi: 10.1021/acssynbio.7b00308. URL https://pubs.acs.org/doi/10.1021/acssynbio.7b00308.

Susan Michaelis, Leonard Guarente, and Jon Beckwith. In vitro construction and characterization of phoA-lacZ gene fusions in Escherichia coli. *Journal of Bacteriology*, 154(1):356–365, 1983. doi: 10.1128/jb.154.1.356-365.1983. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC217467/.

Xiaotian Ming, Kristina Smith, Hiroaki Suga, and Ya-Ming Hou. Recognition of tRNA Backbone for Aminoacylation with Cysteine: Evolution from Escherichia coli to Human. *Journal of Moecular Biology*, 318(5):1207–1220, 2002. doi: 10.1016/S0022-2836(02)00232-2. URL https://www.sciencedirect.com/science/article/pii/S0022283602002322.

Bijoy K Mohanty and Sidney R Kushner. Analysis of post-transcriptional RNA metabolism in prokaryotes. *Methods*, 155:124–130, 2019. doi: 10.1016/j.

ymeth.2018.11.006. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6568318/.

Dino Moras. Proofreading in translation: dynamics of the double-sieve model. *Proceedings of the National Academy of Sciences of the United States of America*, 107(51):21949–50, 2010. ISSN 1091-6490. doi: 10.1073/pnas.1016083107. URL http://www.ncbi.nlm.nih.gov/pubmed/21149735.

Rana C Morris and Mark S Elliott. Queuosine Modification of tRNA: A Case for Convergent Evolution. *Molecular Genetics and Metabolism*, 74(74):147–159, 2001. doi: 10.1006/mgme.2001.3216. URL https://www.sciencedirect.com/science/article/pii/S1096719201932160?via%3Dihub.

Takahito Mukai, Atsushi Yamaguchi, Kazumasa Ohtake, Mihoko Takahashi, Akiko Hayashi, Fumie Iraha, Satoshi Kira, Tatsuo Yanagisawa, Shigeyuki Yokoyama, Hiroko Hoshi, Takatsugu Kobayashi, and Kensaku Sakamoto. Reassignment of a rare sense codon to a non-canonical amino acid in Escherichia coli. *Nucleic Acids Research*, 43(16):8111–8122, 2015. ISSN 13624962. doi: 10.1093/nar/gkv787. URL https://academic.oup.com/nar/article/43/16/8111/1077583.

Martin E Mulligan, Jürgen Brosius, and William R McClure. Characterization in vitro of the effect of spacer length on the activity of Escherichia coli RNA polymerase at the TAC promoter. *The Journal of Biological Chemistry*, 260(6):3529–3538, 1985. URL https://pubmed.ncbi.nlm.nih.gov/3882710/.

Richard S Mursinna, Keun Woo Lee, James M Briggs, and Susan a Martinis. Molecular dissection of a critical specificity determinant within the amino acid editing domain of leucyl-tRNA synthetase. *Biochemistry*, 43(1):155–165, 2004. ISSN 0006-2960. doi: 10.1021/bi034919h.

Masahiro Naganuma, Shun-ichi Sekine, Yeeting Esther Chong, Min Guo, Xiang-Lei Yang, Howard Gamper, Ya-Ming Hou, Paul Schimmel, and Shigeyuki Yokoyama. The selective tRNA aminoacylation mechanism based on a single

G•U pair. *Nature*, 510(7506):507–11, 2014. ISSN 1476-4687. doi: 10.1038/nature13440. URL http://www.ncbi.nlm.nih.gov/pubmed/24919148.

Nobukazu Nameki, Haruichi Asahara, Mikio Shimizu, Norihiro Okada, and Hyouta Himeno. Identity elements of Saccharomyces cerevisiae tRNAHis. *Nucleic Acids Research*, 23(3):389–394, 1995. ISSN 03051048. doi: 10.1093/nar/23.3.389.

Paolo Natale, Thomas Brüser, and Arnold J M Driessen. Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane—Distinct translocases and mechanisms. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1778(9):1735–1756, 2008. doi: 10.1016/j.bbamem.2007.07.015. URL https://www.sciencedirect.com/science/article/pii/S0005273607002775.

Heinz Neumann, Adrian L. Slusarczyk, and Jason W. Chin. De Novo generation of mutually orthogonal aminoacyl-TRNA synthetase/ tRNA pairs. *Journal of the American Chemical Society*, 132(7):2142–2144, 2010a. ISSN 00027863. doi: 10.1021/ja9068722. URL https://pubs.acs.org/doi/10.1021/ja9068722.

Heinz Neumann, Kaihang Wang, Lloyd Davis, Maria Garcia-Alai, and Jason W. Chin. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature*, 464(7287):441–444, 2010b. ISSN 0028-0836. doi: 10.1038/nature08817. URL http://www.nature.com/doifinder/10.1038/nature08817.

Chrysa Ntountoumi, Panayotis Vlastaridis, Dimitris Mossialos, Constantinos Stathopolulos, Ioannis Iliopoulos, Vasilios Promponas, Stephen G Oliver, and Grigoris D Amoutzias. Low complexity regions in the proteins of prokaryotes perform important functional roles and are highly conserved. *Nucleic Acids Research*, 47(19):9998–10009, 2019. doi: 10.1093/nar/gkz730. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6821194/.

Cédric Orelle, Erik D. Carlson, Teresa Szal, Tanja Florin, Michael C. Jewett, and Alexander S. Mankin. Protein synthesis by ribosomes with tethered subunits.

Nature, 524:119–124, 2015. ISSN 0028-0836. doi: 10.1038/nature14862. URL http://www.nature.com/doifinder/10.1038/nature14862.

Syozo Osawa, Thomas H Jukes, Kimitsuna Watanabe, and Akira Muto. Recent Evidence for Evolution of the Genetic Code. *Microbiological Reviews*, 56(1): 229–264, 1992. ISSN 0146-0749. doi: 10.1128/mr.56.1.229-264.1992. URL https://pubmed.ncbi.nlm.nih.gov/1579111/.

Patrick O'Donoghue, Laure Prat, Ilka U Heinemann, Jiqiang Ling, Keturah Odoi, Wenshe R Liu, and Dieter Söll. Near-cognate suppression of amber, opal and quadruplet codons competes with aminoacyl-tRNA^{Pyl} for genetic code expansion. *FEBS Letters*, 9(21):3931—3937, 2012. doi: 10.1016/j.febslet.2012.09.033. URL https://www.sciencedirect.com/science/article/pii/S1074552102000947.

Andrés Palencia, Thibaut Crépin, Michael T Vu, Tommie L Lincecum, Susan A Martinis, and Stephen Cusack. Structural dynamics of the aminoacylation and proofreading functional cycle of bacterial leucyl-tRNA synthetase [S1]. *Nature structural & molecular biology*, 19(7): 677–84, 2012. ISSN 1545-9985. doi: 10.1038/nsmb.2317. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3392462&tool=pmcentrez&rendertype=abstract.

George H Patterson, Susan M Knobel, Wallace D Sharif, Steven R Kain, and David W Piston. Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy. *Biophysical Journal*, 73(5):2782–2790, 1997. doi: 10.1016/S0006-3495(97)78307-3. URL https://www.sciencedirect.com/science/article/pii/S0006349597783073.

Jean-Denis Pédelacq, Stéphanie Cabantous, Timothy Tran, Thomas C Terwilliger, and Geoffrey S Waldo. Engineering and characterization of a superfolder green fluorescent protein. *Nature biotechnology*, 24(1):79–88, 2006. ISSN 1087-0156. doi: 10.1038/nbt1172. URL https://www.nature.com/articles/nbt1172.

John J Perona. *Glutaminyl-tRNA Synthetases*, chapter 1, pages 1–x. Austin (TX): Landes Bioscience, 2013. URL https://www.ncbi.nlm.nih.gov/books/NBK6506/.

Valérie Pezo, David Metzgar, Tamara L Hendrickson, William F Waas, S Hazebrouck, Volker Döring, Philippe Marlière, Paul Schimmel, and Valérie de Crécy-Lagard. Artificially ambiguous genetic code confers growth yield advantage. *Proceedings of the National Academy of Sciences of the United States of America*, 101(23):8593–7, 2004. ISSN 0027-8424. doi: 10.1073/pnas.0402893101. URL https://www.pnas.org/content/101/23/8593/.

Kathleen Postle and Rhonda F Good. DNA sequence of the *Escherichia coli tonB* gene. *Proceedings of the National Academy of Sciences*, 80(17):5235–5239, 1983. doi: https://doi.org/10.1073/pnas.80.17.5235. URL https://www.pnas.org/doi/10.1073/pnas.80.17.5235.

Kathleen Postle and Rhonda F Good. A bidirectional Rho-independent transcription terminator between the *E. coli tonB* gene and an opposing gene. *Cell*, 41(2):577–585, 1985. doi: https://doi-org/10.1016/S0092-8674(85)80030-1.

Douglas C Prasher, Virginia K Eckenrode, William W Ward, Frank G Prendergast, and Milton J Cormier. Primary structure of the Aequorea victoria green-fluorescent protein. *Gene*, 111(2):229–233, 1992. doi: 10.1016/0378-1119(92)90691-h. URL https://www.sciencedirect.com/science/article/abs/pii/037811199290691H.

Oliver Rackham and Jason W Chin. Cellular logic with orthogonal ribosomes. *Journal of the American Chemical Society*, 127(50):17584–17585, 2005a. ISSN 0002-7863. doi: 10.1021/ja055338d. URL http://www.ncbi.nlm.nih.gov/pubmed/16351070.

Oliver Rackham and Jason W Chin. A network of orthogonal ribosome·mRNA pairs. *Nature Chemical Biology*, 1(3):159—166, 2005b. doi: 10.1038/nchembio719. URL https://www.nature.com/articles/nchembio719z.

Markus Ralser, Robert Querfurth, Hans-Jörg Warnatz, Hans Lehrach, Marie-Laure Yaspo, and Sylvia Krobitsch. An efficient and economic enhancer mix for PCR. *Biochemical and Biophysical Research Communications*, 347(3):747–751, 2006. doi: 10.1016/j.bbrc.2006.06.151. URL https://www.sciencedirect.com/science/article/abs/pii/S0006291X06014707.

Wenqi Ran and Paul G. Higgs. The influence of anticodon-codon interactions and modified bases on codon usage bias in bacteria. *Molecular Biology and Evolution*, 27(9):2129–2140, 2010. ISSN 07374038. doi: 10.1093/molbev/msq102.

Benjamin J. Rauch, Joseph J. Porter, Ryan A. Mehl, and John J. Perona. Improved Incorporation of Noncanonical Amino Acids by an Engineered tRNATyr Suppressor. *Biochemistry*, 55(3):618–628, 2016. ISSN 15204995. doi: 10.1021/acs.biochem.5b01185. URL https://pubs.acs.org/doi/10.1021/acs.biochem.5b01185.

Brian G Reid and Gregory C Flynn. Chromophore formation in green fluorescent protein. *Biochemistry*, 36(22):6786–6791, 1997. doi: 10.1021/bi970281w. URL https://pubs.acs.org/doi/10.1021/bi970281w.

Noah M Reynolds, Beth a Lazazzera, and Michael Ibba. Cellular mechanisms that control mistranslation. *Nature reviews. Microbiology*, 8(12):849–856, 2010. ISSN 1740-1526. doi: 10.1038/nrmicro2472. URL http://dx.doi.org/10.1038/nrmicro2472.

Lluís Ribas de Pouplana and Paul Schimmel. Aminoacyl-tRNA synthetases: Potential markers of genetic code development. *Trends in Biochemical Sciences*, 26 (10):591–596, 2001. ISSN 09680004. doi: 10.1016/S0968-0004(01)01932-6.

Wesley E Robertson, Louise F H Funke, Daniel de la Torre, Julius Fredens, Kaihang Wang, and Jason W Chin. Creating custom synthetic genomes in *Escherichia coli* with REXER and GENESIS. *Nature Protocols*, 16(5):2345–2380, 2021. doi: 10.1038/s41596-020-00464-3. URL https://www.nature.com/articles/s41596-020-00464-3.

Rita Rocha, Pedro José Barbosa Pereira, Manuel A S Santos, and Sandra Macedo-Ribeiro. Unveiling the structural basis for translational ambiguity tolerance in a human fungal pathogen. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 108(34):14091–14096, 2011. doi: 10.1073/pnas.1102835108. URL https://pubmed.ncbi.nlm.nih.gov/21825144/.

Matthew G Romei and Steven G Boxer. Split Green Fluorescent Proteins: Scope, Limitations, and Outlook. *Annual Review of Biophysics*, 48:19–44, 2019. doi: 10.1146/annurev-biophys-051013-022846. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6537611/.

Nathaniel Roquet, Ava P Soleimany, Alyssa C Ferris, Scott Aaronson, and Timothy K Lu. Synthetic Recombinase-Based State Machines in Living Cells. *Science*, 353(6297):aad8559, 2016. doi: 10.1126/science.aad8559.

Abbey E Rosen and Karin Musier-Forsyth. Recognition of G-1:C73 atomic groups by Escherichia coli histidyl-tRNA synthetase [S1]. *Biomacromolecules*, 126(1):64–65, 2004. ISSN 01606891. doi: 10.2307/4016193.

Alexander C Roth. Decoding properties of tRNA leave a detectable signal in codon usage bias. *Bioinformatics*, 28(18):340–348, 2012. ISSN 13674803. doi: 10.1093/bioinformatics/bts403. URL https://academic.oup.com/bioinformatics/article/28/18/i340/248836.

Paul A Rowley, Matthew C A Smith, Ellen Younger, and Margaret C M Smith. A motif in the C-terminal domain of ΦC31 integrase controls the directionality of recombination. *Nucleic Acids Research*, 36(12):3879–3891, 2008. doi: https:/doi.org/10.1093/nar/gkn269. URL https://academic.oup.com/nar/article/36/12/3879/1125990.

Karen Rutherford and Gregory D Van Duyne. The ins and outs of serine integrase site-specific recombination. *Current Opinion in Structural Biology*, 24(1):125–131, 2014. ISSN 0959440X. doi: 10.1016/j.sbi.2014.01.003. URL http://dx.doi.org/10.1016/j.sbi.2014.01.003.

Karen Rutherford, Peng Yuan, Kay Perry, Robert Sharp, and Gregory D Van Duyne. Attachment site recognition and regulation of directionality by the serine integrases. *Nucleic Acids Research*, 41(17):8341–8356, 2013. doi: https:/doi.org/10.1093/nar/gkt580. URL https://academic.oup.com/nar/article/41/17/8341/2411091.

Manuel A Santos, Gerard Keith, and Mick F Tuite. Non-standard translational events in *Candida albicans* mediated by an unusual seryl-tRNA with a 5'-CAG-3' (leucine) anticodon. *The EMBO Journal*, 12(2):607–616, 1993. doi: 10.1002/j.1460-2075.1993.tb05693.x. URL https://www.embopress.org/doi/abs/10.1002/j.1460-2075.1993.tb05693.x.

Manuel A Santos, Victoria M Perreau, and Mick F Tuite. Transfer RNA structural change is a key element in the reassignment of the CUG codon in *Candida albicans*. *The EMBO Journal*, 15(18):5060–5068, 1996. doi: 10.1002/j.1460-2075.1996.tb00886.x. URL https://www.embopress.org/doi/abs/10.1002/j.1460-2075.1996.tb00886.x.

Paul Schimmel. Development of tRNA synthetases and connection to genetic code and disease. *Protein science : a publication of the Protein Society*, 17(10):1643–1652, 2008. ISSN 1469-896X. doi: 10.1110/ps.037242.108.

Emmanuelle Schmitt, Michel Panvert, Sylvain Blanquet, and Yves Mechulam. Transition state stabilization by the 'high' motif of class I aminoacyl-tRNA synthetases: The case of Escherichia coli methionyl-tRNA synthetase. *Nucleic Acids Research*, 23(23):4793–4798, 1995. ISSN 03051048. doi: 10.1093/nar/23.23.4793.

Stephan Scholtissek and Frank Grosse. A cloning cartridge of λ t_0 terminator. *Nucleic Acids Research*, 15(7):3185, 1987. doi: https://doi.org/10.1093/nar/15.7.3185. URL https://academic.oup.com/nar/article/15/7/3185/1149849.

Ilham Ayub Shahmuradov, Rozaimi Mohamad Razali, Salim Bougouffa, Aleksandar Radovanovic, and Vladimir Bajic. bTSSfinder: a novel tool for the prediction of promoters in cyanobacteria and Escherichia coli. *Bioinformatics*, 33 (3):334–340, 2017. doi: https://doi.org/10.1093/bioinformatics/btw629. URL https://pubmed.ncbi.nlm.nih.gov/27694198/.

João Simões, Ana R Bezerra, Gabriela R Moura, Hugo Araújo, Ivo Gut, Mónica Bayes, and Manuel A S Santos. The Fungus Candida albicans Tolerates Ambiguity at Multiple Codons. *Frontiers in Mircobiology*, 7(401):1–11, 2016. doi: 10.3389/fmicb.2016.00401. URL https://www.frontiersin.org/articles/10.3389/fmicb.2016.00401/full.

E V Smirnova, V A Lakunina, I Tarassov, I A Krasheninnikov, and P A Kamenski. Non-canonical functions of aminoacyl-tRNA synthetases. *Biochemistry. Biokhimiia*, 65(8):888–897, 2000. ISSN 0006-2979. doi: 10.1134/S0006297912010026.

Margaret C M Smith, William R A Brown, Andrew R McEwan, and Paul A Rowley. Site-specific recombination by ΦC31 integrase and other large serine recombinases. *Biochemical Society Transactions*, 38 (2):388–394, 2010. ISSN 0300-5127. doi: 10.1042/BST0380388. URL https://portlandpress.com/biochemsoctrans/article-abstract/38/2/388/67007/Site-specific-recombination-by-C31-integrase-and?redirectedFrom=fulltext.

Temple F. Smith and Hyman Hartman. The evolution of Class II Aminoacyl-tRNA synthetases and the first code. *FEBS Letters*, 589(23):3499–3507, 2015. ISSN 18733468. doi: 10.1016/j.febslet.2015.10.006. URL http://dx.doi.org/10.1016/j.febslet.2015.10.006.

Victor V Solovyev and Asaf Salamov. Automatic Annotation of Microbial Genomes and Metagenomic Sequences. In Richard W Li, editor, *Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies*, chap-

ter 28, pages 61–78. Nova Science Publishers, 2011. ISBN 978-1-61668-682-6. URL http://www.softberry.com/berry.phtml?topic=bprom&group=programs&subgroup=gfindb.

Akiko Soma and Hyouta Himeno. Cross-species aminoacylation of tRNA with a long variable arm between Escherichia coli and Saccharomyces cerevisiae. *Nucleic Acids Research*, 26(19):4374–4381, 1998. ISSN 03051048. doi: 10.1093/nar/26.19.4374.

Michael A Sørensen, Johan Elf, Elli Bouakaz, Tanel Tenson, Suparna Sanyal, Glenn R Björk, and Måns Ehrenberg. Over expression of a tRNALeu isoacceptor changes charging pattern of leucine tRNAs and reveals new codon reading. *Journal of Molecular Biology*, 354(1):16–24, 2005. ISSN 00222836. doi: 10.1016/j.jmb.2005.08.076. URL https://pubmed.ncbi.nlm.nih.gov/16236318/.

Kathryn E. Splan, Michael E. Ignatov, and Karin Musier-Forsyth. Transfer RNA modulates the editing mechanism used by class II prolyl-tRNA synthetase. *Journal of Biological Chemistry*, 283(11):7128–7134, 2008. ISSN 00219258. doi: 10.1074/jbc.M709902200.

Mathias Sprinzl and Konstantin S Vassilenko. Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Research*, 33(DI):D139–D140, 2005. doi: 10.1093/nar/gki012. URL https://pubmed.ncbi.nlm.nih.gov/15608164/.

Gayathri Srinivasan, Carey M James, and Joseph A Krzycki. Pyrrolysine Encoded by UAG in Archaea: Charging of a UAG-Decoding Specialized tRNA. *Science*, 296(5572):1459–1463, May 2002. doi: 10.1126/science.1069588. URL https://pubmed.ncbi.nlm.nih.gov/12029131/.

François St-Pierre, Lun Cui, David G Priest, Drew Endy, Ian B Dodd, and Keith E Shearwin. One-step cloning and chromosomal integration of DNA. *ACS Synthetic Biology*, 2(9):537–541, 2013. doi: 10.1021/sb400021j. URL https://pubs.acs.org/doi/10.1021/sb400021j.

Peter Strazewski, Ewa Biala, Kay Gabriel, and William H McClain. The relationship of thermodynamic stability at a G × U recognition site to tRNA aminoacylation specificity. , 5(11):1490–1494, 1999. doi: 10.1017/s1355838299991586. URL https://europepmc.org/article/med/10580477.

Oksana M. Subach, Illia S. Gundorov, Masami Yoshimura, Fedor V. Subach, Jinghang Zhang, David Grüenwald, Ekaterina A. Souslova, Dmitriy M. Chudakov, and Vladislav V. Verkhusha. Conversion of Red Fluorescent Protein into a Bright Blue Probe. *Chemistry and Biology*, 15(10):1116–1124, 2008. ISSN 10745521. doi: 10.1016/j.chembiol.2008.08.006.

Doug Sweetser, Michael Nonet, and Richard A Young. Prokaryotic and eukaryotic RNA polymerases have homologous core subunits (Saccharomyces cerevisiae/RPB2). *Biochemistry*, 84(March):1192–11, 1987. ISSN 0027-8424. doi: 10.1073/pnas.84.5.1192.

Kay Terpe. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology*, 72:211–222, 2006. doi: 10.1007/s00253-006-0465-8. URL https://link.springer.com/article/10.1007/s00253-006-0465-8.

S. Thorbjarnardottir, T. Dingermann, T. Rafnar, O. S. Andrésson, D. Söll, and G. Eggertsson. Leucine tRNA family of Escherichia coli: Nucleotide sequence of the supP(Am) suppressor gene. *Journal of Bacteriology*, 161(1):219–222, 1985. ISSN 00219193.

Hong Tian, Danni Deng, Jie Huang, Dongning Yao, Xiaowei Xu, and Xiangdong Gao. Screening system for orthogonal suppressor tRNAs based on the species-specific toxicity of suppressor tRNAs. *Biochimie*, 95(4):881–888, 2013. ISSN 03009084. doi: 10.1016/j.biochi.2012.12.010. URL http://dx.doi.org/10.1016/j.biochi.2012.12.010.

Giuseppe Tocchini-Valentini, Margaret E Saks, and John Abelson. tRNA leucine

identity and recognition sets. *Journal of Molecular Biology*, 298(5):779–793, 2000. ISSN 0022-2836. doi: 10.1006/jmbi.2000.3694.

Leticia Torres, Antje Krüger, Eszter Csibra, Edoardo Gianni, and Vitor B Pinheiro. Synthetic biology approaches to biological containment: pre-emptively tackling potential risks. *Essays in Biochemistry*, 60(4):393–410, 2016. doi: 10.1042/EBC20160013. URL https://portlandpress.com/essaysbiochem/article/60/4/393/78400/Synthetic-biology-approaches-to-biological.

Roger Y Tsien. The green fluorescent protein . *Annual Review of Biochemistry*, 67: 509–544, 1998. doi: 10.1146/annurev.biochem.67.1.509. URL https://www.annualreviews.org/doi/full/10.1146/annurev.biochem.67.1.509.

Agnes Ullmann, François Jacob, and Jacques Monod. Characterization by in vitro complementation of a peptide corresponding to an operator-proximal segment of the beta-galactosidase structural gene of Escherichia coli . *Journal of Molecular Biology*, 24(2):339–343, 1967. doi: 10.1016/0022-2836(67)90341-5. URL https://www.sciencedirect.com/science/article/abs/pii/0022283667903415.

Gabriele Varani and William H McClain. The G × U wobble base pair. A fundamental building block of RNA structure crucial to RNA function in diverse biological systems. *EMBO Reports*, 1(1):18–23, 2000. doi: 10.1093/embo-reports/kvd001. URL https://www.embopress.org/doi/full/10.1093/embo-reports/kvd001.

Geoffrey S Waldo, Blake M Standish, Joel Berendzen, and Thomas C Terwilliger. Rapid protein-folding assay using green fluorescent protein . *Natural Biotechnology*, 17(7):691–695, 1999. doi: 10.1038/10904. URL https://www.nature.com/articles/nbt0799_691.

Kathryn P Wall, Rebecca Dillon, and Michelle K Knowles. Fluorescence quantum yield measurements of fluorescent proteins: a laboratory experiment for

a biochemistry or molecular biophysics laboratory course . *Biochemistry and Molecular Biology Education*, 43(1):52–59, 2014. doi: 10.1002/bmb.20837. URL https://iubmb.onlinelibrary.wiley.com/doi/pdf/10.1002/bmb.20837.

Harris H Wang, Farren J Isaacs, Peter A Carr, Zachary Z Sun, George Xu, Craig R Forest, and George M Church. Programming cells by multiplex genome engineering and accelerated evolution. *Nature*, 460(7257):894–898, 2009a. ISSN 0028-0836. doi: 10.1038/nature08187. URL https://www.nature.com/articles/nature08187.

Harris H Wang, Farren J Isaacs, Peter a Carr, Zachary Z Sun, George Xu, Craig R Forest, and George M Church. Programming cells by multiplex genome engineering and accelerated evolution. *Nature*, 460(7257):894–898, 2009b. ISSN 0028-0836. doi: 10.1038/nature08187. URL http://dx.doi.org/10.1038/nature08187.

Kaihang Wang, Julius Fredens, Simon F Brunner, Samuel H Kim, Tiongsun Chia, and Jason W Chin. Defining synonymous codon compression schemes by genome recoding. *Nature*, 539(7627):59–64, 2016. ISSN 0028-0836. doi: 10.1038/nature20124. URL http://dx.doi.org/10.1038/nature20124.

L Wang, A Brock, B Herberich, and Peter G Schultz. Expanding the genetic code of Escherichia coli. *Science (New York, N.Y.)*, 292(5516):498–500, 2001. ISSN 0036-8075. doi: 10.1126/science.1060077.

Lei Wang and Peter G Schultz. A general approach for the generation of orthogonal tRNAs. *Chemistry and Biology*, 8(9):883–890, 2001. ISSN 10745521. doi: 10.1016/S1074-5521(01)00063-1.

Qian Wang, Angela R. Parrish, and Lei Wang. Expanding the Genetic Code for Biological Studies. *Chemistry and Biology*, 16(3):323–336, 2009c. ISSN 10745521. doi: 10.1016/j.chembiol.2009.03.001. URL http://dx.doi.org/10.1016/j.chembiol.2009.03.001.

A Warshel, G Naray-Szabo, F Sussman, and J-K Hwang. How do serine proteases really work? *Biochemistry*, 28(9):3629–3637, 1989. doi: 10.1021/bi00435a001. URL https://pubs.acs.org/doi/10.1021/bi00435a001.

I Weygand-Durasević, N Ban, D Jahn, and D Söll. Yeast seryl-tRNA synthetase expressed in Escherichia coli recognizes bacterial serine-specific tRNAs in vivo. *European journal of biochemistry / FEBS*, 214(3):869–77, 1993. ISSN 0014-2956. URL http://www.ncbi.nlm.nih.gov/pubmed/7686490.

Kristoffer Skovbo Winther, Michael Askvad Sørensen, and Sine Lo Svenningsen. Polyamines are Required for tRNA Anticodon Modification in Escherichia coli. *Journal of Molecular Biology*, 433(15):167073–167086, 2021. doi: 10.1016/j.jmb.2021.167073. URL https://www.sciencedirect.com/science/article/pii/S0022283621002916.

C R Woese. On the evolution of the genetic code. *Proceedings of the National Academy of Sciences of the United States of America*, 54(6):1546–1552, 1965. doi: 10.1073/pnas.54.6.1546. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC300511/.

Han Xiao, Abhishek Chatterjee, Sei-hyun Choi, Krishna M Bajjuri, Subhash C Sinha, and Peter G Schultz. Genetic incorporation of multiple unnatural amino acids into proteins in mammalian cells. *Angewandte Chemie International Edition*, 1:epub, 2013. ISSN 14337851. doi: 10.1002/anie.201308137. URL http://dx.doi.org/10.1002/anie.201308137.

Wen Yan and Christopher Francklyn. Cytosine 73 is a discriminator nucleotide in vivo for histidyl-tRNA in Escherichia coli. *Journal of Biological Chemistry*, 269 (13):10022–10027, 1994. ISSN 00219258. URL https://pubmed.ncbi.nlm.nih.gov/8144499/.

Wen Yan, John Augustine, and Christopher Francklyn. A tRNA identity switch mediated by the binding interaction between a tRNA anticodon and the accessory

domain of a class II aminoacyl-tRNA synthetase. *Biochemistry*, 35(21):6559–6568, 1996. ISSN 00062960. doi: 10.1021/bi952889f. URL https://pubs.acs.org/doi/10.1021/bi952889f.

Tatsuo Yanagisawa, Ryohei Ishii, Ryuya Fukunaga, Takatsugu Kobayashi, Kensaku Sakamoto, and Shigeyuki Yokoyama. Crystallographic Studies on Multiple Conformational States of Active-site Loops in Pyrrolysyl-tRNA Synthetase. *Journal of Molecular Biology*, 378(3):634–652, 2008. ISSN 00222836. doi: 10.1016/j.jmb.2008.02.045.

C Yanofsky, T Platt, L P Crawford, B P Nichols, G E Christie, H Horowitz, M Van Cleemput, and A M Wu. The complete nucleotide sequence of the tryptophan operon of *Escherichia coli*. *Nucleic Acids Research*, 9(24):6647–6668, 1981. doi: https://doi.org/10.1093/nar/9.24.6647. URL https://academic.oup.com/nar/article/9/24/6647/2375514.

Daniel G Yansura and Dennis J Henner. [5] Use of *Escherichia coli* trp promoter for direct expression of proteins. *Methods in Enzymology*, 185:54–60, 1990. doi: https://doi.org/10.1016/0076-6879(90)85007-B. URL https://www.sciencedirect.com/science/article/abs/pii/007668799085007B?via%3Dihub.

Travis S Young and Peter G Schultz. Beyond the canonical 20 amino acids: Expanding the genetic lexicon. *Journal of Biological Chemistry*, 285(15):11039–11044, 2010. ISSN 00219258. doi: 10.1074/jbc.R109.091306.

Jing Yuan, Tasos Gogakos, Arianne M. Babina, Dieter Söll, and Lennart Randau. Change of tRNA identity leads to a divergent orthogonal histidyl-tRNA synthetase/tRNA His pair. *Nucleic Acids Research*, 39(6):2286–2293, 2011. ISSN 03051048. doi: 10.1093/nar/gkq1176.

Yu Zeng, Wei Wang, and Wenshe R Liu. Towards reassigning the rare AGG codon in Escherichia coli. *ChemBioChem*, 15(12):1750–1754, 2014. ISSN

14397633. doi: 10.1002/cbic.201400075. URL https://chemistry-europe.onlinelibrary.wiley.com/doi/10.1002/cbic.201400075.

Yunan Zheng, Marc J Lajoie, James S Italia, Melissa A Chin, George M Church, and Abhishek Chatterjee. Performance of optimized noncanonical amino acid mutagenesis systems in the absence of release factor 1. *Mol. BioSyst.*, 12(6):–, 2016. ISSN 1742-206X. doi: 10.1039/C6MB00070C. URL http://dx.doi.org/10.1039/C6MB00070C.