Aminoacyl-tRNA synthetases:

investigations of tRNA specificity

for application in ProxiMAX/synthetic biology

Marta Maria Ferreira Amaral

Doctor of Philosophy

Aston University

May 2020

Aston University

Aminoacyl-tRNA synthetases: investigations of tRNA specificity for application in ProxiMAX/synthetic biology

Marta Maria Ferreira Amaral

Doctor of Philosophy

May 2020

Thesis summary

ProxiMAX randomisation is a nondegenerate saturation mutagenesis technology that offers control over identity, location and relative ratio of amino acids within a protein library. It employs a maximum of 20 codons in saturated positions to encode 20 amino acids. This eliminates bias, degeneracy and provides maximal diversity. Since a maximum of 20 codons are used, many codons remain available to encode unnatural amino acids (UAAs). Incorporation of UAAs is essential tool in protein engineering to expand protein repertoires. The aim of this project is to establish fundamentals for the development of an *in vitro* transcription/translation system that will ultimately be coupled with ProxiMAX randomisation to produce synthetic gene libraries simultaneously encoding multiple UAAs.

The focus of the current project was *E. coli* Alanyl-tRNA synthetase (AlaRS), an enzyme that aminoacylates $tRNA^{Ala}$ with L-Alanine. The aim was to engineer the active site of the synthetase to incorporate UAAs and to develop an assay by which the resulting AlaRS protein library might be screened. ProxiMAX randomisation was employed to produce two variant libraries of AlaRS, each had removed native amino acid editing function. Six positions from amino acid binding pocket were selected for randomisation to encode 18 natural amino acids. Each resulting library encoded $>10^7$ novel variant proteins.

The initial development of screening assay for synthetase libraries involved using eGFP with stop codon as a reporter gene to assess aminoacylation of tRNA suppressor by AlaRS enzyme. The assay was designed in *S. cerevisiae* to prevent exogenous *E. coli* AlaRS and tRNA from working with endogenous system. The controls designed for the assay revealed that yeast read through the stop codon in eGFP and therefore, aminoacylation of tRNA suppressor by AlaRS proved to be difficult to assess. Alternative stratagems for future screens are considered in light of the results contained within the current study.

Keywords: aminoacyl-tRNA synthetases, alanyl-tRNA synthetase, unnatural amino acids, nondegenerate mutagenesis, protein engineering

To mama and tata who always support me in my education and personal life.

Acknowledgements

I would like to thank my supervisor Prof. Anna Hine for her excellent guidance, ongoing support and constant encouragement throughout my project. You have taught me so many valuable lessons and I am beyond grateful for that. These thanks are extended to my co-supervisor Dr Andrew Sutherland for his help and guidance.

I would also like to thank my sponsor, Isogenica Ltd., especially Dr. Chris Ullman, Dr. Laura Frigotto, Dr. Guy Hermans as well as Dr. Matt Smith, Dr. Simon Cooper, Dr. Chris Hodson, Dr. Gabriela Ivanova-Berndt, and everyone else, especially Margarida and Marianne, who made me feel very welcome during my industrial placement and offered their help and support.

Great thanks go to my research group members including Dr. Mohammed Ashraf, Dr. Anupama Chembath, Ben Wagstaffe and Dr. Andrew Poole. Thank you for your training, technical help, helpful discussions, endless favours, tasty cakes and that you were always there for me. I would not be here without you. I would like to thank Prof. Roslyn Bill and Dr. Stephane Gross for kindly sharing with me yeast strains and vectors. A special thanks go to all past and present members of lab 347, especially to Pinar, Michelle, Sarah, Anj and Romez for your help and suggestions. Lastly, thank you to Allan and Laura for help with flow cytometry.

On a personal note, a big thank you goes to my friends Allan and David who have been there for me from the very beginning to the end. My thanks are also extended to Shaun, Laura, Floren, Madara and Ross, who I shared many laughs and trips to the pub. Thank you all for your words of support and encouragement in the past few years.

A considerable thank you to my family who gave me the best prep for life I could ask for and who were cheering for me through the whole process and never stopped believing in me. Not to forget my biggest cheerleaders of all time: Patrycja, Justyna and Renata. The three of you mean the world to me!

Finally, I would like to thank my dearest husband Tiago for the best support he gave me. It was not always easy but you have always taken care of me and made it possible for me to make it through to the end. I would not be here without you. Special thanks also go to my sweetest daughter Olivia for being the reason I would smile every day and gave me the motivation and power to complete this project.

List of Contents

List of Abbreviations

°C	degree Celsius
+ ctrl	positive control
A	adenine
A. fulgidus	*Archaeoglobus fulgidus*
AA	natural amino acid
aaRS	aminoacyl-tRNA synthetase
A. victoria	*Aequorea victoria*
Ala	alanine
AlaRS	alanyl-tRNA synthetase
alaS	AlaRS gene
Amp	ampicillin
AMP	adenosine monophosphate
ATP	adenosine triphosphate
bp	base pair
BSA	bovine serum albumin
C	cytosine
C666A_AlaRS	full sequence AlaRS with point mutation C666A
cDNA	complementary DNA
CSM	Complete Supplement Mixture
ctrl	control
D	guanine or adenine or thymine
D-AA	D forms of amino acids
ddH_2O	double distilled water
DEPC	Diethylpyrocarbonate
dH_2O	distilled water
DMF	dimethylformamide
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide
DTT	dithiothreitol
dsDNA	double-stranded DNA
E. coli	*Escherichia coli*
EF-Tu	Elongation factor Tu

EDTA	ethylenediaminetetraacetic acid
eGFP	enhanced green fluorescent protein
$eGFP_{amber39}$	eGFP with amber stop codon in position 39 of the protein
FACS	Fluorescence-activated cell sorter
FRET	fluorescence resonance energy transfer
FW (e.g. FW1)	framework fragment
FW4_N461	framework 4 of N461 library
FW4_C666A_875	framework 4 of full sequence with point mutation C666A library
G	guanine
GFP or wtGFP	green fluorescent protein or wild type green fluorescent protein
Gly	glycine
I	inosine
IPTG	isopropyl-β-D-thiogalactoside
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HF	high fidelity
HpL	HypperLadder, eg HpL I
HRP	horseradish peroxidase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
K	guanine or thymine
kb	kilobase
KCl	potassium chloride
kDa	kilodalton
L	DNA ladder/marker
L-AA	L forms of amino acids
LAM	lower alignment marker
LB	lysogeny broth
LiAc	lithium acetate
Lib (e.g. Lib1)	library fragment
Lib_C666A_AlaRS	library for full-length AlaRS with point mutation C666A
Lib_N461_AlaRS	library for aminoacylation domain of AlaRS
M. barkeri	*Methanosarcina barkeri*
MARS	multi-tRNA synthetase complex
$MgCl_2$	magnesium chloride
$MgSO_4$	magnesium sulfate

mRNA	messenger RNA
MW	molecular weight
N or nt	nucleotide
N461_AlaRS	truncated AlaRS (residues 1-461)
NaCl	sodium chloride
NGS	next-generation sequencing
Ni-NTA	nickel-nitrilotriacetic acid
NTC	non-template control
OD_{600}	optical density at a wavelength of 600 nm
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PEG	polyethylene glycol
pET45b(+)	pET vector
pET45b(+):*alaS*	wild type AlaRS (full sequence) in pET
pET45b(+):C666A_875	full-length AlaRS with point mutation C666A in pET
pET45b(+):N461	aminoacylation domain only of AlaRS in pET
PBS	phosphate-buffered saline
PBS-T	phosphate-buffered saline with Tween™ 20
Pol III	RNA Polymerase III
pUC19	pUC19 vector
PVDF	polyvinylidene difluoride
Pyl	pyrrolysine
PylRS	Pyrrolysyl-tRNA synthetase
RNA	ribonucleic acid
rpm	revolutions per minute
rRNA	ribosomal RNA
RT-PCR	Reverse Transcriptase PCR
S	guanine or cytosine
S. cerevisiae	*Saccharomyces cerevisiae*
SD	Synthetic defined
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
Ser	serine
snRNA	small nuclear RNA

SOC	super optimal broth
SOP	standard operating procedure
ssDNA	single-stranded DNA
T	thymine
TAE	Tris-Acetate-EDTA
TBE	Tris-Borate-EDTA
tRNA	transfer RNA
UAA	unnatural amino acid
v/v	volume by volume
w/v	weight by volume
WCL	whole cell lysate
wt	wild type
wtAlaRS	wild type AlaRS
wt875_AlaRS	wild type full sequence AlaRS
x *g*	times g-force
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
YE	yeast extract medium
YNB	yeast nitrogen base
YPD	yeast peptone dextrose

List of Tables

List of Figures

Chapter 1: Introduction

1.1. The importance of translation

During the early studies on protein translation, Francis Crick in his paper "On protein synthesis" (Crick, 1958) gathered all up-to-date research and developments towards understanding the process of protein generation. He outlines the early experimental evidence proving former theoretical speculations surrounding the subject and considers the universal formation of proteins from amino acids across all living organisms and the control of the process. At this time in the past, little was also known about the site of protein synthesis with possibilities including the nucleus, cytoplasm or both locations. Nevertheless, the vague idea of the process was forming into what is known today as the basic principles of translation. Crick writes about microsomal particles, their association with amino acids and abundance in cells, they are later to be named as ribosomes. At this stage, the family of "activating enzymes" and their role in binding and activating its cognate amino acids for protein synthesis were already known, as illustrated in Figure 1.1.

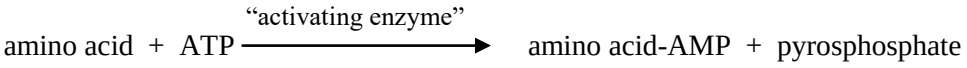


Figure 1.1. The reaction of amino acid activation.

The reaction is catalysed by aminoacyl-tRNA synthetase, here called "activating enzyme".

Crick also accurately reports that the activated amino acid has phosphorylated carboxyl group and in this form is only found when bound with its enzyme. The discovery of aminoacyl-tRNA synthetases (aaRSs) in the cell-sap fraction of rat liver cells was published as early as 1955 by Hoagland (1955) and colleagues, and by Berg in 1956 in yeast cells (1956). Subsequently, more publications reported the discovery of the same enzymes in bacterial cells and finally in animals.

Translation is the process of protein synthesis which takes place in the cytoplasm and uses amino acids as monomers for building a polypeptide chain. The generation of new polymer is very complex with many different components and careful control by various translation-associated factors. The fundamental role in protein generation belongs to different types of ribonucleic acids (RNA): messenger RNA (mRNA) – a carrier of genetic information for the sequence of protein, transfer RNA (tRNA) – a carrier of amino acid to the site of protein synthesis, and ribosomal RNA (rRNA) which makes up ribosome, an organelle which coordinates addition of amino acids (Figure 1.2.). Translation process is precisely controlled and regulated by different factors such as initiation or elongation factors. A very important molecule that also facilitates the process is enzyme aminoacyl-tRNA synthetase which catalyses the addition of amino acid onto 3'-end of tRNA.

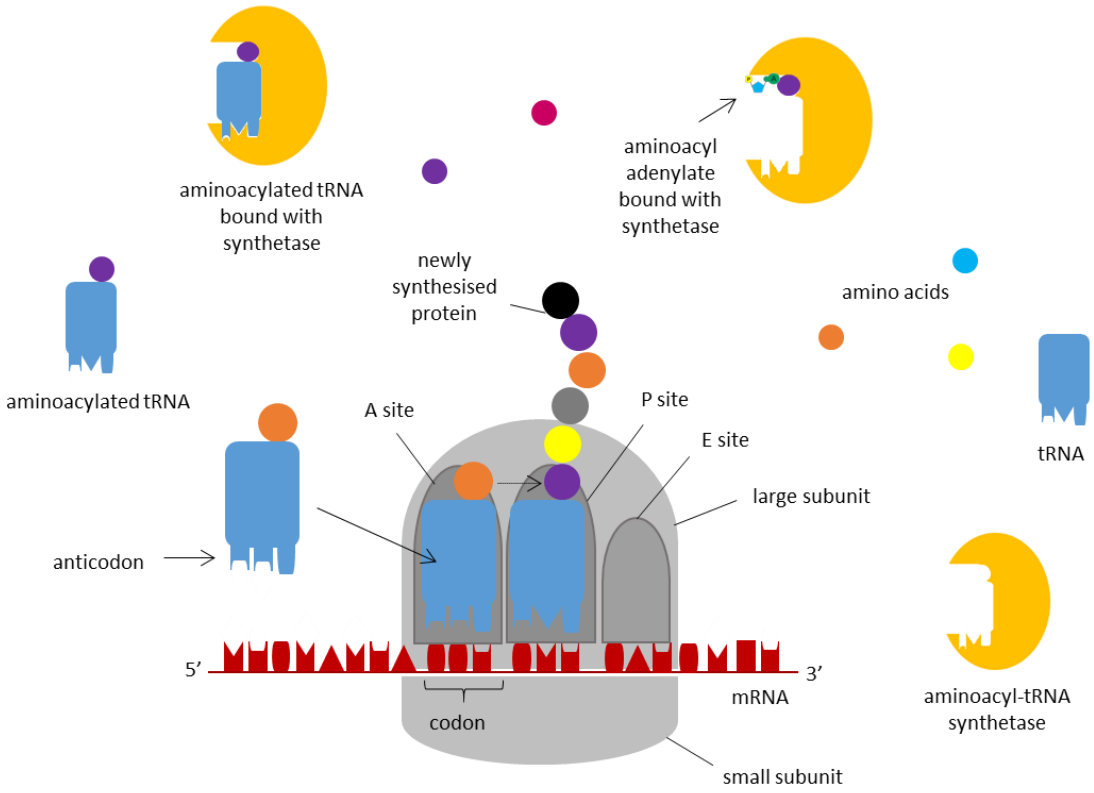


Figure 1.2. The process of protein synthesis.

Aminoacyl-tRNA synthetase (in yellow) activates amino acids to aminoacyl adenylate intermediates that are then attached to tRNAs (in blue). Aminoacylated tRNAs transport attached amino acids to the ribosome (in grey) which reads the sequence of mRNA and translates it into protein. Charged tRNA occupies A site in the ribosome, followed by P site and E site. Translation takes place in the cytoplasm.

The cellular translational machinery uses only natural amino acids (AAs) in their L-forms or achiral amino acid (Glycine only) to build polypeptide chain. It would be very valuable to use unnatural amino acids (UAAs) in their D-forms (mirror images to AAs as seen in Figure 1.3.) to expand the protein repertoire during synthetic protein synthesis for different applications in life sciences. D-AAs are not naturally recognised by aaRSs since evolution favoured L-AAs. This means that proteases might not be able to recognise D-AAs, so potentially preventing cleavage of proteins where the unnatural D-form is found. This gives the potential for new drug candidates with the possibility of longer half-life and better fit to the target receptors. Using D-AAs in proteins would also be useful in synthetic biology, protein engineering and other life science areas.

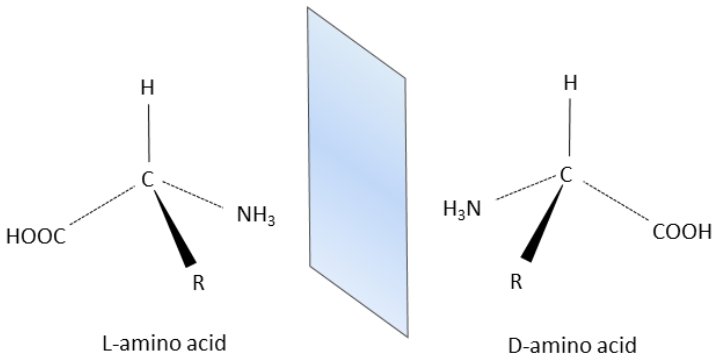


Figure 1.3. Enantiomers of amino acids.

Enantiomers are amino acids that are mirror images of each other such as naturally occurring L-amino acid and unnatural D-amino acid. R- functional group.

1.2. tRNA

1.2.1. History of discovery

The initial speculations about a single-stranded RNA molecule that carries attached amino acids to a complementary template strand for protein synthesis were proposed in 1955 in the letter by Francis Crick "On Degenerate Templates and the Adaptor Hypothesis" (Berg, 1956; Barciszewska, Perrigue, and Barciszewski, 2016). More work towards the discovery of tRNA was later done by Paul Zamecnik in 1958 (Hoagland, Stephenson, Scott, Hecht, and Zamecnik, 1958). The very first experimental studies proving that monomers of proteins, amino acids, were activated to aminoacyl adenylates involved in catalytic reactions by enzymes (later discovered as aaRSs), were conducted as early as 1955 by different research groups. These experiments involved charging single-stranded small RNAs with radioactive amino acids, what ultimately, lead to the discovery that enzymes associated required ATP-AMP exchange during catalysis and that there are many specific enzyme species needed for activation of different amino acids and binding to again specific tRNAs. The discovery of the family of aaRSs was published in 1956 by Paul Berg and his co-workers (Berg, 1956; Barciszewska *et al.*, 2016).

1.2.2. The primary, secondary and tertiary structure of tRNA

The studies on tRNA structure commenced in 1956 by Robert Holley (Holley, 1963, 1968; Barciszewska *et al.*, 2016). The very first tRNA that he was working with was $tRNA^{Ala}$ from *Saccharomyces cerevisiae* and the research was focused on resolving its sequence and relationship between nucleotides in terms of Watson-Crick base pairing. It took nine years to identify the sequence of $tRNA^{Ala}$ yet it still had to be corrected a few years later. Robert Holley was awarded the Nobel Prize in 1968 for his findings (Holley, 1968). Based on the first sequence of $tRNA^{Ala}$, the secondary structure of the cloverleaf model was suggested by Elisabeth Keller. The model was later confirmed by others including Tom RajBhandary for $tRNA^{Phe}$ from *S. cerevisiae* (Figure 1.4.) (RajBhandary, Stuart, Faulkner, Chang, and Khorana, 1966). After many years of research, it was

established that all tRNA molecules share the same cloverleaf model of secondary structure based on common Watson-Crick base pairing. Moreover, all tRNAs have conserved CCA-3' sequence at the end of the acceptor arm, an extension of double-stranded acceptor stem, as well as four hairpin helices including three major loops and one variable loop (Figure 1.4.). One of the major loops is called the anticodon loop responsible for interaction with mRNA that determines next amino acid in the protein sequence.

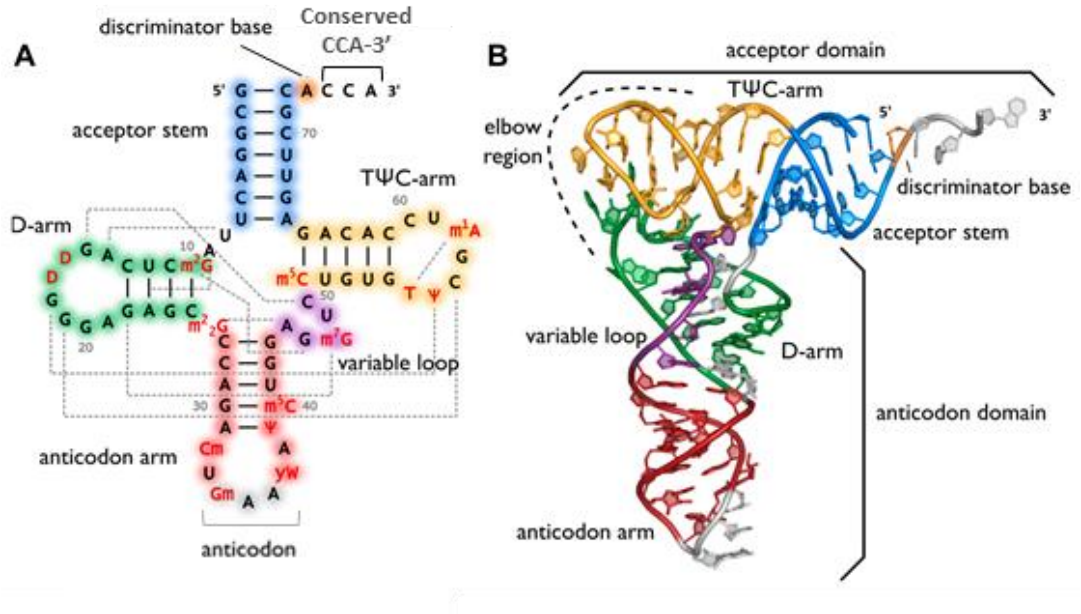


Figure 1.4. The secondary and tertiary structure of tRNA.

The illustrated tRNA is the first-ever sequenced $tRNA^{Phe}$ from S. cerevisiae showing the colour-coded details of its structure. A. The cloverleaf model of secondary structure $tRNA^{Phe}$. The discriminator base for this tRNA is highlighted in orange, anticodon in grey and post-transcriptional modification are in red. The grey dashed lines show interactions within tRNA in the tertiary structure. B. The L-shaped tertiary structure of $tRNA^{Phe}$ with corresponding colour-coded structures to the secondary structure (A.). Adapted from Lorenz, Lunse and Morl (2017).

The first tertiary structure of tRNA, specifically $tRNA^{Phe}$, was resolved in 1973 by Alexander Rich and his group at a resolution of 4 Å (Kim *et al.*, 1973). The work continued and allowed to obtain a better resolution in later years. The structural research confirmed that all tRNAs have two helices in the tertiary structure with L-shape folding (Figure 1.4.). The three-dimensional structure is stabilised by base triplets and non-Watson-Crick base pairing. In this L-shaped form, the distance between the anticodon loop and 3' attachment site of amino acid is of over 70 Å, providing a binding domain for aaRSs. The tRNA generally remains in this form upon binding with aaRSs and other proteins with some tRNAs undergoing slight bending (Barciszewska *et al.*, 2016).

1.2.2.1. Watson-Crick and wobble base pairing

Crick and Watson first described the molecular shape of DNA as double helix which is two strands of DNA running opposite to each other and forming a double-stranded molecule (Watson and Crick, 1953). DNA strands are antiparallel and while one runs in 5' → 3' direction, the opposite one runs in 3' → 5'. DNA comprises long stretches made of a combination of four nucleotides: two purines – guanine (G) and adenine (A) and two pyrimidines – cytosine (C) and thymine (T). They bind together via sugar-phosphate backbones forming hydrogen bonds in a complementary interaction between nucleotides to form base pairs. So-called Watson-Crick base pairing includes G binding to C and A to T. There are three hydrogen bonds between G and C, and two between A and T. In RNA strand, uracil (U) is substituted for thymine. Exceptions to the above rules are defined by wobble pairing (non-Watson-Crick pairing) (Eun, 1996).

Complementary base pairing occurs during DNA replication, transcription and translation. During protein synthesis, nucleotides form Watson-Crick base pairs during anticodon-codon recognition between tRNA and mRNA. As seen in Figure 1.5., complementary base pairing rules are strictly followed in the first and second position of a codon with the third and second position of an anticodon interaction. The third and the last position of the codon can also form a normal pair with the first position of the anticodon. However, it often does not follow this rule and instead forms a wobble pair

(Topal and Fresco, 1976). The most common non-Watson-Crick pair is G•U, a consequence of amino-imino tautomerism (Eun, 1996). It is a type of tautomerism in which the amino and the imino forms are in equilibrium and the migration of a hydrogen atom takes place. Other common wobble pairs include those with inosine (I), I•C, I•A, or I•U (Eun, 1996).

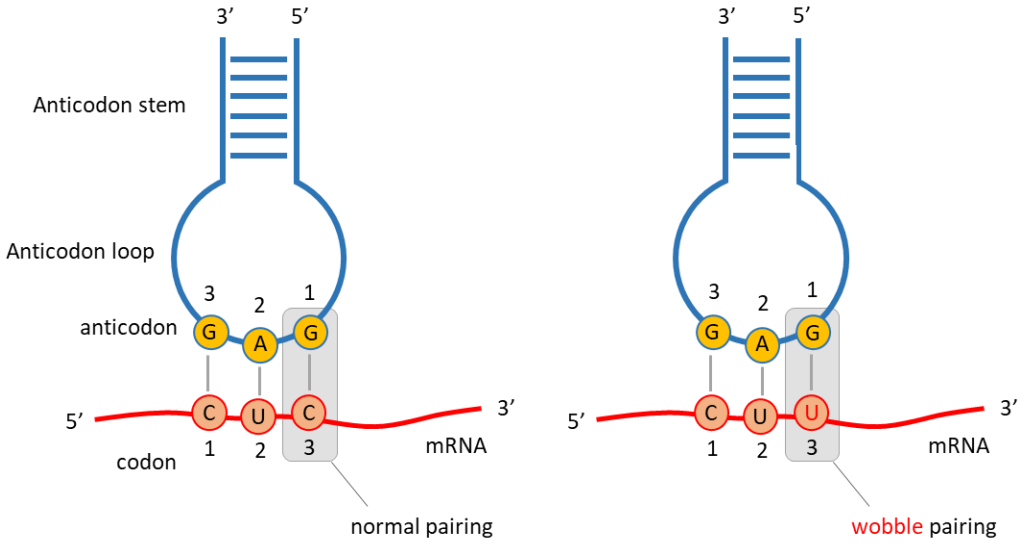


Figure 1.5. Wobble base pairing during an interaction between anticodon of tRNA and codon of mRNA.

Normal pairing refers to Watson-Crick base pairs: G with C and A with T. Wobble pairing is allowed at the third codon position and the first anticodon position.

Wobble pairs are important in the codon-anticodon interaction. They allow for fewer tRNA molecules to recognise multiple codons in the mRNA. However, this process is carefully controlled. During translation, elongation factor Tu (EF-Tu) first selects for cognate tRNA to bind at the A site of the ribosome during initial binding. At this stage, the wobble pairing may occur in the first or second position of the codon. However, later on, the ribosome proofreads the attached tRNA at the A site and if the base paring does not follow Watson-Crick rules at these positions, the ribosome

discards the tRNA. The conserved nucleotides of 16S rRNA G530, A1492 and A1493 are known to participate in this checkpoint. Once the cognate tRNA is accepted, the ribosome undergoes a conformational change. This change closes the ribosomal domain that discriminates against wobble pairs in the first or second codon positions but still accepts wobble in the third position. (Demeshkina, Jenner, Westhof, Yusupov, and Yusupova, 2012)

1.2.2.2. G•U wobble base pair

A G•U wobble base pair is also found in the $tRNA^{Ala}$ acceptor stem where it plays a critical role as a major determinant in recognition by alanyl-tRNA synthetase – a relationship that is going to be discussed in more details in Section 1.5.2. This base pair was initially discovered upon first sequencing of tRNA – yeast $tRNA^{Ala}$ (Holley *et al.*, 1965; Varani and McClain, 2000). During this discovery, it was shown that the primary sequence of $tRNA^{Ala}$ was folding into the secondary structure including the G•U base pair. The role of the same base pair was later discovered as the major discriminator in the recognition of $tRNA^{Ala}$ by the synthetase (Hou and Schimmel, 1988, 1989) and that the occurrence of this base pair is widely found in all tRNA species (Varani and McClain, 2000). In Figure 1.4., the G•U base pair is present in the acceptor stem of $tRNA^{Phe}$ as G4•U69, while in Figure 1.11. of $tRNA^{Ala}$ as G3•U70.

1.2.3. Generation of tRNAs

In eukaryotes, RNA Polymerase III (PolIII) is responsible for the transcription of tRNAs genes. PolIII also synthesises 5S rRNA and other small cytoplasmic and nuclear RNAs species. In short, this transcription is controlled by internal control regions which comprise different types of discontinued structures with essential sequences (including box A and box B) separated by nonessential ones. Alternative gene organisations can include an upstream promoter where box A and box B are found either within tRNA itself or are located in various places nearby (Guffanti *et al.*, 2006). The location of box A and B depends on the type of RNA transcribed by PolIII and what genes are involved in transcription. After tRNA transcription, the maturation process involves cutting

3’-trailer sequence with endo- and exonucleases and adding CCA-3’ conserved sequence. Moreover, the maturation includes trimming 5’-end nonessential sequence with RNase P, intron splicing, ligation and finally, post-transcriptional modifications of nucleosides within tRNA (Barciszewska *et al.*, 2016).

In bacteria, tRNA genes are often localised in clusters on the chromosomes. The transcription units for tRNA can be divided into tRNA genes only, tRNA genes within rRNA operons (specifically in the spacer sequence) and last but not the least, tRNA operons – with some of these also carrying genes like EF-Tu (Toh, Hori, Tomita, Ueda, and Watanabe, 2009). RNA polymerase recognises these different gene organisations and allows for transcription of tRNA with a single promoter. The first transcript of tRNA is called precursor and it has extra sequences that need to be trimmed to achieve fully mature tRNA that is ready to participate in translation (Altman, 1975). In *Escherichia coli* (*E. coli*), initially the 3’-end additional sequence is cut with several nucleases and after that, the 5’-end is processed by RNase P. Also, CCA-3’ is already present in the acceptor’s arm, unlike in eukaryotes. Due to a different copy number of tRNA genes, the number of transcripts respectively varies and can be influenced by promoter activity, abundant or deprived amino acid level, the growth rate of organism etc. (Toh *et al.*, 2009). Just like eukaryotic tRNAs, prokaryotic tRNAs undergo post-transcriptional modifications.

1.2.4. Function of tRNAs

The main and the most important function of tRNA is its participation and key role in protein synthesis. Up to 10% of the total RNA in the cell corresponds to tRNAs, making it the most abundant RNA. As mentioned, its key role is to transport amino acids from the cell to the site of protein translation in the ribosome. The anticodon of tRNA determines binding of the molecule with the attached amino acid to the complementary sequence within mRNA. In regards to nontranslational functions, tRNA participates in the regulation of gene expression during deficiency of amino acids. It also has a role in amino acid biosynthesis and energy metabolism. Moreover, mutations to tRNA

sequence can lead to disease and dysfunctions in protein synthesis as well as other cellular processes (Barciszewska *et al.*, 2016). The examples include encephalopathies such as MELAS (mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke-like episodes) or MERRF (myoclonus epilepsy associated with ragged-red fibres). While other diseases reduce respiratory capacity, leading to increases in the level of reactive oxygen species and consequently, leading to hypertension (Kirchner and Ignatova, 2014).

1.3. The family of aminoacyl-tRNA synthetases

Aminoacyl-tRNA synthetase (aaRS) is a key enzyme in protein synthesis. It catalyses the addition of a cognate amino acid to a corresponding tRNA (Maršavelski, Lesjak, Močibob, Weygand-Đurašević, and Tomić, 2014). This process is called aminoacylation and it occurs in a two-step reaction. Its initial phase involves binding the amino acid and ATP, followed by activating the amino acid with ATP to form an intermediate aminoacyl-AMP (Figure 1.1.). Subsequently, the enzyme recognises its cognate tRNA, and the aminoacyl moiety is attached by the covalent bond to adenosine of the 5'-CCA-3' in the tRNA acceptor arm. Once tRNA is charged with the amino acid, it leaves the enzyme and goes to the site of protein synthesis. This process is illustrated in Figure 1.6.

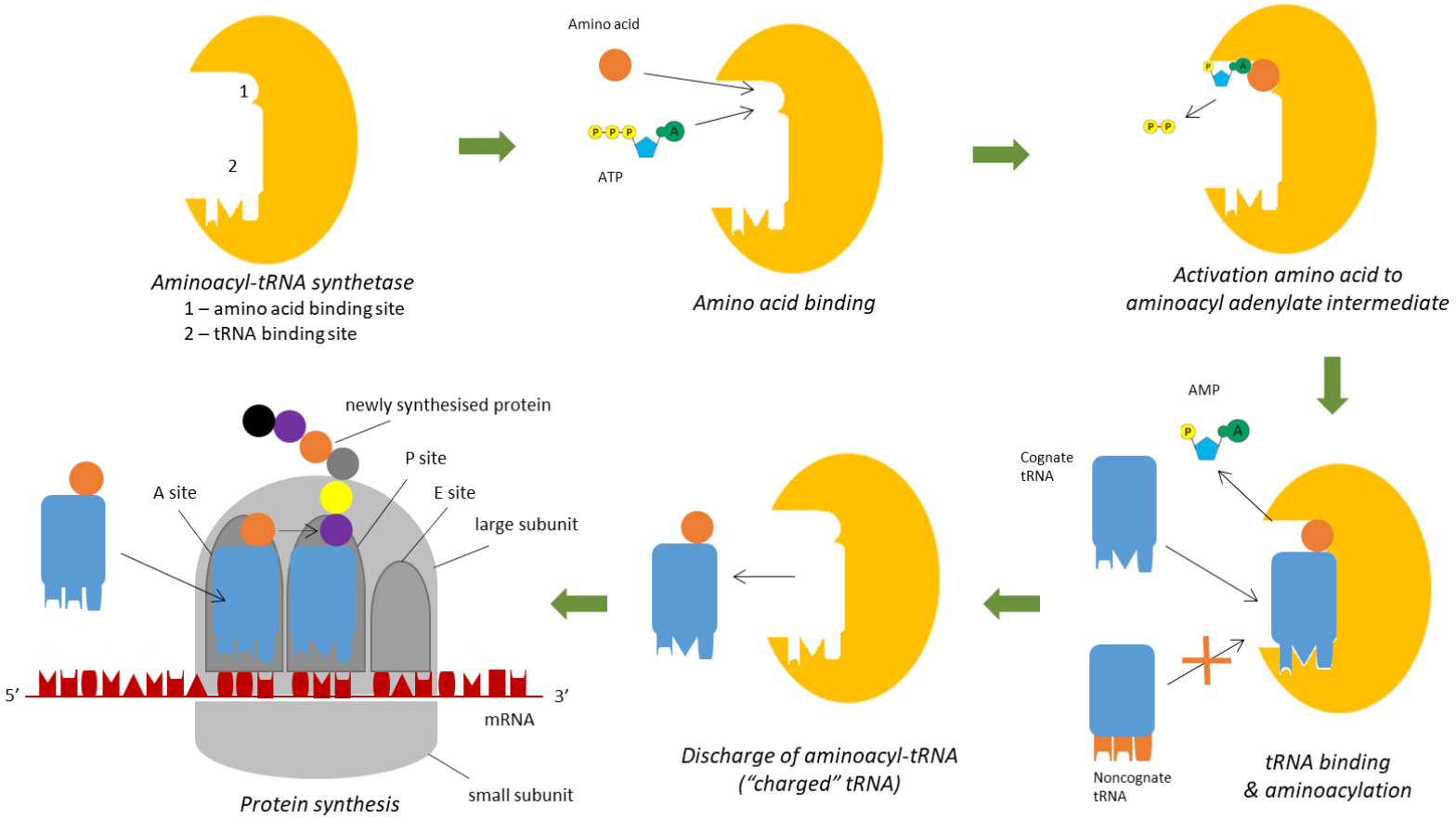


Figure 1.6. Aminoacylation reaction catalysed by aminoacyl-tRNA synthetase.

Firstly, an amino acid is activated to form aminoacyl adenylate. Secondly, the amino acid is attached to tRNA what is catalysed by aminoacyl-tRNA synthetase.

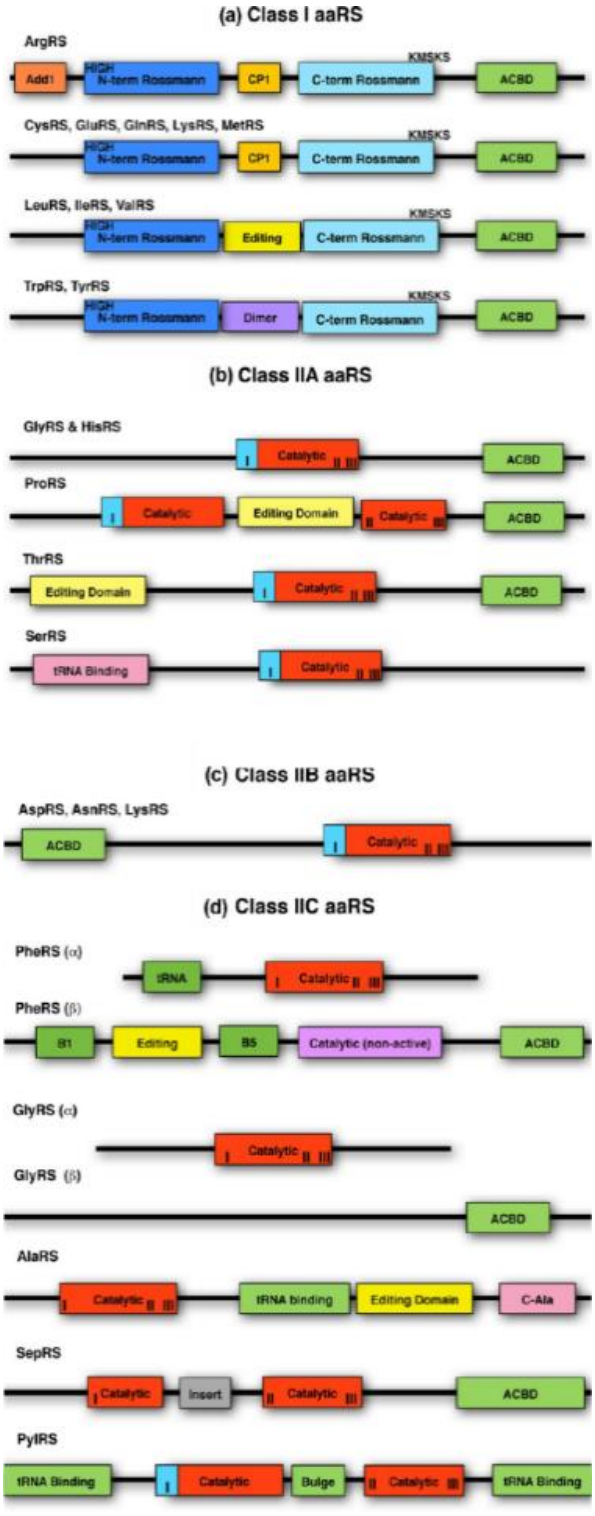
In *E. coli*, the full set of enzymes includes twenty different aaRSs, with one enzyme per amino acid. Over a number of decades, the research on these catalysts from various organisms was focused on understanding the structure-function correlation, nontranslational functions and other related questions. The research on these enzymes proved to be more challenging than first anticipated. Although all enzymes are involved in aminoacylation and share common domains and motifs, overall their structure varies between enzymes activating different amino acids but also between enzymes activating the same amino acids found in different organisms. This phenomenon is illustrated in Figure 1.7. Besides, eukaryotic aaRSs differ from bacterial and often are made of more subunits since they may interact with larger complexes such as multi-tRNA synthetase complex (MARS) (Perona and Hadd, 2012). The family of enzymes is divided into class I and II, based on the two distinct types of the aminoacylation domain related to the attachment of amino acid. In class I, the amino acid is attached by its carbonyl group to the 2'-OH of the ribose of the terminal adenosine A76 of tRNA. While in class II enzymes, the site of amino acid's attachment is 3'-OH. The two different sites of attachments are caused by various amino acids' features such as hydrophobicity of large amino acids catalysed by class I aaRSs (Tamura and Hasegawa, 1999).

The enzymes from class I are more structurally related to each other than enzymes from class II. They are mostly monomers. The general structure of class II enzymes (the interest of this study) includes seven-stranded β-sheet flanked by α-helices and they can be homodimers (class IIA and B) or the mixture of homo- and heterodimers (class IIC) (Perona and Hadd, 2012). Enzymes from class II have three conserved motifs in the active site domain. Motif 1 is made of long α-helix with a conserved proline at its end and is involved in dimer formation in most enzymes. The other two conserved motifs are involved in the reactions of the active site and are arranged as antiparallel β-strands connected by a loop. While motif 3 binds ATP required for activation of amino acid, motif 2 is responsible for coupling this ATP with the amino acid, and binding the resulting aminoacyl adenylate to the 3'-end of tRNA. In class II enzymes, the major groove of the tRNA acceptor stem is positioned facing the aminoacylation domain. While the process of tRNA binding is complex in terms of structural changes, understanding what is happening during aminoacylation is even more

challenging, with incomplete information for most of the enzymes or little information at all for others. As such, the process of aminoacylation varies between different enzymes. During aminoacylation, with class II enzymes, the tRNA is first recognised and bound with the enzyme via its major groove. Conversely, class I enzymes bind their cognate tRNA via the minor groove of the acceptor stem (Ruff *et al.*, 1991; Penny J. Beuning, Gulotta, and Musier-Forsyth, 1997). The differences between the organisation of protein domains between two classes but also within class II enzymes are shown in Figure 1.7a. As this was not complicated enough the same type of enzyme, for example, Phenylalanyl-RS (PheRS), in different organisms also show major differences in the crystal structures (Figure 1.7b.).

Difficulties in understanding the aminoacylation process are restricting the progress of synthetases in research areas including protein engineering and synthetic biology. Through extensive studies on this family of enzymes, it was discovered that apart from the role in translation, aaRSs have many other functions. For instance, Tyrosyl-RS is also involved in inflammation reactions (Lang *et al.*, 2014).

a.


(a) Class I aaRS
ArgRS
Add1
HIGH
N-term Rossmann
CP1
KMSKS
C-term Rossmann
ACBD
CysRS, GluRS, GlnRS, LysRS, MetRS
HIGH
N-term Rossmann
CP1
KMSKS
C-term Rossmann
ACBD
LeuRS, IleRS, ValRS
HIGH
N-term Rossmann
Editing
KMSKS
C-term Rossmann
ACBD
TrpRS, TyrRS
HIGH
N-term Rossmann
Dimer
KMSKS
C-term Rossmann
ACBD
(b) Class IIA aaRS
GlyRS & HisRS
Catalytic
ACBD
ProRS
Catalytic
Editing Domain
Catalytic
ACBD
ThrRS
Editing Domain
Catalytic
ACBD
SerRS
tRNA Binding
Catalytic
(c) Class IIB aaRS
AspRS, AsnRS, LysRS
ACBD
Catalytic
(d) Class IIC aaRS
PheRS (α)
tRNA
Catalytic
PheRS (β)
B1
Editing
B5
Catalytic (non-active)
ACBD
GlyRS (α)
Catalytic
GlyRS (β)
ACBD
AlaRS
Catalytic
tRNA binding
Editing Domain
C-Ala
SepRS
Catalytic
Insert
Catalytic
ACBD
PylRS
tRNA Binding
Catalytic
Bulge
Catalytic
tRNA Binding

b.

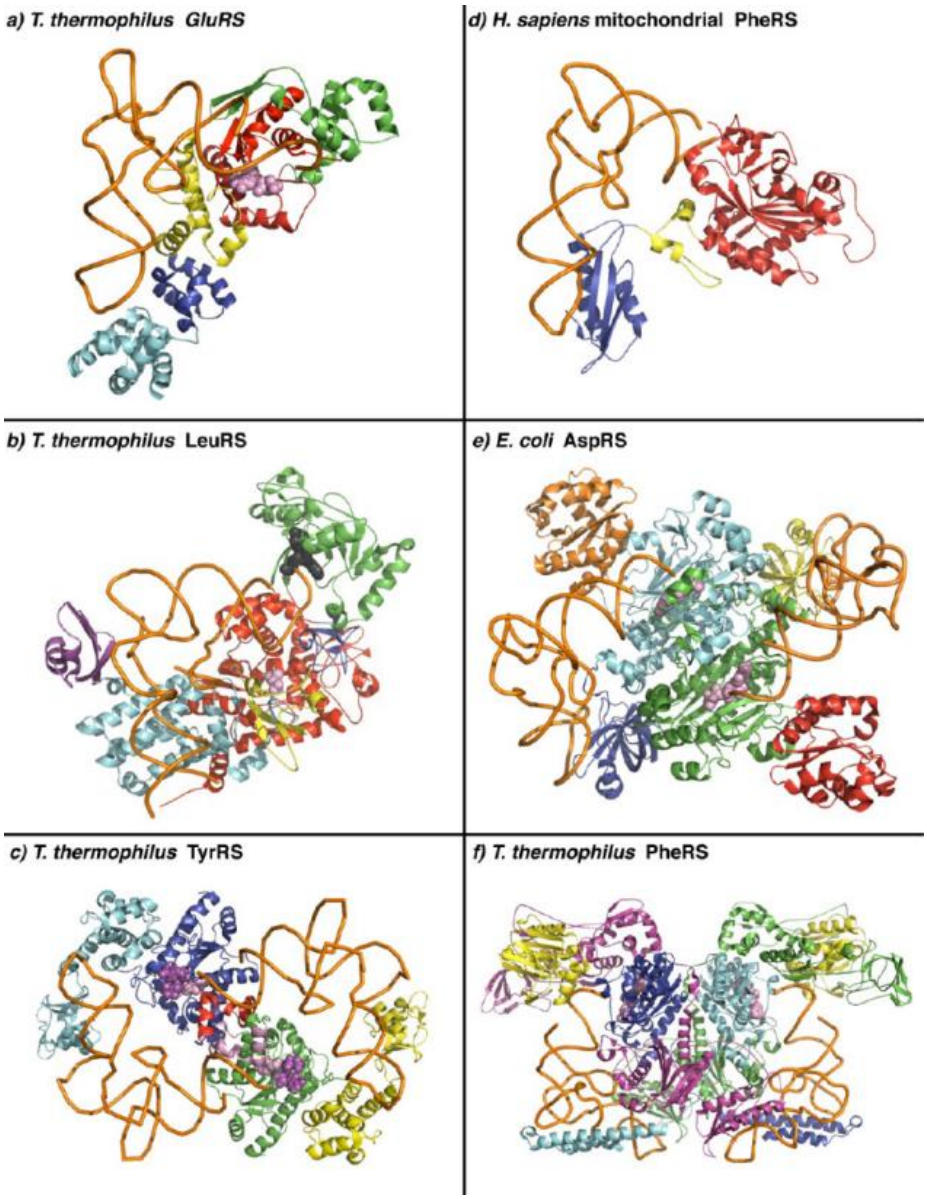


Figure 1.7. The family of aminoacyl-tRNA synthetases: division into classes.

a. The schematic image of the family of aaRSs show differences between a set of enzymes in terms of organisation of the same domains in the protein, extra domains in some enzymes probably associated with an extra function of this enzyme The family is divided into class I and II, and subclasses within class II. b. The variety of structures from aaRS family across different species bound with tRNA (in orange) and ATP (in pink). The colour coding is connected to the same domains (Perona and Hadd, 2012).

1.4. *E. coli* Alanyl-tRNA synthetase

1.4.1. Structure and function

The focus of this project is an enzyme alanyl-tRNA synthetase (AlaRS) from the bacterium *E. coli*. AlaRS belongs to the class IIC family of aaRSs. In *E. coli*, it is composed of 875 amino acids (Beebe, Mock, Merriman, and Schimmel, 2008) and its molecular weight is 96 000 DA (Sood, Slattery, Filley, Wu, and Hill, 1996; Sood, Hill, and Slattery, 1997).

1.4.1.1. Domains of *E. coli* AlaRS

E. coli AlaRS has three domains: the N-terminal catalytic domain, the editing domain and the C-terminal domain (Figure 1.8.). The catalytic domain is 461 amino acids long and is responsible for aminoacylation and tRNA recognition via G3:U70 base pair recognition in the acceptor stem of tRNA. Sometimes this domain is additionally divided based on its aminoacylation function and tRNA recognition as shown in Figure 1.8. (Beebe *et al.*, 2008). The editing domain is responsible for the removal of mischarged amino acids by hydrolytic editing. The core editing centre is found in the region of residues 553-705. AlaRS can incorrectly activate serine (Ser) and glycine (Gly) because of their structural similarity to cognate alanine (Ala) as shown in Figure 1.9. (Guo *et al.*, 2009). Once Ser or Gly is activated, the enzyme can even charge them onto $tRNA^{Ala}$. The last domain of AlaRS is the C-terminal domain. One of its functions is oligomerisation but others are not well described because it is a poorly-defined region of the enzyme. However, in mutated AlaRS without the C-terminal domain, both aminoacylation and the editing activity are reduced. Unfortunately, the exact mechanism of that reduction is still unknown (Dignam *et al.*, 2011).

Unlike eukaryotic AlaRSs which are monomers, in archaea and bacteria including *E. coli*, the enzyme forms a dimer of identical subunits in solution. Sedimentation experiments (Dignam *et al.*, 2011) and results from other studies using dynamic light scattering (Sood *et al.*, 1996) and analytical ultracentrifugation (Sood *et al.*, 1997) confirm dimerisation of AlaRS in *E. coli*.

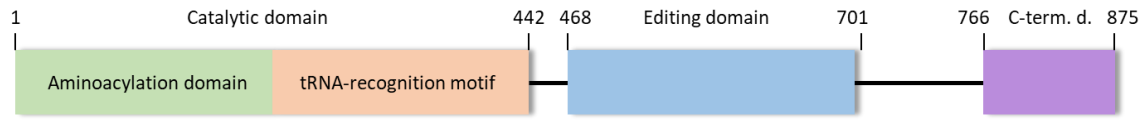


Figure 1.8. Three domains of full-length E. coli AlaRS.

Full-length E. coli AlaRS includes three domains: catalytic domain, 1-442, compromising aminoacylation domain and tRNA-recognition motif; editing domain, 468-701; C-terminal domain, 766-875. Adapted from Guo, Shapiro, Schimmel and Yang (2010).

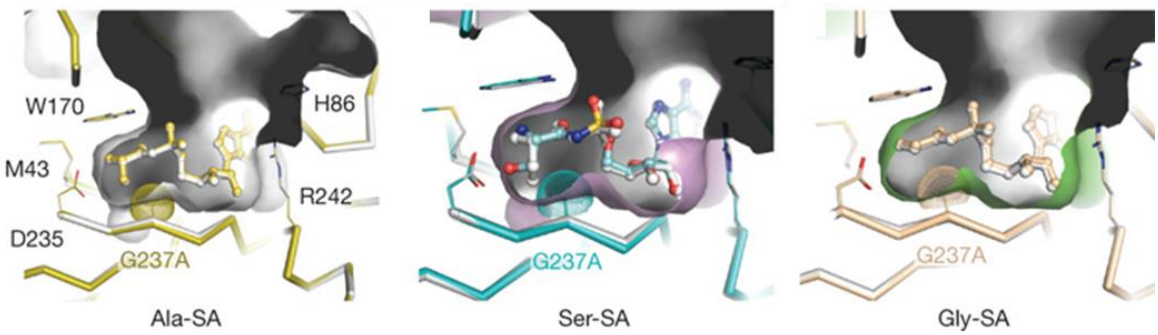


Figure 1.9. Recognition of Ala, Ser and Gly by E. coli AlaRS catalytic domain.

The figure shows the amino acid binding pocket of E. coli AlaRS catalytic domain with mutation G237A. The mutated catalytic domain still maintains binding and activation of three amino acids: Ala, Ser and Gly. White schematics represent the wild type catalytic domain of AlaRS, while coloured (gold, teal, peach) mutated enzyme. The grey active site cavity belongs to the wild type enzyme for simplicity of understanding the diagrams (Guo et al., 2009).

The full crystal structure of AlaRS in *E. coli* has not yet been resolved. This is common for all enzymes from this family and one of the reasons is that the enzyme undergoes structural changes when it binds Ala, ATP and $tRNA^{Ala}$ (Swairjo *et al.*, 2004). These are very dynamic changes taking place in a short time, and consequently, are difficult to characterise. For instance, complex interactions between editing and aminoacylation domains lead to a variety of geometrical changes that cannot be captured in the crystal structure of AlaRS (Dignam *et al.*, 2011). However, the crystal structure of the catalytic domain AlaRS of *E. coli* is available, both in free form and bound with activated Ala (Guo *et al.*, 2009).

Figure 1.10. demonstrates the crystal structure of the catalytic domain with Ala-SA (5'-*O*-[*N*-(L-alanyl)-sulfamoyl]adenosine), an adenylate analogue. The Ala-SA is bound in the middle of the active site of the aminoacylation domain (darker shade of blue) with tRNA recognition motif on the other site (lighter shade of blue).

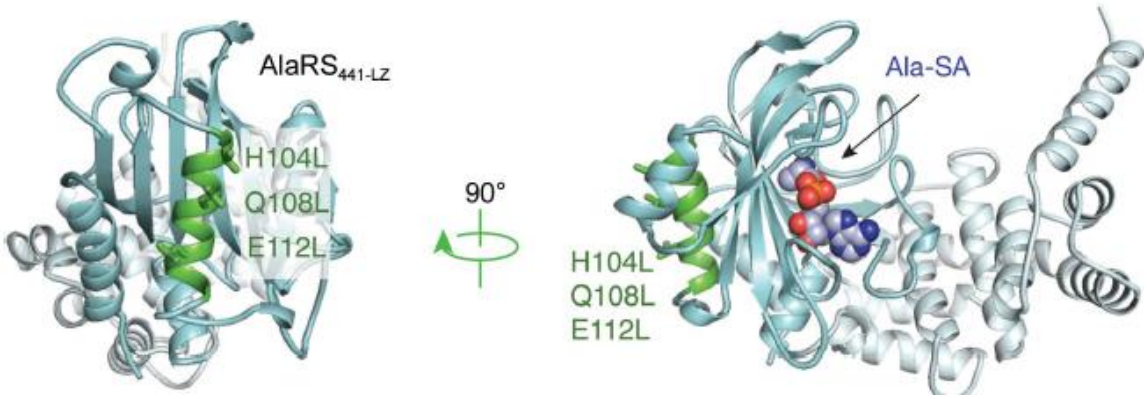


Figure 1.10. The crystal structure of the catalytic domain of E. coli AlaRS$_{441\text{-}LZ}$ with attached Ala-SA (an adenylate analogue).

The aminoacylation domain is in blue and the tRNA recognition domain in light blue. The Ala-SA is bound by the aminoacylation domain in the amino acid binding site. The green helix represents three leucine substitutions that were crucial for crystallisation (Guo et al., 2010).

1.5. The system of *E. coli* AlaRS and $tRNA^{Ala}$

1.5.1. The relationship between AlaRS and $tRNA^{Ala}$

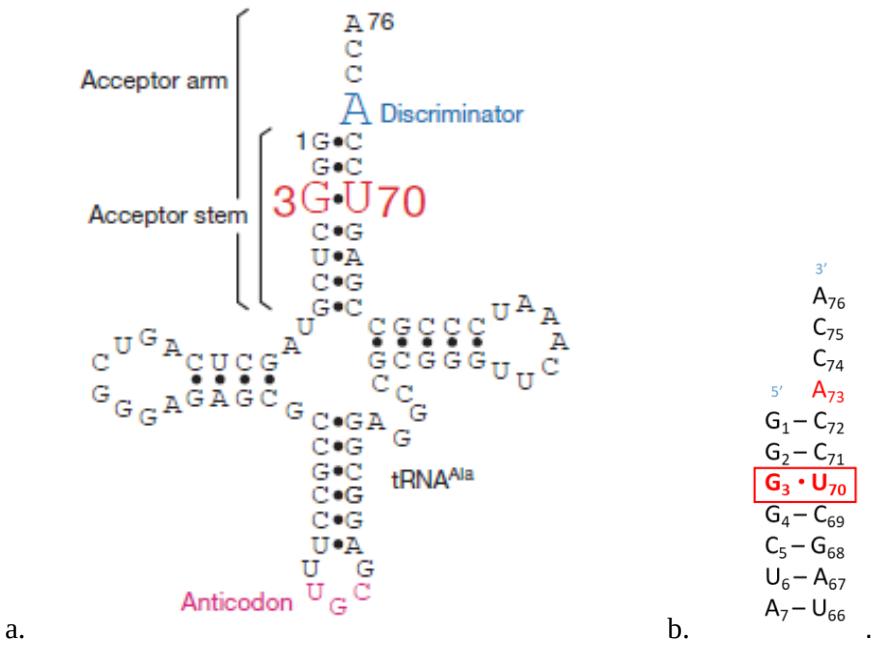
The cognate aaRS and tRNA pairs have sophisticated and precise recognition mechanisms that are unique for individual pairs to select correct molecules from the pool of enzymes, tRNAs and amino acids. The structure of tRNA contains identity determinants that are responsible for the recognition by cognate aaRS. Similarly, enzymes have conserved residues that are of higher or lower importance in the recognition of tRNAs. Both determinants cannot be changed because this would have adversely affected the tRNA recognition and the aminoacylation process (Cavarelli and Moras, 1993).

The orientation of $tRNA^{Ala}$ on the enzyme surface is influenced by the attachment of Ala and ATP to the binding sites (Dignam *et al.*, 2011). This is caused by the changes in the conformational shape of AlaRS which depend on the type of ligand bound. The 3'-end of $tRNA^{Ala}$ interacts with Lys74 found in the aminoacylation domain of AlaRS (Dignam *et al.*, 2011). From mutation studies, it was confirmed that this amino acid is not responsible for catalysing aminoacylation but it is important for alignment of tRNA with AlaRS (Hill and Schimmel, 1989). Therefore, even if Lys74 does not participate in aminoacylation directly, its substitution led to a 98% decrease in the enzyme catalytic efficiency (Hill and Schimmel, 1989).

1.5.2. The importance of the G3:U70 base pair and other determinants in $tRNA^{Ala}$

It has been long established that AlaRS recognises only the $tRNA^{Ala}$ acceptor arm and that there is no interaction with the $tRNA^{Ala}$ anticodon. The same recognition pattern is also known for a few other aaRSs, including LeuRS and SerRS (P. J. Beuning and Musier-Forsyth, 1999). Naganuma *et al.* (2014) confirmed that AlaRS recognises G3:U70 wobble base pair found in the acceptor stem of $tRNA^{Ala}$. Although the experiments are on the AlaRS-$tRNA^{Ala}$ system from *Archaeoglobus fulgidus* and not from *E. coli*, the regions of interests are conserved between these species (Figure 1.11.).

There are several base pairs from the acceptor stem which get involved in the interaction with the aminoacylation domain of AlaRS. The study found that the nucleotide G3 most likely forms hydrogen bonds with Asp450, while U70 with Asp450, Tyr449 and Gly453 in AlaRS from *A. fulgidus*. Generally, the region of residues 449-453 (Tyr-Asp-Ser-His-Gly) at the α14 tip fits tightly with the nucleotides in the acceptor stem of $tRNA^{Ala}$ (Figure 1.12.). This region is conserved within species and in *E. coli* corresponds to residues 398-402: Tyr-Asp-Thr-Tyr-Gly (Naganuma *et al.*, 2014).



a. b.

Figure 1.11. The cloverleaf model of $tRNA^{Ala}$ from A. fulgidus and acceptor arm of $tRNA^{Ala}$ from E. coli.

a. The structure has a distinguished anticodon loop, major discriminator G3:U70 and A73 discriminator (Naganuma et al., 2014). b. Acceptor arm from E. coli showing the same discriminators as molecule from A. fulgidus.

For several years, many studies confirmed that the wobble pair G3:U70 in $tRNA^{Ala}$ is a major determinant in AlaRS binding (Figure 1.11b.) (Hou and Schimmel, 1989). Moreover, the recognition of this base pair is so specific that the enzyme can recognise and aminoacylate different isolated tRNA fragments, for instance, a "minihelix" or a "microhelix", as long as they have the wobble pair in their sequence (Musier-Forsyth and Schimmel, 1999; Naganuma, Sekine, Fukunaga, and Yokoyama, 2009). Concerning $tRNA^{Ala}$, the microhelix and minihelix are synthetic, truncated species of tRNA molecule comprising either the acceptor arm only (microhelix) or the acceptor arm plus the TΨC arm and loop (minihelix). Importantly, both can be aminoacylated although neither has an anticodon loop. They are used in the studies with aaRSs, for example, to prove the importance of determinants found in the acceptor stem of $microhelix^{Ala}$ (Francklyn and Schimmel, 1989; Francklyn, Shi, and Schimmel, 1992). Mutations within $microhelix^{Ala}$ were also tested for aminoacylation by AlaRS. The first substitution of U2:A71 reduced but did not stop aminoacylation, while C2:G71 prevented it. Whereas the introduction of U1:A73 resulted in a very low but reproducible aminoacylation, the substitution G1:C73 blocked it (Francklyn *et al.*, 1992). Determinants specific for recognition of G3:U70 of $tRNA^{Ala}$ are found within the N-terminal domain of AlaRS, AlaRS(1-461) and include: Asp235, Asp285, Arg314, Ala409 (Beebe *et al.*, 2008). Moreover, it was found that for *in vitro* translation, the fragment AlaRS(1-461) consisting of the catalytic domain is satisfactory for the aminoacylation of $tRNA^{Ala}$ because it contains one of the tRNA recognition motif (Swairjo *et al.*, 2004). Also, the fragments of 76 or 93 fewer amino acids, 385N and 368N, respectively, can activate Ala to its aminoacyl adenylate intermediate but they do not have the power to aminoacylate $tRNA^{Ala}$. Therefore, a motif responsible for the tRNA recognition and aminoacylation must be present between 385-461 residues (Hill and Schimmel, 1989).

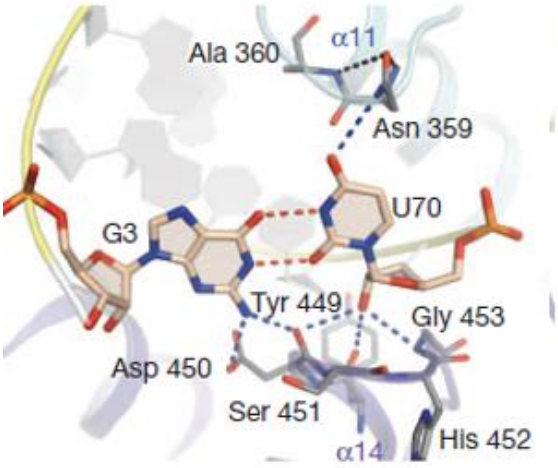


Figure 1.12. The interaction between the A. fulgidus AlaRS and tRNA^{Ala}.

The G3:U70 base pair hydrogen bonds with the residues in the α14 tip in the tRNA recognition domain (Naganuma et al., 2014).

Thermodynamic studies (Strazewski, Biala, Gabriel, and McClain, 1999) confirmed that the mutated tRNAs in which the wobble pair G3:U70 was substituted for Watson-Crick base pairs have higher stability but lower acceptor activity for Ala. Moreover, these mutated tRNAs have different geometric conformations from a wild type (wt) tRNA that prevent them from forming alanine:enzyme complex and so from charging tRNA with Ala (Strazewski *et al.*, 1999). It transpires that particular base pair (G3:U70) generates a deformation point in tRNA^{Ala} so it can fit within the enzyme. The mutated tRNAs do not have this property (Chang, Varani, Bhattacharya, Choi, and McClain, 1999). In *E. coli*, an additional determinant is the nucleotide A73, of which mutation does not stop aminoacylation as long as the wobble pair is present, but it decreases the efficiency of reaction considerably (Fischer, Beuning, and Musier-Forsyth, 1999). This nucleotide is also responsible for the stabilisation of the G1:C72 base pair (Fischer *et al.*, 1999). The above nucleotide and base pairs are illustrated in Figure 1.11.

1.5.3. $tRNA^{Ala}$ recognition elements within *E. coli* AlaRS

There are several independent $tRNA^{Ala}$ recognition sites found across different domains of *E. coli* AlaRS (Beebe *et al.*, 2008). In the studies using AlaRS enzymes without the catalytic domain, AlaRS(438-730) and AlaRS(438-875), both mutants were able to recognise $tRNA^{Ala}$ even when they lacked identified specific determinants supposedly responsible for binding to the G3:U70 base pair: Asp235, Asp285, Arg314 and Ala409. This leads to the conclusion that the specificity of mutated AlaRS was overseen by the acceptor stem of $tRNA^{Ala}$, and in fact, the specific determinants that are involved in the editing role are found in residues 438-730. Taking this further, the substitution of a conserved Arg693 for Lys693 found in the strand-loop-strand motif, increased specificity for $tRNA^{Thr}$ and loss of selectivity for $tRNA^{Ala}$ with about 20% lower rate of editing activity. This indicates that Arg693 is a critical determinant in the tRNA recognition for editing function of the enzyme, and it is highly conserved in AlaRS.

The recognition of tRNA by domains other than the aminoacylation domain is important, for example, for the editing function of the enzyme. In *E. coli*, the AlaRS editing centre is in the region of residues 553-705 and this is conserved through evolution. Those recognition sites located within the editing and C-terminal domains recognise misacylated Gly-$tRNA^{Ala}$ and Ser-$tRNA^{Ala}$. Since AlaRS readily misacylates Ser and Gly, the editing mechanism for the hydrolysis of mischarged amino acids from $tRNA^{Ala}$ has to be reliable. The problem of a relatively significant rate of misacylation by AlaRS was solved by adding free-standing AlaXps during evolution (Guo *et al.*, 2009). Thus, the recognition elements for $tRNA^{Ala}$ found within both the editing domain and AlaXps function as checkpoints for aminoacylation of correct amino acid in order to prevent mutations in proteins and their consequences. These checkpoints complement each other so if the editing mechanism of AlaRS does not clear Ser-$tRNA^{Ala}$, AlaXps will act instead. The selection pressure of AlaXps is really focused on deacylation of Ser as they do not always pick up misacylation of Gly-$tRNA^{Ala}$ during editing (Guo *et al.*, 2009).

AlaXps are free-standing fragments found in the cells alongside AlaRS. Thus, Type I AlaXp is homologous to *E. coli* AlaRS(438-730) including the core editing domain and extra flanking sequence. When used in the studies for clearance of Ser-tRNA^{Ala}, Type I AlaXp was inactive. Its inactivity was associated with the lack of tRNA^{Ala} binding and reduction in editing activity resulted from a loss of affinity for tRNA^{Ala}. However, at much higher concentration of Ser-tRNA^{Ala} (2000 x), the free-standing editing domain was capable of deacylating misacylated tRNA^{Ala}. Type II AlaXp consists of Type I AlaXp and an extended carboxy domain like in *E. coli* AlaRS, making it homologous to *E. coli* AlaRS(438-875). This protein is fully active during the clearance of Ser-tRNA^{Ala}. Even though AlaRS(438-875) lacks G:U specific determinants that are found within the aminoacylation domain, it still manages to clear mischarged tRNA^{Ala} with strong activity. As an example, this portion of the enzyme fails to deacylate Ser-tRNA^{Thr} but upon insertion of G3:U70 into tRNA^{Thr} it clears serine from Ser-tRNA$^{Thr}_{G3:U70}$. Additionally, R693K mutation of AlaRS(438-875) results in loss of selectivity for tRNA^{Ala} and about 20% lower rate of editing activity.

Summing up, the tRNA^{Ala} recognition elements are found across different motifs in AlaRS such as:

1) the aminoacylation domain: Asp235, Asp285, Arg314, Ala409,
2) the editing domain between 680-699 with key residue Arg693,
3) the C-terminal domain with nonspecific binding between 808-875.

Recognition in 1) and 2) are independent of each other because these domains can fulfil their functions when isolated from AlaRS enzyme.

For the purpose of the current study, the only required function of AlaRS is the aminoacylation of tRNA^{Ala}. The editing domain function is undesirable, since the aim of the project is precisely to engineer AlaRS to misacylate tRNA^{Ala} with UAAs. Fortunately, it has been reported previously that the catalytic domain is sufficient for tRNA charging (Swairjo *et al.*, 2004; Guo *et al.*, 2010). Therefore, the catalytic domain fragment of AlaRS(1-461) alone may be sufficient for the current

study. Alternatively, the editing function of AlaRS can be inactivated by mutating residue 666 from Cys to Ala (Guo *et al.*, 2009).

1.6. Engineering novel aaRSs via saturation mutagenesis

The key to engineering novel enzymes (here aaRSs) is to change key residues in the active site of the enzymes such that different substrates may be accepted. Where structural information concerning the enzyme is available, those key residues have already been identified. In that case, the next stage is to make as many changes as possible to those key residues. The resulting DNA library encodes a mixture of proteins (a protein library) which ideally contains every possible combination of amino acids in those key locations. The library is then screened against a substrate of interest with the hope that one or more of the variant enzymes will have specificity for the required, novel substrate.

Traditionally, mutations were made one at a time by a variety of techniques, but a change in substrate specificity normally requires mutations at several positions and multitudes of individual genes each containing multiple, single changes would be prohibitive in terms of time. Therefore, many years ago, the concept of saturation mutagenesis was introduced. Here, key codons are each replaced with a degenerate codon (typically NNN or NNK/NNS), to encode all 20 amino acids at each of the key locations. Within the resulting mixture, any one gene will be unique, but collectively, the library should include all possible combinations of codons and thus encode all possible combinations of amino acids at these key locations.

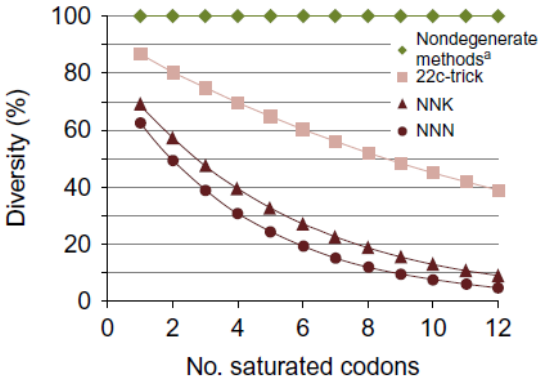
Unfortunately for the subsequent screening, the genetic code is degenerate. For NNN, each codon or nucleotide triplet results from a different combination of four bases: adenine (A), guanine (G), cytosine (C), and thymine (T). Altogether, there are 64 possible combinations of codons divided into 61 codons for 20 amino acids and 3 codons for termination codons. This means that one amino acid can be encoded by two codons (Cys, Asp, Glu, Phe, His, Lys, Asn, Gln, Tyr), three codons (Ile), four

codons (Ala, Gly, Pro, Thr, Val), or even six codons (Leu, Ser, Arg). Only two amino acids, Trp and Met are encoded by one codon. The degeneracy of the codon, when considered in making a gene library, can limit its diversity and consequently introduce bias. As such, a disproportionate content of codons has a direct influence on the library size (higher degeneracy increases the number of variants), which is a very important factor when choosing the method of the library screening in the subsequent steps because each method has its own limitation in terms of the number of variants. Additionally, the bias that is carried when using saturated codons will have an impact during the screening process, especially with ligand-based methods that assume an equal concentration of proteins in order to function properly.

1.6.1. Reducing degeneracy in saturated libraries

Several different approaches have been introduced in order to overcome the consequences of using degenerate codons and among these are NNK and NNS saturation codons (K=T/G, S=G/C). They aim to reduce redundancy by limiting the third base in the codon which decreases the number of combinations from 64 to 32 codons for 20 amino acids and one termination codon. This not only decreases redundancy but also reduces the numbers that are so important both in the library construction and screening processes. The next approach was the development of the 22c-Trick that uses only 22 codons for 20 amino acids (Kille *et al.*, 2013), greatly decreasing the redundancy and library size. The 22c-Trick relies on the subsets of oligonucleotides with limited degeneracy at specified codons (Kille *et al.*, 2013). There are three primers that aid generation of saturated codons in PCR: first contains NDT (A/C/G/T; A/G/T; T), second VHG (A/C/G; A/C/T; G), third codon TGG. If a combination of these primers is used, valine and leucine appear twice what gives the name to this methodology as 22c stands for 22 codons. In particular, it is important to optimise the annealing temperature of the primers to accomplish a good codon saturation (Kille *et al.*, 2013). Otherwise, the diversity of the library will be compromised. The consequences of these methods in comparison with fully nondegenerate saturation (exactly 20 codons encoding 20 amino acids) are compared in Figure 1.13.

A Diversity



B Encoded bias

No. saturated codons	Ratio most common:rarest codon combinations[b] NNN	NNK/NNS	22c-trick	Nondegenerate methods[a]
3	216:1	27:1	8:1	1:1
6	4.7×10^4:1	729:1	64:1	
9	1.0×10^7:1	2.0×10^4:1	512:1	
12	2.2×10^9:1	5.3×10^5:1	4096:1	

C Encoded truncation

No. saturated codons	NNN	NNK/NNS	22c-trick	Nondegenerate methods[a]
3	13	9		
6	25	17		0%
9	35	25		
12	44	32		

Figure 1.13. Comparison of performance of common saturation mutagenesis techniques.

Green colour shadowing indicates ideal performance, pale pink colouration indicates tolerable performance, and dark pink indicates unacceptable performance, where nondegenerate methods[a] may be created via various methodologies as described in Section 1.6.2. (A) Diversity was calculated using the formula $d = 1/(N\sum_k p_k^2)$ *(Makowski & Soares, 2003) and is in agreement for a 12-mer peptide saturated with codon NNN (Krumpe, Schumacher, McMahon, Makowski, & Mori, 2007). (B) Ratios[b] represent the theoretical relative concentrations of each individual gene combining any of the most common codons (Leu/Arg/Ser, NNN/NNK; or Leu/Val, 22c-trick) vs each individual gene containing any combination of the rarest codons (Met/Trp, NNN; Cys/Asp/Glu/Phe/His/Ile/Lys/Met/Asn/Gln/Trp/Tyr, NNK; or 18 codons (omitting Leu/Val), 22c-trick). (C) Truncation is calculated as the percentage of sequences that contain one or more termination codons within the saturated region (Ferreira Amaral, Frigotto, and Hine, 2017).*

When both constructing and screening a DNA library, its diversity, which is defined as the percentage of unique species, is imperative. However, when mutating only a single codon using a degenerate (NNN or NNK/NNS) method, the diversity is immediately decreased to 60-70%. With every one codon targeted, diversity drops drastically, resulting in only up to 10 % diversity when 12 codons are mutated. Even with near-nondegenerate saturation, the 22c-Trick, the diversity decreasing pattern follows degenerate methods with the number of targeted codons, resulting in 40 % diversity only for 12 targeted codons. Nevertheless, the diversity remains unchanged and equals 100 %, at least at a theoretical level, regardless number of targeted codons in nondegenerate mutagenesis (Figure 1.13A). Diversity leads directly to a second important measure in assessing library, namely encoded bias.

Bias is introduced to the protein library by encoding some amino acids with multiple codons and others with just one or two codons, according to the nature of the genetic code. However, in preparation for a gene library, there is a need for equal representation of amino acids and this can only be delivered by using nondegenerate saturation methods. The bias is also being enriched with the number of amino acids introduced. The ratios in the bias table (Figure 1.13B.) give the theoretical proportion for mutated codons found in a gene between combinations of the most abundant codons (such as six codons for Ser), and combination of the least abundant codons (such as Trp). Although Figure 1.13B. gives results of a bias with the maximum codon per amino acid, pointing out the worst possible consequence, it is apparent that major bias is introduced while using NNN/NNK/NNS or the 22c-Trick. Such bias is totally removed in nondegenerate methods because there is only one codon assigned for one amino acid during randomisation for example in MAX, ProxiMAX or Slonomics®. Because amino acids are not equally represented by the genetic code, if using all codons in the mutagenesis, gene sequences are unevenly represented. The last but not least, there is the risk of protein truncation while opting for degenerate methods of mutagenesis because they do not exclude the termination codons (Figure 1.13C.). This is especially important as nonfunctional truncated protein may aggregate inside the cells (Ferreira Amaral *et al.*, 2017).

In summary, nondegenerate saturation mutagenesis gives the possibility to include all twenty codons in theoretically equal ratio and some methods like ProxiMAX allow the user to select specific combinations of codons for targeted positions. At the same time, they exclude termination codons eliminating the risk of introducing gene sequences that can potentially give rise to truncated proteins. Moreover, the library size is kept to the minimum as it removes degenerate codons and instead of 64 uses only 20. This also means that the diversity or the value of unique DNA sequences is maximised to 100 %. All of these characteristics of nondegenerate methods allow for generation of diverse, high-quality libraries with a reduced pool of variants (all unique) which in turn results in improved efficiency of screening techniques (Tang *et al.*, 2012).

1.6.2. Nondegenerate saturation mutagenesis methodologies

1.6.2.1. TRIM Technology: Trinucleotide Phosphoramidites

The first-ever found method of nondegenerate mutagenesis was TRIM technology which uses trinucleotide phosphoramidites (Virnekas *et al.*, 1994). TRIM adds a triplet onto a growing nucleotide in a single reaction that is more efficient than adding one base. Ahead of reaction, a mixture of trinucleotide phosphoramidites is produced and used later during mutagenesis. Although the method is, at least in theory, controlled, in practice it has risks of introducing a biased mixture of phosphoramidites because single bases are known to have various coupling efficiency during oligo synthesis (Ho *et al.*, 1996). Later this bias can be further enriched during the mutagenesis step. To prevent this from happening, commercial kits with TRIM technology are available such as GeneArt Combinatorial DNA Libraries (Thermo Fisher Scientific). (Ferreira Amaral *et al.*, 2017)

1.6.2.2. MAX randomisation

One of the first published nondegenerate saturation mutagenesis that did not require specialised equipment or reagents but could be replicated in any lab was developed in Hine lab and called MAX randomisation (Hughes, Nagel, Santos, Sutherland, and Hine, 2003). This technology relies on delivery of selection oligos that hybridise to complementary DNA template while at the same time delivering saturation mutagenesis at the selected codons (Figure 1.14.). After hybridisation of selection oligos, they are ligated to seal the gaps between individual ones, followed by amplification in PCR to generate a randomisation cassette. Selection oligonucleotides are short, usually made of nine nucleotides (six nucleotides are conserved with the template while three correspond to one MAX codon) and can form a mixture of twenty in order to saturate defined position. Besides selection oligos, there are also two-terminal oligos that serve as primer sites for the last step of MAX randomisation which is amplification. (Ferreira Amaral *et al.*, 2017)

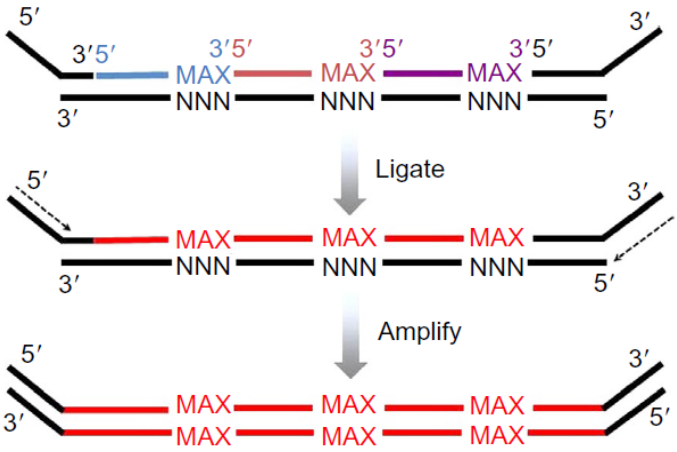


Figure 1.14. Overview of the MAX randomisation technique.

In MAX randomisation (Hughes et al., 2003) a single template oligonucleotide is synthesised that is fully degenerate at the designated, saturated codons. Meanwhile, a set of up to 20 small selection oligonucleotides are synthesised individually, for each saturated position. Each selection oligonucleotide consists of a short (typically in the order of 6 bp) addressing region that is fully complementary to the template and one MAX codon, where a MAX codon is the favoured codon for a single amino acid in the organism of interest. The selection oligonucleotides are mixed as required and, alongside two terminal oligonucleotides, are hybridised with the template and ligated together. The ligated strand is then selectively amplified with primers complementary to the terminal oligonucleotides, to generate a randomisation cassette (Ferreira Amaral et al., 2017).

The randomisation cassette generated in the process can later be used in the production of gene libraries via overlap PCR or be employed for site-directed mutagenesis as a double-stranded primer in QuikChange® mutagenesis (Agilent) or Q5® Site-Directed Mutagenesis Kit (NEB). Even though MAX randomisation offers saturation mutagenesis of multiple codons, only two contiguous codons can be targeted at the same time what is a limiting factor while using this technology for example for antibody libraries.

1.6.2.3. ProxiMAX randomisation

To permit nondegenerate saturation of multiple contiguous codons, the Hine lab next invented ProxiMAX randomisation. ProxiMAX is a nondegenerate saturation technology that like MAX randomisation, uses one codon only per amino acid (Ashraf *et al.*, 2013). The technology itself can be replicated and adapted for a small scale library production in any lab because it relies on common molecular techniques such as restriction digest with Type IIS enzymes, amplification and blunt-ended ligation (Figure 1.15.). The fundamental components of ProxiMAX are oligonucleotides that play a key role in delivering MAX codon at the site of randomisation. The process of randomisation consists of ligation, amplification and restriction digestion that form one cycle if more contiguous codons are to be changed. In ProxiMAX, one codon, called MAX, is added per cycle onto a conserved DNA section. This codon is a combination of a number of amino acids that are introduced at the mutation site and the ratio of amino acids as well as codon optimisation is fully controlled and decided on beforehand. As this is nondegenerate mutagenesis, one amino acid can only be expressed by one codon to eliminate degeneracy. MAX codons are delivered at 3'-end of double-stranded oligonucleotides which sequence is largely-conserved and can even be only partially double-stranded, or be self-complementary hairpins. The positions of randomisation, scattered or contiguous within a gene are controlled and can achieve full saturation. The technology was until recently available under commercial platform Colibra™ at Isogenica Ltd, which was partially automated giving outstanding results in gene libraries and advantage of adding two codons at the time (Frigotto *et al.*, 2015) instead of one in a manual setting. Nevertheless, fully manual ProxiMAX can be achieved in any lab as it does not require specialised reagents and it also offers a good result for the library (Poole, 2015). The advantage of using hexamer nucleotides as two MAX codons per cycle than just one is more efficient and allows for high-throughput experiments. (Ferreira Amaral *et al.*, 2017).

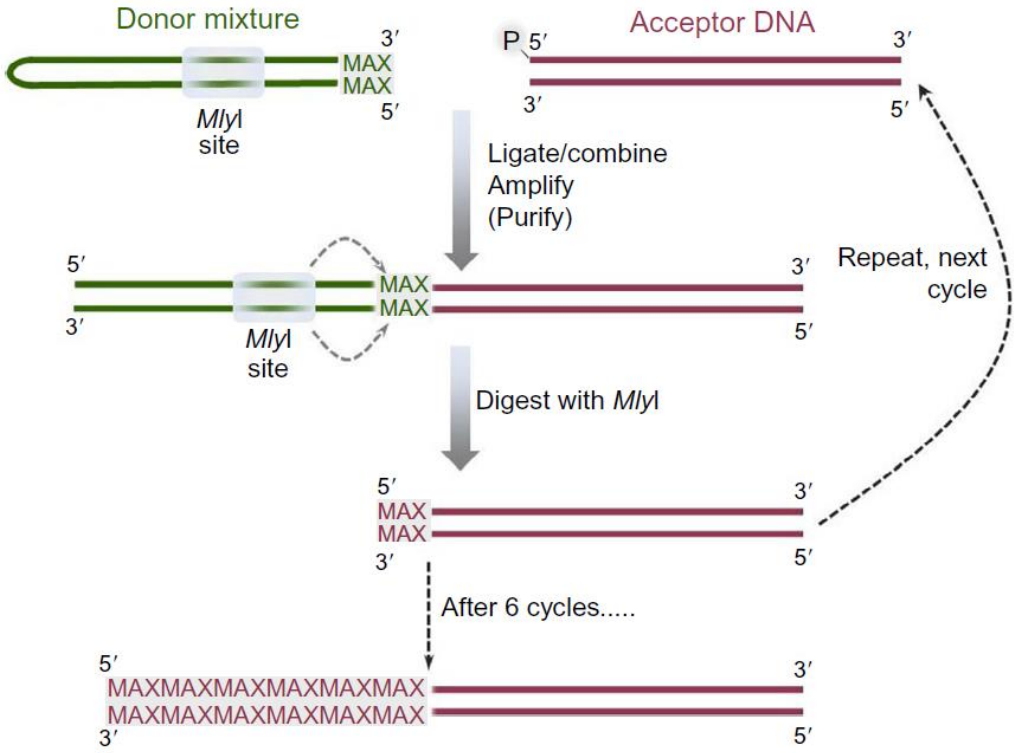


Figure 1.15. A schematic of ProxiMAX randomisation.

Double-stranded DNA donors, carrying the required "MAX" codons at their termini, are ligated onto a double-stranded DNA acceptor sequence (phosphorylated at the required 5'-end only). The donors can take the form of partially double-stranded DNA, fully double-stranded DNA, or hairpin oligonucleotides (as shown). After ligation, the products are amplified by PCR. Depending on whether the process is performed with automation or manually, the donor oligonucleotides can either be combined before or after ligation, with the automated process substantially reducing the number of steps involved and permitting the use of hexanucleotide donors as required (Frigotto et al., 2015). The amplified, purified product is then digested with MlyI and the process repeated, using the digestion product from round 1 as the acceptor for the next round of ligation. Different sets of donor oligonucleotides (up to 20 independently synthesised, double-stranded oligonucleotides) are cycled to prevent potential carryover from one round of addition to the next (Ferreira Amaral et al., 2017).

1.6.2.4. Slonomics®/SlonoMax™

This method was first published as an automated gene synthesis technology (Van den Brulle *et al.*, 2008; Schatz, Delcher, and Salzberg, 2010). Slonomics® relies on cycles of reactions including in the first instance sticky-end ligation between two hairpin oligonucleotides, a splinker and an anchor (Figure 1.16.). These hairpins have overhangs of usually three bases and once ligated, they are immobilised via a biotin moiety of the anchor. To ensure only immobilised product is taken into the next step, subsequent washes remove any unligated splinker. The next step, digestion with a Type IIS endonuclease enzyme of the immobilised product leaves a three-base, single-stranded overhang anchor which is then discarded, while supernatant contains splinker with a new codon. The elongated splinker can enter the next cycle if required, for up to six codons (18 bp). An extended fragment is called 'elongation fragment' and individual fragments when digested with Type IIS restriction enzymes can be later combined by ligations (Waldmann, 2006, 2013). This technology does not involve PCR amplification step seen with MAX and ProxiMAX randomisations or TRIM technology. Slonomics® was commercialised as SlonoMAX™ and it can be used to make combinatorial libraries. The advantage of Slonomics® to other traditional gene synthesis methodologies is relying on a universal set of building blocks, a total of 64 splinkers and 4096 anchors, rather than single-stranded oligos that require to be generated separately (Waldmann and Fuhrmann, 2006; Ferreira Amaral *et al.*, 2017).

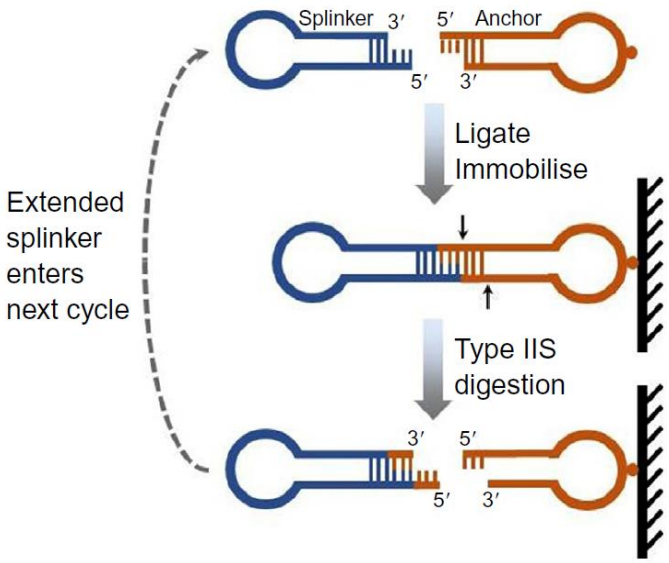


Figure 1.16. Overview of the Slonomics® process.

Slonomics® (Waldmann, 2006; Van den Brulle et al., 2008; Waldmann, 2013) begins with joining hairpin splinker oligonucleotides via sticky-ended ligation, to a mixture of hairpin anchor oligonucleotides. The ligated product is then immobilised via a biotin moiety contained within the anchors, washed and digested with Type IIS restriction enzyme Eam1104I, which generates a three-base sticky-ended overhang. The resulting extended splinker, which is now free in solution, enters the next round of addition. Up to six rounds of addition are performed to generate an "elongation block" and several elongation blocks may be joined together via further Type IIS digestion and subsequent ligation (Van den Brulle et al., 2008). No PCR amplification is involved in the process (Ferreira Amaral et al., 2017).

1.7. Methods for screening enzyme libraries

Once the library of variants is produced, the next step is to subject this library to a reliable screening or selection method. The difference between those two is that screening considers every protein in terms of anticipated function, while selection excludes nonfunctional protein variants in the first instance (Xiao, Bao, and Zhao, 2015). Since some of the libraries have a high number of variants, once expressed, they cannot be examined one by one. Rather, a high-throughput method has to be employed. The choice of library screening technique depends on the nature of the library, properties of the engineered protein, the number of variants and others. Those libraries with variants higher than 10^{10} have to be screened in *in vitro* methods because the transformation efficiency of *in vivo* methods would limit the number of clones. Anything of a smaller size can be fitted in *in vivo* techniques. For those protein libraries which express ligand-binding activities, suitable methods involve ligand immobilisation followed by exposure of the library variants found in solution. Those proteins that do not bind with the ligand are washed away and those bound are later eluted and amplified. This cycle is often repeated 3-5 times and is called enrichment. Due to the binding activity of the proteins, screening methods such as ribosome display (Hanes and Plückthun, 1997) or CIS display (Odegrip *et al.*, 2004) can be employed because when the successful protein is immobilised, it can be isolated and its DNA sequenced.

There is a different challenge with enzyme libraries because they often catalyse the reaction between components via binding but release them shortly after the catalysis is completed. Because of the rapid action of the catalysis and often short binding of the substrate, enzymes cannot be immobilised and their encoding sequences resolved. For this reason, any of the above display methods cannot be used for screening enzyme libraries, unless the activity of the enzyme is a permanent inhibition through binding. Instead of a display technique, the enzyme activity is often tested in a positive and negative selection based on various selection pressures such as life and death of cells through antibiotic resistance, production of the gene and others.

1.7.1. Double selection: positive and negative

Double selection relies on positive and negative selection markers, to analyse and screen the library that expresses enzyme variants (Liu and Schultz, 1999; M. Pastrnak and Schultz, 2001). The aim of selection is to lead to a discovery of novel enzymes, with a new or enhanced, desired function. The positive and negative steps are designed to select for functional enzymes that exhibit a desired feature. They are often based on a reporter-based selection such as the activity of the enzyme variant that will regulate reporter activity. For example, the reporter can be a gene of which expression ensures the survival of the cells when exposed to the antibiotic. Production of this antibiotic-resistant protein is, therefore, controlled by the desired enzyme variants (Xiao *et al.*, 2015). This type of reporter based selection can be either riboswitch/ribozyme-based or transcriptional-regulator-based strategies. An example of a reporter based selection as a positive marker in double selection is β-lactamase gene. This popular antibiotic resistance gene is often substituted for one with stronger selection pressure, chloramphenicol acetyltransferase (Miro Pastrnak, Magliery, and Schultz, 2000). Negative selection marker may be a toxic gene, for example, barnase or uracil phosphoribosyltransferase.

Schultz and co-workers regularly employ their well-established double selection assays for screening aaRS variants coupled with a tRNA suppressor, to introduce UAAs into proteins and thus expand the genetic code with a variety of UAAs. In the first step, positive selection relies on the production of antibiotic-resistant proteins to identify clones that encode any functional protein. The enzyme variants are expressed in the presence of tRNA suppressor, UAAs and AA and antibiotic-resistant protein gene with an amber stop codon in a permissive position. When the variant aminoacylates the tRNA suppressor with any of the amino acids present (UAA or AA), it reads through a stop codon, successfully expressing an antibiotic-resistant protein that results in cell survival – a cell with a functional aaRS variant. The second step subjects all of the survivors to a negative selection. This time the expression is carried out in the absence of the chosen UAA (i.e. natural AAs only) to express a toxic gene such as barnase, again containing an amber stop codon. Cells with aaRS variants that

are capable of aminoacylating the suppressor tRNA with natural AAs successfully read through the stop codon, whereas those that can only incorporate the UAA do not. This expression (i.e. successful read-through) results in the production of a toxic gene that ultimately, kills the clones encoding those enzyme variants that recognise both natural AAs. The surviving cells carry the desired functional enzyme variants that have passed double selection process and are able to recognise only the UAA of choice (Melancon and Schultz, 2009).

1.7.2. GFP-reporter and FACS screening

Another reporter-based selection system widely used is GFP-reporter that is coupled with Fluorescence-Activated Cell Sorting (FACS) (G. Yang and Withers, 2009). FACS, a type of flow cytometry is able to sort cells based on their fluorescence signals, single or multiple as required. The method itself is highly sensitive and also considered as ultrahigh-throughput screening suitable for directed enzyme evolution. Although FACS can be also used during *in vitro* compartmentalisation, product entrapment in the cells, or active enzymes displayed on the cell surface, for the purpose of this project, only GFP-reporter system is going to be outlined here.

The gene of Green Fluorescent Protein (GFP) and its variants such as enhanced GFP (eGFP) can also be used as reporter genes in screening the family of aaRS enzymes. In an example of this system (Santoro, Wang, Herberich, King, and Schultz, 2002), two termination codons are placed within a T7 RNA polymerase gene, which in turn, if translated, expresses a GFP gene (under control of the T7 promoter). A third termination codon is placed in a chloramphenicol resistance gene. All these genes, along with an amber suppressor tRNA are encoded on the same plasmid. In this complex example, a double selection is also employed. The first step involves expressing the aaRS variants in the presence of UAA and chloramphenicol so that clones that can aminoacylate the suppressor tRNA with any amino acid will read through chloramphenicol resistance gene, survive antibiotic selection and go on to produce both T7 RNA polymerase and thus GFP. The screen is then repeated in a second step in the absence of both UAA (i.e. natural AAs only) and chloramphenicol, so those cells with

fluorescent signal include enzyme variants that recognise only natural AA, while those functional synthetases that aminoacylate tRNA suppressor with UAA were unable to read through T7 RNA polymerase and GFP genes and thus do not have fluorescent signal. The second step is coupled with FACS sorting and fluorescent cells are discarded. In this way, the cells with desired aaRS variants can be isolated and the sequence of mutant enzyme determined.

1.7.2.1. Fluorescent protein– GFP and its variants

GFP is the most well-known and established fluorescent protein with application across different fields of life sciences. It was originally isolated from a jellyfish *Aequorea victoria* and firstly published in 1962 (Shimomura, Johnson, and Saiga, 1962). GFP is made of 238 amino acids and its molecular weight is approximately 27 kDa. It has a β-barrel structure of eleven β-strands with a chromophore found in the centre of this structure as seen in Figure 1.17. (F. Yang, Moss, and Phillips, 1996). The chromophore emits green fluorescence upon exposure to blue light and when oxidisation of α–β carbon bond of residue Tyr66 occurs (Heim, Prasher, and Tsien, 1994). Truncated GFP will not exhibit any fluorescence. Its most common application includes protein localisation, organelle labelling, gene regulation (Kaishima, Ishii, Matsuno, Fukuda, and Kondo, 2016).

Currently, a variety of GFP homologous from different organisms are also available and often used instead of the original one as they offer enhanced properties in terms of signal intensity, stability, different excitation and emission spectra (better suitable for the application), length of emission and others. As an example enhanced GFP (eGFP) is based on wtGFP but with mutations around the chromophore region. Thus mutating residue Ser65 to Thr allowed for stabilisation of hydrogen-bonding network and making it 5-fold brighter than wtGFP and faster to detect when using eGFP in living cells as a fluorescent marker (Heim, Cubitt, and Tsien, 1995). An additional mutation of Phe64 to leucine improved maturation of protein at 37 °C. The disadvantage of using eGFP over wtGFP is its weak predisposition to dimerize and minor sensitivity to pH (Day and Davidson, 2009). To ensure that eGFP remains a monomer, one of the three point mutations, F223R, L221K, and most preferable

A206K, can be introduced to the gene (Nagai *et al.*, 2002; Zacharias, Violin, Newton, and Tsien, 2002). eGFP has a peak excitation wavelength at 488 nm and a peak excitation wavelength at 507 nm. Considering that other fluorescent proteins were developed based on eGFP variant (with slightly different properties), eGFP remains a very reliable and well-used fluorescent market during library screening and selection process.

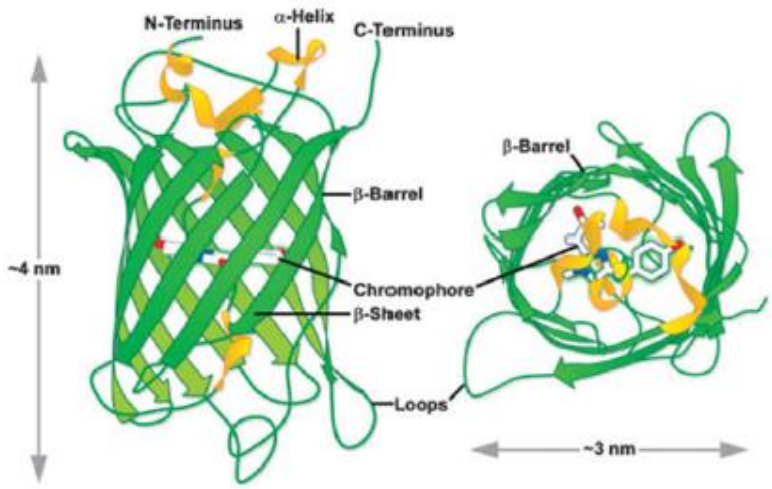


Figure 1.17. The β-barrel structure of A. victoria GFP.

Drawing includes approximate dimensions of A. victoria GFP and it is based on Protein Data Bank ID: 1w7s (Day and Davidson, 2009).

1.8. Current advances in introducing UAAs into proteins and expansion of genetic code

The technologies to incorporate UAAs into proteins were developed for both *in vivo* and *in vitro* applications. The earliest research involved chemical aminoacylation of tRNAs as pre-charged tRNAs that were then used for *in vitro* translation (Hecht, Alford, Kuroda, and Kitano, 1978). A

different approach was largely developed by Schultz and co-workers and involved engineering aaRS and tRNA pairs. This approach had been extensively developed ever since also by other researchers, and currently, there are over 200 UAAs that can be incorporated into proteins *in vivo* or *in vitro* by different mutant pairs.

1.8.1. Chemical aminoacylation

Chemical aminoacylation was first developed by Hecht and colleagues (Hecht *et al.*, 1978). At that time, it was assumed that aaRSs are not efficient in aminoacylating tRNAs with UAAs and an alternative method was sought. Therefore, chemical aminoacylation was developed for the purpose of using pre-charged tRNAs with UAAs during proteins synthesis. In the process, tRNA with deleted 3'-terminal cytidine and adenosine is ligated with 2'(3')-O-acylated pCpA derivatives by T4 RNA ligase (Robertson, Ellman, and Schultz, 1991). The reaction has been developed further for the number of years, but it is generally considered as technically challenging.

1.8.2. Flexizyme system

Suga and co-workers began the research on synthetic ribosomes that can aminoacylate tRNAs in the early 2000s (Lee, Bessho, Wei, Szostak, and Suga, 2000). He later developed Flexizyme system that is well-known for charging UAAs onto tRNAs (Murakami, Ohta, Ashigai, and Suga, 2006; Ohuchi, Murakami, and Suga, 2007a). It comprises of a synthetic ribosome that demonstrates the aaRSs-like function of aminoacylation various tRNAs. The flexizyme distinguishes tRNA by its 3'-end RCCA-3' (R=A/G) and its discriminator base is at position 73. It also recognises benzylic moiety on the leaving group (Murakami *et al.*, 2006). The system is widely used in *in vitro* translation system for incorporation of UAAs into proteins. Flexizyme can charge amino acids onto tRNA in a high-throughput manner and it is far more advantageous to conventional chemical aminoacylation.

1.8.3. Engineering cognate aaRS and tRNA pairs

The most commonly aaRS and tRNA pairs used for genetic code expansion are *Methanocaldococcus jannaschii* Tyrosyl-tRNA synthetase-$tRNA^{Tyr}$ pair (MjTyrRS-$tRNA^{Tyr}$) or the *Methanosarcina barkeri* Pyrrolysyl-tRNA synthetase-$tRNA^{Pyl}$ pair (MbPylRS-$tRNA^{Pyl}$) as wild-type or engineered (Vargas-Rodriguez, Sevostyanova, Söll, and Crnković, 2018). The first methods of incorporation UAAs into proteins involved using amber suppression codon, as this is a rare codon in *E. coli*, and it can be used to encode UAA of choice. However, both pairs allow for anticodon mutation of tRNA without losing the enzyme's specificity. MjTyrRS-$tRNA^{Tyr}$ can only be used in prokaryotic systems such as *E. coli* because *in vivo* it can cross-connect with the endogenous system of eukaryotic TyrRS and $tRNA^{Tyr}$ (Vargas-Rodriguez *et al.*, 2018). Conversely, unique properties of PylRS and its cognate $tRNA^{Pyl}$ (known also as $tRNA_{CUA}$) make the pair an interesting tool for incorporation of UAAs via amber suppression. The pair is known to be orthogonal in bacterial and eukaryotic cells and therefore, it can be used alongside endogenous tRNAs and aaRSs (Wan, Tharp, and Liu, 2014).

The cognate pair PylRS-$tRNA_{CUA}$ (including the active site engineered PylRS) was used in several studies by Chin and colleagues (Elliott, Bianco, and Chin, 2014) as well as Schultz and co-workers (Vargas-Rodriguez *et al.*, 2018). For example, it was employed to incorporate lysine with post-translational modification like methylation, acetylation or ubiquitination into the proteins (Elliott *et al.*, 2014). Other UAAs included biophysical probes such as photo-crosslinkers. Also, the site-specific labelling of proteins for fluorescence resonance energy transfer (FRET) studies was developed. Another technology by Chin and colleagues used this pair in parallel with his orthogonal ribosome ribo-Q1 which recognises quadruplet codons for the incorporation of UAAs (Neumann, Wang, Davis, Garcia-Alai, and Chin, 2010).

1.9. The aims and objectives of the project

This project aimed to provide the foundations for the development of a novel *in vitro* translation system to introduce multiple UAAs via *in vitro* translation of proteins to expand the genetic code, ultimately from 20 to 32 codons. The system would work in parallel with ProxiMAX randomisation and synthetic genes to be utilised during library generation and screening. Since ProxiMAX is a nondegenerate saturation mutagenesis procedure, it only uses 20 codons from the genetic code, 'leaving' many spare codons that could be used in the novel system.

To be able to develop an *in vitro* translation system, there is a requirement to engineer novel aminoacyl-tRNA synthetases which recognises UAAs (in particular D-AAs) and insert them in a protein during *in vitro* translation. Novel tRNA molecules are also essential so they form cognate pairs with novel enzymes.

In this project, the development of new enzymes is based on *E. coli* AlaRS while new tRNAs are based on *E. coli* $tRNA^{Ala}$. The lack of anticodon recognition by AlaRS makes it an ideal target for mutagenesis since tRNA specificity is determined only by the acceptor arm. Thus, it is possible to mutate the anticodon of $tRNA^{Ala}$ whilst leaving the acceptor arm unchanged. Meanwhile, the amino acid binding site of *E. coli* AlaRS will be mutated to promote binding with D-Ala and subsequently, other UAAs. This AlaRS already misacylates both smaller glycine and larger serine in its natural environment (Guo *et al.*, 2009). Therefore, there is a possibility that by mutating residues within the amino acid binding site, new amino acids will be recognised by the enzyme. Obviously, since the acceptor arm of the $tRNA^{Ala}$ remains unchanged, resulting tRNAs would be individually pre-charged, prior to being placed in the in vitro translation reactions, akin to the semi-synthetic approach of Suga and co-workers (Ramaswamy, Saito, Murakami, Shiba, and Suga, 2004; Murakami *et al.*, 2006; Ohuchi, Murakami, and Suga, 2007b). Ultimately, to prevent recycling or editing of mutant tRNAs by native AlaRS, these pre-charged mutant tRNAs along with native pre-charged tRNAs (or possibly only pre-charged $tRNA^{Ala}$) would be used in an *in vitro* translation reaction of synthetic genes in

which only the normal 20 ProxiMAX codons are used to encode natural AAs and the remaining 12 codons are used to encode UAAs in positions of saturation.

The fundamental and central part of the project is to create a gene library based on *E. coli* AlaRS and mutating its amino acid binding site. The library will be generated using ProxiMAX randomisation because it is a nondegenerate, unbiased and achieving maximal diversity method of making gene variant libraries. Two libraries will be produced. The first library would be based on the N-terminal domain of *E. coli* AlaRS only. This domain can aminoacylate tRNAs on its own (Naganuma *et al.*, 2014) and excludes the editing function that is not desired for the application of this project. Meanwhile, the second library would be a full-length enzyme in which the editing function is eliminated by a C666A mutation. As a preparation for the library build, the genes that will be used in the generation of libraries have to be amplified from *E. coli* and cloned into vectors. Their expression and functionality in aminoacylation assays will also be undertaken.

Once the gene libraries are produced, it is vital to screen them using a reliable assay. Previous aaRS mutagenesis has been developed using orthogonal systems such as the PylRS and cognate $tRNA_{CUA}$ that are orthogonal to the cells in which they are produced (Blight *et al.*, 2004). However, such systems are not suitable to translate with multiple UAAs in parallel, since they rely on the use of a single, amber termination codon for UAA incorporation. Therefore, the final objective of this project is to develop an assay for screening libraries of *E. coli* AlaRS variants, conducted in a different host system so that selected *E. coli* AlaRS and $tRNA^{Ala}$ would not interact with native enzymes and tRNAs.

Chapter 2: Materials and Methods

Medium and buffers were prepared with distilled water (dH_2O). Extra purified 18.2 MΩ water also called double distilled water (ddH_2O) was used in all molecular procedures such as Polymerase Chain Reaction (PCR), DNA elution in DNA extraction and others. To any work related to RNA DEPC-Treated Water was used.

2.1. Materials

2.1.1. Risk Assessments

The appropriate risk assessments were undertaken and stored in the laboratory.

2.1.2. Cell lines

Table 2.1. Cell lines for E. coli and S. cerevisiae.

Organism	Strain	Genotype
E. coli	Tuner™(DE3)	F^- ompT hsdS_B (r_B^- m_B^-) gal dcm lacY1(DE3)
E. coli	DH5α	F^- φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169 *rec*A1 *end*A1 *hsd*R17(r_K^-, m_K^+) *pho*A *sup*E44 λ^- *thi*-1 *gyr*A96 *rel*A1
S. cerevisiae	BY4741	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0
S. cerevisiae	BY4741 NMD2Δ	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 *nmd2*Δ

2.1.3. Vectors

Table 2.2. List of vectors used in the study.

Vector	Source	Comments
pUC19	New England Biolabs	Purchased from New England Biolabs (NEB)
***Sma*I-phophatased pUC19**	Thermo Fisher Scientific (formerly Fermentas)	Purchased from Thermo Fisher Scientific
pET45b(+)	Novagen™	Already present in Prof Anna V Hine's lab stock
pYES2	Invitrogen™	A gift from Prof Roslyn Bill's lab stock
pXRH3	Addgene	A gift from Prof Marc Gartenberg (Addgene plasmid #63145; http://n2t.net/addgene:63145; RRID:Addgene_63145) (Chou, Patel, and Gartenberg, 2015)

2.1.4. Markers

2.1.4.1. DNA ladders

2.1.4.1.1. HyperLadder™ 1 kb (Bioline)

2.1.4.1.2. HyperLadder™ 25 bp (Bioline)

2.1.4.1.3. 1 kb DNA ladder (NEB)

2.1.4.1.4. GeneRuler™ 1 kb DNA Ladder (Fermentas)

2.1.4.1.5. GeneRuler™ 1 kb Plus DNA Ladder (Thermo Scientific)

2.1.4.1.6. Low Molecular Weight Marker, 10-100 nt (Affymetrix)

2.1.4.2. Protein Ladders

2.1.4.2.1. PageRuler Plus Prestained Protein Ladder (Thermo Scientific)

2.1.4.2.2. Prestained Protein MW Marker (Thermo Scientific)

2.1.4.2.3. PageRuler Prestained Protein Ladder (Thermo Scientific)

2.1.5. Media recipes

2.1.5.1. Solid medium

The medium was prepared according to the recipe and substituted with 1.5 % (w/v) Difco™ Agar, Granulated (BD Biosciences). The solution was autoclaved at 121 °C for 20 min and left to cool to approximately 55 °C. To prepare solid medium plates, 25-30 mL of agar-media solution was added in a sterile manner onto a petri dish. The plates were then left to set at the room temperature overnight and stored at 4 °C on the following day.

2.1.5.2. LB broth medium

Tryptone	1 % (w/v)
Yeast extract (YE)	0.5 % (w/v)
Sodium chloride (NaCl)	1 % (w/v)

To prepare 2 % (w/v) medium solution, LB broth powder (Fisher Bioreagents) was dissolved in distilled H_2O. The solution was autoclaved at 121 °C for 20 min. LB was supplemented with 100 µg/mL sterile ampicillin solution (when cooled down) or 1.5 % (w/v) agar for solid medium (prior to autoclaving), as required.

2.1.5.3. 2X YT medium (at Isogenica)

Tryptone	1.6 % (w/v)
Yeast extract (YE)	1 % (w/v)
Sodium chloride (NaCl)	0.5 % (w/v)

2.1.5.4. SOC medium (Super Optimal Broth with Catabolite repression)

Tryptone	2 % (w/v)
Yeast extract (YE)	0.5 % (w/v)
Sodium chloride (NaCl)	0.01 M
Potassium chloride (KCl)	0.0025 M
Magnesium chloride ($MgCl_2$)	0.01 M
Magnesium sulfate ($MgSO_4$)	0.01 M
Glucose	0.02 M

2.1.5.5. YPD medium (Rich Medium)

Yeast extract (YE)	1 % (w/v)
Peptone	2 % (w/v)
Dextrose	2 % (w/v)

To prepare YPD medium, yeast extract and peptone were mixed in 90 % of the final volume of distilled H_2O and autoclaved at 121 °C for 20 min. Once it cooled down to approximately 55 °C, filter sterilised 20 % (w/v) dextrose was added. YPD was supplemented with 1.5 % (w/v) agar for solid medium (prior to autoclaving), as required.

2.1.5.6. Synthetic Defined (SD) (Selective/Minimal) Medium

Complete Supplement Mixture (CSM) Single or Double Drop-Out (Formedium) and Yeast Nitrogen Base (YNB) without amino acids (Formedium) were mixed according to the manufacturer's instructions. The solution was autoclaved at 121 °C for 20 min. Once it cooled down to approximately 55 °C, filter sterilised 20 % (w/v) dextrose or 20 % (w/v) galactose was added. SD was supplemented with 1.5 % (w/v) agar for solid medium (prior to autoclaving), as required.

2.1.6. Buffer recipes

2.1.6.1. 1X Tris-Acetate-EDTA (TAE) Buffer (50X) (Fisher Scientific)

Tris	0.04 M
Acetic acid	0.02 M
EDTA	0.001 M

2.1.6.2. 1X Tris-Borate-EDTA (TBE) Buffer (10X) (Gibco™ by Life Technologies)

Tris	0.1 M
Boric acid	0.09 M
EDTA	0.001 M

2.1.6.3. 1X Standard Taq DNA Polymerase Reaction Buffer (10X) (NEB)

Tris-HCl	0.01 M
Potassium chloride (KCl)	0.05 M
Magnesium chloride ($MgCl_2$)	0.0015 M

2.1.6.4. 1X Q5® High-Fidelity DNA Polymerase Reaction Buffer (5X) (NEB)

2.1.6.5. 1X Phusion HF Buffer (5X) (Thermo Scientific)

2.1.6.6. 1X AMV Reverse Transcriptase Reaction Buffer (10X) (NEB)

Tris-Acetate	0.05 M
Potassium acetate	0.075 M
Magnesium acetate	0.008 M
Dithiothreitol (DTT)	0.01 M

2.1.6.7. DNA Gel Loading Dye (6X) (Thermo Fisher Scientific)

Tris-HCl (pH 7.6)	0.00167 M
Bromophenol blue	0.005 %
Xylene cyanol FF	0.005 %
Glycerol	10 %
EDTA	0.010 M

2.1.6.8. GelPilot Loading Dye (5X) (QIAGEN)

2.1.6.9. Gel Loading Dye Purple (6X) (NEB)

2.1.6.10. Blue Gel Loading Dye (6X) (NEB)

Ficoll®-400	2.5 %
EDTA	0.011 M
Tris-HCl	0.0033 M
SDS	0.017 %
Bromophenol blue	0.015 %

2.1.6.11. 1X T4 DNA Ligase Reaction Buffer (10X) (NEB)

Tris-HCl	0.05 M
Magnesium chloride ($MgCl_2$)	0.01 M
ATP	0.001 M
Dithiothreitol (DTT)	0.01 M

2.1.6.12. 1X T4 DNA Ligase Buffer (10X) (Thermo Scientific)

Tris-HCl	0.04 M
Magnesium chloride ($MgCl_2$)	0.01 M
ATP	0.0005 M
Dithiothreitol (DTT)	0.01 M

2.1.6.13. 1X CutSmart Restriction Digest Buffer (10X) (NEB)

Tris-Acetate	0.02 M
Magnesium acetate	0.01 M
Potassium acetate	0.05 M
BSA	100 µg/mL

2.1.6.14. 1X FastDigest Buffer (10X) (Thermo Scientific™)

2.1.6.15. 1X SYBR® Gold Nucleic Acid Gel Stain (10,000X) (Invitrogen™ by Life Technologies)

2.1.6.16. 1X SYBR® Safe DNA Gel Stain (10,000X) (Invitrogen™ by Life Technologies)

2.1.6.17. SDS-PAGE, PAGE, acid-urea PAGE and Protein purification buffers

<u>At Isogenica:</u>

2.1.6.17.1. Binding buffer (A)

Sodium phosphate buffer (pH 7.5)	0.05 M
Sodium chloride (NaCl)	0.5 M
Imidazole	0.02 M

To obtain pH 7.5 of sodium phosphate buffer, 84% of Na_2HPO_4 was mixed with 16% of NaH_2PO_4.The buffer was filter sterilised using the Corning units with 0.22 µm pore filter.

2.1.6.17.2. Elution buffer (B)

Sodium phosphate buffer (pH 7.5)	0.05 M
Sodium chloride (NaCl)	0.5 M
Imidazole	0.25 M

To obtain pH 7.5 of sodium phosphate buffer refer to 2.1.6.17.1.

2.1.6.17.3. 1X NuPAGE™ MOPS SDS Running Buffer (20X) (Novex™ by Life Technologies)

3-(N-morpholino)propanesulfonic acid (MOPS)	0.05 M
Tris	0.05 M
SDS	0.1 %
EDTA	0.01 M
pH	7.7

2.1.6.17.4. 1X NuPAGE™ LDS Sample Buffer (4X) (Novex™ by Life Technologies)

2.1.6.17.5. 1X Bolt® Sample Reducing Agent (10X) (Novex™ by Life Technologies)

<u>At Aston:</u>

2.1.6.17.6. Lysis buffer (pH 8.0)

Sodium phosphate buffer	0.05 M
Sodium chloride (NaCl)	0.3 M
Imidazole	0.01 M

2.1.6.17.7. Native column wash buffer (pH 8.0)

Sodium phosphate buffer	0.05 M
Sodium chloride (NaCl)	0.3 M
Imidazole	0.02 M

2.1.6.17.8. Native column elution buffer (pH 8.0)

Sodium phosphate buffer	0.05 M
Sodium chloride (NaCl)	0.3 M
Imidazole	0.25 M

2.1.6.17.9. 1X SDS-PAGE sample buffer

Tris-HCl (pH 6.8)	0.05 M
Dithiothreitol (DTT)	0.1 M
SDS	2 % (w/v)
Glycerol	8 % (v/v)
Bromophenol blue	0.1 % (w/v)

2.1.6.17.10. 1X Tris-Glycine-SDS PAGE Buffer (10X) (National Diagnostics)

Tris	0.025 M
Glycine	0.192 M
SDS	0.1 % (w/v)

2.1.6.17.11. 1X Acid-Urea PAGE running buffer

Sodium acetate (pH 5.0)	0.1 M

2.1.6.17.12. 1X Acid-Urea PAGE sample buffer

Urea	7 M
Sodium acetate (pH 5.0)	0.1 M
Bromophenol blue	0.05 % (w/v)

2.1.6.17.13. 1X Acid-Urea PAGE staining solution

Sodium acetate (pH 5.0)	0.1 M
SYBR® Gold Nucleic Acid Gel Stain (Invitrogen)	1X

2.1.6.18. 1X Transfer buffer for Western Blot

Tris	0.025 M
Glycine	0.192 M
pH	8.3

2.1.6.19. 1X Aminoacylation buffer

HEPES	0.1 M
Magnesium chloride ($MgCl_2$)	0.02 M
Dithiothreitol (DTT)	0.01 M
ATP	0.01 M
Potassium chloride (KCl)	0.01 M

2.1.7. Other solutions

2.1.7.1. Ammonium Persulfate

Ammonium Persulfate (Sigma-Aldrich)	20 % (w/v)

2.1.7.2. Ampicillin (Amp)

The stock solution of ampicillin (100 mg/mL) was prepared by dissolving 4.5 g ampicillin powder (Sigma) in 45 mL ddH_2O. Solution was sterilised using 0.22 µm syringe filter. Stocks of 1.5 mL were stored at -20 °C.

2.1.7.3. 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal)

To prepare 2 % (w/v) X-gal, 0.2 g powder X-Gal was dissolved in 10 mL DMF. The solution was sterilised using 0.22 µm syringe filter. Stocks of 1 mL aliquots were stored at -20 °C.

2.1.7.4. dNTPs

To prepare dNTP solutions from 100 mM stock solution (NEB, Promega), ddH_2O was used to dilute the stock to required concentration.

2.1.7.5. Ethanol

Ethanol	70 % (v/v)

2.1.7.6. EZ-ECL (Biological Industries)

Solution A contains luminol and enhancer. Solution B contains stable peroxide solution.

2.1.7.7. Galactose

Dextrose [D-(+) Glucose] (Sigma-Aldrich)	20 % (w/v)

The solution was filter sterilised.

2.1.7.8. Glucose

Dextrose [D-(+) Glucose] (Sigma-Aldrich)	20 % (w/v)

The solution was filter sterilised.

2.1.7.9. InstantBlue™ Protein Stain (Expedeon)

2.1.7.10. Isopropanol

Isopropanol	70 % (v/v)

2.1.7.11. Isopropyl β-D-1-thiogalactopyranoside (IPTG)

To prepare 1 M IPTG, 2.38 g powder IPTG was dissolved in 10 mL ddH_2O. Solution was sterilised using 0.22 µm syringe filter. Stocks of 1.5 mL were stored at -20 °C.

2.1.7.12. Lithium Acetate

To prepare 1 M Lithium Acetate (LiAc), 0.102 g was mixed in 10 mL distilled water. The solution was filter sterilised. To prepare 0.1 M LiAc, 1 M LiAc was diluted 10 fold with distilled water.

2.1.7.13. Poly(ethylene glycol) (PEG)

Poly(ethylene glycol) average MW (3015-3685) (Aldrich Chemistry)	50 % (w/v)

2.1.7.14. Sodium Dodecyl Sulfate (SDS)

Sodium Dodecyl Sulfate (Sigma-Aldrich)	10 % (w/v)

2.1.7.15. Salmon Sperm Single-Stranded DNA

It was a gift from Dr. Michelle Clare.

2.2. Methods

2.2.1. Molecular procedures

2.2.1.1. Polymerase Chain Reaction (PCR)

PCR reactions were carried out in 1X reaction buffer (2.1.6.3., 2.1.6.4., 2.1.6.5), 200 μM dNTPs (2.1.7.4.), 100-500 nM forward and reverse primers (Eurofins Genomics or Sigma) and 1 pg – 1 μg of DNA. Enzyme concentrations varied depending on the manufacturer's recommendations (Taq Polymerase vs High-Fidelity Polymerase). PCR conditions, temperature and time were based on the manufactures instructions, but generally the conditions for High-Fidelity Polymerase (Phusion or Q5) were: 98 °C for 30 sec followed by 30 cycles of: 98 °C for 10 sec, primers dependent temperature for 30 sec and 72 °C for 20-30 sec/kb, with a final step at 72°C for 2 min. While the conditions for Taq Polymerase were: 95 °C for 30 sec followed by 30 cycles of: 95 °C for 30 s, primers dependent temperature for 30 sec, and 68 °C for 1 min/kb, with a final step at 68 °C for 5 min. Every set of reaction included negative control without DNA present.

2.2.1.1.1. Gradient PCR

Gradient PCR reactions were set up based on 2.2.1.1. The conditions of cycling were determined by testing a gradient temperature for primers annealing. The volume per reactions was usually 10-25 µL.

2.2.1.1.2. Colony PCR

Colony PCR reactions were set up based on 2.2.1.1. A colony was picked with a sterile pipette tip from a solid media plate and inoculated in an individual PCR reaction tube. PCR was performed with Taq Polymerase. The volume per reactions was usually 10-25 µL. The colonies were saved by inoculating in liquid media or by streaking on solid media plates.

2.2.1.2. RT-PCR

RT-PCR used in this study was two-step RT-PCR using AMV Reverse Transcriptase (10 U/µL) (NEB). Every RT-PCR reaction included negative control without AMV Reverse Transcriptase present. The difference was made up with RNA DEPC-Treated Water.

Step one: 20 ng of total extracted RNA (2.2.1.8.4), 1 µM gene-specific forward and reverse primer, 10 mM dNTPs and RNA DEPC-Treated Water was mixed in a total volume of 10 µL. The reaction mixture was incubated at 65 °C for 5 min, followed by a brief spin and placed on ice. To 10 µL of RNA/primer mixture were added 1X AMV buffer (2.1.6.6.), 10 U AMV Reverse Transcriptase, up to 10 U RNase Inhibitor and RNA DEPC-Treated Water. The total reaction volume was 20 µL. The reaction mixture was incubated at 25 °C for 5 min, followed by 42 °C for 60 min and AMV enzyme was inactivated 80 °C for 5 min.

Step two: cDNA generated in step one served as a template for amplification in step two RT-PCR. Step two RT-PCR was carried out in 1X Taq buffer (2.1.6.3.), 200 µM dNTPs (2.1.7.4.), 200 nM

gene-specific forward and reverse primers (Eurofins Genomics) and 2 µL of cDNA template (step one reaction) in a total volume of 50 µL. The total volume was made up with RNA DEPC-Treated Water. The reaction was incubated at following conditions: 95 °C for 30 sec, followed by 30 cycles of: 95 °C for 30 sec, 55 °C for 30 sec, and 68 °C for 1 min/kb, with a final step at 68 °C for 5 min. Every step two reactions included a negative control without cDNA present.

2.2.1.3. Addition of phosphate group to the 5'-ends of DNA (phosphorylation)

Up to 20 pmol DNA were incubated in 1X ligase buffer (2.1.6.11. or 2.1.6.12.), 1 mM ATP and 10 U T4 Polynucleotide Kinase (NEB or Thermo Scientific) in a total volume of 20 µL at 37 °C for 30 min. The enzyme was heat-inactivated at 65 °C for 20 min.

2.2.1.4. Removal of a phosphate group from the 5'-ends of DNA (dephosphorylation)

Up to 1 pmol DNA were incubated in 1X Antarctic Phosphatase Reaction Buffer (10X) and 5 U Antarctic Phosphatase (NEB) in a total volume of 20 µL at 37 °C for 30 min. The enzyme was heat-inactivated at 80 °C for 2 min. If required, the reaction was scaled up.

2.2.1.5. Ligation

The ligation was performed in the molar ratio of 1:3 vector to insert with up to 100 ng of insert DNA. The reaction was set up in 1X ligase buffer (2.1.6.11. or 2.1.6.12.) and 400 U T4 DNA Ligase (NEB or Thermo Scientific) in a total volume up to 50 µL. The ligation was incubated at 16 °C overnight or at 22 °C for 2 h with inactivation at 65 °C for 10 min.

2.2.1.6. Restriction digest

The restriction digestion was performed using up to 1 µg of DNA, 1X digestion buffer (2.1.6.13. or 2.1.6.14.), 10 U of restriction enzyme in a reaction volume up to 100 µL. Incubation and inactivation

parameters varied depending on the enzyme used.

2.2.1.7. Mutagenesis

2.2.1.7.1. ProxiMAX randomisation

Colibra™ (Isogenica) was employed for ProxiMAX randomisation (Ashraf *et al.*, 2013; Frigotto *et al.*, 2015; Poole, 2015).

2.2.1.7.2. Via PCR

In order to introduce mutation in DNA using PCR, appropriate primers were designed with one carrying a mutation to use in PCR reaction (2.2.1.1.).

2.2.1.8. Nucleic acid purification

2.2.1.8.1. DNA purification including PCR products, enzyme digestions, and other DNA containing samples

Samples were purified using different purification spin columns kits: QIAquick® PCR Purification Kit (QIAGEN), MinElute® PCR Purification Kit (QIAGEN), Wizard® SV Gel and PCR Clean-Up System (Promega) or GENECLEAN® Turbo Kit (MP Biomedicals) according to the manufacturer's instructions. NanoDrop was used to quantify purified DNA (2.2.1.9.).

2.2.1.8.2. DNA plasmids

To extract plasmid DNA from the bacterial culture grown in LB broth (2.1.5.2.), QIAprep® Spin Miniprep Kit (QIAGEN) was used according to the manufacturer's instructions.

To isolate plasmid DNA from yeast cells grown in Selective Medium (2.1.5.6.), a protocol "Isolation of plasmid DNA from yeast using the QIAprep® Spin Miniprep Kit" was followed. It was found on

QIAGEN® website as a User-Developed Protocol by Michael Jones, Chugai Institute for Molecular Medicine, Ibaraki, Japan. The following changes were applied to the methodology. In step 2, 500 µL instead of 250 µL of Buffer P1 were used. In step 3, 600 µL of acid-washed glass beads (425-600 µm) (Sigma-Aldrich) were used to mix with the sample and lyse in TissueLyser LT (QIAGEN), for 5-10 min at 50 Hz. The TissueLyser LT insert adapter for the samples was pre-cooled for 30 min at -80 °C. NanoDrop was used to quantify extracted plasmid DNA (2.2.1.9.).

2.2.1.8.3. Gel extraction

DNA was extracted from the agarose gel using QIAquick® Gel Extraction Kit (QIAGEN) or Wizard® SV Gel and PCR Clean-Up System (Promega) according to the manufacturer's instructions. NanoDrop was used to quantify gel extracted DNA (2.2.1.9.).

2.2.1.8.4. RNA extraction

RNA was isolated from *S. cerevisiae* using RNeasy Mini Kit (QIAGEN) according to the manufacturer's instructions. Prior to extraction, cells had to be lysed following 2.2.7.3. Once cells were lysed, DNase digestion was performed using RNase-Free DNase Set (QIAGEN) with incubation at 22 °C for up to 30 min before adding lysed supernatant onto a column. Instead of 70 % (v/v) ethanol, 70 % (v/v) isopropanol was used. RNA was eluted with RNase-free water provided in the kit. If required, RNA was re-purified with another DNase treatment before column purification. NanoDrop was used to quantify purified RNA (2.2.1.9.).

2.2.1.9. Nucleic acid quantification

NanoDrop 2000c (at Aston) or NanoDrop 8000 (at Isogenica) (Thermo Scientific) was used to measure the concentration and quality of DNA and RNA. A blank solution, for example, 1 µL of ddH_2O or Buffer EB (QIAGEN) was used to calibrate the spectrophotometer. Measurements were

taken by pipetting 1 μL of sample onto the pedestal of NanoDrop. After calibration or measurement, the pedestal was cleaned with a tissue.

2.2.2. Electrophoresis

2.2.2.1. Agarose gel electrophoresis of DNA

Dependent on the size of analysed DNA, gels of different concentrations were prepared by dissolving Hi-Res Standard Agarose powder (AGTC Bioproducts) in 1X TAE (2.1.6.1.) (at Aston) or Molecular Grade Agarose powder (Melford) in 1X TBE (2.1.6.2.) (at Isogenica). The mixture was heated in a microwave until fully dissolved. When the solution cooled down, 1X SYBR® Safe DNA Gel Stain (10,000X) (2.1.6.15.) or 1X SYBR® Gold Nucleic Acid Gel Stain (10,000X) (2.1.6.16.) (Invitrogen™ by Life Technologies) was added for nucleic acid detection. The mixture was purred into a gel form with a comb attached to form well, and it was left to set.

The analysed samples were pre-mixed with 1X sample loading buffer (2.1.6.7.-2.1.6.10.) before being loaded to the wells. DNA markers/ladders (2.1.4.1) were used as a reference. Electrophoresis was performed in 1X TAE (Aston) or 1X TBE (Isogenica) at 10 V/cm. The gel was visualised by SYBR® Safe DNA or SYBR® Gold Nucleic Acid staining under UV light in the transilluminator. GeneSnap software was used to see the photo on the computer.

2.2.2.2. Agarose gel electrophoresis of RNA

Prior to the start, any equipment used for the experiment was sterilised with DNA zap (Ambion). Any used plasticware or sample tubes were RNase free. Water used was double distilled water (for buffers), DEPC-treated or RNase water. Samples were prepared by mixing with purple 6X gel loading dye (New England Biolabs) and heated for 3 min at 70 °C, followed by a snap cool on ice for 1 min. A 1 % (w/v) agarose gel was prepared with 0.5X TAE. Electrophoresis was performed at 80 V for 30 min ensuring that the running buffer, 0.5X TAE did not overheat. A DNA ladder was

loaded in a parallel with the samples to confirm the correct procedure of electrophoresis, not as a size reference.

2.2.2.3. Polyacrylamide gel electrophoresis

2.2.2.3.1. SDS-PAGE (at Isogenica)

Electrophoresis was performed using pre-casted gels NuPAGE 4-12 %, 1.0 mm (Novex™ by Life Technologies), XCell SureLock™ Mini-Cell Electrophoresis System (Thermo Fisher Scientific) and PowerEase® 500 Power Supply (Thermo Fisher Scientific).

Sample preparation: Lysed and purified samples were prepared by mixing with 1X LDS Sample Buffer NuPAGE (4X) and 1X Bolt™ Sample Reducing Agent (10X). Unpurified samples (whole cell lysate) were prepared by mixing 0.5X sample with 1X NuPAGE LDS Sample Buffer (4X), 1X Bolt™ Sample Reducing Agent (10X) and 0.5X ddH_2O. Sample mixtures were incubated at 95 °C for 5 min and cooled down before loaded to the gel.

Samples were loaded to the wells alongside the marker (2.1.4.2.). The electrophoresis was performed in 1X NuPAGE MOPS SDS Running Buffer at 200 V for 50 min. The gel was removed from the plates and incubated in InstantBlue™ (Expedeon) at the rocking platform from 15 min to 1 hour. The image of the gel was captured using a digital camera or UV transilluminator.

2.2.2.3.2. SDS-PAGE (at Aston)

Electrophoresis was performed using handcast gels and Mini-PROTEAN® Tetra Vertical Electrophoresis Cell (BioRad) and PowerPac™ HC High-Current Power Supply (BioRad). If gels could not be handcasted, 4–15% Mini-PROTEAN® TGX™ Precast Gels (Bio-Rad) were used.

Gel preparation: Firstly, 12 % resolving gel was prepared according to the recipe in Table 2.3. The components were added sequentially, well mixed and purred between 0.7 mm glass plates to set. The gel was covered with 70% ethanol to obtain a uniform set. After polymerisation, the ethanol was washed off with dH_2O three times. Secondly, a 4 % stacking gel was prepared according to the recipe in Table 2.3. The components were mixed and added onto set resolving gel. The comb was applied and the gel was left to set.

Table 2.3. Components in resolving and stacking SDS-PAGE gels.

Component	Resolving gel	Stacking gel
Polyacrylamide[1]	12 %	4 %
Tris-HCl *(pH 6.8 and 8.8)	0.4 M *(pH 8.8)	0.3 M *(pH 6.8)
SDS	0.1 %	0.1 %
Ammonium persulfate	0.07 %	0.08 %
TEMED[2]	0.08 %	0.1 %

[1] ProtoGel 30% (w/v), 37.5:1 Acrylamide to Bisacrylamide Solution (National Diagnostics)

[2] Tetramethylethylenediamine (National Diagnostics)

Sample preparation: Lysed and purified samples were prepared by mixing with 1X SDS-PAGE Sample Buffer (2.1.6.17.9.). Unpurified samples (whole cell lysate) were combined with equal volumes of 4X SDS-PAGE Sample Buffer (2.1.6.17.9.). Samples mixed with buffer were incubated at 95 °C for 5 min and cooled down before loaded to the gel.

Samples were loaded to the wells alongside the appropriate standard (2.1.4.2.). The electrophoresis was performed in 1X Tris-Glycine-SDS PAGE Buffer (2.1.6.17.10.) at a constant voltage of 150 V until the dye migrated until the bottom of the gel. It was then removed from the plates and incubated in InstantBlue™ (2.1.7.9.) at the rocking platform for 15 min to 1 hour. The image of the gel was captured using a digital camera or UV transilluminator.

2.2.2.3.3. Acid-urea PAGE (Aston)

Electrophoresis was performed using handcast gels and the same system as in 2.2.2.3.2.

Gel preparation: components and their final concentrations for this type of gel can be found in Table 2.4. Urea was weighed out and mixed with acrylamide solution and sodium acetate (pH 5.0). The mixture was left on the magnetic stirrer until urea was fully dissolved. The solution was prepared with UltraPure™ DEPC-Treated Water (Thermo Fisher Scientific) until 20 mL. Freshly prepared ammonium persulfate and TEMED were added to the concentration given in Table 2.4. Well mixed solution was purred between 0.7 mm glass plates to set. Acid-urea PAGE gels were uniform and no stacking gels were prepared.

Table 2.4. Components in acid-urea PAGE gels.

Component	Final concentration
Polyacrylamide[1]	12-25 %
Urea	7 M
Sodium acetate (pH 5.0)	0.1 M
Ammonium persulfate	0.12 %
TEMED	0.06 %

[1] Accugel 40% (w/v), 19:1 Acrylamide to Bis-acrylamide solution (National Diagnostics)

Sample preparation: samples were prepared by mixing in 1:1 ratio with 1X acid-urea PAGE Sample Buffer (2.1.6.17.12.). When analysed tRNA on its own, samples were incubated at 95 °C for 5 min and cooled down before loaded to the gel.

The gel was pre-run at 4 °C for 30 min. Prior to sample loading, the wells were extensively washed using 1000 µL pipette tip to wash off residual urea in the well. Samples were loaded to the wells alongside the appropriate standard (2.1.4.2.). The electrophoresis was performed in 1X acid-urea

PAGE Running Buffer (2.1.6.17.11.) at a constant voltage of 100 V to do not exceed 10 V/cm until the dye migrated 3/4th of the gel. It was then removed from the plates and incubated in 1X acid-urea PAGE Staining Solution (2.1.6.17.13.) at the rocking platform for 15-30 min. The image of the gel was captured using UV transilluminator.

2.2.3. Transformation

2.2.3.1. Into *E. coli*

The competent cells were thawed on ice. DNA to be transformed was added to competent cells (Table 2.1. DH5α or Tuner™(DE3) cells) and incubated on ice for 10 min. The tube was placed in a water bath or heat block at 42 °C for 30 sec and then placed on ice for 2 min. Following that, 250 µL SOC medium (2.1.5.4.) was added and the mixture was incubated in a shaker (220-250 rpm) at 37 °C for up to 60 min. If white/blue screening was performed, 50-200 µL of transformed cells was first added to 50 µL 2 % X-gal (2.1.7.3.) and 10 µL 1M IPTG (2.1.7.11.). Otherwise, the mixture was plated out directly on LB agar (amp) plates which were incubated at 37 °C overnight.

2.2.3.2. Into *S. cerevisiae*

In order to introduce plasmids into *S. cerevisiae*, lithium acetate method was chosen for yeast transformation (Schiestl and Gietz, 1989). Five ml of liquid YPD was inoculated with a freshly streaked single colony of desired *S. cerevisiae* strain and grown overnight at 30 °C with shaking at 200 rpm. This starting culture was then used the next day to inoculate 50 mL YPD to OD_{600} of 0.2. The culture was grown at 30 °C until OD 0.7 followed by centrifugation at 3,000 rpm for 3 min at room temperature to pellet the cells. The pellets were washed twice with 1 mL distilled water and centrifuged at 7,000 rpm for 30 sec after each wash. Finally, the pellet was resuspended in 300 µL of 100 mM LiAc to achieve competent cells. Once the cells were ready, the following components were mixed for each sample in the given order:

- 50% (w/v) PEG 240 µL
- competent cells 50 µL
- plasmid 800 – 1000 µg
- Salmon Sperm ssDNA 10 µL

The tube was briefly vortexed and at the end 32 µL 1 M LiAc was added. The content of the tube was again briefly vortexed and incubated for 20 min (one-plasmid transformation) or 30 min (two-plasmid transformation) at 30 °C and followed by heat shock for 15 min (one-plasmid transformation) or 20 min (two-plasmid transformation) at 42 °C. After temperature incubations, the samples were centrifuged at 7,000 rpm for 30 sec, then the pellets were washed twice with 1 mL distilled water to ensure complete pellet resuspension. Finally, the cells were resuspended in 1 mL distilled water and 200 µL was plated on selective media plates and grown for 2 to 3 days at 30 °C.

2.2.4. Sequencing

2.2.4.1. Sanger sequencing

The service of Eurofins Genomics was used for sequencing purified plasmid DNA. To sequence samples, the cycle sequencing technology (dideoxy chain termination/cycle sequencing) on ABI 3730XL sequencing machine was used. This method is a modification of the traditional Sanger sequencing method. It uses personally designed primers that are offered to be synthesized by Eurofins Genomics. The sample that was sent for sequencing was purified plasmid DNA (2.2.1.8.2.) of concentration 50-100 ng/µL.

2.2.4.2. Next-Generation Sequencing (NGS)

NGS libraries for analysis with MiSeq platform were prepared according to the manufactures' instructions and SOP at Isogenica Ltd. Data analysis was also performed at Isogenica Ltd.

2.2.5. Aminoacylation

The aminoacylation reaction was performed using the aminoacylation buffer (2.1.6.18.), 19 μM tRNA, 500 nM enzyme and 100 μM amino acid. The reaction was incubated at 37 °C for 30-40 min and the tubes were placed on ice immediately after until analysis.

2.2.6. Protein expression

2.2.6.1. In *E. coli*

Transformed Tuner™(DE3) cells (*E. coli*) were expressed in LB broth medium (2.1.5.2.) substituted with 100 μg/mL ampicillin. Starting culture was inoculated with a single colony from the agar plate and incubated overnight at 37 °C. The overnight culture was added in 1:100 or 1:50 ratio to pre-warmed LB in the culture flasks. The flasks were incubated at 37 °C in the shaking incubator at 220-250 rpm until the achieved growth was at least 0.4 and did not exceed 0.9 at OD_{600}. The growth was monitored by measuring 1 mL of culture at OD_{600} using a spectrophotometer. Once the appropriate growth was achieved, the expression was induced with IPTG if required, and the culture was incubated at a specified temperature and time. Post-expression cells were pelleted by centrifugation at 5,000 x *g* and the resulting pellets stored at -20°C.

2.2.6.2. In *S. cerevisiae*

Cells were freshly transformed using method 2.2.3.2. A single colony from the freshly transformed plate was inoculated in required SD Drop-Out media (2.1.5.6.) and grown at 30 °C at 200 rpm for the required amount of time.

2.2.7. Lysate production

2.2.7.1. Using sonication

Bacterial cell pellets were resuspended in 0.1 original culture volume of native cell lysis buffer (2.1.6.17.6.) and incubated on ice for 10 minutes. Post-incubation the cells were sonicated 10 times at 90% power for 20 sec with 2 min cooling period in between cycles using sonicator IKA®-Labortechnik ULTRA-TURRAX™ T25 (JANKE & KUNKEL). The insoluble fraction was pelleted by centrifugation at 16,000 x *g* for 20 min at 4 °C. The pellet was discarded and the supernatant with soluble fraction was transferred to a fresh tube and stored at -20 °C until required.

2.2.7.2. Using BugBuster®

Bacterial cell pellets were resuspended in BugBuster® MasterMix (Merck Millipore, previously Novagen®) at 5 mL per 1 g cell pellet. If required, BugBuster® could be supplemented with protease inhibitors such as cOmplete™, Mini Protease Inhibitor Cocktail (Roche). The mixture was incubated on the rocking platform for up to 20 min at room temperature. The insoluble fraction was pelleted by centrifugation at 16,000 x *g* for 20 min at 4 °C. The pellet was discarded and the supernatant with soluble fraction was transferred to a fresh tube and stored at -20 °C until required.

2.2.7.3. Using glass beads

Normally, 5 – 10 mL yeast culture was used to collect cells at 3,000 rpm for 5 min. To lyse the cells, pellets were resuspended in 600 µL buffer RLT supplemented with reducing agent, 14.3 M 2-Mercaptoethanol and then mixed with 500 µL acid-washed glass beads (425-600 µm) (Sigma-Aldrich). The mixture was lysed in TissueLyser LT (QIAGEN), for 5-10 min at 50 Hz.

2.2.8. Protein purification

2.2.8.1. Using Ni-NTA resins and chromatography column (at Aston)

Cell lysate (2.2.7.1. or 2.2.7.2.) was mixed with up to 1 mL Ni-NTA His•Bind® Resin (Merck Millipore previously Novagen®) and incubated on a rocking platform or roller mixer for 1 hour at the room temperature. The mixture was loaded onto Poly-prep® Chromatography Column (BioRad) and the cap was removed to allow flow-through, followed by two washes of the column with 4 mL wash buffer (2.1.6.17.7.). The protein was eluted 4 times using 0.5 mL elution buffer (2.1.6.17.8.). All fractions were collected in separate tubes and 20 μL was removed for analysis in SDS-PAGE gel.

2.2.8.2. Using Ni2+ column in ÄKTA Start (at Isogenica)

Bacterial cell lysate (2.2.7.2) was purified according to the manufacturer's protocol. Protein purification was performed on HisTrap™ HP 1 mL column (GE Healthcare Life Sciences) using ÄKTA start system (GE Healthcare Life Sciences). Prepared Binding and Washing buffer (2.1.6.17.1) were used. The protein was eluted over 30 ml in 30 fractions using Elution buffer (2.1.6.17.2.). Flow-through, washes and elutions were collected in separate tubes for analysis in SDS-PAGE gel. Software UNICORN Start was used to monitor the process of purification.

2.2.9. Fluorescence microscopy

For analysis of protein expression (eGFP) under fluorescence microscopy, 1 mL of yeast culture was removed and cells were collected at 7,000 rpm for 3 min. After removing 900 μL supernatant, the pellet was resuspended in the remaining volume. Fifteen μL was pipetted onto a glass plate and covered with a coverslip before using nail varnish to seal around the edges. To view the sample under the microscope, a drop of oil was added on the top of a lens and the inverted glass plate was placed under Leica DMI4000 B inverted microscope (Leica Microsystems) connected to Leica DFC360 FX camera with objective HCX PL FLUOTAR 63.0x1.25 OIL. Green fluorescence emission (508 nm)

images of cells were taken after excitation with 488 nm laser. The images were taken and processed in LAS AF software.

2.2.10. Western Blot

2.2.10.1. Semi-dry method

To allow the analysis in Western Blot, firstly, protein samples were resolved on SDS-PAGE gel (2.2.2.3.2.). To prepare for transfer of proteins from the gel to the membrane, PVDF membrane was pre-wetted in 100% methanol and then washed for 10 min in 1X transfer buffer (2.1.6.18.) with filter paper. The sandwich was assembled on a semi-dry blotter (Thermo Scientific) according to the manufacturer's protocol and transferred at 25 V for 35 min.

2.2.10.2. Protein detection

The membrane was blocked using 5% (w/v) skimmed milk solution in PBS-T (PBS with 0.1% Tween-20) at room temperature for 30 min. After, the membrane was incubated in 5% (w/v) skimmed milk solution in PBS-T with the primary HRP-conjugated antibody (at given concentration) at 4 °C overnight. The next day, the membrane was washed three times for 5 minutes in PBS-T. The membrane was then equilibrated for 2 min with earlier prepared EZ-ECL Enhanced Chemiluminescence Detection Kit for HRP. C-DiGit® Blot Scanner (LI-COR Biosciences) was used to capture the image of the blot using Image Studio Digits program (LI-COR).

2.2.11. Flow Cytometry

Flow cytometry was performed using FC500 (Beckman Coulter). For each sample, 100,000 events were acquired under a medium flow rate, and samples were collected in triplicates. The fluorescence emission for eGFP (508 nm) was measured after excitation at 488 nm. While for DRAQ5 Fluorescent

Probe Solution (Thermo Scientific™) subsequent excitation was at 647 nm followed by emission at 681 nm. The results were analysed using FlowJo program.

Chapter 3: Cloning of native and mutant *alaS* genes

3.1. Cloning of the wild type *E. coli* AlaRS gene

The *E. coli* alanyl-tRNA synthetase gene is called *alaS* and is 2631 base pairs (bp) long. This gene was amplified from *E. coli* by PCR (2.2.1.1.) and its size was confirmed in agarose gel electrophoresis (2.2.2.1.). The resulting product (Figure 3.1.) was phosphorylated (2.2.1.3.) and ligated (2.2.1.5.) into pET45b(+) that had been digested (2.2.1.1.) with *Pml*I, which generates blunt ends. The resulting ligated plasmids were transformed into *E. coli* and plated onto LB (amp) plates (2.2.3.1.). Resulting colonies were screened by colony PCR (2.2.1.1.2.) using a T7 promoter forward primer and an *alaS*-specific reverse primer, to generate a product of the expected size of 2742 bp. Two positive colonies were identified (Figure 3.2.). The correct sequence of native *alaS* in clone pET45b(+):*alaS* was confirmed by Sanger sequencing (2.2.4.1.).

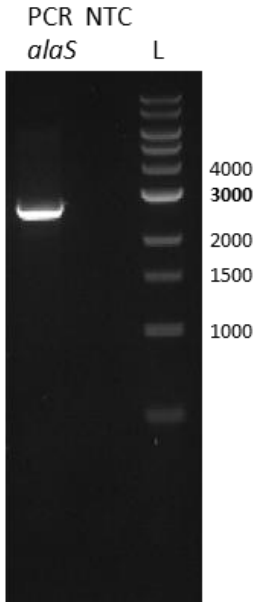


Figure 3.1. Amplification of alaS gene from E. coli Tuner™(DE3) cells.

1 % agarose gel stained with SYBR Safe dye shows results of PCR with an expected size of 2631 bp, amplified with Q5® High-Fidelity DNA Polymerase (NEB). L: 1 kb DNA ladder (NEB). NTC: a non-template control.

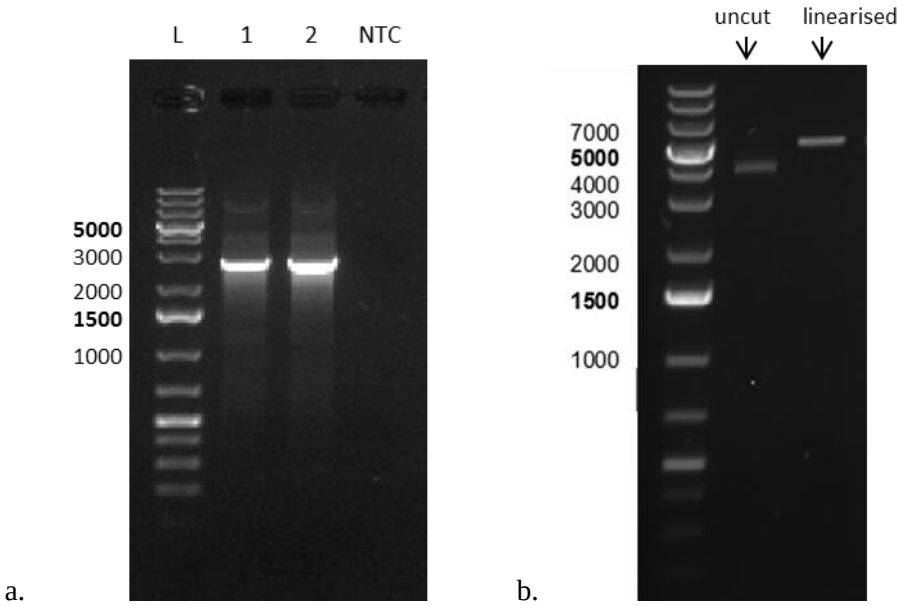


Figure 3.2. Colony screening of clones pET45b(+):alaS.

1 % agarose gels stained with SYBR Safe dye shows results of colony screening with a. colony PCR and b. enzyme digestion. a. Colony PCR for clone 1 and 2 with an expected size of 2742 bp amplified using T7 promoter forward primer and an alaS-specific reverse primer. b. Linearised clone 2 confirms the expected size of 7891 bp as a total plasmid length. L: GeneRuler 1 kb Plus DNA ladder. NTC: a non-template control.

3.2. Cloning of truncated gene enzyme N461_AlaRS

3.2.1. Gene amplification

The N-terminal domain of AlaRS is also called the aminoacylation domain and hence its name; this domain alone is sufficient to catalyse an aminoacylation reaction. As such, this domain can itself be used as a truncated enzyme, with the advantage of removal of editing domain function that is native to intact AlaRS. To clone the aminoacylation domain, the existing clone of native *alaS* described in Section 3.1. was amplified by inverse PCR with the workflow shown in Figure 3.3.

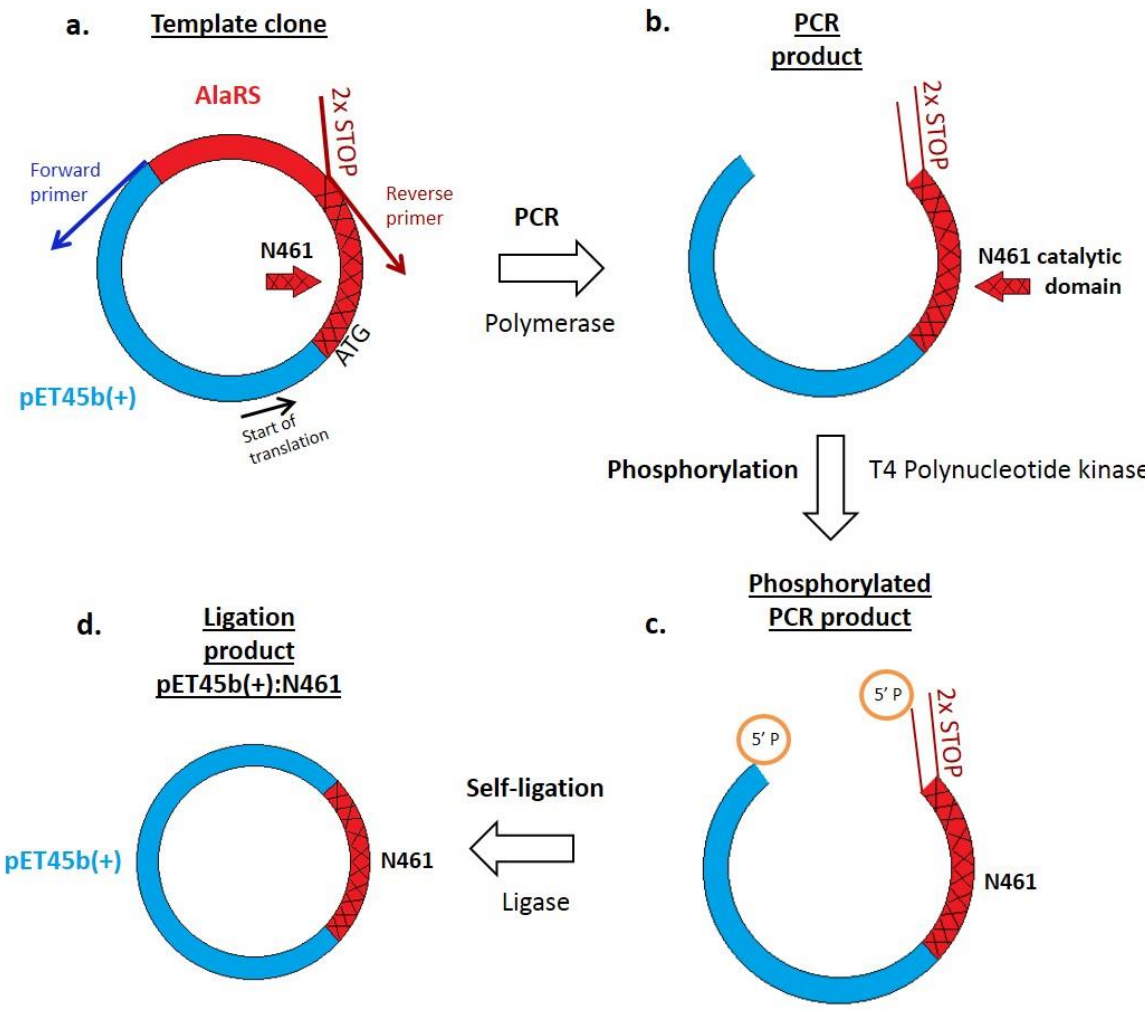


Figure 3.3. The workflow of cloning truncated enzyme N461_AlaRS.

a. Template clonal plasmid pET45b(+):alaS was amplified whilst concomitantly introducing two STOP codons via one of the PCR. b. The PCR product was phosphorylated by T4 Polynucleotide Kinase to aid ligation. c. The phosphorylated PCR product was ligated. d. Final product of the ligation, the truncated N461 AlaRS enzyme as a new clone pET45b(+):N461.

Appropriate primers were designed to amplify the vector and N-terminal domain only (first 461 amino acids) from clone pET45b(+):*alaS*. In addition, two termination codons were introduced in the primer annealing to the 3'-end of N461 domain. The PCR reaction using Phusion Polymerase amplified a product of 6649 bp with annealing temperature at 55 °C (2.2.1.1.). The product was analysed in 1 % TBE agarose gel by electrophoresis (Figure 3.4.). PCR reactions were purified using the QIAgen PCR Purification Kit (2.2.1.8.1.).

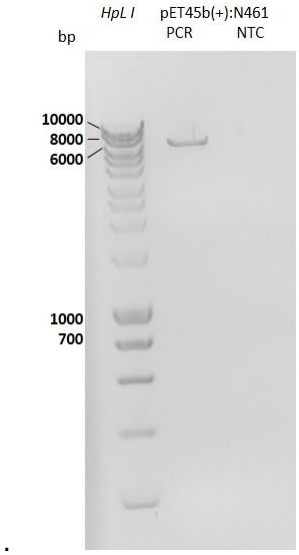


Figure 3.4. Amplification of pET45b(+):alaS for generation of truncated gene N461_AlaRS in pET45b(+) vector.

0.8 % agarose gels stained with SYBR Safe dye shows results of PCR with an expected size of 6649 bp. PCR product was amplified using Phusion Polymerase. HpL I: HyperLadder™ 1kb (Bioline). NTC: non-template control.

3.2.2. Self-ligation

Prior to ligation, phosphate groups were attached to the 5'-end of amplicons by the T4 Polynucleotide Kinase, following manufacturer's instructions, to permit subsequent intramolecular ligation of the PCR product. Self-ligation reactions were performed with varying quantities of the phosphorylated PCR product (65, 78, 98 ng) with 5 units of T4 DNA Ligase (5 U/μL) in a total volume of 30 μL for 1 h incubation at 22 °C, followed by heat inactivation at 72 °C for 5 min (2.2.1.5.).

The self-ligation reaction was assessed by PCR, using two primers: T7 terminal primer which sequence is found within pET45b(+) and a sequencing primer, designed and used previously for sequencing of clone pET45b(+):*alaS*. If self-ligation was successful and 5'- and 3'-ends of linear PCR amplicon were joined, the polymerase would amplify the product passing through the point of ligation and generate a 1090 bp PCR product. Figure 3.5. suggests that self-ligation had been successful, as indicated by the band of the expected size in lanes 1 and 2.

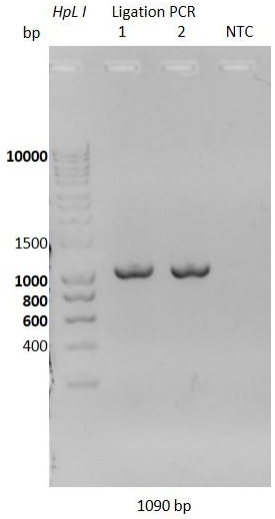


Figure 3.5. Amplification from self-ligation of pET45b(+):N461.

1 % agarose gels stained with SYBR Safe dye shows results of PCR of ligation sample with an expected size of 1090 bp. PCR product was amplified using Phusion Polymerase. HpL I: HyperLadder™ 1kb (Bioline). NTC: non-template control.

3.2.3. Transformation

Ligations products of pET45b(+):N461 were transformed into OneShot Chemically Competent Cells (*E. coli* DH5α strain) (2.2.3.1.). The results are shown Table 3.1. Although the transformation efficiency for the manufacturer's test plasmid was lower than expected (>1x10^9), the number of colonies achieved from the self-ligations was substantial.

Table 3.1. Transformation of self-ligated samples for pET45b(+):N461.

ligation	quantity of self-ligated DNA (ng)	Number of colonies	colony forming unit (CFU)
L1	65	300	6.38 x 10^5
		800	3.42 x 10^5
L2	78	160	2.86 x 10^5
		592	2.11 x 10^5
L3	95	328	4.62 x 10^5
		960	2.72 x 10^5
pUC19 (control)	0.01	51	1.41 x 10^6

3.2.4. Colony PCR

Colony PCR was set up using the same primers as when verifying the success of ligation, T7 terminal primer, sequencing primer, and Taq Polymerase in 30 µL reactions (2.2.1.1.2.). QIAxcel Advanced (QIAGEN) was used to analyse the results of colony PCR (Figure 3.6).

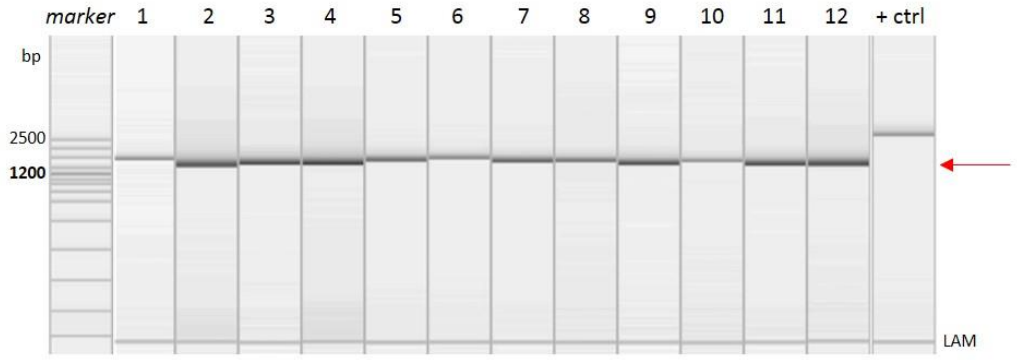


Figure 3.6. Colony screening for pET45b(+):N461 by colony PCR.

The results of colony PCR were analysed using QIAxcel Advanced – a capillary electrophoresis system. Colonies 1 to 12 display an expected size of PCR product at 1090 bp. LAM: lower alignment marker; + ctrl: positive control.

Four clones were selected and used to set up a 3 mL starting culture (2X YT (2.1.5.3.) with 100 µg/mL Amp). These cultures were used for plasmid DNA purifications (2.2.1.8.2.) that were then sent for Sanger sequencing. Results showed that three clones had 100 % correct sequence while one lacked the two STOP codons that were supposed to have been introduced by the C-terminal primer.

3.3. Cloning of a full-length *alaS* gene with a mutation to knock out the native editing function

It has previously been demonstrated that a Cys to Ala mutation at position 666 in the ORF of the full-length of AlaRS is sufficient to prevent editing (Guo *et al.*, 2009). This approach was employed in the current study to knock out the editing function in AlaRS from the existing clone pET45b(+):*alaS*. Primers were designed to amplify the full-length of pET45b(+):*alaS* with one primer introducing Ala instead of Cys at residue 666 (Figure 3.7.).

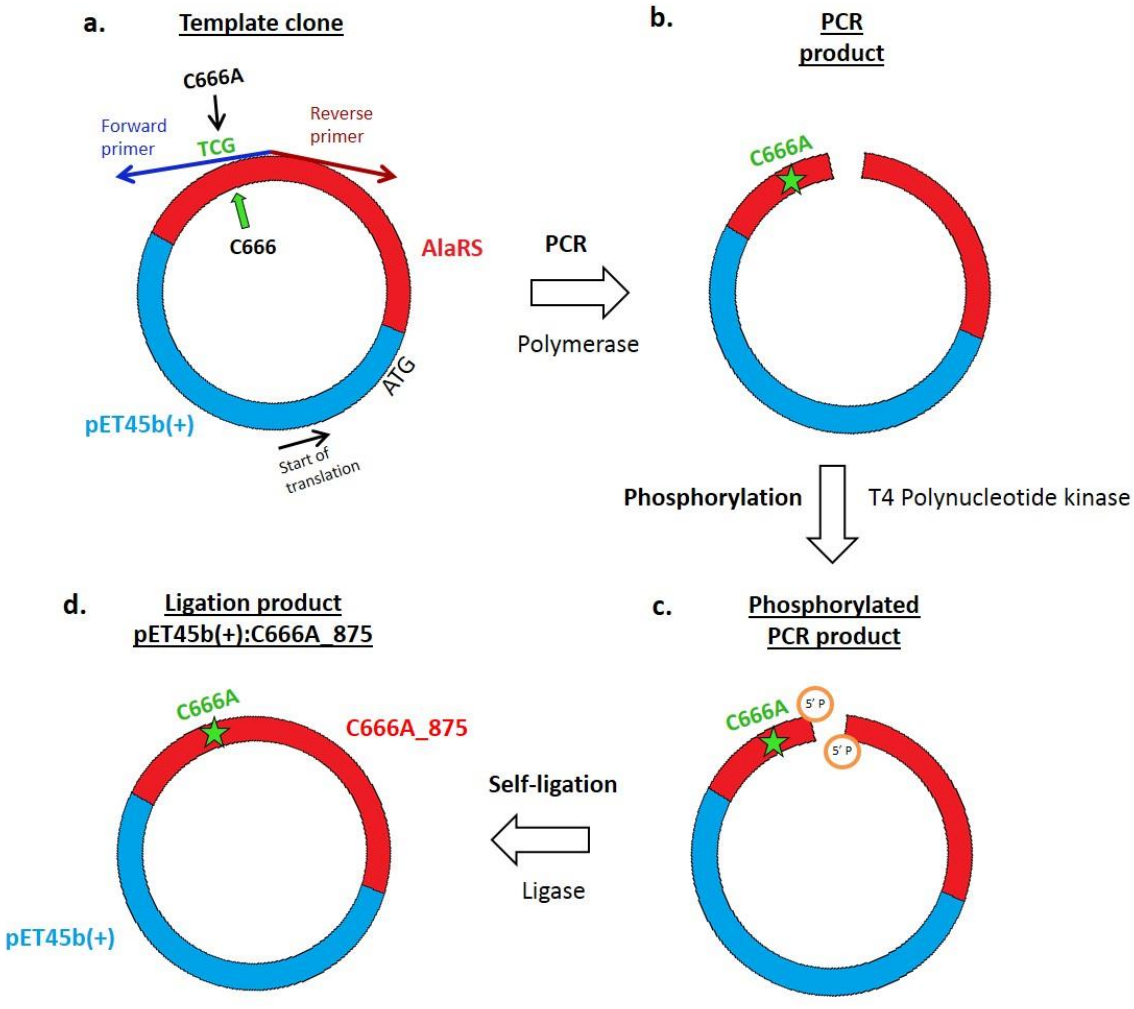


Figure 3.7. The workflow of introducing a point mutation into pET45b(+):alaS.

a. Template clone pET45b(+):alaS was amplified with primers while forward primer carried a C666A mutation to pET45b(+):alaS. b. The resulting PCR product with C666A point mutation was phosphorylated by T4 Polynucleotide Kinase to aid ligation. c. Phosphorylated PCR product was self-ligated. d. The final product of the ligation, a full-length enzyme with knock out of the native editing function C666A_AlaRS as a new clone pET45b(+):C666A_875.

3.3.1. Gene amplification

A gradient PCR with annealing temperatures from 58 °C to 70 °C was set up for 9 reactions of 50 µL each, using Phusion Polymerase (2.2.1.7.2., 2.2.1.1.1.). The results showed that at an annealing temperature of 65.2 °C and below, there was a secondary product of about 2500 bp, likely due to nonspecific annealing of the primer(s) (Figure 3.8.). The expected PCR product is 7891 bp and it corresponds to the top band in Figure 3.8. A single product is observed for the temperature 66.9 °C and above. Therefore, to amplify and mutate pET45b(+):*alaS* as required, a PCR with amplification temperature at 67 °C was set up in the total volume of 100 µL. Amplicons were purified (2.2.1.8.1.) and 50 ng and 100 ng of DNA were visualised on the gel (Figure 3.9.). The results from electrophoresis show a single product of the correct size of 7891 bp.

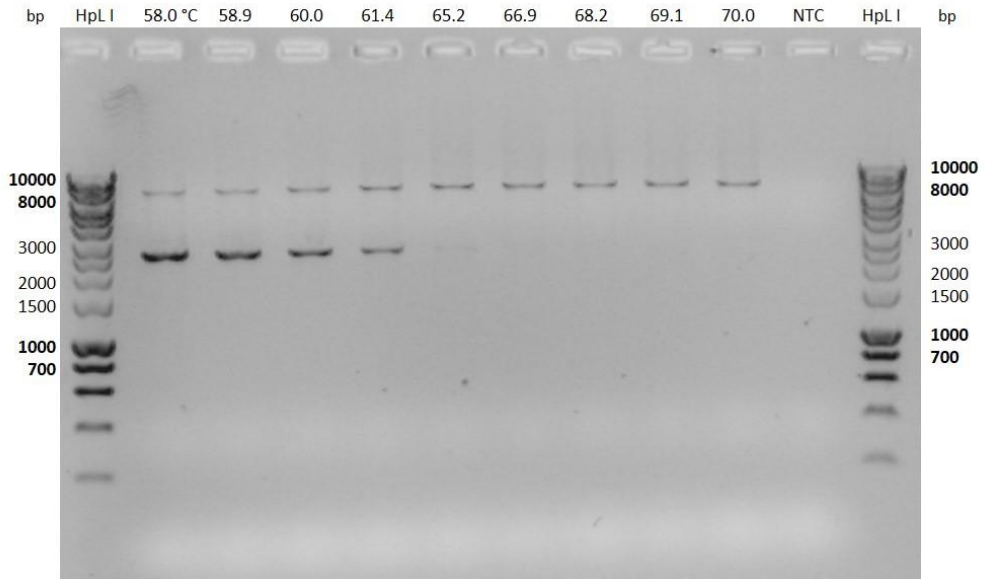


Figure 3.8. Gradient PCR for amplification of pET45b(+):alaS for generation of pET45b(+):C666A_875.

0.5 % agarose gels stained with SYBR Safe dye shows results of gradient PCR with an expected size of 7891 bp. Gradient PCR with annealing temperatures 58-70 °C was set for the amplification with Phusion Polymerase of pET45b(+):alaS to introduce the C666A mutation. The unspecific amplification is found for annealing temperatures between 58.0 and 65.2 °C. HpL I: HyperLadder™ 1kb (Bioline). NTC: non-template control.

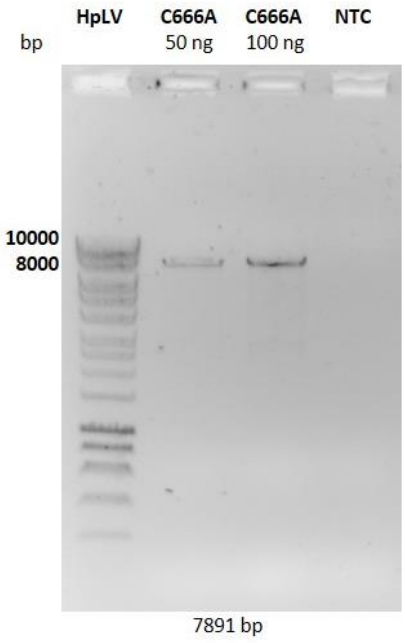


Figure 3.9. Amplification of pET45b(+):alaS for generation of pET45b(+):C666A_875.

0.5 % agarose gels stained with SYBR Safe dye shows results of PCR with an expected size of 7891 bp. PCR products of 50 ng and 100 ng were analysed in the gel. The product of this PCR, pET45b(+):C666A_875, is a full-length AlaRS gene with a point mutation to knock out the native editing function of the enzyme that is inserted in pET45b(+) vector. HpL I: HyperLadder™ 1kb (Bioline). NTC: non-template control.

Linearised PCR product pET45b(+):C666A_875 was then self-ligated (2.2.1.5.). The ligation reaction was preceded with the attachment of phosphate groups to the 5'-end of amplicons (2.2.1.3.). Ligation reactions were set up using a different amount of purified DNA pET45b(+):C666A_875 and they were catalysed with 5 units of T4 DNA Ligase, in the total volume of 40 μL. Incubation for 2 h 15 min at 22 °C was followed by enzyme heat deactivation for 5 min at 72 °C. Ligations were transformed into *E. coli* DH5α cells. Six colonies were selected for growing in liquid culture, and plasmids then extracted, as previously described. Purified plasmids were sent for DNA Sanger sequencing. Five out of 6 clones had 100% correct sequence and they were ready for subsequent use.

Chapter 4: Expression

4.1. Optimisation of expression

Initially, both the full sequence wtAlaRS and the truncated enzyme (N461_AlaRS) were expressed in *E. coli* Tuner™(DE3) cells under various conditions, in order to find out the optimal growth temperatures, time of induction and IPTG concentration for maximal expression of proteins (2.2.6.1.). Protein expression was examined in electrophoresis using SDS-PAGE gels (2.2.2.3.). Cell lysates were prepared with sonication or BugBuster® (2.2.7.).

4.1.1. IPTG concentration

Initially, different IPTG concentrations were examined to determine the optimal concentration for induction of expression (Figure 4.1.). Predicted sizes for enzyme wtAlaRS and N461_AlaRS (including a His tag of 1 kDa) are 97 and 53 kDa, respectively. These sizes were approximately confirmed in Figure 4.1a. and b. (predominant bands are indicated by red arrows), albeit that the PageRuler Plus 100 kDa marker migrates slightly faster in SDS-PAGE gel than the full-length AlaRS enzyme. Figure 4.1a. demonstrates that whilst there is slightly higher production of wtAlaRS in response to the highest (1.0 mM) concentration of IPTG, the expression of the truncated enzyme is similar across all IPTG concentrations tested.

4.1.2. Temperature

Post-induction growth temperature can affect both the quality (soluble or insoluble) and quantity of recombinant protein yield. Therefore, the effect of different temperatures was examined in the overnight expression of both full sequence and truncated enzyme. Both supernatant and cell debris, obtained after cell lysis, were analysed. Figure 4.2. shows that for wt875_AlaRS expression was observed at 25 °C as well as 30 °C but the protein is found in the insoluble fraction of cell debris

after cell lysis. Conversely, Figure 4.2. demonstrates a very good expression of soluble N461_AlaRS when the cells are cultured at 30 °C post-induction.

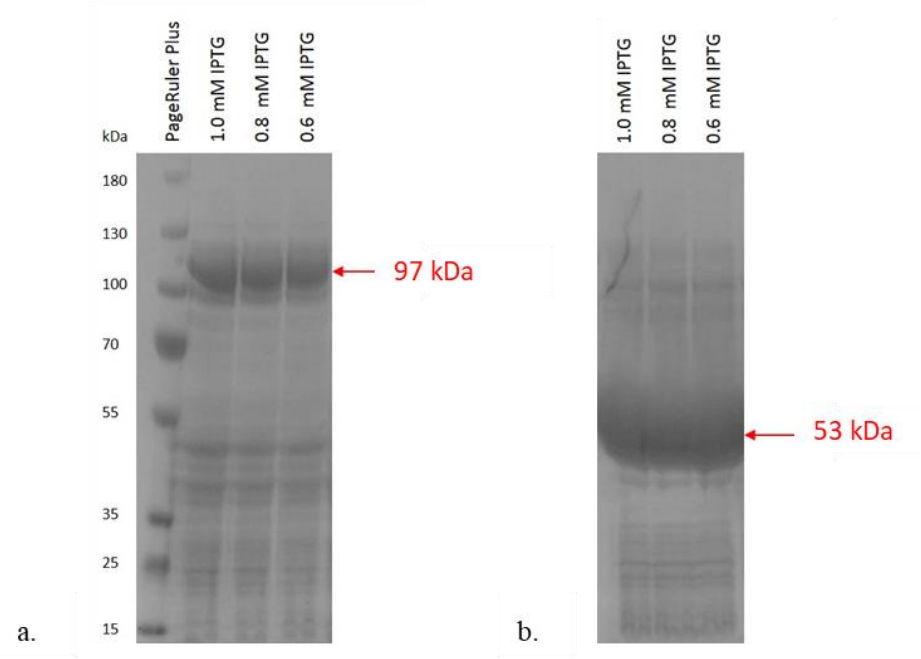


Figure 4.1. Expression of a. wild type full-length AlaRS (wt875_AlaRS) and b. truncated enzyme (N461_AlaRS) using different IPTG concentrations.

Polyacrylamide gel electrophoresis using 4-12% gradient SDS-PAGE shows a. wt875_AlaRS and b. N461_AlaRS expression induced with 1.0, 0.8 and 0.6 mM IPTG for 3 h at 30 °C with shaking (220 rpm). Both gels show whole cell lysate samples with overexpression of full sequence (MW=97 kDa) and truncated (MW=53 kDa) enzymes indicated by red arrows. The gels were stained with InstantBlue.

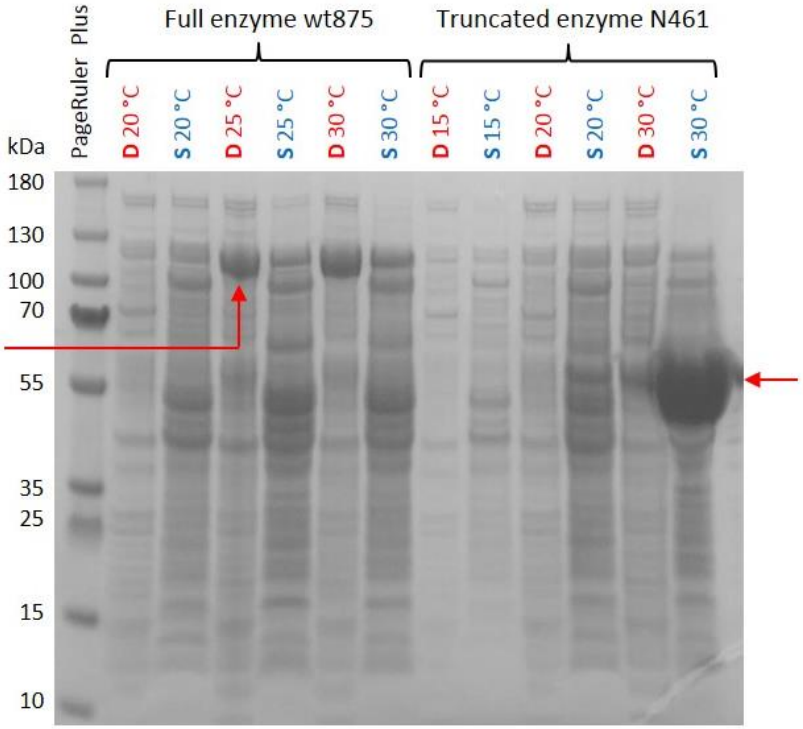


Figure 4.2. Expression of wild type full-length AlaRS (wt875_AlaRS) and truncated enzyme (N461_AlaRS) at different temperatures in cell debris and supernatant.

Polyacrylamide gel electrophoresis using 4-12% gradient SDS-PAGE shows samples from the overnight uninduced expression of full-length and truncated enzyme at different temperatures, in cell debris (labelled D) and supernatant (labelled S) after cell lysis. Red arrows indicate expressed wt875_AlaRS (to the left) and N461_AlaRS (to the right). The gel was stained with InstantBlue.

4.1.3. Reiteration of IPTG concentration

Having selected optimal expression temperatures of 25 and 30 °C, further IPTG induction concentrations were examined to compare production levels of both the full sequence and truncated enzymes (Figure 4.3.).

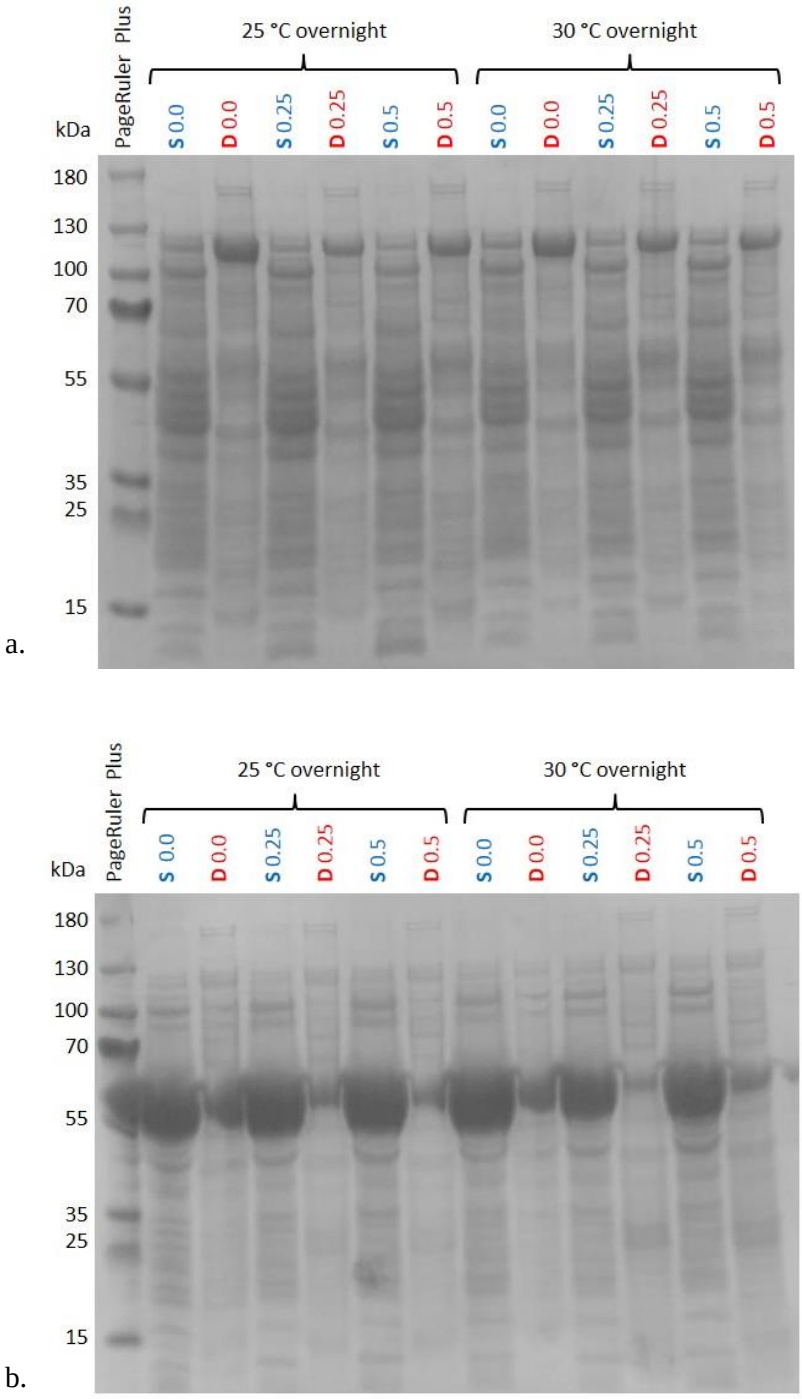


Figure 4.3. Expression of a. wild type full-length AlaRS (wt875_AlaRS) and b. truncated enzyme (N461_AlaRS) samples using different temperatures and IPTG concentrations.

Polyacrylamide gel electrophoresis using a 4-12% gradient SDS-PAGE of samples from overnight expression at 25 and 30 °C in the absence or presence of 0.25 and 0.5 mM IPTG, for a. the full-length and b. truncated enzymes. D: cell debris. S: supernatant. The predominant bands show the expression of the desired enzyme. The gels were stained with InstantBlue.

The determined conditions were then used to express C666A_AlaRS, which showed similar expression characteristics to those observed for wtAlaRS as shown in Figure 4.4. where the protein expression of different samples is compared alongside.

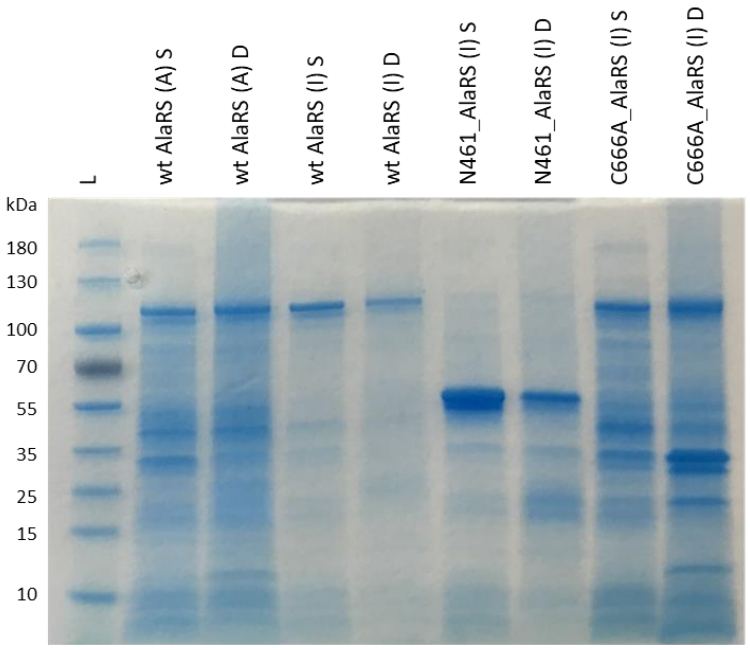


Figure 4.4. Expression of wtAlaRS, C666A_AlaRS and N461_AlaRS in E. coli.

Polyacrylamide gel electrophoresis using a 4-12% gradient SDS-PAGE shows sample of expression wtAlaRS at Aston for (A) and Isogenica for (I), as well as N461_AlaRS at Isogenica, and C666A_AlaRS at Aston. Samples were analysed after lysis but before purification. Conditions for expression of the proteins were as followed: wtAlaRS (Aston): 1 mM IPTG, 3h at 37 °C; wtAlaRS (Isogenica): 0.6 mM IPTG, 4h at 30 °C; C666A_AlaRS (Aston): 1mM IPTG, 4h at 37 °C; N461_AlaRS (Isogenica): 0.6 mM,4h at 30 °C. Each expressed protein compares soluble (S) fractions (supernatant), versus insoluble (D) fractions (cell debris). Predominant bands show the expression of the desired enzyme with the expected size of: 97 kDa for wtAlaRS and C666A_AlaRS, and 53 kDa for N461_AlaRS. The gel was stained with InstantBlue.

4.2. AlaRS purification

4.2.1. AKTA purification

Since the yield of the truncated protein N461_AlaRS was much greater, this protein was purified using the AKTA system (2.2.8.2.), which is a liquid chromatography system using columns in protein purification and allows precise analysis of eluted fractions. Analysis of the starting material (whole cell lysate - WCL), flow-through, wash and eluted fractions by PAGE confirmed a very good level of expression of the truncated protein (1 mM IPTG at 30 °C for 3 h) (Figure 4.5.). The maximal yield of soluble, recombinant protein was observed in elution fractions 9-11, whilst the final fraction (lane 14) shows that the vast majority of the protein had been eluted by this point. The purified samples show a major band around 55 kDa marker band corresponding to N461_AlaRS and a low level of co-purified protein below the size of N461_AlaRS. The results show that the protein could be isolated in high quantities, and would thus be appropriate for future *in vitro* application.

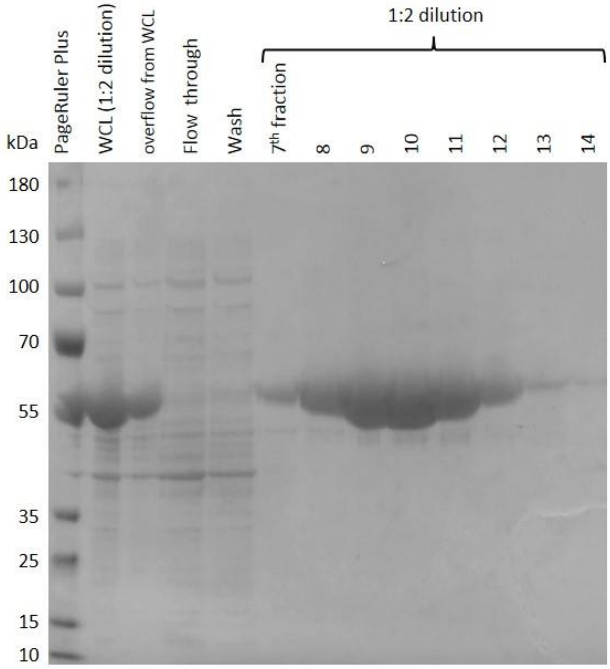


Figure 4.5. Expression and purification of truncated enzyme N461_AlaRS using AKTA start.

Polyacrylamide gel electrophoresis using a 4-12% gradient SDS-PAGE to examine the samples of expression of the truncated enzyme when induced with 1 mM IPTG and incubated at 30 °C for 3 h, followed by AKTA purification of the soluble enzyme. The size of the protein is expected to be 53 kDa (with His tag). The gel was stained with InstantBlue. WCL: whole cell lysate and fractions 7-14 were diluted 1:2.

4.2.2. Laboratory-scale, affinity purification and activity of wtAlaRS, C666A_AlaRS and N461_AlaRS.

Using the determined conditions, each of the proteins was expressed in 50 ml cultures and purified using bench-scale affinity chromatography (2.2.8.1.). As illustrated in Figure 4.6., the truncated protein gave the highest yield, but all were expressed and purified successfully.

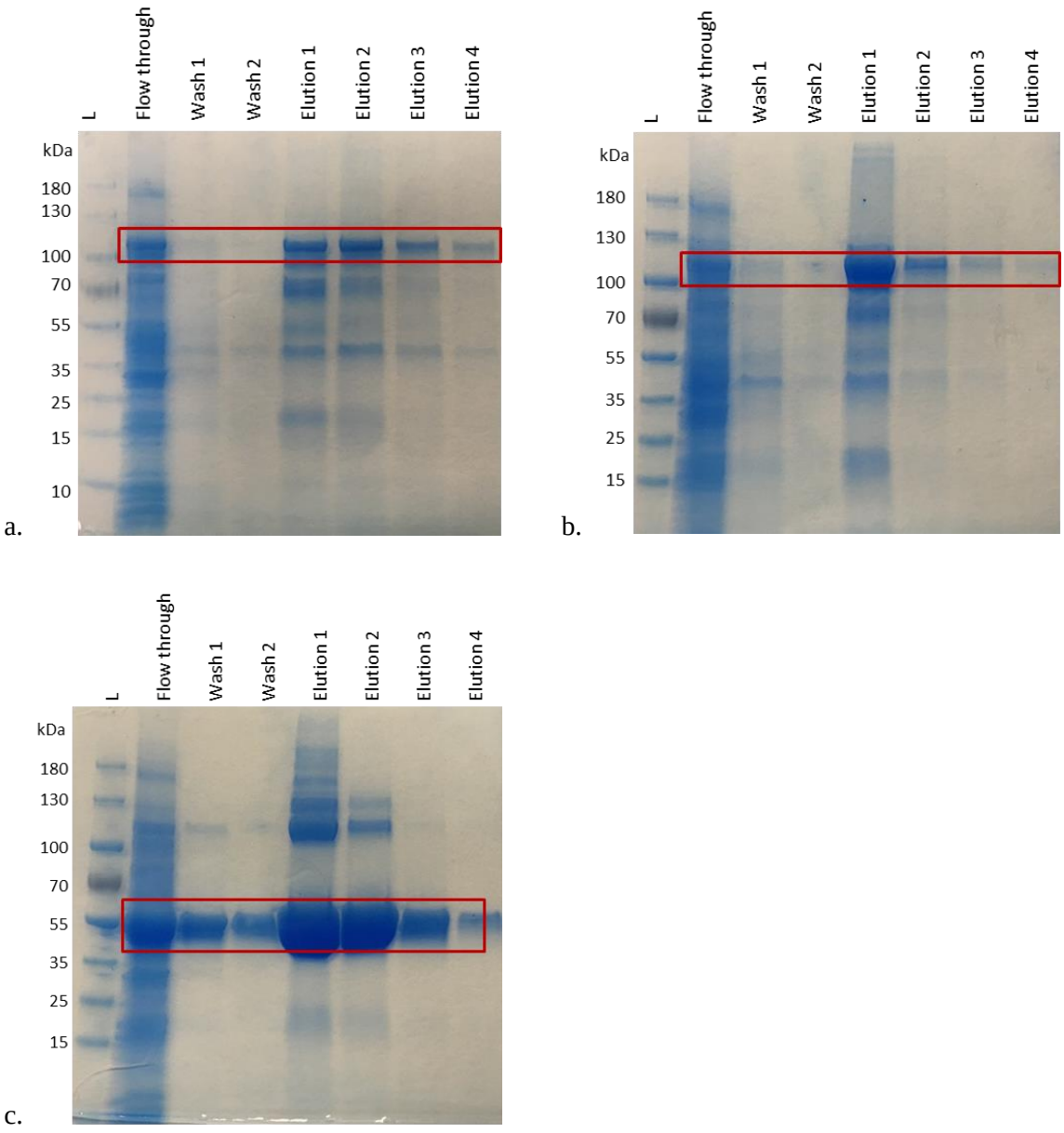


Figure 4.6. Purification of a. wtAlaRS, b. C666A_AlaRS and c. N461_AlaRS expressed in E. coli.

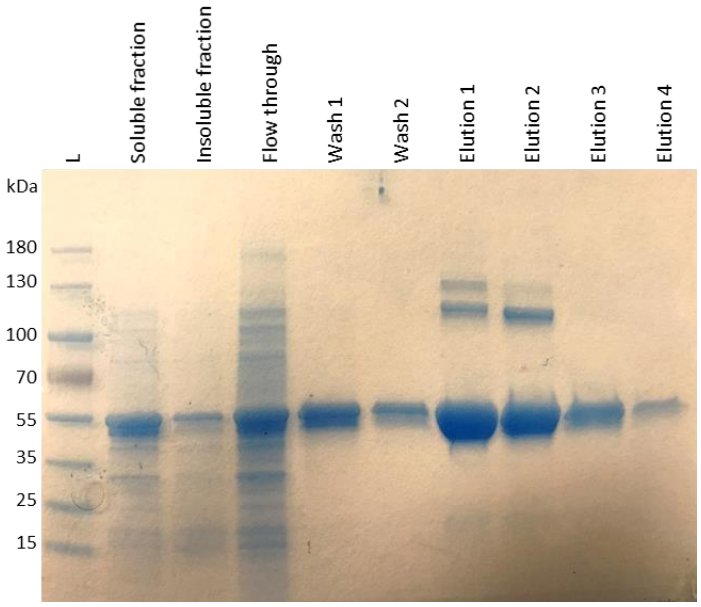
Polyacrylamide gel electrophoresis using a 4-12% gradient SDS-PAGE shows samples of expression and purification also shown in Figure 4.4. with details about induction of expression. The purification was an affinity column chromatography. The expected size of the protein is 97 kDa (wtAlaRS and C666A_AlaRS and 53 kDa (N461_AlaRS). The gel was stained with InstantBlue.

4.3. Aminoacylation activity

Interestingly, with a few exceptions, most publications in the field do not demonstrate aminoacylation of tRNA by purified enzymes *per se*, but rather tend to show the incorporation of UAAa (following aminoacylation) into proteins and/or peptides. To determine whether it was possible to detect aminoacylation reliably, both mass spectrometry of commercial, pre-acylated tRNA and acid-urea PAGE gel (2.2.2.3.3.) analysis of native tRNA, commercial, pre-acylated tRNA and tRNA aminoacylated with one or more of the purified enzymes was attempted (Figure 4.7.). Figure 4.8. shows the result of electrophoresis pre-charged and uncharged $tRNA^{Phe}$ in acid-urea PAGE gel with a difference in migration distance between the samples. Uncharged tRNAs are expected to migrate faster because their mass is lower than those of charged tRNAs. The results of aminoacylation (2.2.5.) of $tRNA^{Ala}$, the $tRNA^{Ala}$ microhelix (Musier-Forsyth and Schimmel, 1999) and the $tRNA^{Ala}$ tetraloop (Musier-Forsyth and Schimmel, 1999) are shown in Figure 4.9.

Unfortunately, none of the methodologies were able to clearly distinguish the difference between acylated and nonacylated tRNAs (Figure 4.8. and Figure 4.9.) and the activities of the purified enzymes could not, therefore, be determined with certainty. As an example, Figure 4.9. shows aminoacylation of $tRNA^{Ala}$ microhelix with L-Ala by truncated enzyme. The difference between lane 5 and 9 that only have unacylated $microhelix^{Ala}$ and lane 6 that should have acylated $microhelix^{Ala}$ as it is a sample of aminoacylation shows somehow slower migration for lane 6, at least partially. The band in lane 6 appears thicker what would indicate that there are two populations of $microhelix^{Ala}$, acylated with L-Ala and unacylated and/or deacylated.

a.



b.

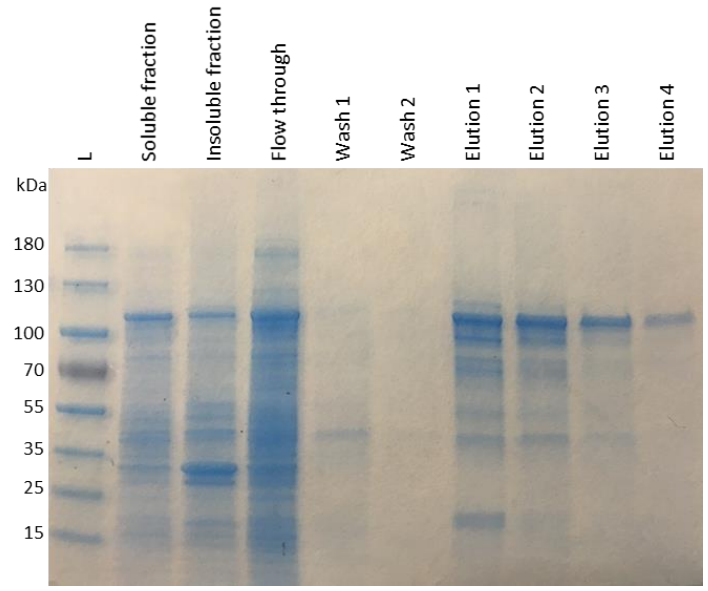


Figure 4.7. Expression and purification of a. N461_AlaRS and b. C666A_AlaRS for aminoacylation studies.

Polyacrylamide gel electrophoresis using a 4-12% gradient SDS-PAGE to show expressed a. N461_AlaRS and b. C666A_AlaRS. The expression was when induced with 1 mM IPTG and incubated at 30 °C for 4 h, followed by lysis and column purification of the soluble enzyme.

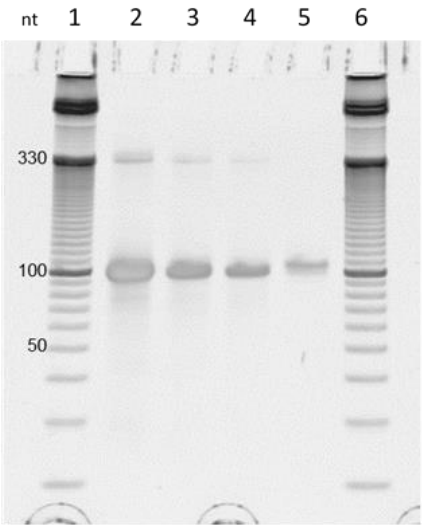


Figure 4.8. Comparison of commercial unacylated and acylated control samples of tRNA^{Phe}.

12 % acid-urea PAGE gel stained with 1X SYBR® Gold Nucleic Acid Gel Stain shows electrophoresis of unacylated and acylated controls tRNA^{Phe}. Samples were commercially acquired. Electrophoresis was at constant 7 W for 2 h. Lanes: 1: DNA ladder; 2: unacylated tRNA^{Phe} 0.1 unit (A260); 3: uncharged tRNA^{Phe} 0.005 unit (A260); 4: uncharged tRNA^{Phe} 0.0025 unit (A260); 5: acylated tRNA^{Phe} 22.5 pmol; 6: DNA ladder. Note: DNA ladder used was not to reference the size of tRNA^{Phe} but to confirm the electrophoresis experiment was successful.

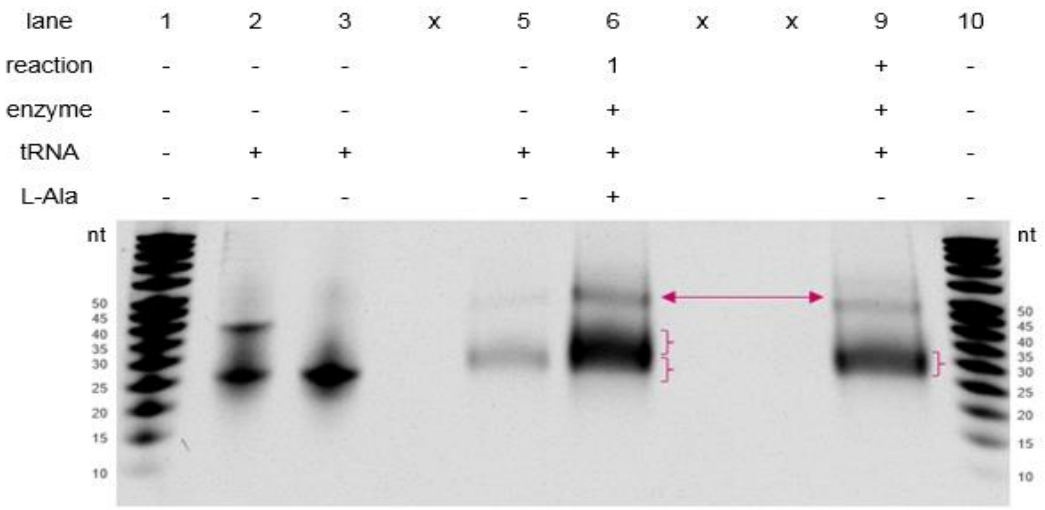


Figure 4.9. Aminoacylation reactions with controls and microhelix^{Ala}.

20 % acid-urea PAGE gel stained with 1X SYBR® Gold Nucleic Acid Gel Stain shows electrophoresis of aminoacylation reaction which involved truncated enzyme, microhelix^{Ala} and L-Ala. The electrophoresis was at 100 V, 4 °C until bromophenol blue reached the bottom of the gel. Lanes: 1: Low Molecular Weight Marker, 10-100 nt (Affymetrix); 2: unprocessed 100 pmol microhelix^{Ala}; 3: boiled 100 pmol microhelix^{Ala}; 5: folded and ethanol precipitated microhelix^{Ala}; 6: ethanol precipitated aminoacylation reaction 1 (5 µL); 9: ethanol precipitated aminoacylation reaction 2 without L-Ala (5 µL), 10: Low Molecular Weight Marker, 10-100 nt (Affymetrix). "x" indicates wells where samples were not loaded. Pink braces indicate the presence of bands.

4.4. Discussion

Both truncated and full-length AlaRS enzymes with removed editing function were effectively expressed in *E. coli* Tuner™(DE3) cells. The optimisation included changing IPTG concentration and temperature. Different IPTG concentration of 0.6 mM, 0.8 mM and 1.0 mM used for induction of expression wt875_AlaRS and N461_AlaRS did not change the expression levels. However, during temperature optimisation, it was demonstrated that protein was expressed at 25 °C and 30 °C but did not express at 20 °C for both, wt875_AlaRS and N461_AlaRS. During expression of wt875_AlaRS, it was shown that a large proportion of protein goes into the insoluble fraction (debris). This was not observed with the truncated protein suggesting that perhaps the full-length protein is harder for *E. coli* to fold when overexpressed.

Despite multiple approaches, methods used to detect acylated tRNAs were unable to show the difference between acylated and nonacylated tRNAs convincingly (Figure 4.8. and Figure 4.9.). In order to achieve better resolution on acid-urea PAGE gels, shorter species of $tRNA^{Ala}$, $microhelix^{Ala}$ and even $tetrahelix^{Ala}$ (data not shown) were used in aminoacylation reactions. Unfortunately, even these failed to show a clear difference. Inability to distinguish between acylated and nonacylated tRNAs also applied to commercial, pre-acylated tRNA, thus, suggesting that the problem was not with aminoacylation reaction but with detection methods. Having said that, the aminoacylation reaction could not be evaluated and therefore, its outcome remains unknown. Nevertheless, aminoacylation reactions are generally not shown in the studies of incorporating UAAs in the protein but instead, the protein is directly used *in vivo* or *in vitro* experiment and the enzyme's function is assessed in different methods, for example, by generation of a fluorescent product.

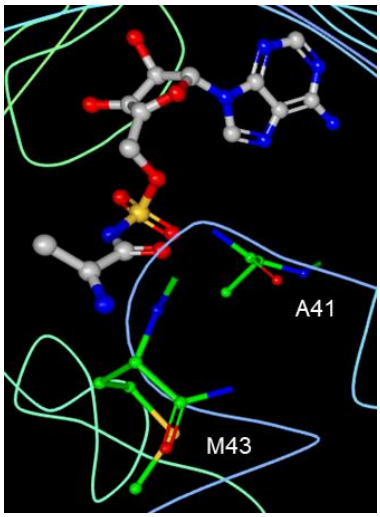
Chapter 5: Library generation

5.1. Library design: Modelling of the *E. coli* AlaRS amino acid binding site

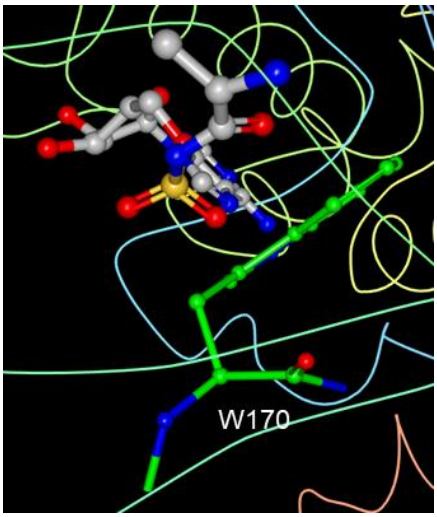
Library generation was a key component of this study which determined the direction of engineering novel enzymes aaRSs. The fundamental component was the choice of enzymes and their site of mutagenesis. The first library was based on the truncated enzyme – aminoacylation domain that fulfils the requirements of tRNA acylation and works on its own without other domains (Section 3.2.). The second library was constructed on the full-length gene with a knock out of the native editing function via C666A point-mutation (Section 3.3.). In both libraries, the target for mutations was the amino acid binding pocket found in the N-terminal domain for which the crystal structure had been previously resolved for *E. coli* and is available in Protein Bank Database: reference 3HXU. This structure also exists with bound alanyl-AMP analogue (Ala-SA) which allowed for a close examination of residues found in close proximity to aminoacyl moiety. Upon assessment of the crystal structure, those amino acids whose functional groups interact directly with L-Ala or lie in close proximity were carefully chosen for mutations. Codon saturation was introduced by ProxiMAX randomisation (2.2.1.7.1; controlled saturation mutagenesis, Ashraf *et al*., 2013). In total, six positions from the amino acid binding site were targeted in an effort to promote binding of selected UAA and expand the scope of the site (Figure 5.1.). These positions were: alanine 41 (A41), methionine 43 (M43), tryptophan 170 (W170), asparagine 212 (N212), valine 214 (V214), and methionine 216 (M216). In order to produce the most variable library possible (facilitated by ProxiMAX), each position was targeted with 18 amino acids (excluding methionine and cysteine which are prone to oxidation and often excluded from libraries by ProxiMAX). Consequently, each library size was 18^6, equal to 34,012,224 different genes and it remains within maximal capacity in terms of variant numbers of many screening methods.

This specific randomisation of six positions is unique in the production of aaRS library for engineering purposes. If the aims of this engineering studies were to be met, the diverted selectivity of *E. coli* AlaRS towards UAA would be paired with $tRNA^{Ala}$ that has mutated anticodon to accommodate new amino acid in ProxiMAX randomisation platform.

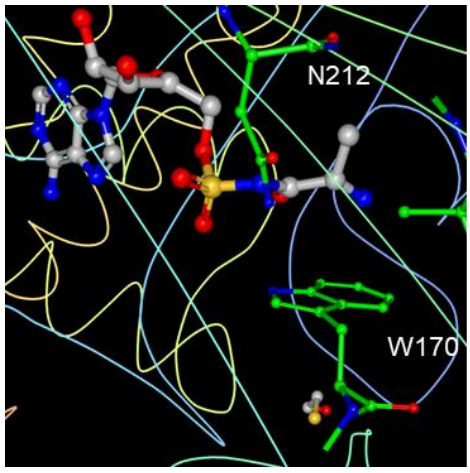
a.



b.



c.



d.

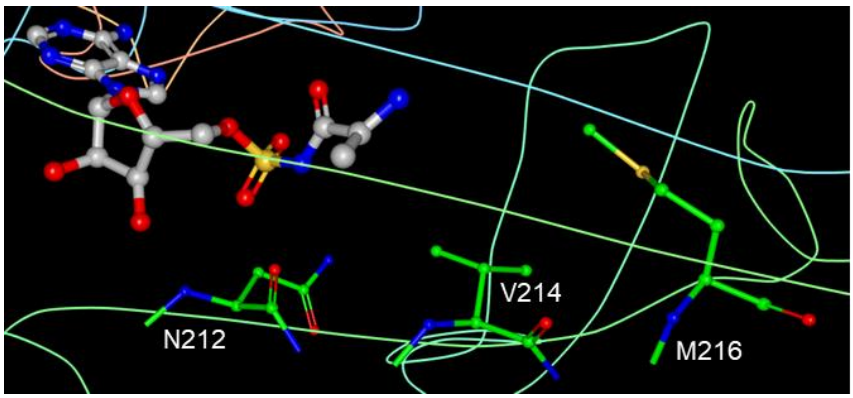


Figure 5.1. Randomisation position in the amino acid binding pocket of AlaRS.

Program RCSB - Ligand Explorer Viewer was used to view PBD entry 3HXU. Six positions in the amino acid binding site were chosen for randomisation and they are: a. alanine 41 (A41) and methionine 43 (M43); b. tryptophan 170 (W170); c. asparagine 212 (N212), valine 214 (V214), and methionine 216 (M216). Each of these six positions was mutated with 18 amino acids (excluding methionine and cysteine) what generated a library of 34,012,224 genetic variants. Two libraries were generated based on this randomisation: first including truncated enzyme with aminoacylation domain only, while second had a full-length enzyme but with a knock out of the native editing function.

5.2. Library construction

Two variants of the AlaRS enzyme were used for library generation: the full-length AlaRS containing the C666A point mutation to knock out the editing function (C666A_AlaRS) (Guo *et al.*, 2009) and the N-terminal aminoacylation domain only of AlaRS, comprising amino acids 1-461 (N461_AlaRS). The genes were divided into segments of 'framework' (FW – conserved regions of the gene) and 'library' (Lib – sections of the genes that would incorporate codon randomisation, Figure 5.2.). Codon saturation was introduced by ProxiMAX randomisation (controlled saturation mutagenesis; Ashraf *et al.*, 2013) at residues 41 and 43 (Lib1) and 212, 214 and 216 (Lib2), whilst residue 170 was instead saturated simply by employing multiple PCR primers (2.2.1.7.2.). To distinguish its method of construction, this section was named Framework 3 (FW3). All fragments but the last framework, FW4, were the same for both libraries. All residues were randomised with an equal mix of 18 amino acids, excluding methionine and cysteine, since these residues are prone to oxidation within proteins. Schematics for these gene libraries are shown in Figure 5.2.

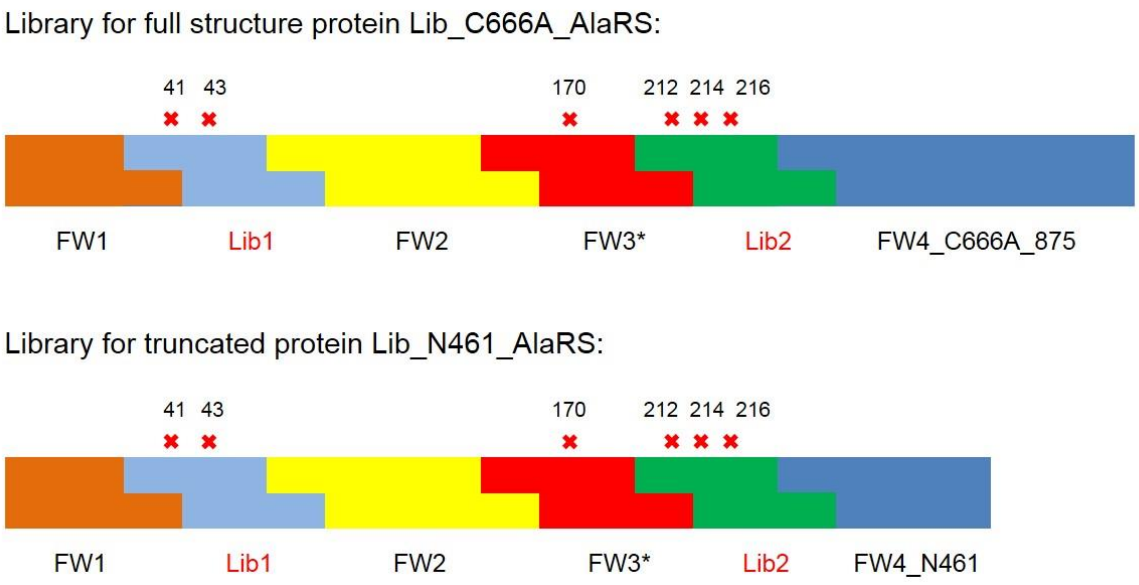


Figure 5.2. The strategy of mutagenesis for libraries Lib_C666A_AlaRS and Lib_N461_AlaRS.

The figure shows a breakdown of libraries as a mixture of framework and library fragments with the location of randomised positions 41, 43, 170, 212, 214 and 216. FW: framework, a fragment without randomisation. Lib: library, a fragment with randomisation by ProxiMAX. Note: FW3 is a framework fragment with a randomisation position that was introduced by PCR primers.*

5.3. Library assembly design

The libraries were then assembled by step-wise ligation of *Bsa*I-treated fragments with compatible overhangs as demonstrated in Figure 5.3. a. and b.

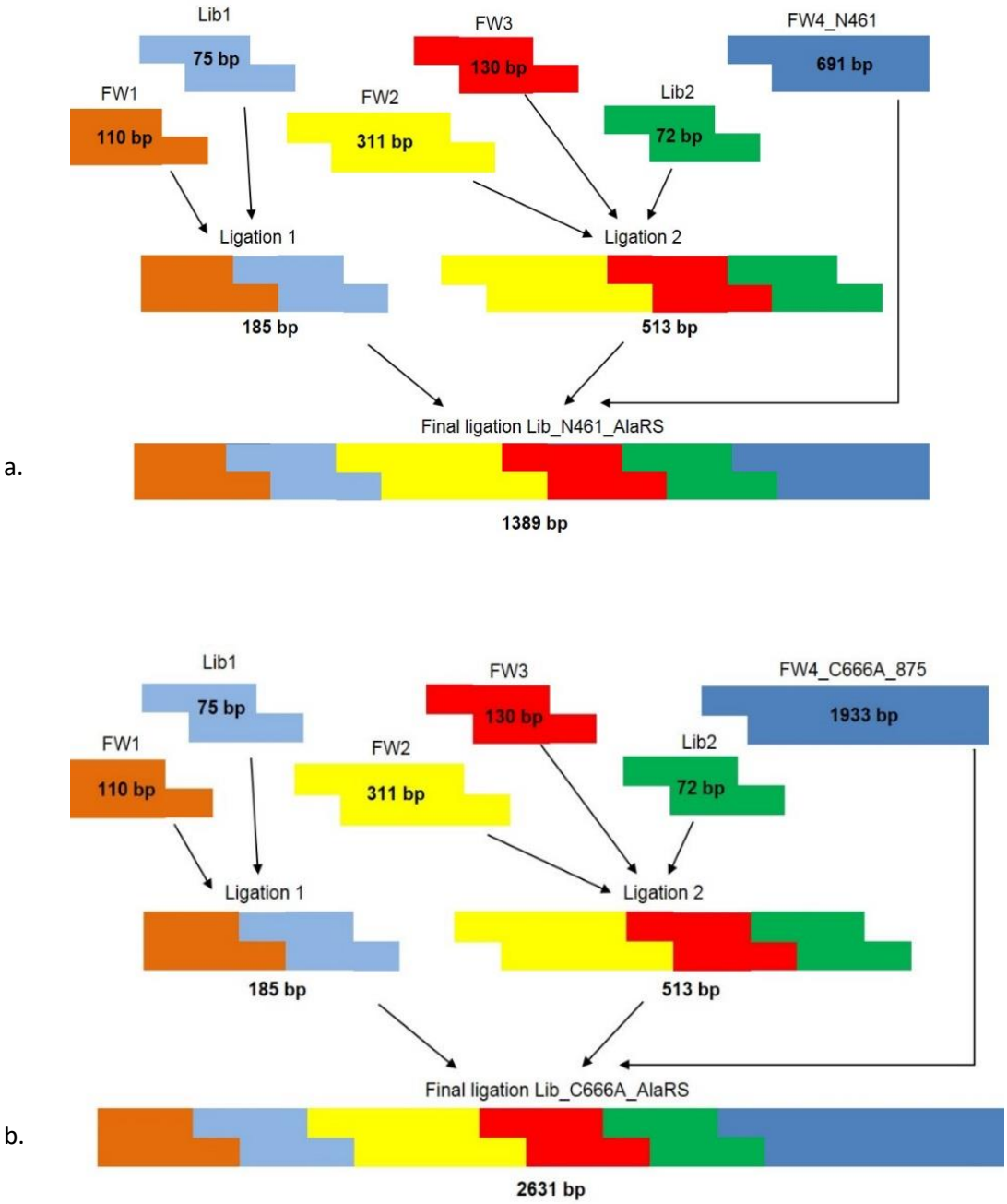


Figure 5.3. The workflow of assembly for a. Lib_N461_AlaRS and b. Lib_C666A_AlaRS.

Ligation of library fragments was performed in the particular order and stages as illustrated: Ligation 1, Ligation 2, etc.

*Bsa*I is a type IIS restriction endonuclease that recognises asymmetric DNA sequence outside of their recognition sequence and the cleavage produces a 5‘-sticky-ended overhang of a variable sequence as shown in Figure 5.4. When producing FW and Lib fragments the primers were used to attach *Bsa*I recognition site. Using *Bsa*I enzyme allowed for flexibility in choosing the cleavage site and the fragment size. The sticky-ends of all fragments had to be carefully designed in order to discourage ligation of fragments that are not positioned next to each other.

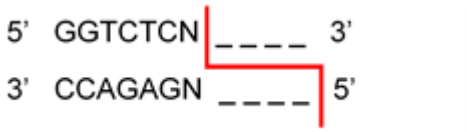


Figure 5.4. BsaI restriction enzyme cleavage site demonstrating 5'-overhang.

5.4. Generation of library fragments

The conserved library fragments FW1, 2, and FW4_N461 were amplified from clone pET45b(+):N461. PCRs were performed with Phusion Polymerase in 100 µL reactions (2.2.1.1.) and an annealing temperature of 60 °C. PCR products were purified (2.2.1.8.1.) and digested with *Bsa*I enzyme (2.2.1.6.). The resulting digested fragments were gel-purified (2.2.1.8.3.) and examined by agarose gel electrophoresis (2.2.2.1.; Figure 5.5.). The difference in size between digested and undigested samples is observed for FW1 and FW2. However, the difference between undigested and digested FW4_N461 was only 10 bp and with a fragment size of about 700 bp it was impossible to discern.

As already described, the saturation of codon 170 in FW3 (Figure 5.2.) was introduced by PCR using multiple primers. In fact, it was a mixture of 18 forward primers in equimolar quantities, each encoding one of the desired 18 amino acids (all amino acids excluding methionine and cysteine). A PCR was set up using pooled forward primers and a reverse primer, Phusion Polymerase and clone

pET45b(+):*alaS* as a template in 100 µL reactions. The annealing temperature was 60 °C. The product of this reaction is shown in Figure 5.6. Once FW3 was amplified it was purified using column (2.2.1.8.1.), digested with *Bsa*I and gel purified. Fragment FW4_C666A_875 was prepared in the same manner, using pET45b(+):C666A_875 as a template and 60 °C for the annealing temperature in the PCR reaction.

Generation of library fragments was performed as described previously via ProxiMAX randomisation (2.2.1.7.1.; Ashraf *et al.*, 2013). When the fragments were randomised, the samples were purified, followed by digestion with *Bsa*I and gel extraction of digests. Figure 5.7. confirms that the sizes of Lib 1 & 2 fragments before and after digestion were as required.

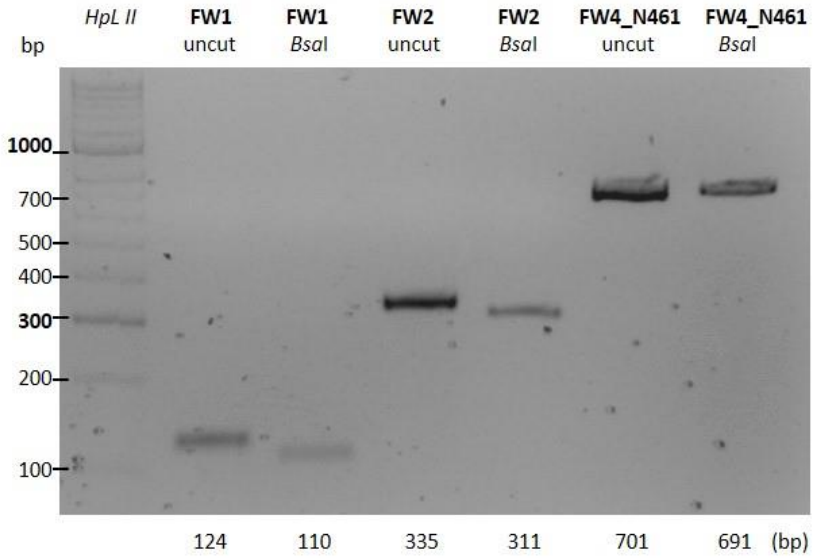


Figure 5.5. Enzyme digestion of fragments FW1, FW2 and FW4_N461.

2 % agarose gel stained with SYBR Safe dye show results of enzyme digestion for fragments FW1, FW2 and FW4_N461. The digestion was performed with BsaI and the gel shows fragments before and after digestion with an indication of expected sizes below the gel. HpL II: HyperLadder II.

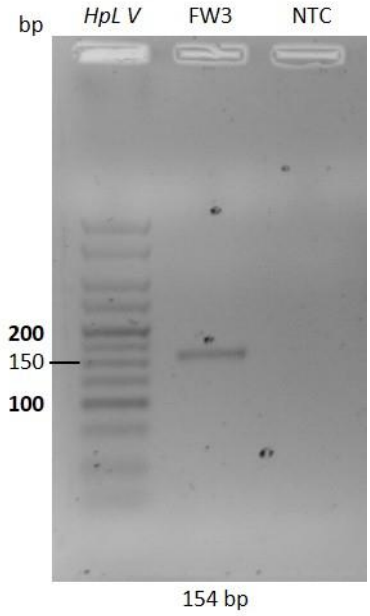


Figure 5.6. Amplification of fragment FW3.

2 % agarose gel stained with SYBR Safe dye shows results of PCR product for fragment FW3 with an expected size of 154 bp. HpLV: HyperLadder V. NTC: non-template control.

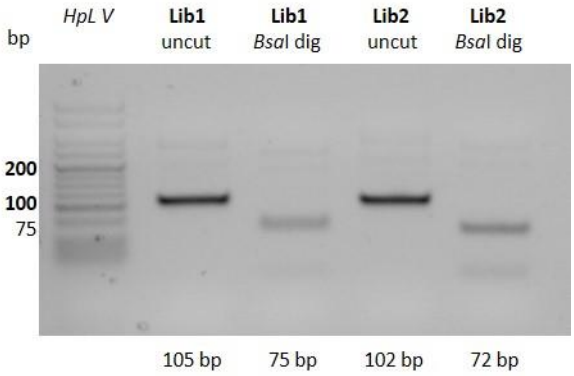


Figure 5.7. Enzyme digestion of fragments Lib1 and Lib2.

2 % agarose gel stained with SYBR Safe dye show results of enzyme digestion for Lib1 and Lib2. The digestion was performed with BsaI and the gel shows fragments before and after digestion with an indication of expected sizes below the gel. HpL V: HyperLadder II (Bioline).

5.5. The library assembly

As illustrated in Figure 5.3., the library assembly was performed in stages. Figure 5.8. shows the final product of the assembly for Lib_N461_AlaRS and Lib_C666C_AlaRS. The major product in the "ligation" lanes correspond to the final assemblies, which are expected to be 1389 bp for Lib_N461_AlaRS and 2631 bp for Lib_C666C_AlaRS. The bands below correspond to the partial ligation of the fragments used in that reaction or unused components. The "pre-lig" lanes show the input material (no ligase was added to these samples), while the last lanes, "ctrl", show the size control samples of FW4_N461 and FW4_875. All expected sizes are indicated below the figure and correspond with the sizes of the products obtained.

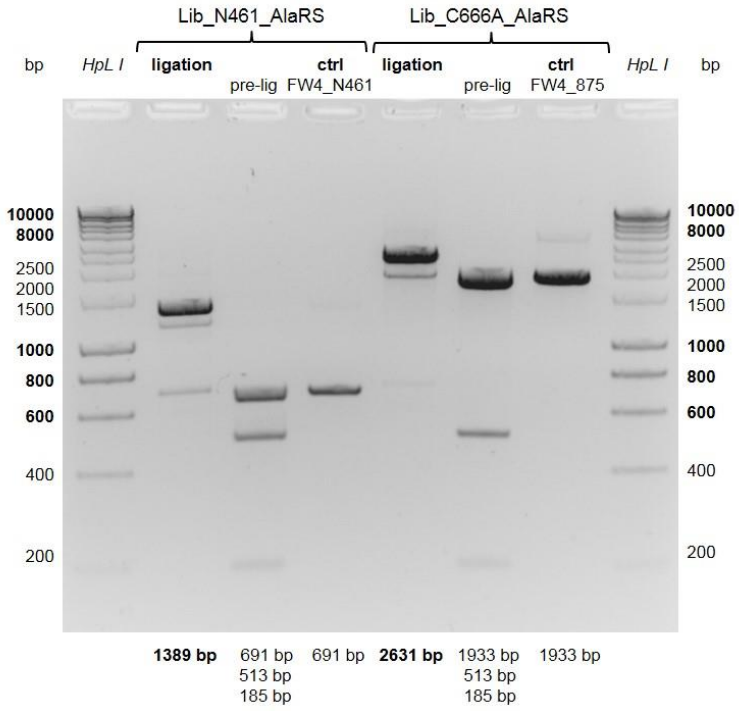


Figure 5.8. Final assembly of fragments for Lib_N461_AlaRS and Lib_C666A_AlaRS.

2 % agarose gel stained with SYBR Safe dye show results of the final library assembly for Lib_N461_AlaRS and Lib_C666A_AlaRS. The sizes of the fragments used in the assembly are given below the gel image. Above the gel, 'ligation' indicates the resulted product of fragments' ligation (full and partial ligation), whilst 'pre-lig' indicates the components of the ligation reactions. Ctrl: control fragment FW4. HpL I: HyperLadder I (Bioline).

To confirm that the major bands visible on the gel in the ligation lanes in Figure 5.8. were indeed the expected products, PCRs were performed using the forward primer used for the generation of FW1 and the reverse primer used for the generation of either FW4_N461 or FW4_C666A_875, with Phusion Polymerase in 100 µL total reaction volumes. As shown in Figure 5.9., products of expected sizes were generated in both libraries. The PCR reactions were then purified and sent for Sanger sequencing which confirmed the presence of randomised sites at the expected locations in the gene, appearing as mixed peaks in the chromatograms.

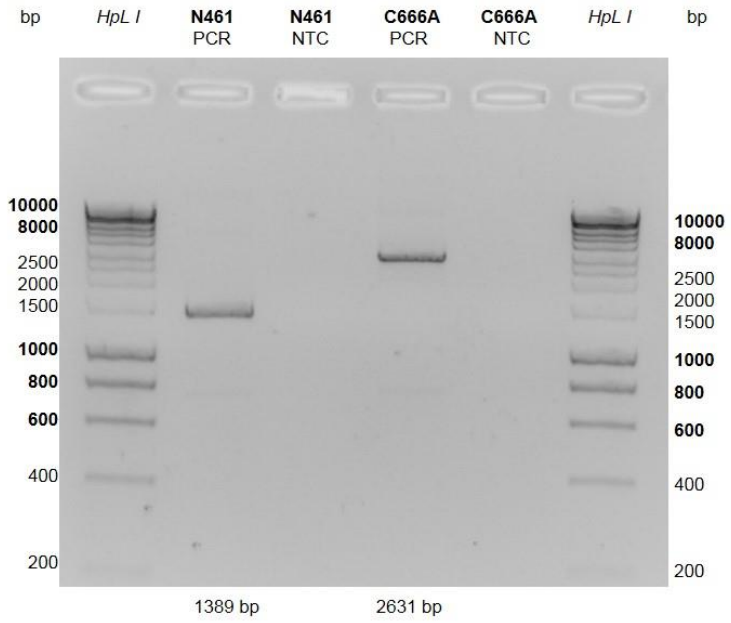


Figure 5.9. Amplification of the final assembly ligation product for Lib_N461_AlaRS and Lib_C666A_AlaRS.

1 % agarose gel stained with SYBR Safe dye show results of final PCR of ligations for libraries Lib_N461_AlaRS and Lib_C666A_AlaRS with an expected size of 1389 bp and 2631 bp, respectively. NTC: non-template control. HpL I: HyperLadder I (Bioline).

5.6. Next-generation sequencing

Residues chosen for randomisation were substituted with 18 amino acids, excluding cysteine and methionine. Amino acids were evenly randomised, therefore each amino acid substitute should account for 5.56 % of the total sequences (100 % /18 = 5.56 %).

In order to confirm even distribution at randomisation sites, library fragments were analysed by next-generation sequencing using the Illumina MiSeq Sequencing System (2.2.4.2.). The results presented in Figure 5.10. and Figure 5.11. are based on sequencing of the libraries of truncated structure enzyme (Lib_N461_AlaRS) and full-length enzyme (Lib_C666A_AlaRS), respectively. Since the fragments of the libraries used in the assembly came from the same sample (apart from diverse FW4s), it is expected that the sequencing results will demonstrate a nearly identical distribution of amino acids among the two libraries. As shown in Figure 5.10. and Figure 5.11, there are insignificant differences that might be down to wrong base recall during sequencing. Observed and expected distribution of amino acids in randomised positions for fragments Lib1, Lib3 and FW3 are illustrated in Figure 5.10. and Figure 5.11. Observed frequencies matched very closely the expected values for most of the amino acids at all six randomised positions. It was also found that certain amino acids such as tyrosine (Y) or glutamate (E) were under-represented across all substitutions, whereas other amino acids, such as leucine (L) or valine (V), demonstrated over-representation. The mechanism of these under- and over-representations is unknown but it could be related to T4 DNA Ligase specificity in blunt-end ligations (Ashraf *et al.*, 2013).

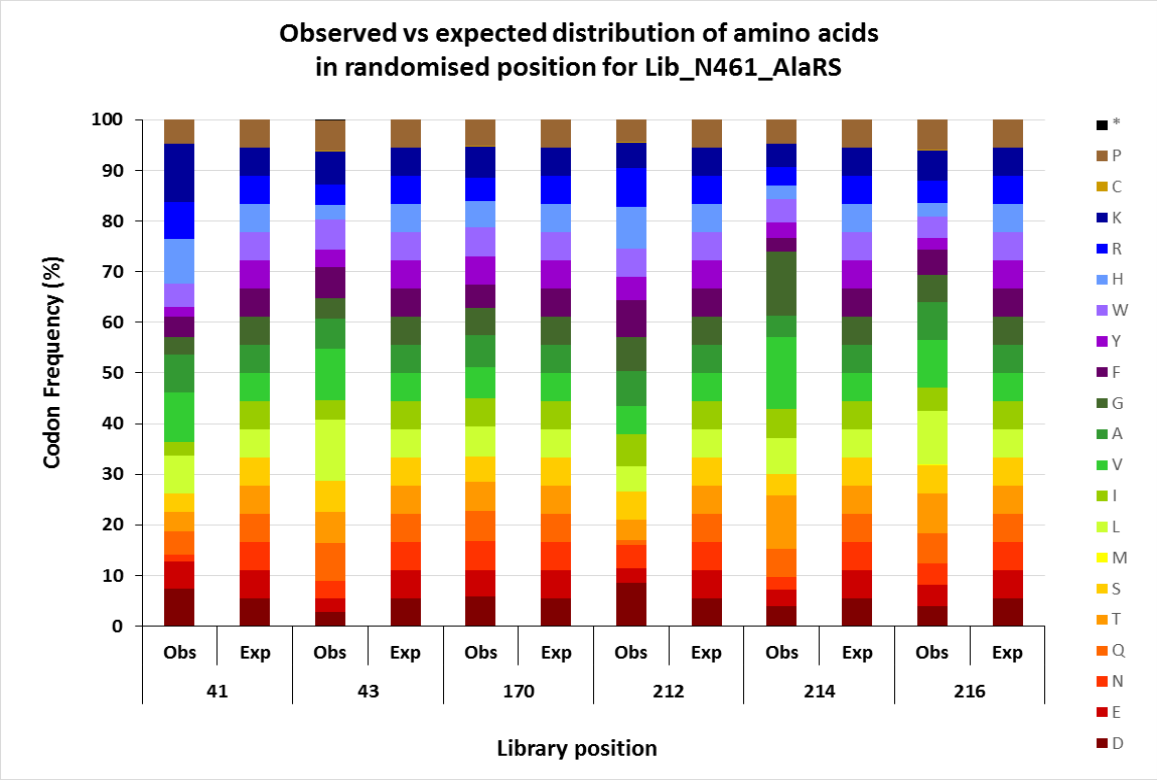


Figure 5.10. Observed vs expected distribution of amino acids in randomised positions of the truncated enzyme library (Lib_N461_AlaRS).

*Letters in the legend correspond to universal amino acids abbreviations in the genetic code: P: proline; C: cysteine; K: lysine; R: arginine; H: histidine; W: tryptophan; Y: tyrosine; F: phenylalanine; G: glycine; A: alanine; V: valine; I: isoleucine; L: leucine; M: methionine; S: serine; T: threonine; Q: glutamine; N: asparagine; E: glutamic acid; D: aspartic acid; *: stop codon.*

Data from MiSeq 185, 24th May 2016 Lib_N461_AlaRS.

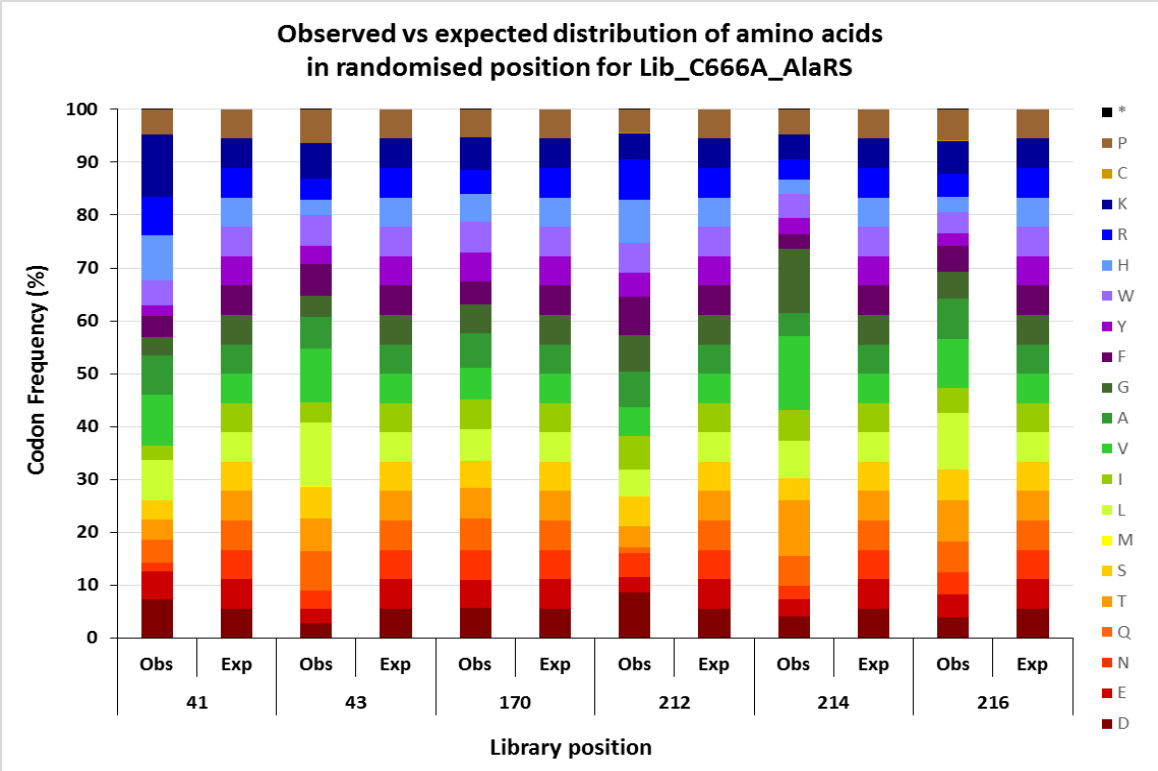


Figure 5.11. Observed vs expected distribution of amino acids in randomised positions of the full-length enzyme library (Lib_C666A_AlaRS).

*Letters in the legend correspond to universal amino acids abbreviations in the genetic code: P: proline; C: cysteine; K: lysine; R: arginine; H: histidine; W: tryptophan; Y: tyrosine; F: phenylalanine; G: glycine; A: alanine; V: valine; I: isoleucine; L: leucine; M: methionine; S: serine; T: threonine; Q: glutamine; N: asparagine; E: glutamic acid; D: aspartic acid; *: stop codon.*

Data from MiSeq 185, 24th May 2016 Lib_C666A_AlaRS.

5.7. Discussion

ProxiMAX is marketed as Colibra™ and it is an automated and proprietary version commercialised at Isogenica Ltd. Although the name ProxiMAX was used throughout, the libraries of AlaRS were actually produced using the Colibra™ platform. Due to the license agreement, the full methodology of generating particular fragments of the library could not be disclosed and comprehensive data from library production could not be displayed here.

There were two randomisation methods used in the construction of libraries: ProxiMAX and via primers in PCR. When comparing the frequency of observed vs expected amino acids using these methods, the randomisation introduced with primers shows a closer match to expected values than with ProxiMAX. It had proved that careful combining and mixing of primers can be used for single randomisation because it shows an excellent result in terms of amino acids distribution (Figure 5.10. and Figure 5.11.). However, for the purpose of randomisation of contiguous codons or even three near-contiguous such as positions 212, 214 and 216 in this study, using ProxiMAX codon saturation was required. ProxiMAX randomisation offers controlled codon saturation of multiple contiguous codons with the choice on amino acid ratio in each of mutated position. Moreover, ProxiMAX uses one codon per amino acid what gives a substantial advantage in terms of control of the library content. It also eliminates bias in the produced library, with each gene variant being unique what eliminates over-representation and under-representation of an amino acid in the mutated position. This decreases the total size of the library generated what increases the chances for a suitable screening method.

The results herein prove that ProxiMAX randomisation was a suitable choice for library production. The results of even representation of codons in the randomised positions correlate well with other studies using the same method ensuring maximal diversity of the variants (Frigotto *et al.*, 2015; Poole, 2015). Regardless of how close the match of observed to expected proportion of codon per position was, some amino acids were under-represented or over-represented. This could have been

the result of not equimolar mix of ProxiMAX oligonucleotides which was later strengthened and reinforced in amplifications. Another explanation is that T4 DNA ligase used during library generation has sequence specificity during blunt end ligation of two DNA sequences. The consequence of these deviations to even representation could be the introduction of bias into the library and resulted in less unique variants unable to fill maximal coverage.

Chapter 6: Initial development of a screening system for mutant AlaRS genes

After library generation, library screening and selection are the next two crucial steps in protein engineering. They respectively allow for evaluation of the library and elimination of nonfunctional variants (Xiao *et al.*, 2015). Thus the aim of screening and selection is to find a new, optimal, engineered protein from within a pool of variants as a successful hit in the well-designed assay. Library screening and selection can be challenging and often requires optimisation or even a change of methodology. In the end, a successful library screening process will work only if it has carefully-selected controls, if it is to lead to the discovery of a novel protein with required activity.

The library generated in Chapter 5 has over 34 million variants and as such, in terms of numbers, is not a challenging library to work with. It can be analysed *in vivo* since transformation efficiency is able to cover the numbers of clones. The size of the library does however impose a high-throughput method for screening phenotypes. The library contains variants for *E. coli* alanyl-tRNA synthetase for engineering a change in substrate specificity and as being an enzyme library, the screening has to be based on the phenotypic properties (Ferreira Amaral *et al.*, 2017).

AlaRS variant libraries have not previously been created and since this project will ultimately require mutation of the tRNA anticodon as well as the AlaRS charging enzyme, it was necessary to develop a new assay. As an adaptation of the assay described by Kuhn, Rubini, Fuhrmann, Theobald, and Skerra (2010) employed for selecting amber suppression mutants, Fluorescence-Activated Cell Sorting (FACS) was selected as a screening method. The AlaRS activity was coupled with expression levels of eGFP which serves as a reporter. Since eGFP has a weak tendency for dimerisation (Day and Davidson, 2009), a point mutation A206K was introduced to ensure that the protein remains in the monomer form (Nagai *et al.*, 2002) (refer to Section 1.7.2.1.). As described previously (Kuhn *et*

al., 2010), to select for the functional synthetases, the assay was designed to contain both negative and positive selections.

6.1. Screening design

The aim of the screening was to select for functional *E. coli* AlaRS synthetases that are able to aminoacylate *E. coli* $tRNA^{Ala}$ with an UAA, for ultimate use in an *in vitro* translation system. Consequently, since both the enzyme and tRNA originate from *E. coli*, the library of enzymes could not be screened in *E. coli* because of cross-reaction between synthetic and native bacterial tRNAs. Similarly, *E. coli* tRNA that was used in the assay would interact with bacterial synthetases. To overcome this challenge, the assay had to be developed in a different host to exclude interaction between native and orthogonal enzymes and tRNAs. The organism *S. cerevisiae* has been employed in previous amber suppression studies to develop orthogonal yeast tRNAs and was therefore selected for use in the current study.

For the purpose of screening assay development, a novel amber suppression system employing the *E. coli* $tRNA^{Ala}$ gene was chosen, so that subsequent mutation of the $tRNA^{Ala}$ anticodon to multiple alternatives would be possible.

The core components of this *in vivo* fluorescence screening assay are:

1) an *E. coli* $tRNA^{Ala}$ suppressor gene in which the anticodon is mutated to bind the amber (5'-TAG-3') termination codon,

2) an eGFP gene ($eGFP_{amber39}$) containing the amber codon introduced at the permissive position 39 (this position can be mutated without affecting the fluorescence of eGFP (Kuhn *et al.*, 2010),

3) an AlaRS gene that is capable of aminoacylating the mutant $tRNA^{Ala}$ with an amino acid (either native or unnatural).

Ultimately, as described previously by Kuhn *et al.* (2010), the screening process should be divided into two parts: a positive and negative selection. In the current study, the positive selection would involve co-expression of the AlaRS gene (or gene library), the suppressor $tRNA^{Ala}$ and the mutated $eGFP_{amber39}$ in the presence of both natural AAs and the required UAA, followed by FACS selection. This step selects for the functional synthetases and so reduces library size by "weeding out" nonfunctional AlaRS enzymes. The population of cells with functional synthetases is then a subject for negative selection to further differentiate enzyme variants. The negative selection process repeats the co-expression, but in the absence of the required UAA. Providing that these cells express only functional enzymes, mutants that can incorporate only the required UAA can no longer read through the amber termination codon and as such, these mutants can no longer fluoresce. Thus, following FACS selection, the pool with cells expressing fluorescent eGFP contains variants that aminoacylate natural AAs, while the pool that expresses nonfluorescent truncated eGFP has variants of an enzyme that aminoacylate only with the required UAA. The details of the screening process are illustrated in Figure 6.1.

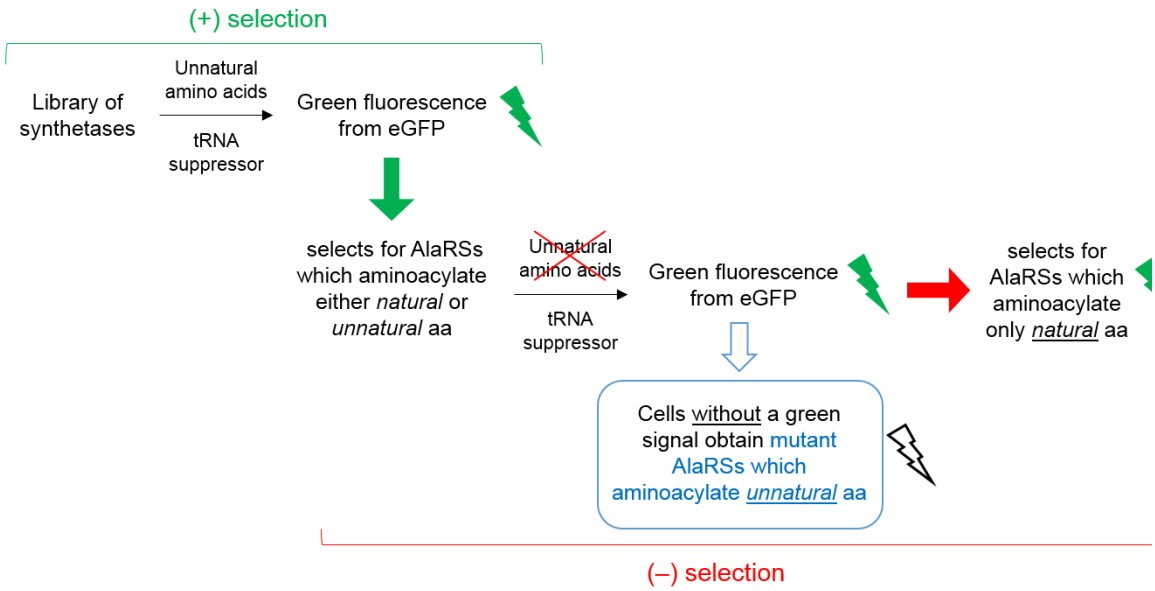


Figure 6.1. Positive and negative selection process of library screening.

The positive (+) selection (in green) selects for synthetases that recognise both AAs and UAAs and yields a pool of functional synthetases. Negative (–) selection (in red) selects for synthetases that recognise only AAs and yields a pool with nonfluorescent cells with a mutant enzyme that aminoacylates only UAAss (in blue).

Co-expression of three separate recombinant genes is challenging and previous studies have involved both, one and two-plasmid systems. Firstly, herein, a design involving a one-plasmid system using pYES2 with uracil selection was attempted (Figure 6.2.). pYES2 is a shuttle vector that can be transformed into *E. coli* or *S. cerevisiae* which is useful to complete all molecular procedures.

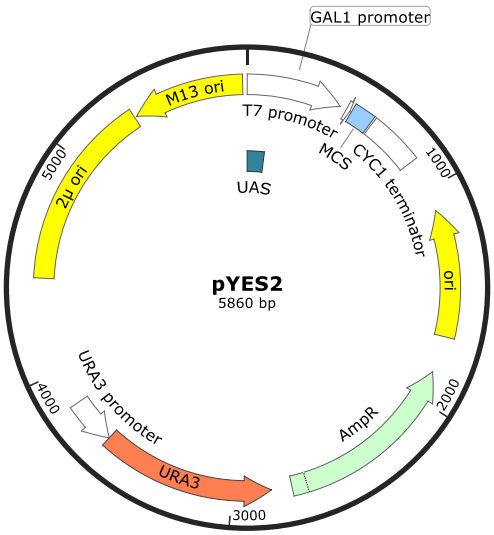


Figure 6.2. Plasmid map of pYES2, uracil yeast selection plasmid.

Among other features, the plasmid includes: URA3 gene for selection of yeast transformants in uracil-deficient medium, Gal1 promoter for inducible expression of cloned genes, pUC origin and 2µ origin for high copy replication in E. coli and yeast organisms, respectively. pYES2 also contains multiple cloning site (MCS) with a range of unique restriction sites to aid gene cloning. All maps were prepared using SnapGene software. (Biotech, SnapGene software)

6.2. The controls for the library selection assay

To evaluate the concept of the assay in terms of functionality, it would be essential that C666A_AlaRS and/or N461_AlaRS could aminoacylate the $tRNA^{Ala}$ suppressor with Ala and that *S. cerevisiae* could then use the resulting orthogonal, charged tRNA to read through the amber termination codon of $eGFP_{amber39}$. Failure to achieve any of these steps would lead to production of only truncated (nonfluorescent) eGFP and would thus render FACS incapable of separating

functional from nonfunctional synthetases. Accordingly, several controls were designed, as detailed in Table 6.1. There were also other checkpoints (Table 6.2.) that were purposely designed to confirm that the expressed components are truly orthogonal and do not cross-react with yeast's translation machinery, a scenario outlined above (Section 6.1.). Hence, the work for the library screening began with the preparation of all constructs.

Table 6.1. Negative and positive controls for library screening assay.

The details of all constructs needed as a control for the screening with details of what genes are found in vector pYES2.

construct	**$tRNA^{Ala}$ suppressor**	**eGFP**	**eGFP with a STOP codon**	**wild type N461_ AlaRS**	**wild type C666A_ AlaRS**	**colour**
1	$tRNA^{Ala}$ suppressor	eGFP	-	-	-	green
2	-	eGFP	-	-	-	green
3	$tRNA^{Ala}$ suppressor	-	$eGFP_{amber39}$	-	-	no fluorescence
4a	$tRNA^{Ala}$ suppressor	-	$eGFP_{amber39}$	N461_AlaRS	-	green
4b	$tRNA^{Ala}$ suppressor	-	$eGFP_{amber39}$	-	C666A_AlaRS	green

Table 6.2. The constructs details and the purpose of their design.

Construct / components	**Check point**	**Expected colour**
eGFP	Is eGFP expressed correctly under short Gal promoter?	green
$tRNA_{sup}$, enzyme, $eGFP_{amber39}$	Does the screening assay work?	green
$tRNA_{sup}$, $eGFP_{amber39}$	Does yeast aminoacylate the suppressor $tRNA^{Ala}$?	no fluorescence
$eGFP_{amber39}$	Does yeast read through the stop codon?	no fluorescence
$tRNA_{sup}$, enzyme	Do the cells have any native/auto fluorescence?	no fluorescence
Empty plasmid pYES2	Do the cells have any native/auto fluorescence?	no fluorescence

6.2.1. Construction of control plasmids

6.2.1.1. Construct 1 – tRNA^{Ala} suppressor and eGFP

Synthetic genes encoding the tRNA^{Ala} suppressor under the control of the Gal1 promoter and eGFP, codon-optimized for expression in *S. cerevisiae* (Kaishima *et al.*, 2016) and under control of the short Gal promoter (Redden and Alper, 2015) were obtained from Eurofins Genomics (Ebersberg, Germany), pre-cloned into the pYES2 vector (Figure 6.2.) to form construct 1 shown in Figure 6.3.

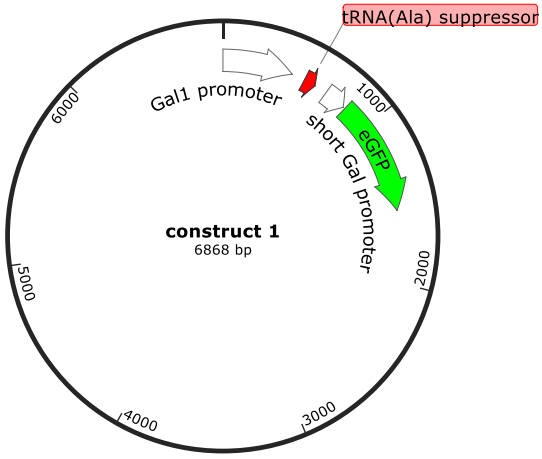


Figure 6.3. The plasmid map of construct 1 with tRNA^{Ala} suppressor and eGFP gene.

6.2.1.2. Construct 2 – eGFP only

This control aimed to check both that the synthetic eGFP gene was functional and that it was expressed as expected under control of the short Gal promoter. Accordingly, the tRNA^{Ala} suppressor gene and proceeding Gal1 promoter were removed from construct 1. Firstly, primers were designed to cut out tRNA^{Ala} suppressor and Gal1 promoter by amplifying around the plasmid, leaving eGFP and associated components in the amplified region. To determine optimal annealing temperature, a gradient PCR was performed (2.2.1.1.1). As seen in Figure 6.5. a. the expected product of size 6195 bp can be found in most samples. However, reaction 4 with the annealing temperature of 47.5 °C has

only one single band at the expected size and therefore, a temperature 48 °C was used to amplify the product in a new, scaled-up PCR (2.2.1.1.; Figure 6.5. b). Following amplification, the PCR reactions were pooled to be purified (2.2.1.8.1.), quantified (2.2.1.9.), phosphorylated (2.2.1.3.) and self-ligated (2.2.1.5.). The ligations were transformed into *E. coli* (2.2.3.1.) and transformants plated on LB Amp plates for incubation overnight at 37 °C. The following day, three colonies were selected and grown in LB (Amp) overnight. Plasmids were then extracted using QIAGEN® QIAprep Spin Miniprep Kit (2.2.1.8.2.) and upon purification, 3 clones were sent for sequencing to Eurofins Genomics. The results received later showed correct sequences for all three clones. This workflow is summarized in Figure 6.4.

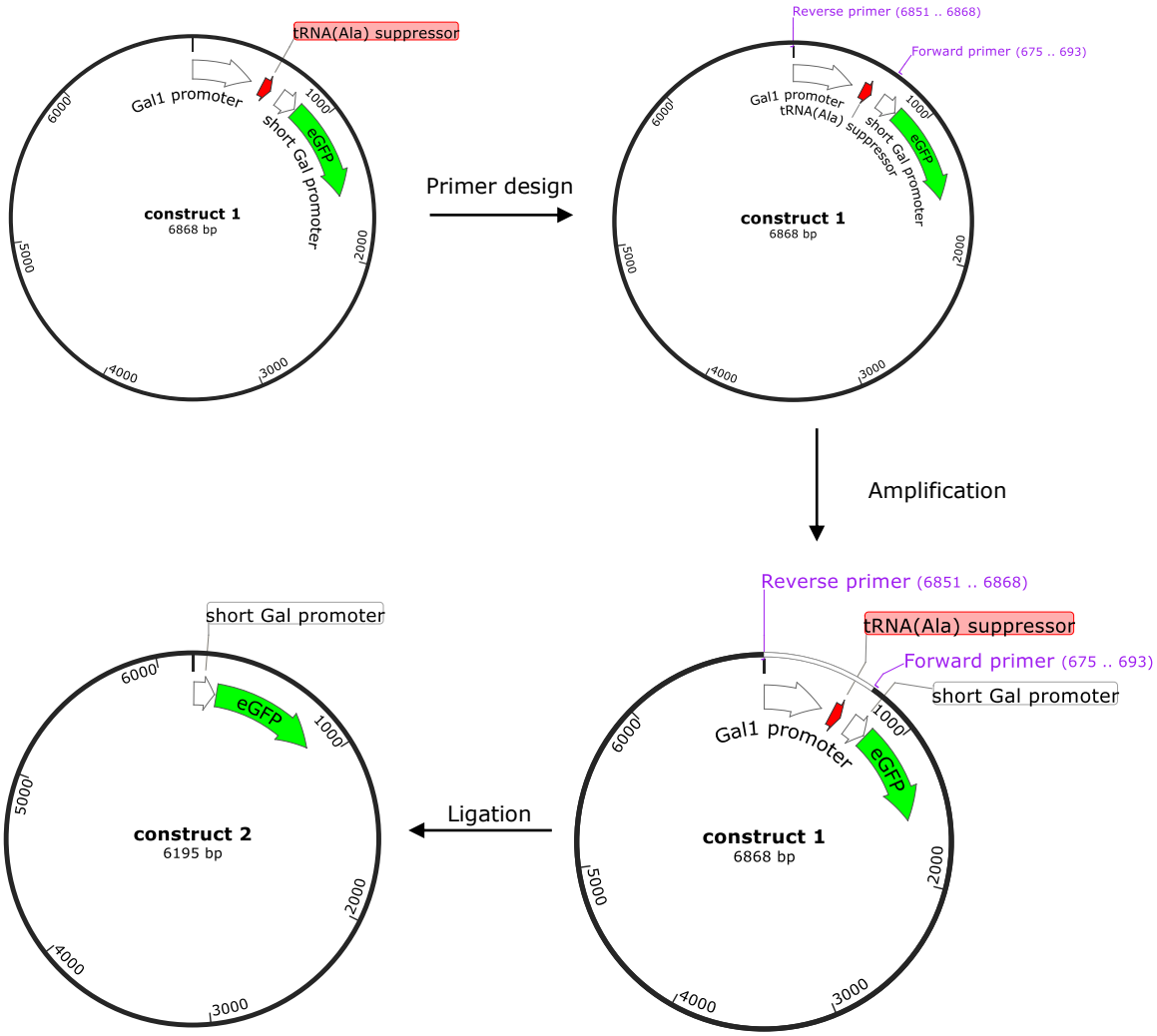
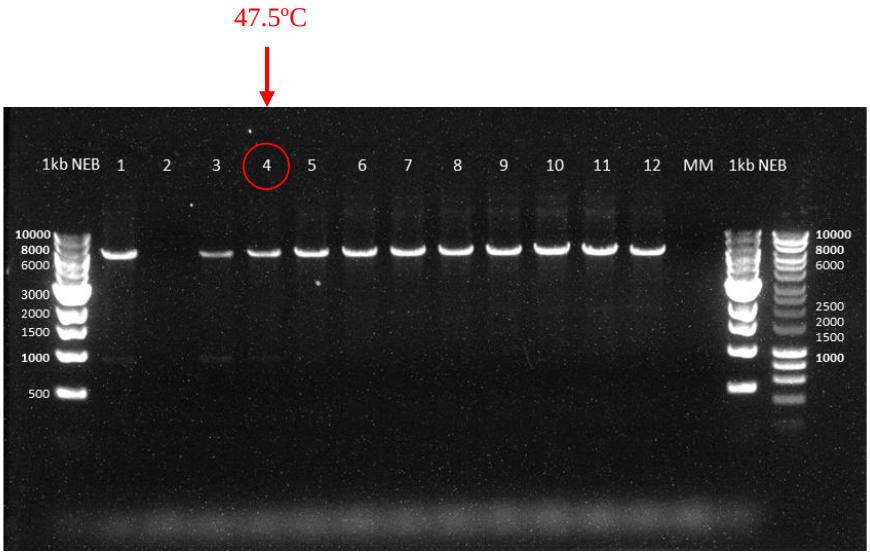


Figure 6.4. The workflow for generation construct 2 from construct 1.

a.



b.

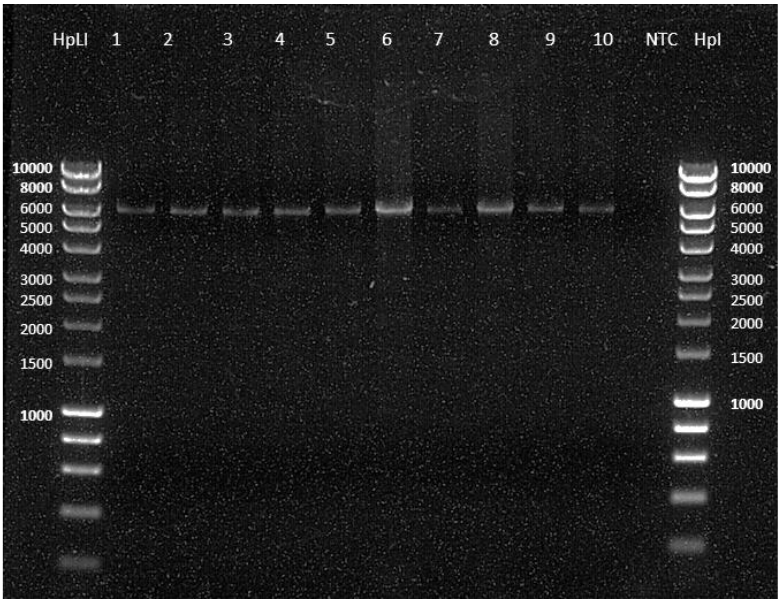


Figure 6.5. Amplification of plasmid backbone for construct 2.

0.8 % agarose gels stained with SYBR Safe dye show results of PCRs with an expected size of 6195 bp. a. Gradient PCR amplifying plasmid backbone for construct 2. Lanes number 1 to 12 represent reactions for temperature gradient 45-60 °C amplified with Q5® High-Fidelity DNA Polymerase (NEB). 1 kb DNA ladder (NEB) is present on both sides of the gel. The marker on the far right side is HyperLadder 1kb (Bioline). MM: a non-template control. b. Scaled-up PCR with amplification at 48 °C representing lanes 1 to 10 as ten samples of 100 µL reactions. HpLI: HyperLadder 1kb. NTC: a non-template control. Construct 2: eGFP only (refer to Figure 6.4.).

6.2.1.3. Construct 3 – $tRNA^{Ala}$ suppressor and $eGFP_{amber39}$

This control was constructed as a preparation for subsequent analysis and to ensure that yeast cannot aminoacylate the $tRNA^{Ala}$ suppressor and so suppress $eGFP_{amber39}$ in the absence of an orthogonal AlaRS gene. To prepare construct 3, a mutation with a TAG stop codon was introduced to the eGFP gene at codon 39 of construct 1. The mutation was introduced with PCR amplification (2.2.1.7.2.) by employing a reverse primer containing the appropriate mutation with a workflow as described in Figure 6.6. The results of the gradient PCR and scaled-up PCR are found in Figure 6.7. Three clones were sent for sequencing to Eurofins Genomics and the results confirmed a mutation of eGFP to all of them.

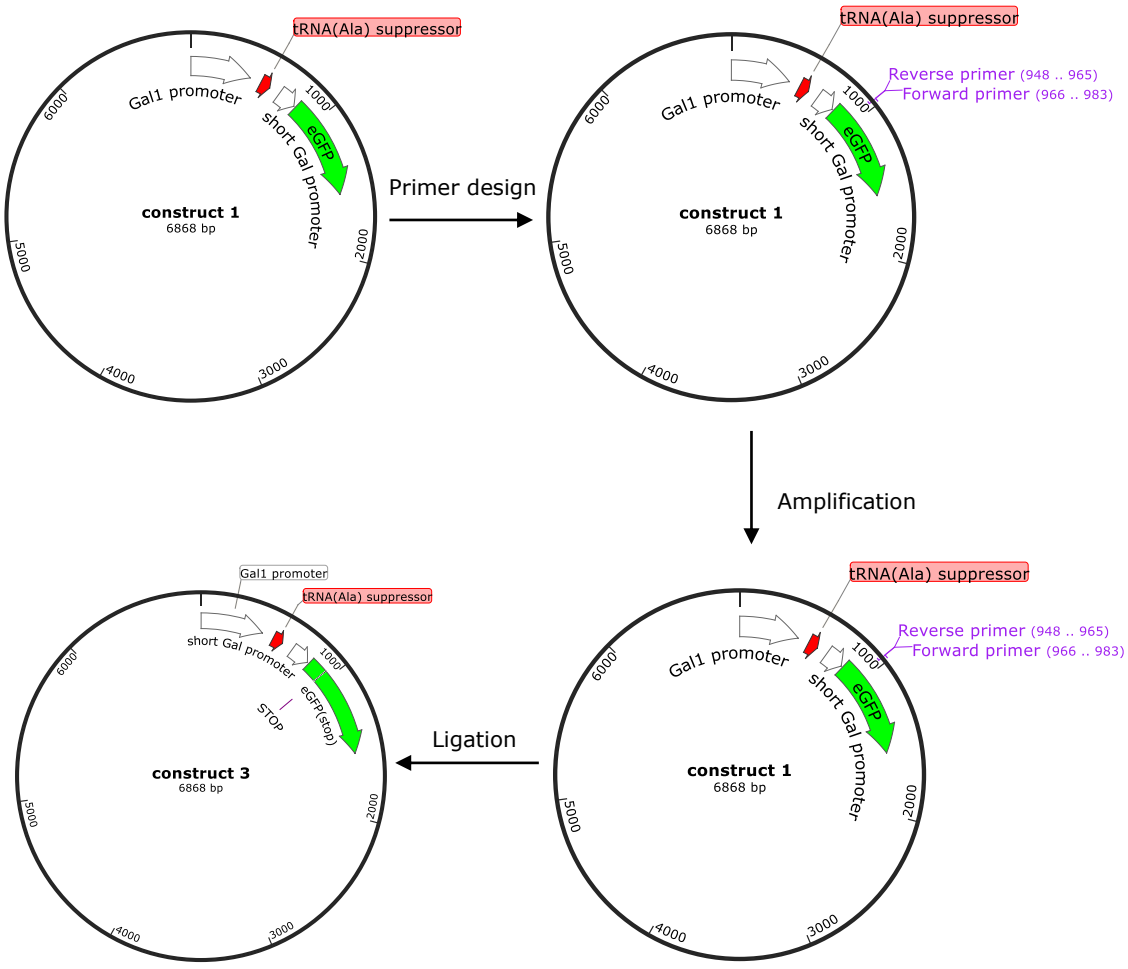
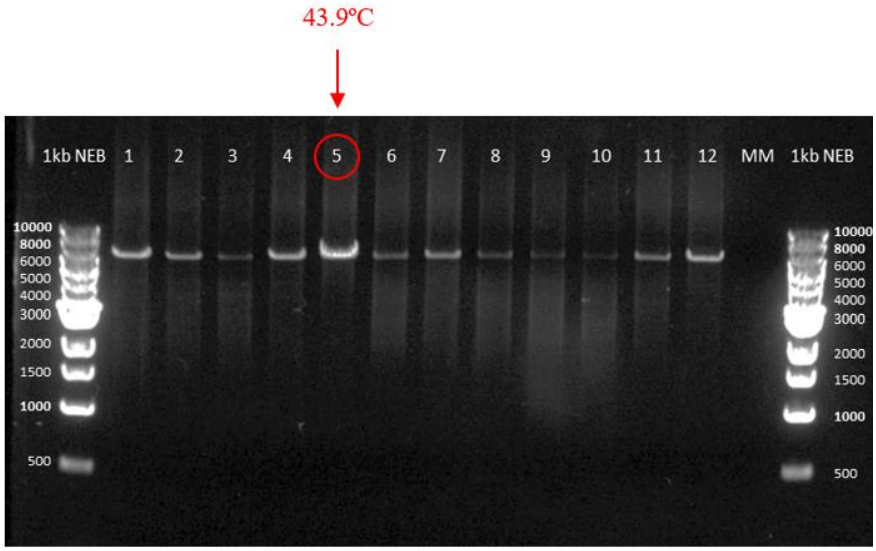


Figure 6.6. The workflow for generation construct 3 from construct 1.

a.



b.

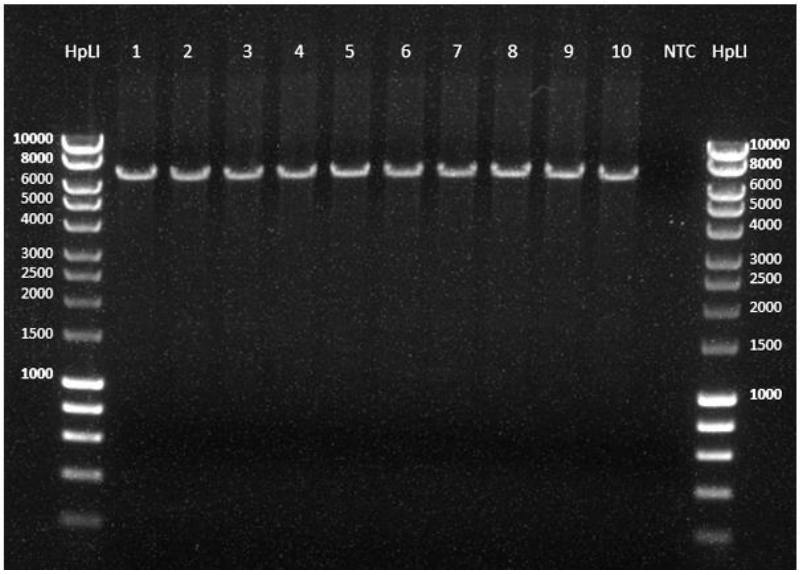


Figure 6.7. Amplification of plasmid backbone for construct 3.

0.8 % agarose gels stained with SYBR Safe dye show results of PCRs with an expected size of 6868 bp. a. Gradient PCR amplifying plasmid backbone for construct 3. Lanes number 1 to 12 represent reactions for temperature gradient 40-54 °C amplified with Q5® High-Fidelity DNA Polymerase (NEB). 1 kb DNA ladder (NEB) is present on both sides of the gel. The marker on the far right side is HyperLadder 1kb (Bioline). MM: a non-template control. b. Scaled-up PCR with amplification at 44 °C representing lanes 1 to 10 as ten samples of 100 µL reactions. HpLI: HyperLadder 1kb. NTC: a non-template control. Construct 3: $eGFP_{amber39}$, $tRNA^{Ala}$ suppressor (refer to Figure 6.6.).

6.2.1.4. Construct 4a and 4b – $tRNA^{Ala}$ suppressor, $eGFP_{amber39}$ and either truncated (construct 4a) or full-length (construct 4b) AlaRS

Constructs 4a and 4b were created to perform the aminoacylation assay with the truncated and full-length *E. coli* AlaRS genes respectively and as such, serve as aminoacylation controls for each form of the enzyme. To prepare constructs 4a and 4b, genes of a truncated and full-length enzyme (Sections 3.2. and 3.3.) were added to construct 3, to be expressed as an operon under the control of the short Gal promoter which is located upstream of the $eGFP_{amber39}$ gene. Firstly, restriction sites had to be introduced at the end of the genes to be able to ligate them with the vector backbone. As illustrated in Figure 6.8., a forward primer introduced both a *Not*I restriction site and Kozak sequence while the reverse primer added an *Sph*I restriction site to both genes. Also, each primer had seven additional nucleotides at 5'-end to enable restriction digestion of PCR amplicons. The primers were used to amplify native genes from constructs in pET45b(+) vector (Sections 3.2. and 3.3.)

A gradient PCR was set up for the truncated and full-length genes (N461_AlaRS and C666A_AlaRS) with results in Figure 6.9. a. and Figure 6.10. a., followed by scaled-up PCR (Figure 6.9. b. and Figure 6.10. b.) and purification of PCR products (2.2.1.8.1.). After purification, double digestion with *Sph*I and *Not*I was set up for construct 3 (vector backbone) and the gene inserts, N461_AlaRS and C666A_AlaRS (2.2.1.6.). The reactions were incubated overnight at 37 °C. Additionally, construct 3 was dephosphorylated (2.2.1.4.). Subsequently, all digests were electrophoresed in low-melt agarose and the fragments of the correct size were gel extracted (2.2.1.8.3.). Once the purified fragments were quantified, they were ligated (2.2.1.5.) and transformed into *E. coli* (2.2.6.1.). A number of colonies of construct 4a were screened in colony PCR with an example illustrated in Figure 6.11. where clones 3, 6, 7, 8, 10, 12, 13 had the correct size of PCR product. Chosen clones were then further assessed by digestion with an enzyme of a unique site to linearise the plasmid and analyse its total size in gel electrophoresis. After successfully passing both colony screening methods, several clones were sent for sequencing. For construct 4b, clones 2, 3, 4, 8, 11 showed a correct PCR product size in colony PCR (Figure 6.12. a.). These plasmids were later digested with

*Sph*I to linearise and since all showed a correct size of 9510 bp were sent for sequencing. The results of sequencing for construct 4a and 4b confirmed correct sequencing in clones 10 and 15 for construct 4a and clones 3 and 4 for construct 4b.

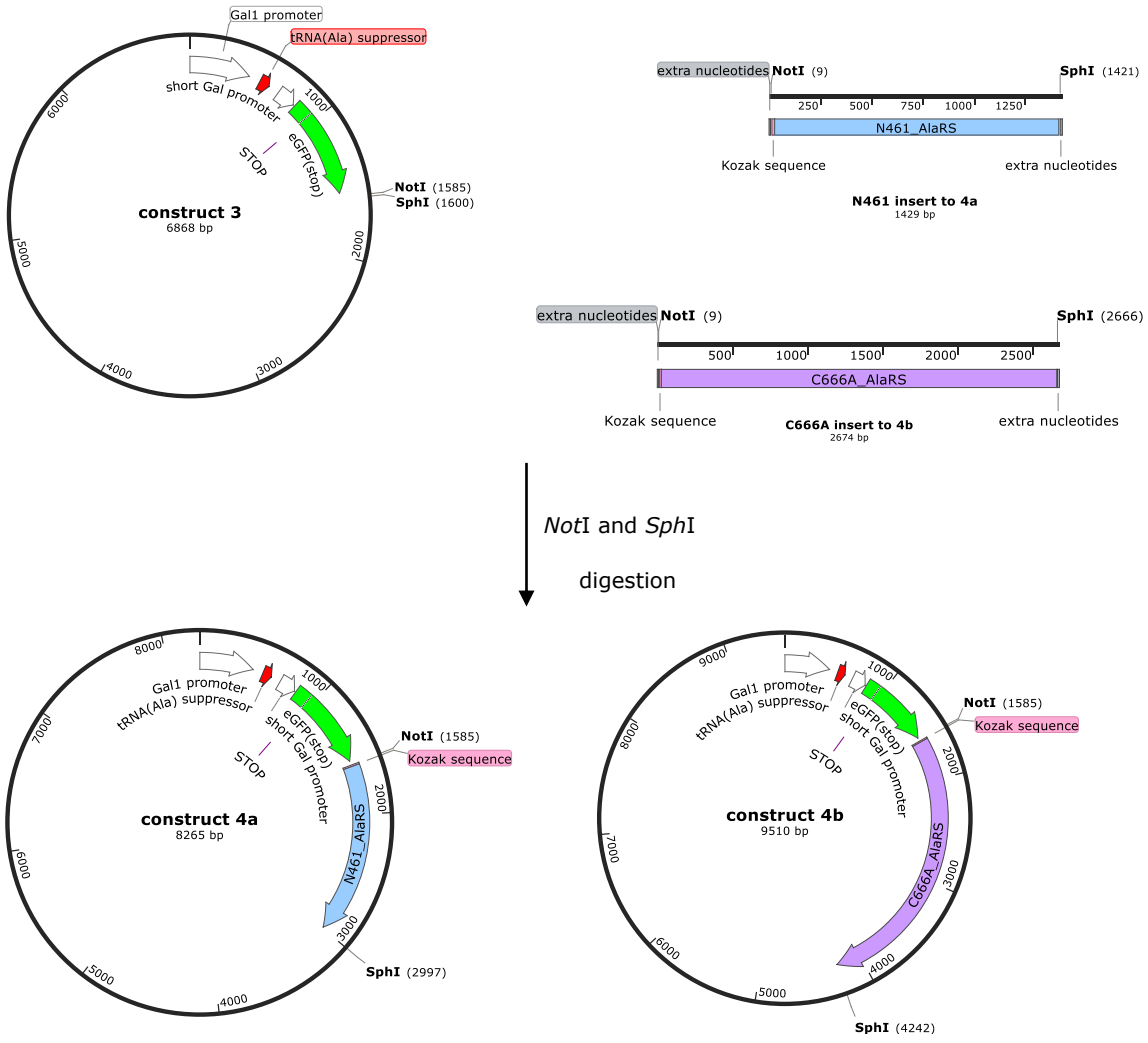
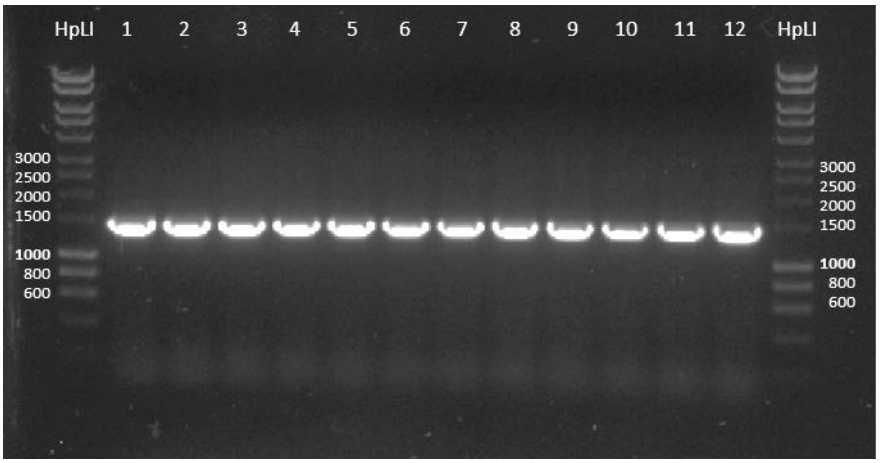


Figure 6.8. The workflow for generation construct 4a and construct 4b.

a.



b.

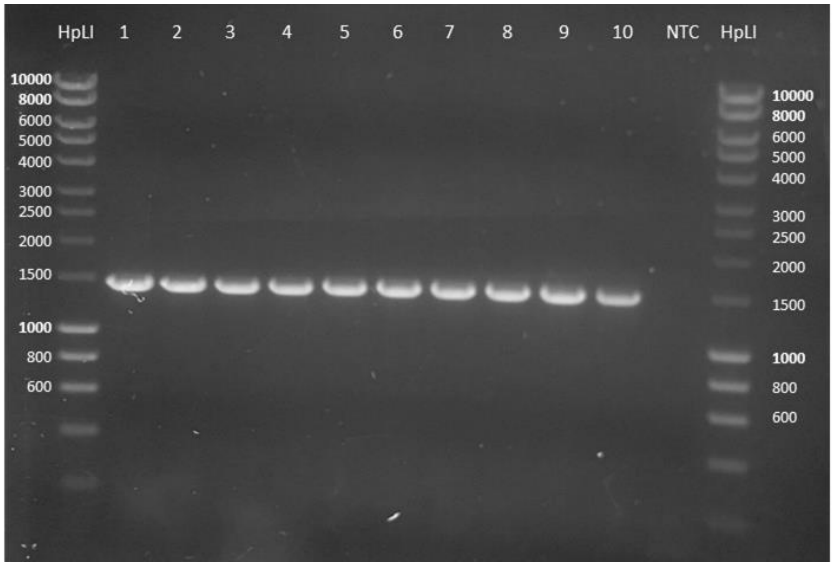
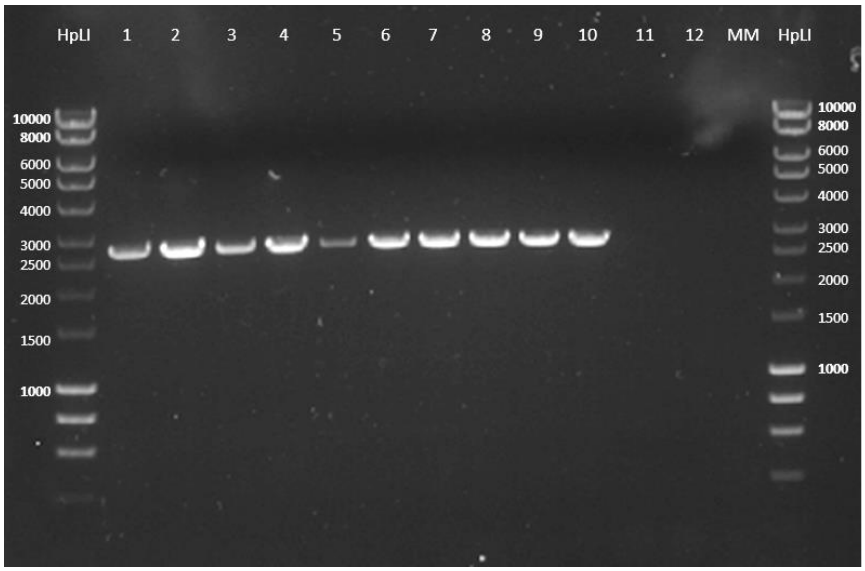


Figure 6.9. Amplification of gene insert N461_AlaRS for construct 4a.

1 % agarose gels stained with SYBR Safe dye show results of PCR with an expected size of 1429 bp. a. Gradient PCR amplifying insert gene for a truncated enzyme for construct 4a. Lanes number 1 to 12 represent reactions for temperature gradient 45-65 °C amplified with Q5® High-Fidelity DNA Polymerase (NEB). b. Scaled-up PCR with amplification at 50 °C representing lanes 1 to 10 as ten samples of 100 µL reactions. HpLI: HyperLadder 1kb. NTC: a non-template control. Construct 4a: $eGFP_{amber39}$, $tRNA^{Ala}$ suppressor, N461_AlaRS (refer to Figure 6.8.).

a.



b.

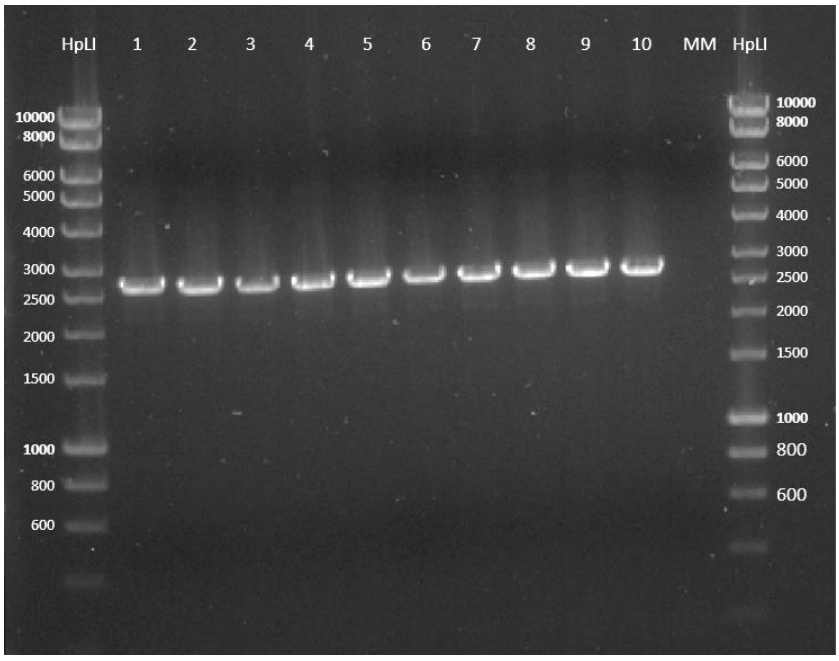
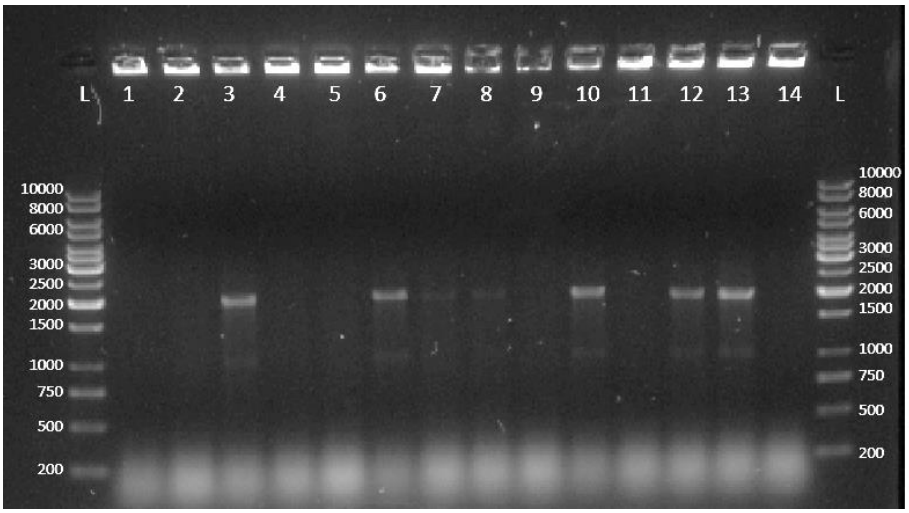


Figure 6.10. Amplification of gene insert C666A_AlaRS for construct 4b.

1 % agarose gels stained with SYBR Safe dye show results of PCR with an expected size of 2674 bp. a. Gradient PCR amplifying gene for truncated enzyme as an insert for construct 4b. Lanes number 1 to 12 represent reactions for temperature gradient 45-65 °C amplified with Q5® High-Fidelity DNA Polymerase (NEB). b. Scaled-up PCR with amplification at 57 °C representing lanes 1 to 10 as ten samples of 100 µL reactions. HpLI: HyperLadder 1kb. MM: a non-template control. Construct 4b: eGFP$_{amber39}$, tRNA^{Ala} suppressor, C666A_AlaRS (refer to Figure 6.8.).

a.



b.

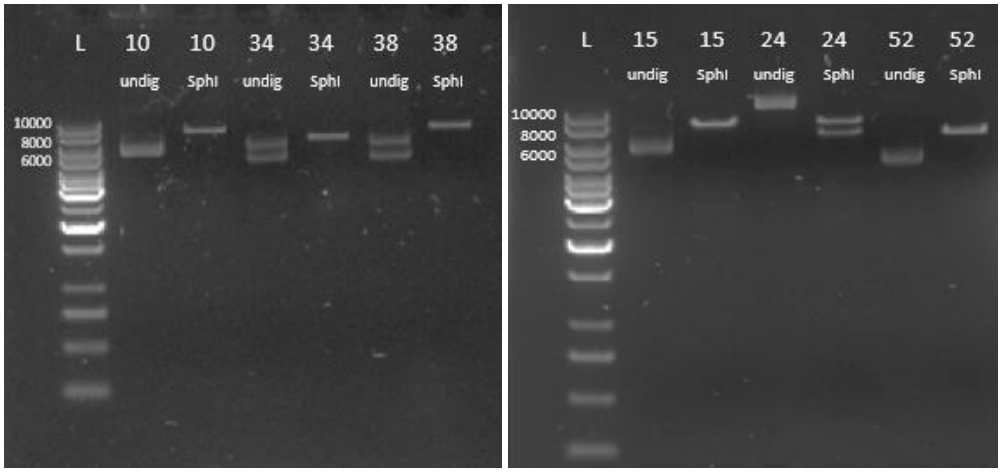
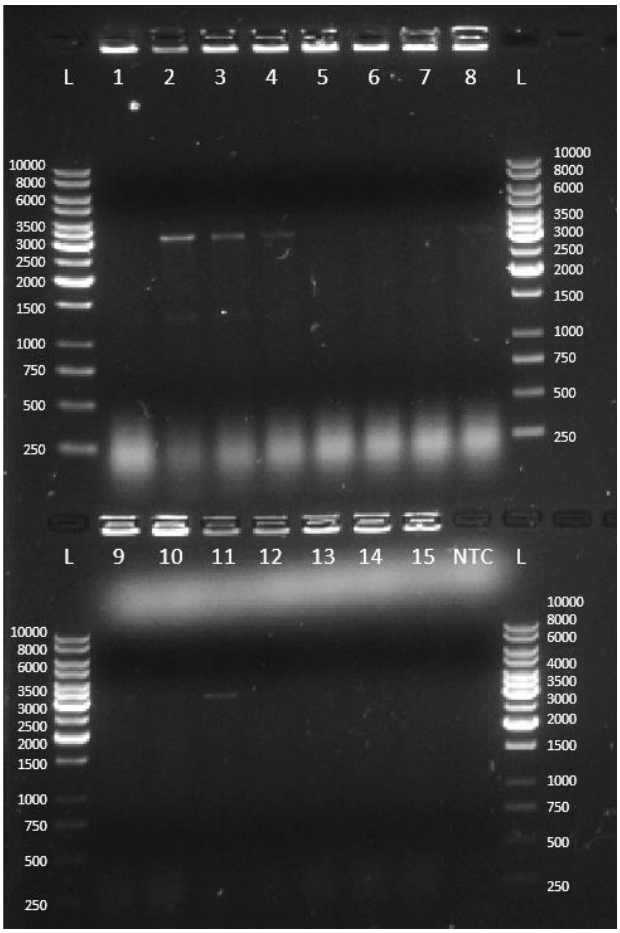


Figure 6.11. Colony screening for construct 4a.

0.8 % agarose gels stained with SYBR Safe dye show results of colony screening a. by colony PCR and b. enzyme digestion. a. An example of colony screening by colony PCR for clones 1 to 14 with an expected size of 2040 bp. b. Diagnostic enzyme digestion for colonies with correct size amplicons in colony PCR. SphI is the unique restriction site in the plasmid and the expected size of the linearised plasmid is 8265 bp. Colony screened are 10, 34, 38, 15, 24, 52. Undig: undigested plasmids. SphI: plasmid linearised with SphI. L: GeneRuler 1kb DNA ladder (Thermo Fisher). Construct 4a: $eGFP_{amber39}$, $tRNA^{Ala}$ suppressor, N461_AlaRS (refer to Figure 6.8.).

a.



b.

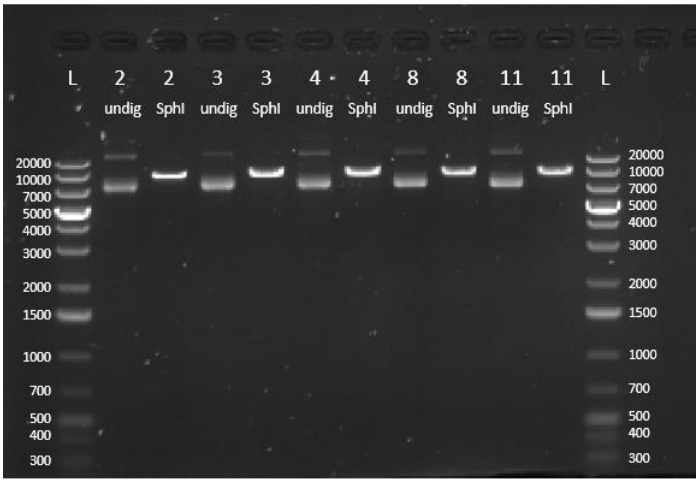


Figure 6.12. Colony screening for construct 4b.

0.8 % agarose gels stained with SYBR Safe dye show results of colony screening a. by colony PCR and b. enzyme digestion. a. An example of colony screening by colony PCR for clones 1 to 15 with an expected size of 3265 bp. The correct size was found for clones 2, 3, 4, 8, 11. NTC – non-template control. L – GeneRuler 1kb DNA ladder (Thermo Fisher). b. Diagnostic enzyme digestion for colonies with correct size amplicons in colony PCR. SphI is the unique restriction site in the plasmid and the expected size of the linearised plasmid is 9510 bp. Undig: undigested plasmids. SphI: plasmid linearised with SphI. L: GeneRuler 1kb plus DNA ladder (Thermo Fisher). Construct 4b: $eGFP_{amber39}$, $tRNA^{Ala}$ suppressor, C666A_AlaRS (refer to Figure 6.8.).

6.3. The functionality of the selection assay

Having confirmed the DNA sequences in each of the constructs, the control constructs were then assessed by both inverted fluorescence microscopy and then FACS.

6.3.1. Analysis of expression under fluorescence microscope

All constructs were transformed into *S. cerevisiae* strain BY4741 (2.2.6.2.). Individual colonies were selected and grown and induced in selection media containing galactose, overnight at 30˚C. Slides were prepared and the cells examined by fluorescence microscopy (2.2.9.; Figure 6.13.).

As illustrated in Figure 6.13., the native gene of eGFP is effectively expressed under the short Gal Promoter in both constructs 1 and 2. In both pictures, it is clear that expression under Short Gal Promoter is effective in SD-U media with supplementation of 2% (w/v) Galactose. The eGFP is bright and clearly visible in yeast cells. Similarly, construct 3 ($tRNA^{Ala}$ suppressor and $eGFP_{amber39}$) shows no fluorescence, as expected. However, constructs 4a and 4b also showed little/no fluorescence. The green colour indicating expression of eGFP is limited and found in only few cells, especially for construct 4b. When compared with intensity for construct 1 and 2, the successful suppression of the amber termination codon at position 39 is very poor. Expression of full-length eGFP from the $eGFP_{amber39}$ gene in constructs 4a and 4b relies on both productions of functional $tRNA^{Ala}$ suppressor and the enzyme alanyl-tRNA synthetase. The enzyme is expected to aminoacylate $tRNA^{Ala}$ suppressor with Ala so that the yeast's ribosomes can then add Ala in translation of $eGFP_{amber39}$ at position 39, so suppressing the stop codon. Even though fluorescence from construct 4a and 4b was never expected to have intensity as high as that of construct 1 or 2, the level observed was insufficient to be of use for library screening assays. Conversely, there is no autofluorescence at all from the negative control, untransformed *S. cerevisiae* BY4741 cells, suggesting that the assay forms the basis further improvement.

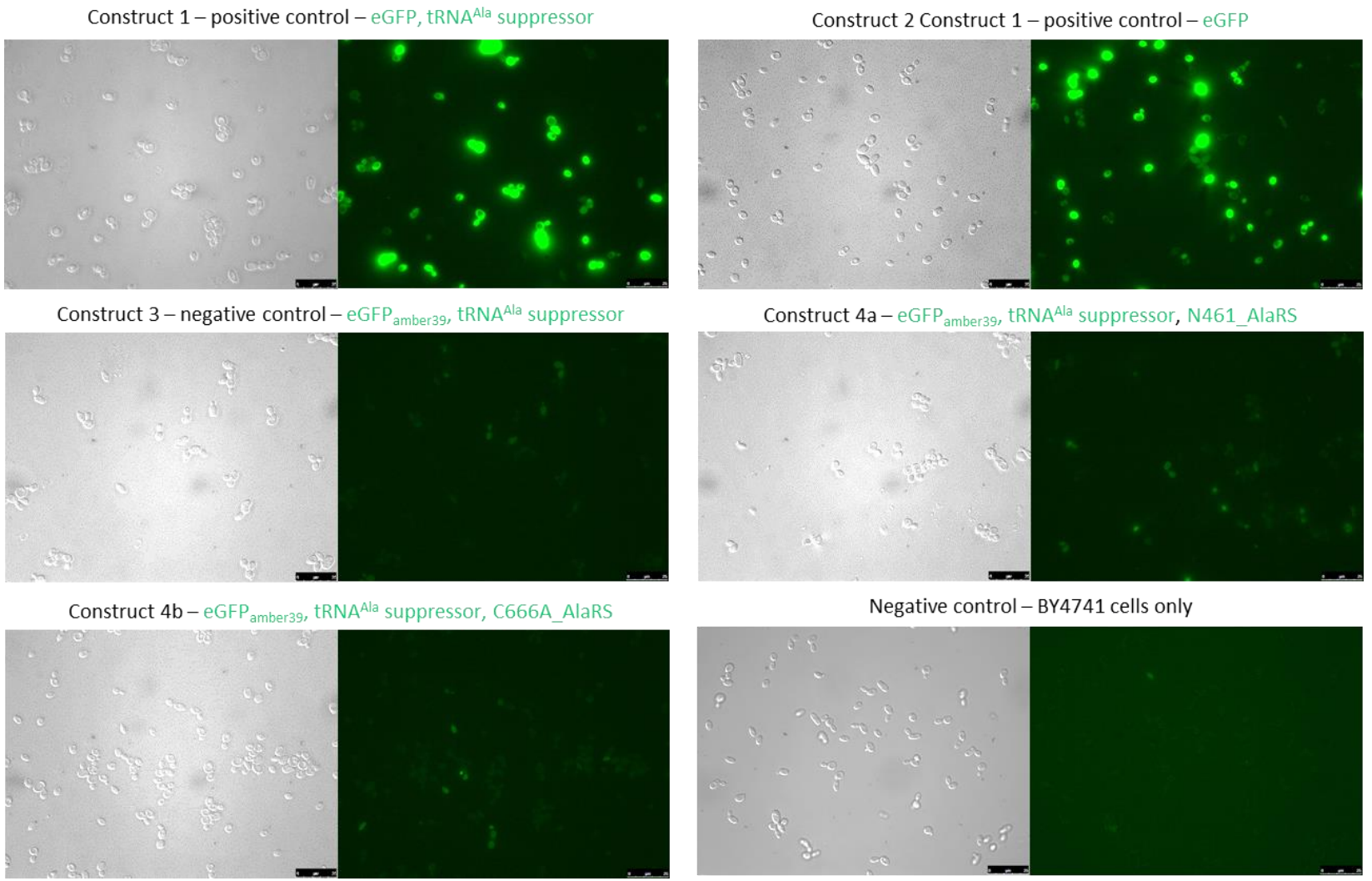


Figure 6.13. Expression of construct 1, 2, 3, 4a and 4b in S. cerevisiae BY4741.
The fluorescence microscopy images of yeast cells expressing eGFP induced overnight in SD-U 2% Gal medium. Negative control: BY4741 cells only and construct 3. Parameters: eGFP: Exposure time: 100 ms; Gain: 5.9; Intensity: 5. Grey: Exposure time: 179 ms; Gain: 2; Intensity: 87. Expression of BY4741 cells only: Parameters: eGFP: Exposure time: 60 ms; Gain: 5.9; Intensity: 5. Grey: Exposure time: 179 ms; Gain: 2; Intensity: 56.

6.4. Introduction of SNR52 promoter and removal of CCA-3'

The controls of construct 4a and 4b were designed to determine if expressed enzyme (N461_AlaRS or C666A_AlaRS) was able to aminoacylate $tRNA^{Ala}$ suppressor to start expressing full-length GFP from $eGFP_{amber39}$. The lack of clear green fluorescence in these constructs resulted from the unsuccessful suppression of the termination codon in $eGFP_{amber39}$ and as such, changes in the assay design were essential.

A manuscript "New Methods Enabling Efficient Incorporation of Unnatural Amino Acids in Yeast" by Wang and Wang (2008) outlined how to efficiently produce *E. coli* tRNA in *S. cerevisiae*. The authors suggested that in order to produce functional *E. coli* tRNA in *S. cerevisiae,* it was mandatory to follow and mimic yeast system of tRNA production. First, the triplet of nucleotides on 3'-end of tRNA that is present in *E. coli* is not found in yeast tRNA genes because it is added enzymatically in post-translational modifications. Therefore, for this study, CCA-3' had to be removed from all $tRNA^{Ala}$ suppressor constructs. Second, the manuscript explained the role of promoter elements A- and B-box sequences that are found within yeast's tRNA gene but are absent in most *E. coli* tRNAs. The solution for this problem was to place a promoter that contains A- and B-box sequences upstream of *E. coli* tRNA. Consequently, the existing promoter Gal1 ahead of *E. coli* tRNA had to be replaced with the SNR52 promoter, one of two promoters tested in that manuscript.

In summary, in order to express functional *E. coli* tRNA in yeast, two sets of primers were designed: one to remove CCA-3' from tRNA in all constructs and another to amplify the SNR52 promoter sequence from *S. cerevisiae* genome in order to clone in into amended constructs without CCA-3' in tRNA gene.

6.4.1. Removal of CCA-3' from *E. coli* tRNA^{Ala} suppressor genes

To remove CCA-3' from the tRNA^{Ala} suppressor gene, a set of inverse primers was designed to amplify the constructs. After PCR and self-ligation of the constructs, transformation into *E. coli* and colony screening, the results of amplification using construct 1 as a template are shown in Figure 6.14. Phosphorylated primers were used in this PCR. Despite a lot of efforts to optimise the PCR reaction, only a small yield of the product was obtained (Figure 6.14). Nevertheless, several reactions were set up to achieve enough product post-purification. Purified PCR reactions were self-ligated. The same methodology was repeated to construct 3, 4a and 4b. The new version of these construct had a suffix 'v2' for 'version 2' of controls. Clones for all constructs were confirmed with sequencing.

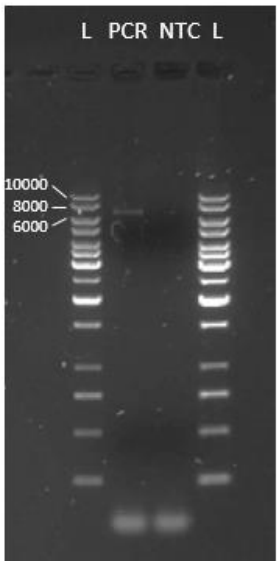


Figure 6.14. Amplification of construct 1 excluding CCA-3' from tRNA gene.

0.8 % agarose gel stained with SYBR Safe dye shows result of a PCR with an expected size of 6865 bp amplified with Q5® High-Fidelity DNA Polymerase (NEB). The amplification temperature was 44 °C. L: GeneRuler 1kb DNA ladder (Thermo Fisher). NTC: a non-template control. Construct 1: eGFP, tRNA^{Ala} suppressor.

6.4.2. Change of promoters for tRNA^{Ala} suppressor gene

To exchange promoters upstream of tRNA^{Ala} suppressor gene within version 2 constructs, the required SNR52 promoter sequence was amplified from the *S. cerevisiae* genome by yeast colony PCR which generated a product of the correct size of 288 bp (Figure 6.15. a). After PCR purification, the product was digested with *Bcl*I (Figure 6.15. b.).

Meanwhile, the Gal1 promoter was removed and concurrently, a *Bcl*I restriction site was introduced into constructs 4a and 4b by PCR as described previously (Figure 6.16.). The gradient PCR showed the correct product for annealing temperatures of 61.8 – 66.0 °C (2.2.1.1.1). For both constructs, 66 °C was used for annealing in the scale-up PCR (2.2.1.1.). After PCR purification, the enzyme digestion with *Bcl*I was completed (2.2.1.8.1.; 2.2.1.6.). Having digested both insert and vector, the two were ligated as shown in Figure 6.17. (2.2.1.5.), transformed and resulting colonies screened by PCR (2.2.3.1.; 2.2.1.1.2.; Figure 6.18.). Several colonies were sent for sequencing (2.2.4.1.) and the correct sequences were found in colony 25 for construct 4bv3 and colonies 1, 5, 32 and 35 for construct 4av3. A suffix ‘v3’ for ‘version 3’ of controls was added to indicate an exchange of Gal1 promoter for SNR52 promoter for expression of tRNA. The plasmid maps for construct 4av3 and 4bv3 are illustrated in Figure 6.19. The process was then repeated for constructs 1 and 3 (data not shown).

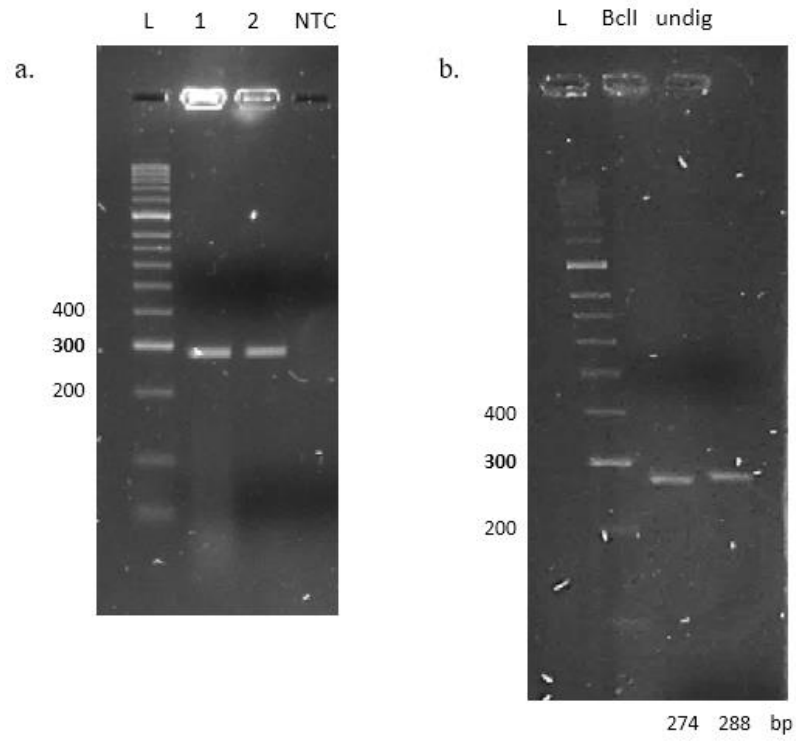


Figure 6.15. Amplification and enzyme digestion of SNR52 promoter.

*3 % agarose gels stained with SYBR Safe dye show results of amplification and enzyme digestion. a. Amplification of SNR52 promoter from S. cerevisiae genome with an expected size of 288 bp with Q5® High-Fidelity DNA Polymerase (NEB). 1, 2: PCR reactions. NTC: non-template control. b. Enzyme digestion of PCR amplicons with BclI with the expected size of 274 bp upon digestion (lane BclI). Undig: undigested PCR. L: HyperLadder™ 50bp (*Bioline*).*

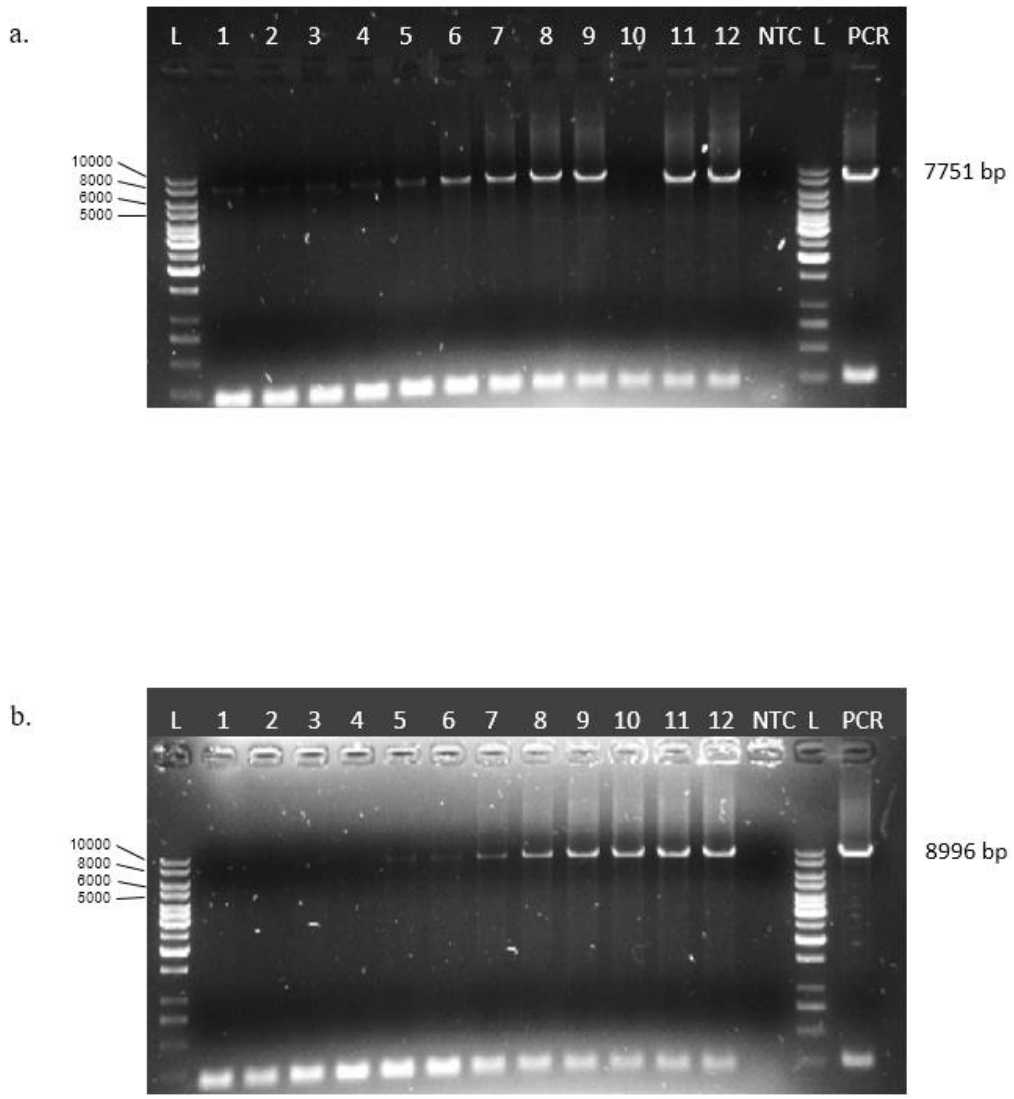


Figure 6.16. Amplification of construct 4a and 4b vector backbone.

0.8 % agarose gels stained with SYBR Safe dye show results of PCR. a. Gradient PCR amplifying the vector backbone of construct 4a with an expected size of 7751 bp. b. Gradient PCR amplifying the vector backbone of construct 4b with an expected size of 8996 bp. Lanes number 1 to 12 represent reactions for temperature gradient 50-66 °C amplified with Q5® High-Fidelity DNA Polymerase (NEB). On the far right of both gels lane PCR represent scaled-up PCR with amplification at 66 °C. L: GeneRuler 1kb DNA ladder (Thermo Fisher). NTC: a non-template control. Construct 4a/b: $eGFP_{amber39}$, $tRNA^{Ala}$ suppressor, N461_AlaRS/C666A_AlaRS.

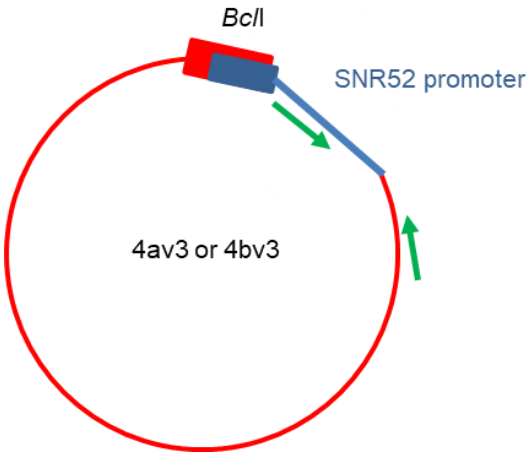


Figure 6.17. Design of colony PCR screening for clones.

Primers (green arrows) were designed to amplify through ligation site. Forward primer anneals to the SNR52 promoter sequence, while reverse primer to vector backbone. Suffix 'v3' indicate 3rd generation of constructs where SNR52 promoter replaced Gal1.

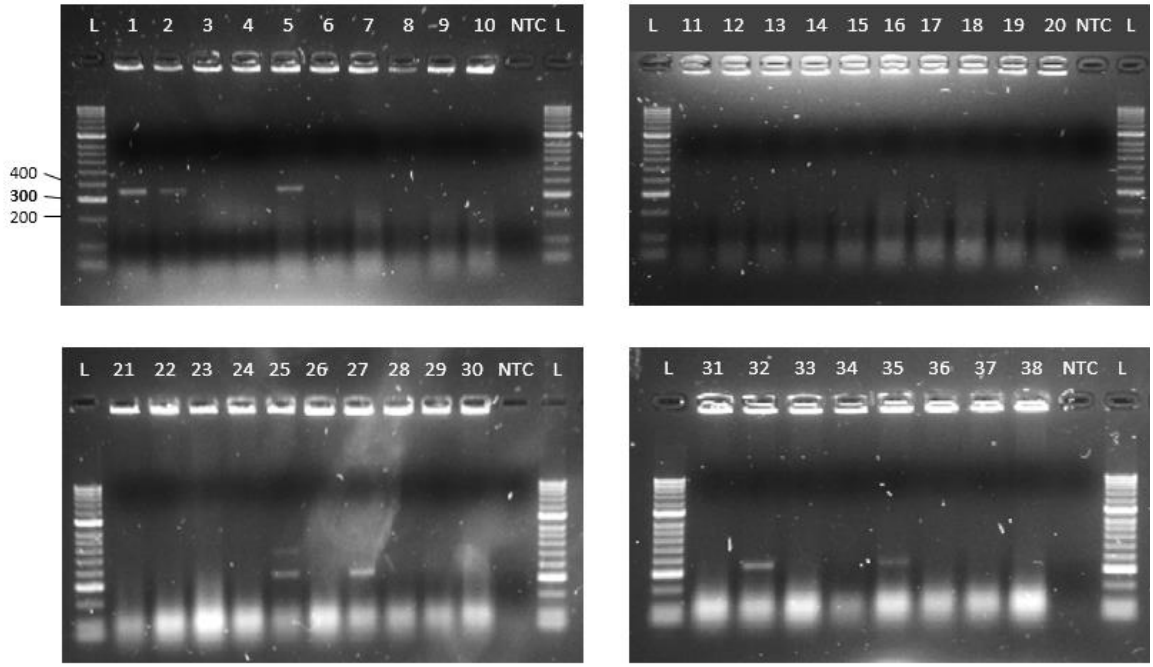


Figure 6.18. Colony PCR for constructs 4av3 and 4bv3.

2 % agarose gels stained with SYBR Safe dye show results of colony screening. for construct 4av3 (colonies 1-10 and 31-38) and for construct 4bv3 (colonies 11-30). The product of colony PCR with an expected size of 355 bp was detected in colonies 1, 2, 5, 25, 27, 32, and 35. NTC: non-template control. L: HyperLadder™ 50bp (Bioline). Construct 4av3/bv3: $eGFP_{amber39}$, $tRNA^{Ala}$ suppressor, N461_AlaRS/C666A_AlaRS (refer to Figure 6.19.).

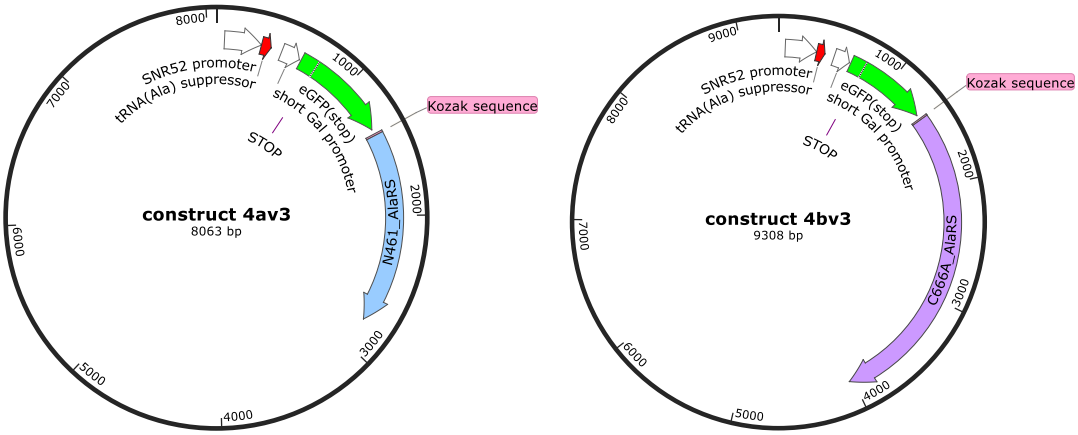


Figure 6.19. The plasmid maps of construct 4av3 and 4bv3.

6.4.3. Analysis of expression under a fluorescence microscope

To verify if the changes to the gene elements of $tRNA^{Ala}$ suppressor improved its expression and in the resulting expression of full-length eGFP from $eGFP_{amber39}$, the yeast cells with appropriate constructs were analysed under the fluorescence microscope (2.2.9.). With these changes, it was anticipated that the $tRNA^{Ala}$ suppressor should now be produced efficiently when the constructs were transformed into yeast strain BY4741 (2.2.6.2.). As shown in Figure 6.20., construct 2, the positive control for expression of eGFP, gave a bright and fluorescent green colour. However, the cells for both constructs, 4av3 and 4bv3, remained of the same weak intensity green colour, just as with constructs 4a and 4b prior to removal of the CCA-3' sequence and addition of the recommended promoter (Figure 6.13.). As part of the experiment, constructs 4av3 and 4bv3 without tRNA promoter serving as negative controls for 4av3 and 4bv3 were expressed alongside. In theory, these controls should not express any $eGFP_{amber39}$ because tRNA suppressor is not expressed due to lack of promoter. From the images in Figure 6.20., it can be observed that fluorescence levels between both samples are very similar.

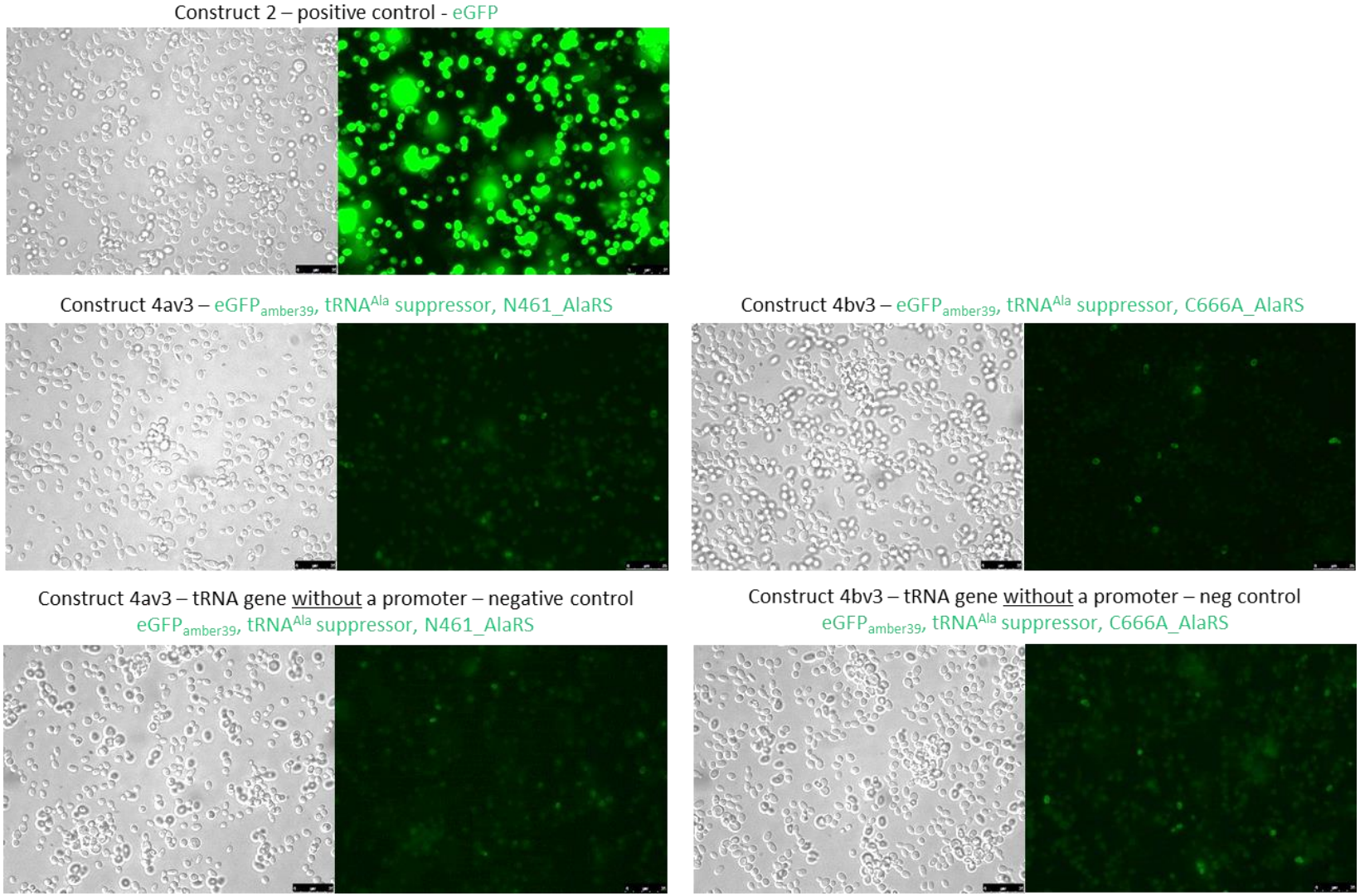


Figure 6.20. Expression of construct 2, 4av3 and 4bv3 in S. cerevisiae BY4741.
The fluorescence microscopy images of yeast cells expressing eGFP induced overnight and 24h in SD-U 2% Gal medium. Parameters: eGFP: Exposure time: 120 ms; Gain: 5; Intensity: 5. Grey: Exposure time: 60 ms; Gain: 5; Intensity: 120.

To determine whether the nonsense-mediated decay pathway was interfering with expression, an alternative strain of yeast was used. The strain BY4741 NMD2Δ has a deletion of protein NMD2 which has a critical role in the Nonsense-Mediated mRNA Decay (NMD) pathway. The same controls were expressed in BY4741 NMD2Δ like in Figure 6.20. From the analysis of images under the fluorescence microscope (Figure 6.21.), it was observed that the expression of full-length eGFP was improved in this new strain, that even the exposure in the fluorescent/GFP channel had to be reduced from 120 ms to 20 ms. In order to keep the relative exposure of other controls within the same experiments as close as possible, the exposure in the fluorescent/GFP channel was reduced from 120 ms to 60 ms, when comparing to images in Figure 6.20. Yeast cells with pYES2 vector only as a negative control remained not fluorescent with only a few false-positive cells. When comparing assay controls ($eGFP_{amber39}$, tRNA suppressor, enzyme), even at this lower exposure, the fluorescence levels for controls 4av3 and 4bv3 was higher in BY4741 NMD2Δ strain (Figure 6.21.) than for the same controls in BY4741 (Figure 6.20.). This indicates that using strain BY4741 NMD2Δ for expression of constructs improved reading through termination codon in $eGFP_{amber39}$.

As part of the experiment, constructs 4av3 and 4bv3 without tRNA promoter serving as negative controls for 4av3 and 4bv3 were expressed alongside. In theory, these controls should not express any $eGFP_{amber39}$ because tRNA suppressor is not expressed due to lack of promoter. When analysing the images obtained in Figure 6.21. and comparing 4av3/4bv3 with and without promoter for tRNA, there is more fluorescence in full constructs 4av3 and 4bv3 with the promoter for tRNA. To be accurate, there is some background fluorescence that is very similar to the one in control with pYES2 vector only.

Although the levels of $eGFP_{amber39}$ expression in constructs 4av3 and 4bv3 had improved when compared with constructs 4a and 4b (Figure 6.20.), this level of fluorescence still remained insufficient for the purposes of screening AlaRS libraries.

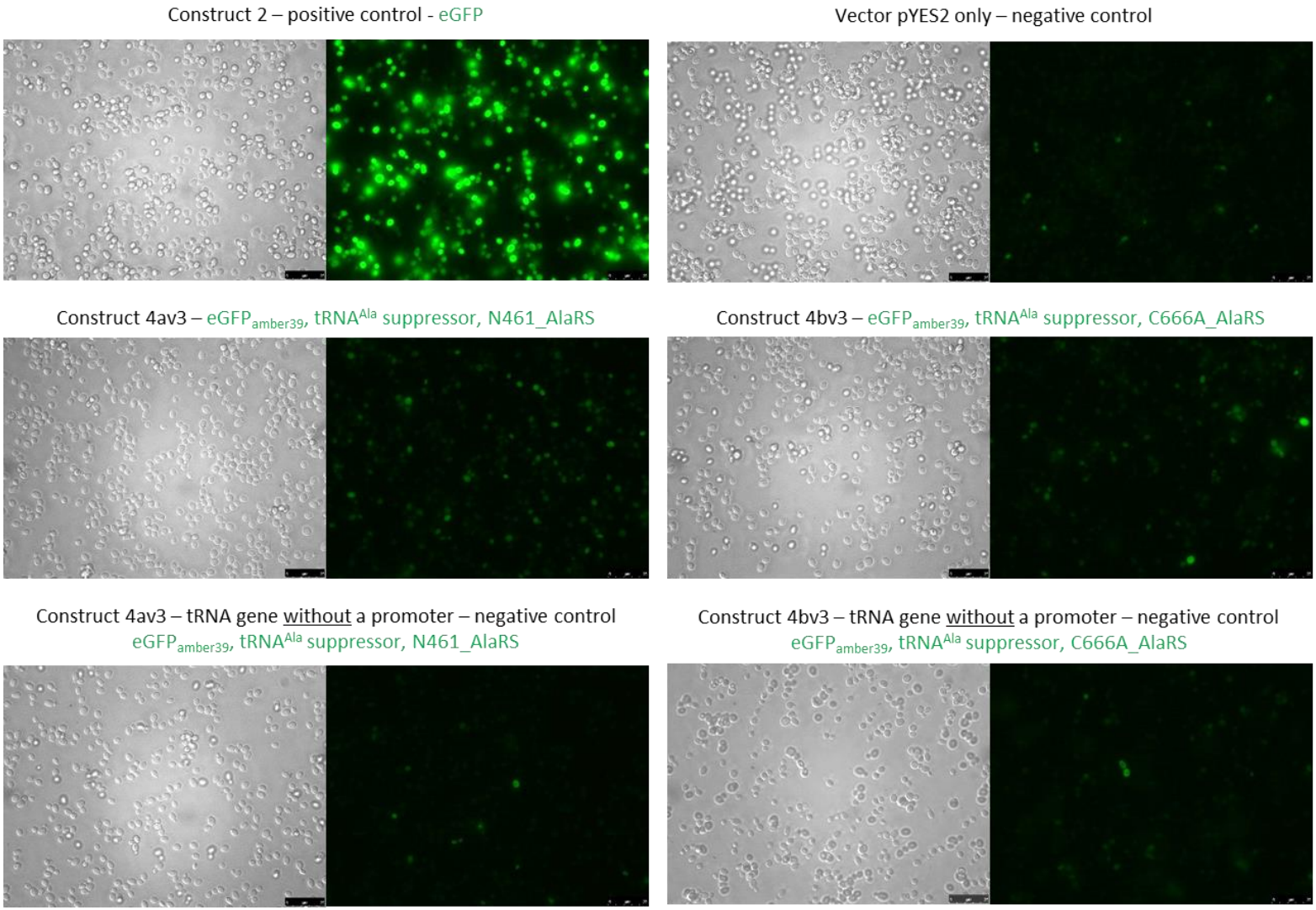


Figure 6.21. Expression of construct 2, 4av3, 4bv3 and vector pYES2 in S. cerevisiae BY4741 NMD2Δ.
The fluorescence microscopy images of yeast cells expressing eGFP induced overnight and 24h in SD-U 2% Gal medium. Parameters: eGFP: Exposure time: 60 ms (20 ms for construct 2); Gain: 5 (3 for construct 2); Intensity: 5 (3 for construct 2). Grey: Exposure time: 60 ms; Gain: 5; Intensity: 120.

6.5. Two-plasmid system – transfer of the eGFP gene to a different vector

The next rational step was to turn into the optimisation of the AlaRS enzyme expression. In previous constructs 4a/4av3 and 4b/4bv3, the AlaRS gene was part of an operon with the $eGFP_{amber39}$ gene. If expression of either was adversely affected, poor expression of full-length GFP would result. Consequently, a two-plasmid system was tested, in which the $eGFP_{amber39}$ gene was transferred to vector pXRH3 (Figure 6.22. a.), which is compatible with both pYES2 and *S. cerevisiae* BY4741 NMD2Δ.

6.5.1. Preparation of new constructs for the two-plasmid system

In the new constructs (v4), the short Gal promoter and the eGFP gene were moved to plasmid pXRH3 (Figure 6.22. b.). In this design, each gene had its own promoter and terminator: in pYES2 vector there was SNR52 promoter for tRNA and the short Gal promoter for the AlaRS enzyme, while in pXRH3 vector, the short Gal promoter was also used for eGFP. This promoter was used again, since it had already proved to be effective in eGFP expression from construct 2 (eGFP only). Vector pXRH3 is selected by histidine and can be used for simultaneous transformation into BY4741 strains with existing clones in the pYES2 vector.

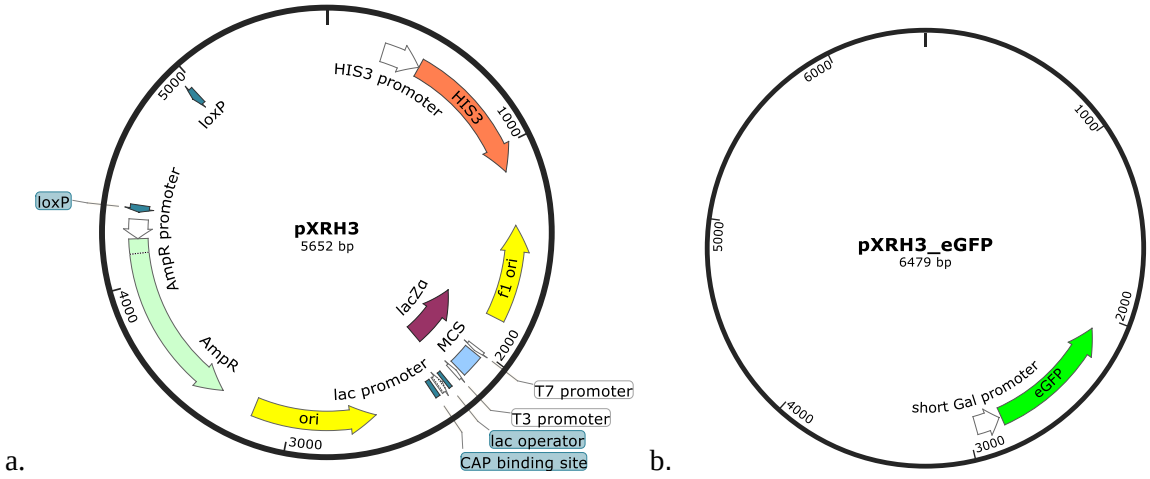


Figure 6.22. Plasmid map of a. pXRH3 histidine yeast selection plasmid and b. cloned short Gal promoter and eGFP.

6.5.1.1. Preparation of construct 4av4 and 4bv4

To prepare construct 4av4 and 4bv4, the $eGFP_{amber39}$ gene had to be removed from the existing construct 4av3 and 4bv3. Firstly, primers were designed to remove this gene and amplify around the plasmid to achieve a pYES2 vector without $eGFP_{amber39}$ gene, as shown in Figure 6.23 (2.2.1.1.). Despite efforts to optimise PCR reaction, for both constructs, there were some additional unspecific bands other than the band of interest (7294 bp for 4av3 and 8539 bp for 4bv3). Instead of PCR optimisation to achieve a single band, the bands of interest were gel extracted (2.2.1.8.3.; Figure 6.24.) and used in a subsequent self-ligation (2.2.1.5.). Once transformed into *E. coli* (2.2.3.1.), colonies were screened in colony PCR (2.2.1.1.2.) with the expected size of 299 bp for both, construct 4av4 (colonies 21-27) and 4bv4 (colony 28). All colonies showed the correct size of colony PCR as shown in Figure 6.25. Colonies 21, 22, 28 were sent for sequencing and all had correct sequence of $eGFP_{amber39}$. However, clone 28 had some additional sequence after $eGFP_{amber39}$ gene, hence the additional bands in colony PCR. The additional diagnostic enzyme digestion with *Sph*I as a unique

restriction site for colonies 21, 22 and 28 aimed to linearise the plasmids to analyse their total size.

Figure 6.26. shows correct sizes for all the colonies analysed.

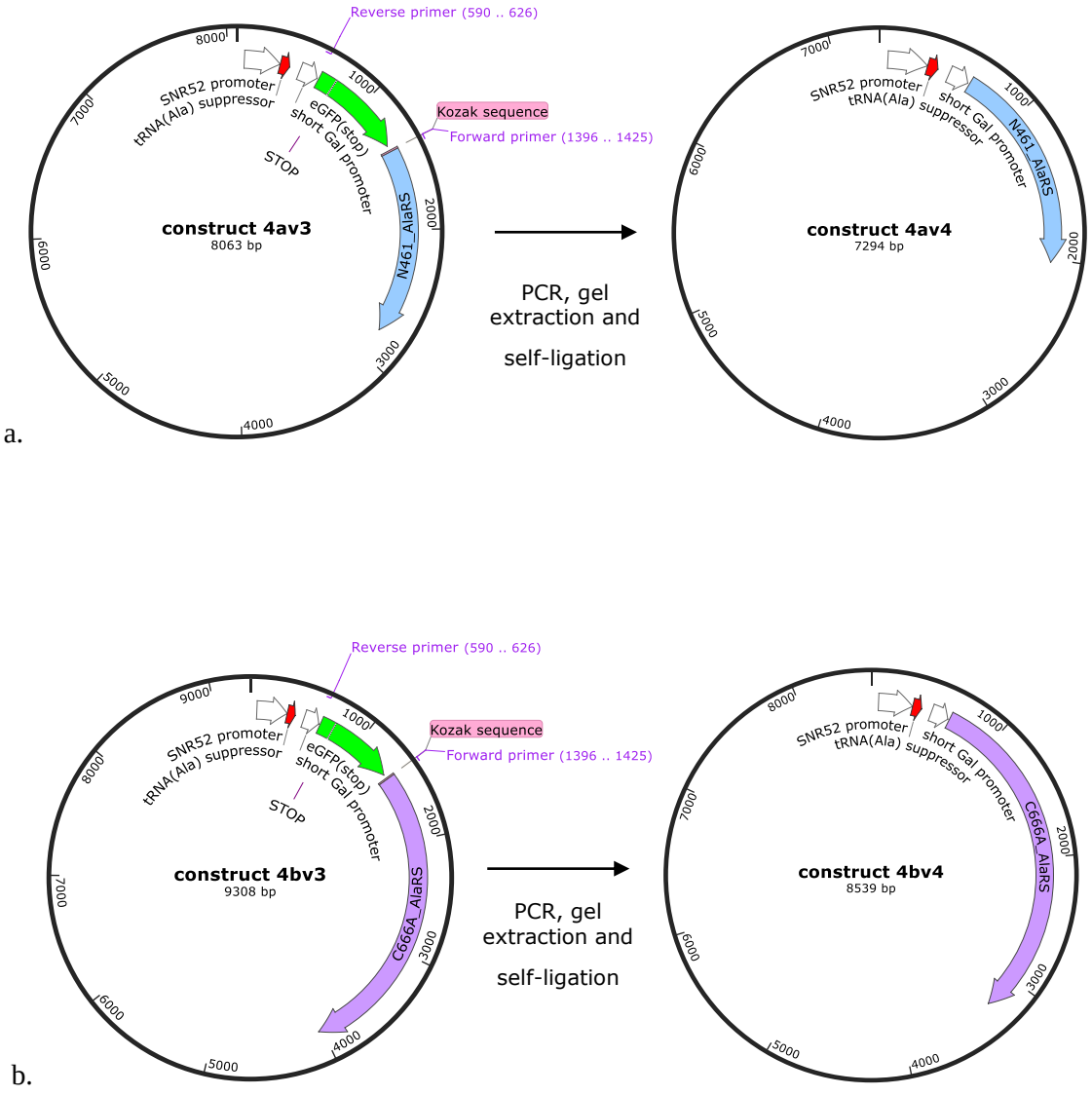


Figure 6.23. The workflow for generation new constructs a. 4av4 and b. 4bv4.

The plasmid maps of construct 4av3 (a.) and 4bv3 (b.) with annealed primers to remove $eGFP_{amber39}$ gene by amplification and self-ligation of PCR reaction to generate construct 4av4 (a.) and 4bv4 (b.). New constructs have the $tRNA^{Ala}$ suppressor gene under SNR52 promoter and the AlaRS gene under short Gal promoter.

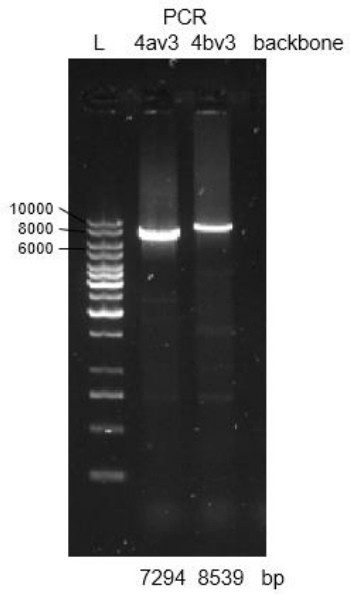


Figure 6.24. Amplification of plasmid backbone 4av3 and 4bv3 for construct 4av4 and 4bv4.

0.8 % agarose gel stained with SYBR Safe dye shows results of PCRs with an expected size of 7294 bp for 4av3 backbone and 8539 bp for 4bv3 backbone. The reactions were amplified using Q5® High-Fidelity DNA Polymerase (NEB). For both lanes, there are visible additional unspecific amplification fragments of shorter lengths than an expected product of PCR. L: GeneRuler 1kb DNA ladder (Thermo Fisher). Construct 4av3/bv3 or 4av4/bv4: $eGFP_{amber39}$, $tRNA^{Ala}$ suppressor, N461_AlaRS/C666A_AlaRS (refer to Figure 6.23.).

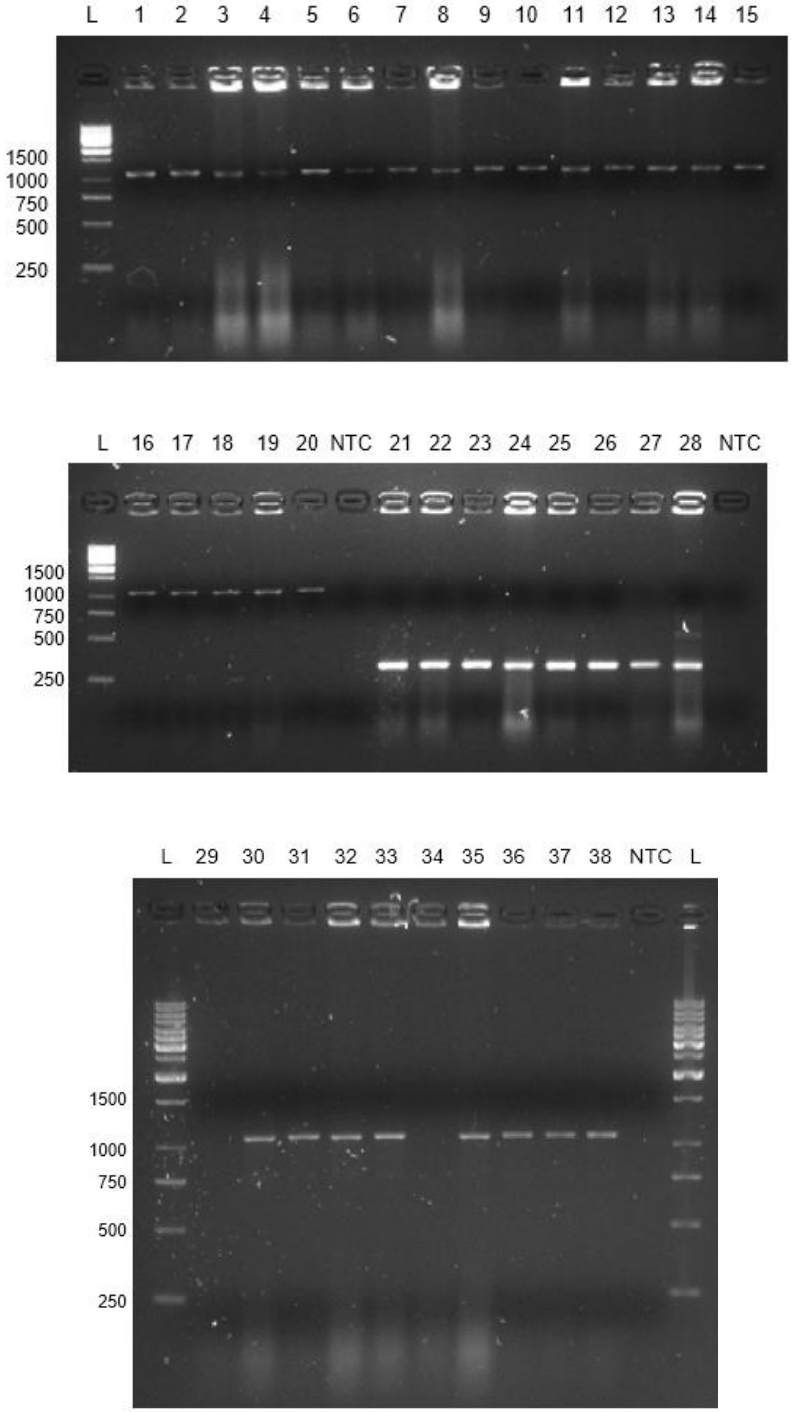


Figure 6.25. Colony PCR for construct pXRH3_eGFP$_{amber39}$, pXRH3_eGFP, 4av4 and 4bv4.

2 % agarose gels stained with SYBR Safe dye show results of colony screening. for construct pXRH3_eGFP$_{amber39}$ (colonies 1-20), pXRH3_eGFP (colonies 29-38), 4av4 (colonies 21-27) and for construct 4bv3 (colony 28). The product of colony PCR for colonies 1-20 with an expected size of 1099 bp was detected in all 20 colonies. The same product was detected in colonies 30-33 and 35-38 for construct pXRH3_eGFP. The product of colony PCR for colonies 21-28 with an expected size of 299 bp was also detected in all colonies. Lane 28 showed other two unspecific bands corresponding to longer DNA than desired PCR product. Other colonies had a single band product in the experiment. NTC: non-template control. L: HyperLadder™ 50bp (Bioline). Construct pXRH3_eGFP$_{amber39}$: GFP$_{amber39}$ (refer to Figure 6.22). Construct 4av4/bv4: tRNA^{Ala} suppressor, N461_AlaRS/C666A_AlaRS (refer to Figure 6.23.).

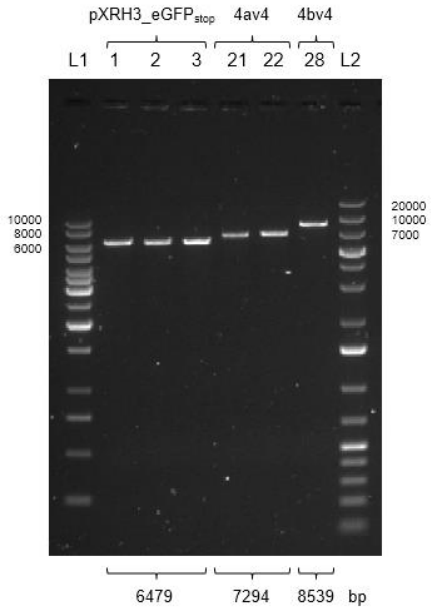


Figure 6.26. Plasmid screening by enzyme digestion.

0.8 % agarose gel stained with SYBR Safe dye shows results of enzyme digestion for colonies of construct pXRH3_eGFP$_{amber39}$, 4av4 and 4bv4. Enzyme EcoRI was used for pXRH3_eGFP$_{amber39}$ and SphI for 4av4 and 4bv4. L1: GeneRuler 1kb DNA ladder (Thermo Fisher). L2: GeneRuler 1kb plus DNA ladder (Thermo Fisher). Construct pXRH3_eGFP$_{amber39}$: eGFP$_{amber39}$ (refer to Figure 6.22). Construct 4av4/bv4: tRNA^{Ala} suppressor, N461_AlaRS/C666A_AlaRS (refer to Figure 6.23.).

6.5.1.2. Preparation of construct pXRH3_eGFP$_{amber39}$ and pXRH3_eGFP

To prepare a new vector with eGFP$_{amber39}$ and another with eGFP under control of the short Gal promoter, the relevant sequences were digested from construct 4bv3 with *Sac*I and *Eco*RI as shown in Figure 6.27. (2.2.1.6.). Resulting fragments were gel extracted and ligated into pXRH3 that had been similarly digested and transformed into *E. coli* (2.2.1.8.3., 2.2.1.5., 2.2.3.1.). The colonies were screened in colony PCR using M13 primers and the results of screening colonies 1 to 20 are shown in Figure 6.25. (2.2.1.1.2.). All 20 samples showed the correct size of PCR at 1099 bp. Colonies 1, 2 and 3 were sent for sequencing and the results showed correct insertion of eGFP$_{amber39}$ into pXRH3. The vector was called pXRH3_eGFP$_{amber39}$. As a control to determine the expression of eGFP in the new vector, the same new vector was similarly-prepared, but with the wt gene of eGFP, from

construct 2 (Figure 6.28.). This vector was called pXRH3_eGFP. The colony screening was done using colony PCR and colonies 30-33 and 35-38 showed correct product in Figure 6.25. Colonies 30, 31 and 32 were sent for sequencing with all showing correct insertion of the eGFP gene.

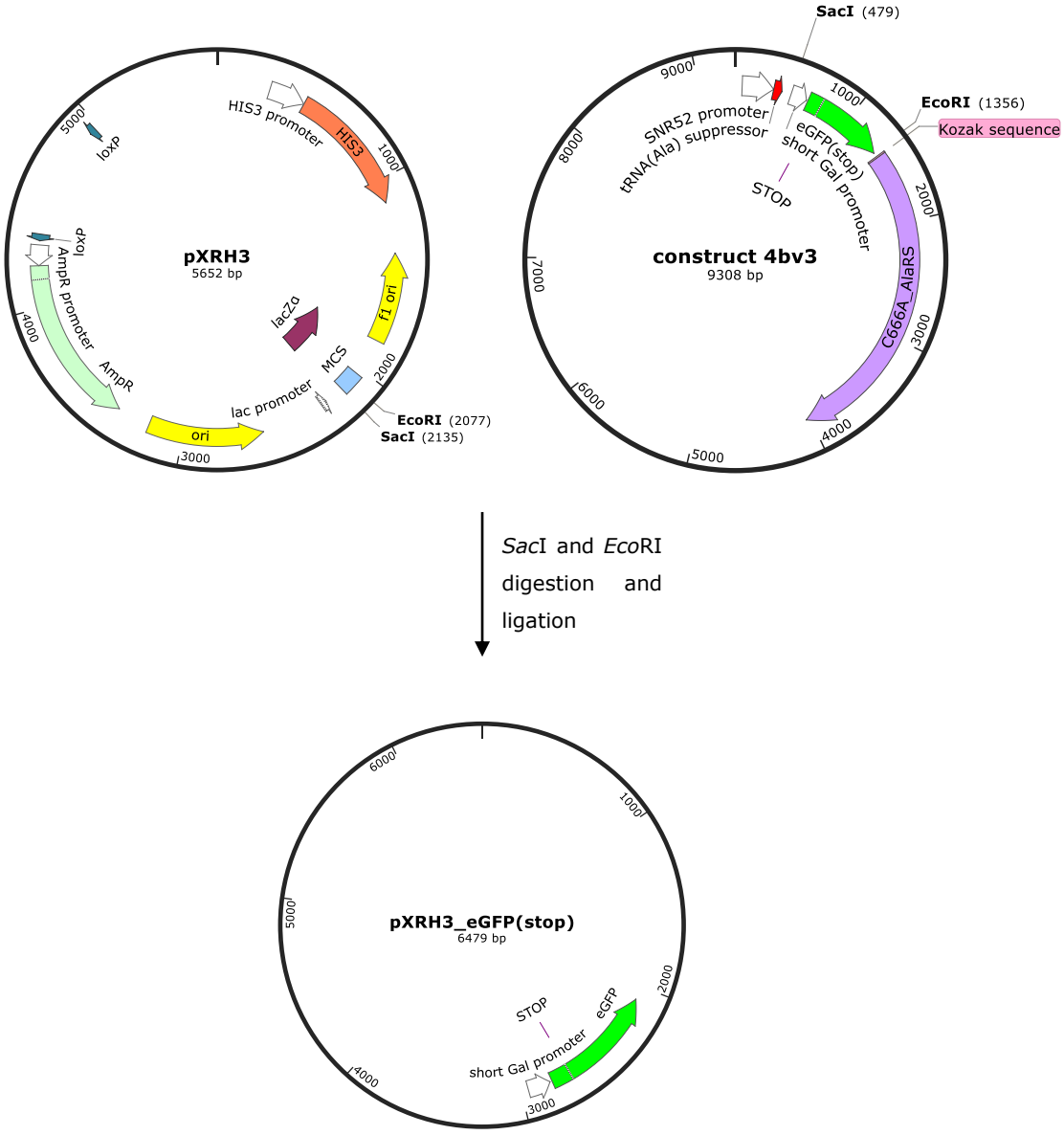


Figure 6.27. The workflow for generation new plasmid, pXRH3 with eGFP$_{amber39}$.

A new vector pXRH3 was cloned with the insert, the eGFP$_{amber39}$ gene under short Gal promoter, from construct 4bv3.

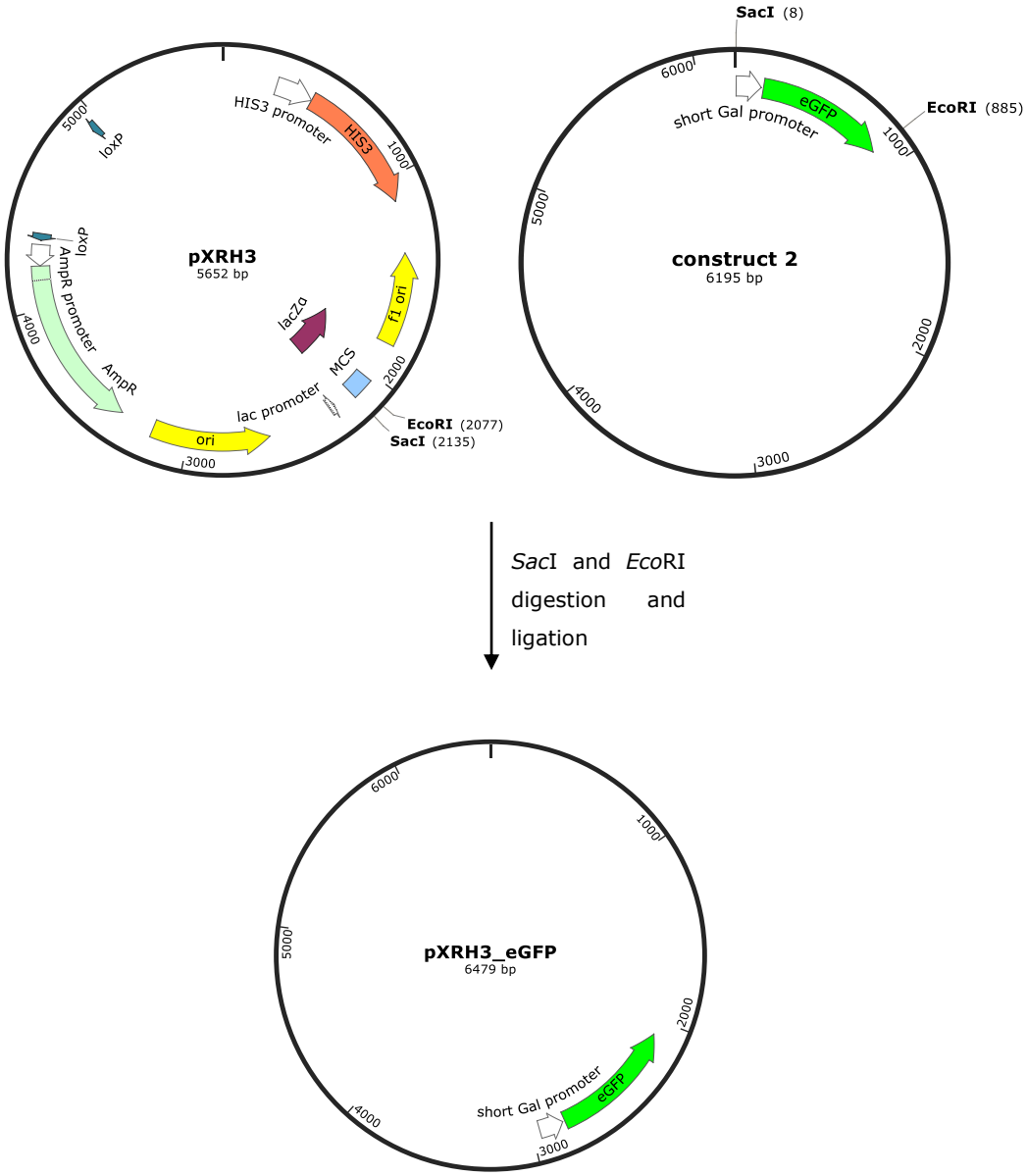


Figure 6.28. The workflow for generation new plasmid, pXRH3 with eGFP.

A new vector pXRH3 was cloned with the insert, the eGFP gene under short Gal promoter, from construct 2.

6.5.2. Analysis of expression under a fluorescence microscope for two-plasmid system.

As it had been performed previously in Sections 6.3.1. and 6.4.3., the dual constructs were transformed into *S. cerevisiae* BY4741 NMD2Δ strain (2.2.3.2.) and were analysed by fluorescence microscopy (2.2.9.).

Upon analysis of the fluorescence microscopy images, it was confirmed that eGFP gene that was moved to vector pXRH3 to form pXRH3_eGFP, expressed the protein and yeast cells exhibited green phenotype. A negative control shown in Figure 6.30., sample with only cells remains without fluorescence Two-plasmid system was analysed. Cells expressing construct 4av4 and pXRH3_eGFP$_{amber39}$ as well as construct 4bv4 and pXRH3_eGFP$_{amber39}$ were analysed. Nevertheless, changing from one-plasmid to two-plasmid system did not elicit changes in the fluorescence levels. Additionally, cells with eGFP$_{amber39}$ only show comparable fluorescence. Clearly, the limited fluorescence is dependent on the presence of the eGFP$_{amber39}$ gene, since no fluorescence at all is visible for cells lacking this gene. Interestingly, there is still some fluorescence in the cells that contain only pXRH3_eGFP$_{amber39}$, suggesting that the yeast cells may inherently be able to read through the termination codon at position 39 to a low level. However, close examination of Figure 6.30. also suggests that fluorescence is a little brighter when the AlaRS and tRNA^{Ala} suppressor genes are present. Since the fluorescence is clearly dependent on the presence of the eGFP$_{amber39}$ gene, it was decided to further analyse the cells by flow cytometry.

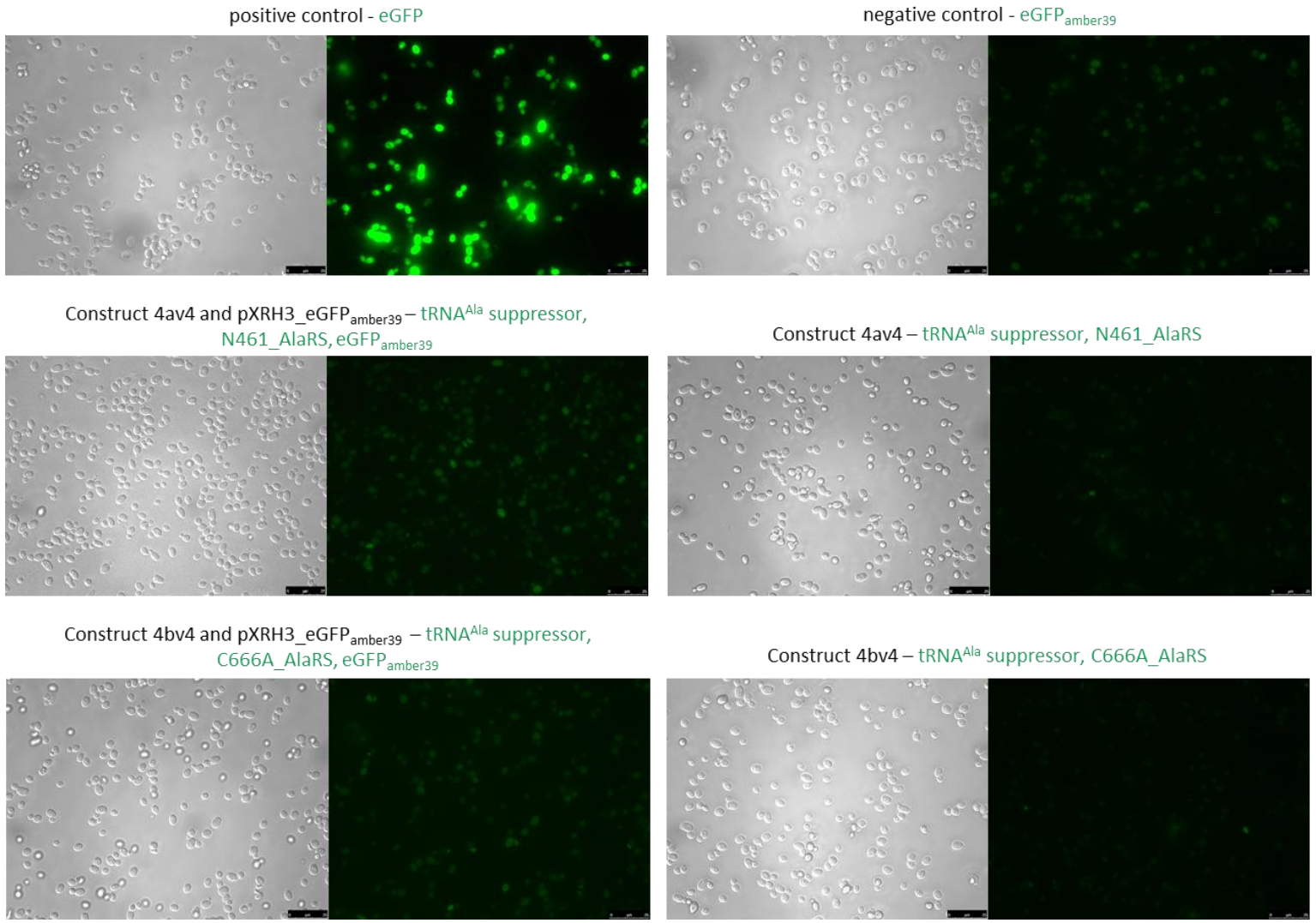


Figure 6.29. Expression of constructs 4av4 and 4bv4, and controls in S. cerevisiae BY4741 NMD2Δ.

The fluorescence microscopy images of yeast cells expressing eGFP induced overnight and 24h in SD-U 2% Gal medium. Parameters: eGFP: Exposure time: 60 ms; Gain: 5; Intensity: 5. Grey: Exposure time: 60 ms; Gain: 5; Intensity: 120.

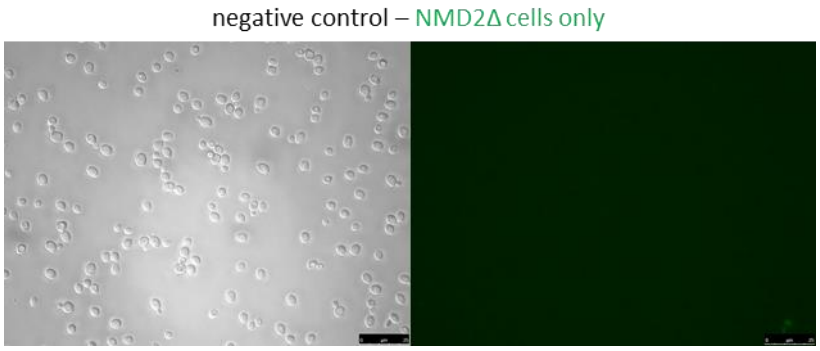


Figure 6.30. The fluorescence microscopy of yeast cells S. cerevisiae BY4741 NMD2Δ.

The fluorescence microscopy images of yeast cells. Parameters: eGFP: Exposure time: 60 ms; Gain: 5; Intensity: 5. Grey: Exposure time: 60 ms; Gain: 5; Intensity: 120.

6.5.3. Analysis of expression by Flow Cytometry for the two-plasmid system.

Flow cytometry was employed to further quantify fluorescence in *S. cerevisiae* (2.2.11.). To measure fluorescence resulting from eGFP rather than other sources such as cell debris, dye DRAQ5 which is a cell-permeable far-red fluorescent dye that intercalates in DNA was used in conjunction with GFP fluorescence, to detect nucleated cells that are also expressing GFP protein. Gates were set to exclude fluorescence resulting from untransformed cells. Cells containing the various combinations of constructs were then examined using this gating, as illustrated in Figure 6.31.

The negative sample containing cells only was used to set the gate and so has 0% events in the gated region. Conversely, as expected in the positive sample with eGFP only, there are 83.7 % events in the gated region. While controls with tRNA, enzyme, $eGFP_{amber39}$, have only 0.55 % (with N461_AlaRS) and 0.32 % (with C666A_AlaRS) of cell population exhibiting fluorescence. At the same time, negative control with $eGFP_{amber39}$ only (no suppressor tRNA or AlaRS) has 1.61 % events with fluorescence. This latter result indicates that even though other components (like $tRNA^{Ala}$

suppressor and enzyme) are not present to aid expression of $eGFP_{amber39}$, the expression, although minimal, still occurs. These findings support results from the fluorescence microscopy and suggest that the yeast cells are inherently able to suppress the termination codon at position 39 of $eGFP_{amber39}$, without the help of orthogonal tRNA or AlaRS. Moreover, the flow cytometry results further determine that these orthogonal genes do not, as was previously suggested by fluorescence microscopy, enhance the expression of full-length eGFP from the $eGFP_{amber39}$ gene.

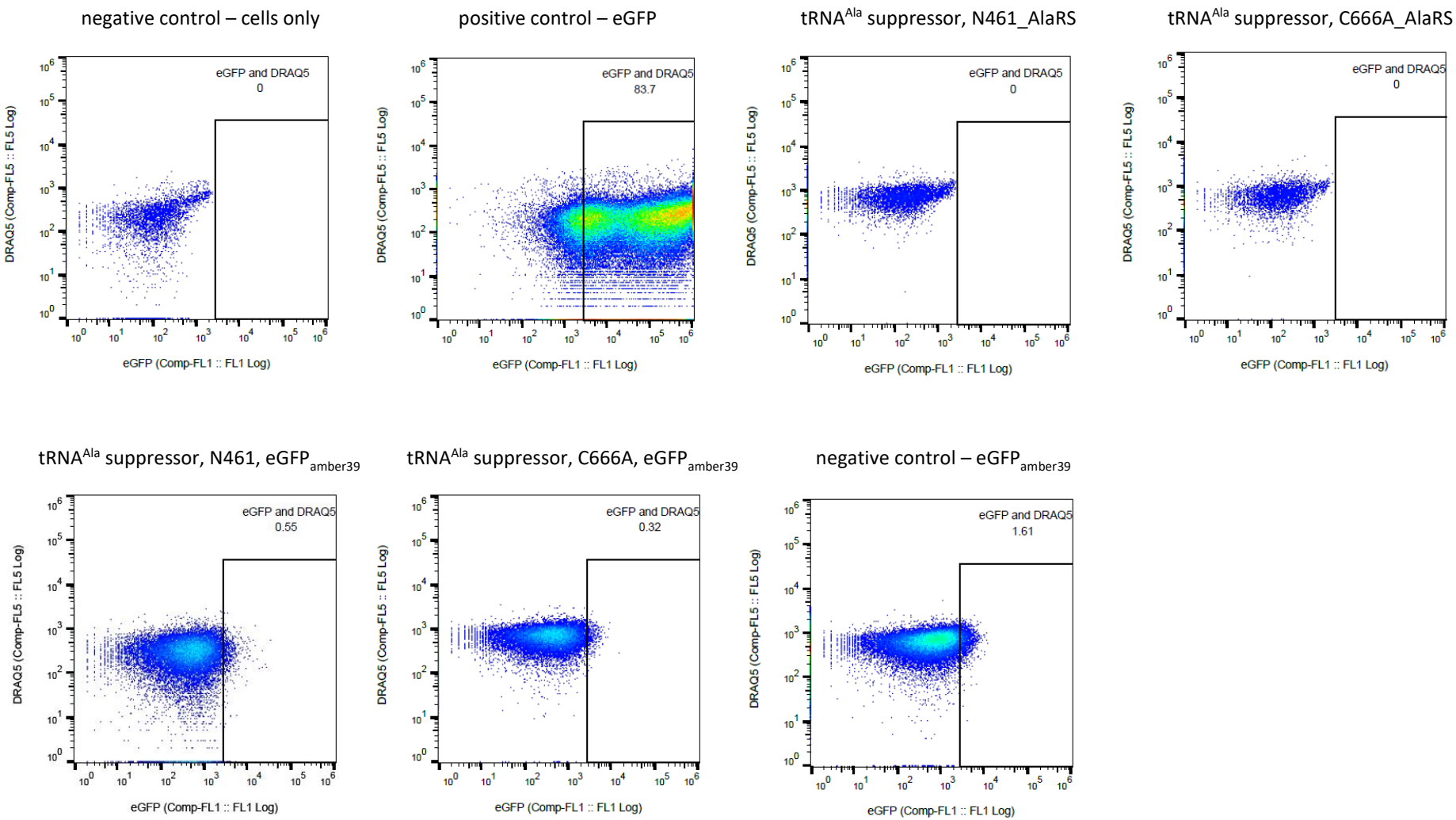


Figure 6.31. Flow cytometry of control samples for eGFP and DRAQ5.

6.5.4. RT-PCR

Since the orthogonal tRNA and AlaRS genes do not appear to enhance expression of full-length GFP, the obvious question is therefore whether or not these genes are being expressed at all. To determine whether the respective genes $eGFP_{amber39}$, $tRNA^{Ala}$ suppressor and either full-length or truncated AlaRS were all being transcribed by the *S. cerevisiae* cells in constructs 4av4 and 4bv4, reverse-transcriptase PCR (RT-PCR) analysis was employed (2.2.1.2.).

This experiment was particularly important to exclude uncertainties about expressing the AlaRS enzymes under control of short Gal promoter downstream of $eGFP_{amber39}$ as an operon. Although it had been confirmed that this promoter works efficiently in expressing eGFP, up to this point it was still unclear whether downstream placed gene of enzyme also get an efficient transcription and translation in constructs 4a and 4b (all versions).

For an effective RT-PCR experiments, a reliable RNA extraction and purification is critical. The RNA extraction is the most determinant process for RT-PCR. RNA extracted has to be of the highest quality, free of carry-over DNA and RNases. DNA can interfere with subsequent amplification step and serve as a template while RNases present will have a damaging effect on the RNA template. The easiest way to extract RNA is using commercial RNA extraction kits. However, in order to use the kit, the sample has to be lysed. The chosen method of lysis was using glass beads to break the yeast cells in a bead mill TissueLyser LT (QIAGEN) (2.2.7.3.). Essential when working with these samples was to purify RNA twice, with DNase treatments during each purification. This was critical for obtaining the RNA templates free from DNA.

6.5.4.1. Assessment of RNA integrity

The integrity and intactness of high quality purified RNA samples should follow a pattern of: 60 % 28S rRNA, 30 % 18S rRNA, 10 % tRNA/snRNA (Park *et al.*, 1996). Once the RNA was purified and quantified (2.2.1.8.4.), it was assessed in agarose gel electrophoresis, which also demonstrated the importance of heat treating RNA samples to prevent aggregation by base pairing, as is reflected in sharp and clear bands for heat-treated RNA fractions in Figure 6.32. Moreover, the ratio of 28S to 18S is 2:1 as expected (Figure 6.32.), confirming that the quality of the prepared RNA was suitable for subsequent RT-PCR analysis (2.2.1.2.).

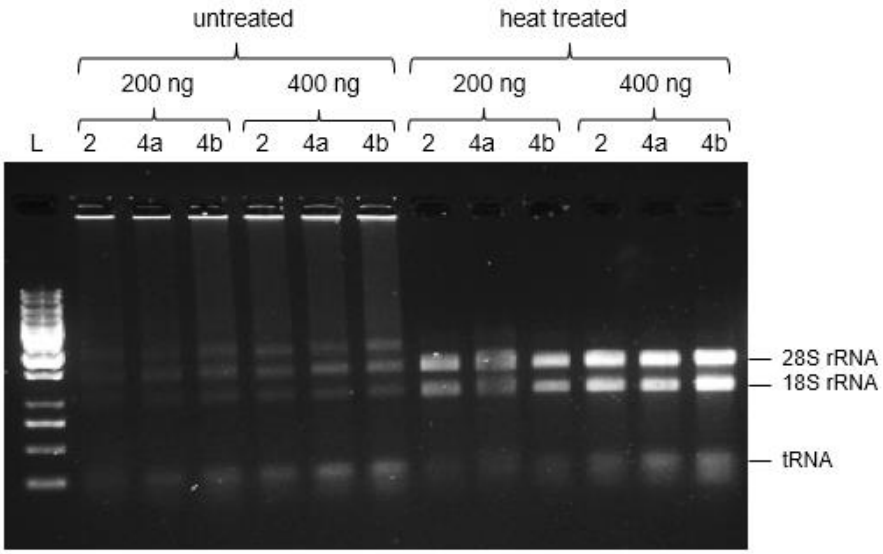


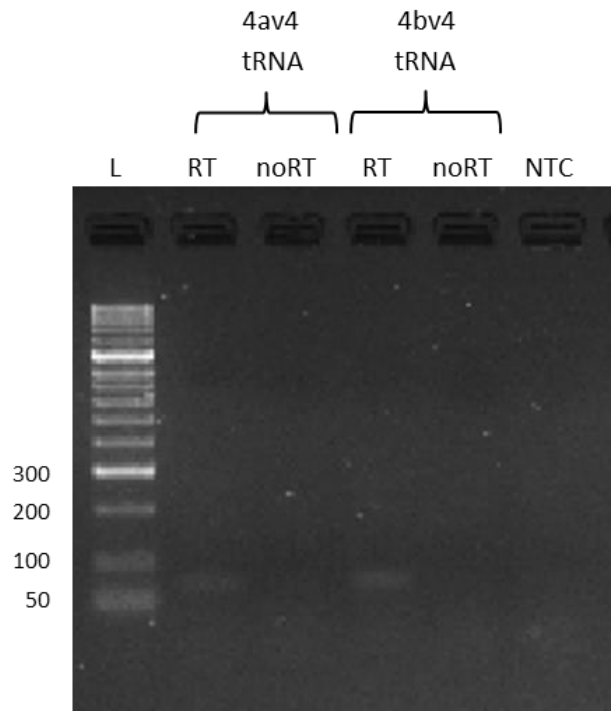
Figure 6.32. Analysis of RNA integrity post-purification.

1 % agarose gel stained with SYBR Safe dye shows results of electrophoresis of RNA samples that were extracted using RNeasy column (QIAGEN) including DNase treatment. The experiment compares heat-treated samples vs untreated, loaded in two amounts, 200 ng and 400 ng. Ladder was run to confirm correct electrophoresis procedure and not for the size reference. RNA was extracted from S. cerevisiae transformed with construct 2, 4a or 4b.

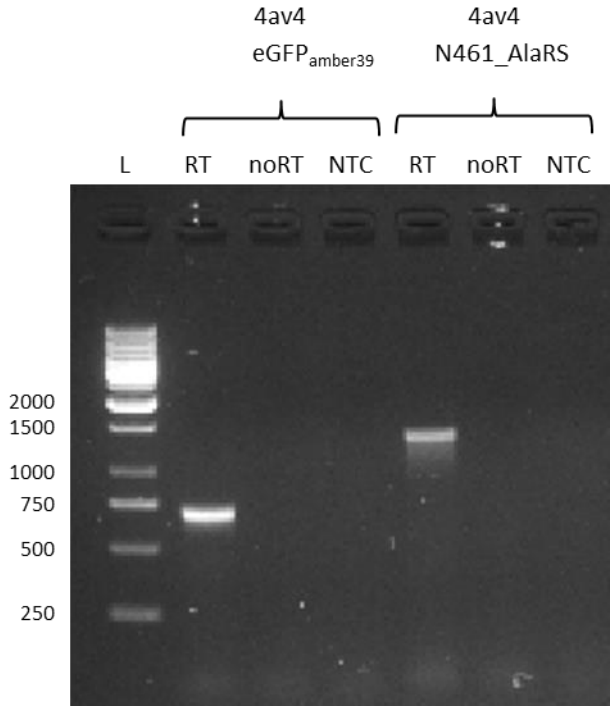
6.5.4.2. Analysis of RT-PCR results

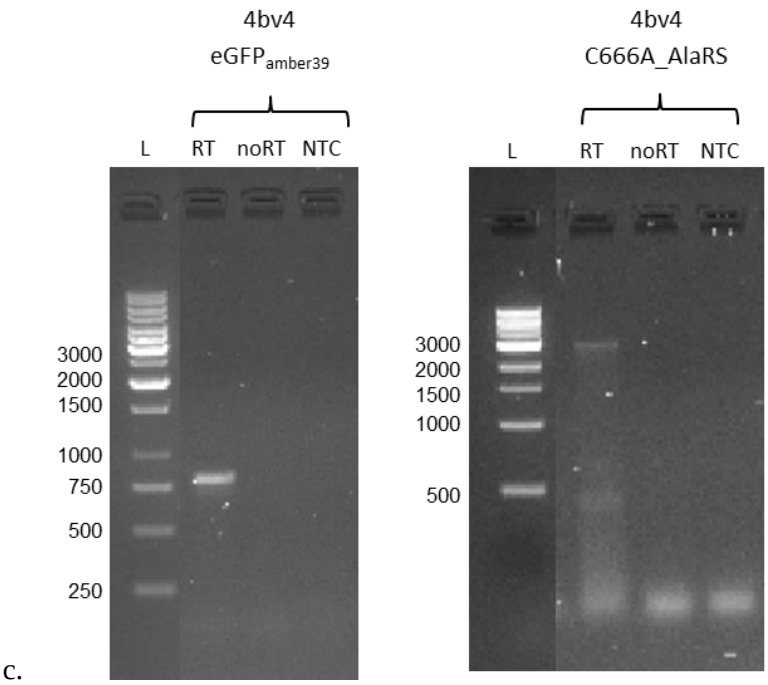
To find out if genes in construct 4av4 and 4bv4 are expressed under the control of their promoters, two-step RT-PCR was employed. One-step RT-PCR was also employed earlier (data not shown) but it was quickly changed to two-step process since reaction without reverse transcriptase (noRT) repeatedly gave a product in one-step RT-PCR, and optimisation did not improve the results. Two-step RT-PCR worked well and all noRT controls were free of product as shown in Figure 6.33. Similarly, non-template controls for step two, which was amplification of cDNA, were also free of product, as expected. Most importantly, expression of all genes present in constructs 4av4 and 4bv4 was successful in RT-PCR. $eGFP_{amber39}$ expression gave the strongest signal when analysed in agarose electrophoresis confirming that expression of this gene. The same was observed for tRNA and enzymes genes. Particularly reassuring was expression of the enzyme genes, here already separated from $eGFP_{amber39}$ unlike in previous versions of constructs 4a and 4b. Weaker signal was observed for C666A_AlaRS (Figure 6.33c.) most likely do to not completely optimised conditions for cDNA amplification of the longest of gene fragments in this experiment. Expression of tRNA gene in both constructs was also demonstrated, although not a strong signal was recorded on the gel as shorter fragments of DNA can appear blurred in electrophoresis. This is normal for tRNA gene with PCR reaction at 70 bp and the same can be observed for the lowest two bands of the ladder at 50 bp and 100 bp (Figure 6.33a.).

a.


4av4
tRNA
4bv4
tRNA
L
RT
noRT
RT
noRT
NTC
300
200
100
50

b.


4av4
eGFP amber39
4av4
N461_AlaRS
L
RT
noRT
NTC
RT
noRT
NTC
2000
1500
1000
750
500
250



c.

Figure 6.33. RT-PCR of 4av4 and 4bv4 samples in the context of gene expressions of $eGFP_{amber39}$, AlaRS and $tRNA^{Ala}$ suppressor.

3 % (a.) and 1.5 % (b., c.) agarose gels stained with SYBR Safe dye show results of electrophoresis of two-step RT-PCR reaction samples of 4av4 and 4bv4. The gene expressions that were analysed included a. $tRNA^{Ala}$ suppressor for 4av4 and 4bv4; b. $eGFP_{amber39}$ and N461_AlaRS for 4av4; c. $eGFP_{amber39}$ and C666A_AlaRS for 4bv4. noRT – negative control without reverse transcriptase enzyme. The expected RT-PCR sizes are 70 bp for tRNA, 720 bp for $eGFP_{amber39}$, 1386 bp for N461_AlaRS, and 2631 bp for C666A_AlaRS. NTC: non-template control. L: for a. HyperLadder™ 50 bp (Bioline); b. GeneRuler 1kb DNA ladder (Thermo Fisher); c. left: GeneRuler 1kb DNA ladder (Thermo Fisher) and right: HpLI: HyperLadder 1kb.

6.5.5. Western blot

Since the RT-PCR had determined that all genes were being transcribed successfully, similar samples were examined by Western blot (2.2.10.) for translation of eGFP (Figure 6.34.). No eGFP was detectable in protein samples from cells containing the $eGFP_{amber39}$ gene, regardless of other combinations of proteins. This result is consistent with the very low levels of fluorescence seen by both the fluorescence microscopy and flow cytometry analyses, where protein levels would likely be too low to be detected by Western blot.

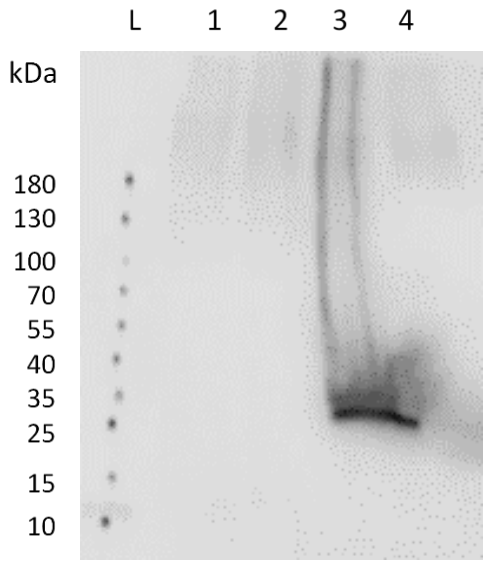


Figure 6.34. eGFP expression in S. cerevisiae BY4741 NMD2Δ.

Western blot showing whole cell lysate expression levels of eGFP in samples 1: $tRNA^{Ala}$ suppressor, N461_AlaRS, $eGFP_{amber39}$; 2: $tRNA^{Ala}$ suppressor, C666A_AlaRS, $eGFP_{amber39}$; 3: eGFP; 4: $eGFP_{amber39}$. The protein was detected in sample 3 only, at the expected size of 27 kDa. An anti-GFP Polyclonal Antibody, HRP conjugated (rabbit) (Invitrogen™) was used.

6.5.6. Discussion

A suitable screening method had to be employed in order to find out subsequently if engineering studies resulted in the generation of novel aaRSs that accept and activate UAA (D-Ala) as well as aminoacylating the tRNA. The assay was designed to be conducted in a different system than *E. coli*, so selected *E. coli* AlaRS and $tRNA^{Ala}$ would not interact with native enzymes and tRNAs from *S. cerevisiae*. Just like with previous screening methods (developed by Schultz), the assay was based on a phenotype of a reporter gene, eGFP, with a termination codon in the permissive position along with expression of *E. coli* AlaRS and a $tRNA^{Ala}$ amber suppressor. As a proof of concept, both truncated and full-length with C666A mutation AlaRS were used, since both had been previously proved effective in aminoacylating $tRNA^{Ala}$ with L-Ala (Swairjo *et al.*, 2004; Guo *et al.*, 2010).

Unfortunately, the optimisation efforts brought little effect. For instance, after applying necessary changes recommended in the manuscript "New Methods Enabling Efficient Incorporation of Unnatural Amino Acids in Yeast" by Wang and Wang (2008) including removal of CCA-3' in tRNA gene and exchange of promoter to achieve efficient expression of *E. coli* tRNA in yeast, there was nearly no difference in expressing $eGFP_{amber39}$ found in constructs 4av3 and 4bv3. Moreover, a different yeast strain with deletion of protein NMD2 which plays a critical role in Nonsense-mediated mRNA decay (NMD) pathway had little effect. Similarly, the expression of proteins in a new deletion strain, BY4741 NMD2Δ, did not meet the expected results of an efficient $eGFP_{amber39}$ expression.

Taken together, the results presented in this chapter demonstrate that the anticipated assay, developed herein, is not suitable to analyse variant AlaRS libraries since the expression of full-length eGFP from the $eGFP_{amber39}$ gene as a result of orthogonal $tRNA^{Ala}$ and AlaRS expression was not achieved. Since both genes were transcribed successfully (as shown in RT-PCR experiments), this is not a result of the constructs or their promoters, but rather must, therefore, result from lack of orthogonal functionality within the yeast cells. Moreover, the fluorescence microscopy experiments demonstrated that the low level of $eGFP_{amber39}$ expression without other components (AlaRS or

tRNA) present was achieved. Although it had not been proved, it appears that yeast was reading-through termination codon, albeit at low levels.

Chapter 7: Discussion

7.1. Summary of results

The initial aim of this project was to clone genes of AlaRS that will be used during library generation. The gene *alaS* was successfully amplified from *E. coli* Tuner™(DE3) strain and cloned into pET45b(+). From the existing construct, the two required genes were prepared. The first included the aminoacylation domain of AlaRS only (N461_AlaRS), whilst the second incorporated the full-length AlaRS with mutation C666A to remove the editing function of the enzyme (C666A_AlaRS). Once all three constructs were successfully prepared and confirmed in Sanger sequencing, expression experiments were undertaken. Expression and purification of both N461_AlaRS and C666A_AlaRS in *E. coli* were successful. Better expression was achieved for the truncated enzyme (N461_AlaRS) while the full-length with mutation (C666A_AlaRS) was also expressed at sufficient levels. The expressed proteins were used in aminoacylation reactions but the difference between acylated and nonacylated tRNA could not be confirmed unequivocally.

After cloning and expression studies, engineering studies commenced. They began with the analysis of the crystal structure of the AlaRS aminoacylation domain bound with Ala-SA and included analysis of the amino acid binding site. Those residues that were in the direct binding with the functional group of Ala were targeted for mutation. In total, six residues were randomised in two gene libraries of AlaRS using ProxiMAX randomisation. The first library involved truncated AlaRS of N-terminal domain only (N461_AlaRS) and the second library had full-length AlaRS with a knock out of editing function (C666A_AlaRS). Genes used as a template for the library build came from initial constructs in pET45b(+). The gene was divided into conserved regions and randomised regions called also library fragments. Each fragment was separately prepared in a series of amplification, digestion and ligation steps. The randomisation with 18 amino acids (excluding Met and Cys) at all six positions in three different fragments was successfully completed. The results of next-generation sequencing showed a close match of observed vs expected encoding of the required 18 amino acids

in an even distribution. Each library had over 34 million of unique gene variants of N461_AlaRS and C666A_AlaRS.

Another aim of the project was to design a novel screen for use with the libraries of aaRSs, specific to the system being developed within this project. Design included positive and negative selection. Before the screening, various controls were designed to ensure that each component of the assay was working correctly. It was proved that the short Galactose promoter (Redden and Alper, 2015) ahead of the eGFP gene prompted effective eGFP expression. However, when co-expressing $eGFP_{amber39}$, wt enzyme (truncated or full-length) and $tRNA^{Ala}$ suppressor, the phenotype of eGFP was not as green as expected, indicating that the assay was not working properly. In the efforts to make it works, a number of things were changed including different promoters for tRNA; removal of CCA from the 3'-end of tRNA; moving the gene for $eGFP_{amber39}$ into a different plasmid so that $eGFP_{amber39}$ and AlaRS genes were placed under two promoters rather than one (they were first expressed as an operon), which ultimately led to swapping of a one-plasmid system for a two-plasmid system, and finally, use of different yeast strains. Nevertheless, the applied changes did not produce desired results and the most that was achieved was low expression of $eGFP_{amber39}$. Moreover, under fluorescence microscopy analysis it was observed that the level of expression was very similar when $eGFP_{amber39}$ was expressed alone (i.e. without *E. coli* $tRNA^{Ala}$ suppressor or *E. coli* AlaRS). Therefore, it could not be concluded that the low level of expression was due to the wt *E. coli* AlaRS aminoacylating the $tRNA^{Ala}$ suppressor and adding amino acid in the stop codon inserted at 39^{th} position of $eGFP_{amber39}$. To confirm the observations from microscopy, the samples were analysed in the flow cytometer. The experiment revealed similar findings. To further investigate the low level of $eGFP_{amber39}$ expression, RT-PCR confirmed that all genes, tRNA, $eGFP_{amber39}$, enzyme were expressed, though eGFP levels were too low to be detected by Western blot, except in the positive control.

7.2. Future directions

7.2.1. Towards a novel *in vitro* translation system

The system for which this project anticipates to provide a fundamental background would be used in the expansion of genetic code to include multiple D-forms of amino acids. The main application of this novel system would be to encode multiple D-AAs or indeed other UAAs in preparation of gene libraries using nondegenerate saturation mutagenesis – specifically, via ProxiMAX randomisation, since the current method of ProxiMAX selects only 20 codons to cover 20 amino acids (Ashraf *et al.*, 2013) and thus many codons could be used to encode UAAs simultaneously.

As such, the core part of this project was to create a library of AlaRSs that could be screened repeatedly with different D-AAs and even other UAAs. Resulting *in vitro* translation would require engineering up to 19 aminoacyl-tRNA synthetases and cognate tRNAs based on *E. coli* AlaRS and *E. coli* $tRNA^{Ala}$ native molecules. Ultimately, within a fully-synthetic system, any twelve of these could be used in translating a synthetic gene that used only 20 codons for the native amino acids in all positions. These numbers result from the rules governing wobble in the third position of the codon, further developed from Crick's original rules by Murphy and Ramakrishnan (2004). When no base modifications are permitted (because the tRNA is transcribed *in vitro*), the codon-anticodon wobble rules allow base pairing between anticodon•codon: A•U/C/G, C•G, G•C/U and U•A/G/U in the first position of the anticodon and third position of the codon, respectively.

While the library was constructed successfully, the developed screen was found wanting. Thus, although the ambition of the project was to set the foundation for a novel system, it may be more prudent to reduce future ambition and to use enzymes that have already been produced elsewhere. For example, Schultz has produced many variants of the PylRS that many incorporated UAAs (Young and Schultz, 2018). PylRS from *M. barkeri*, like *E. coli* AlaRS, also ignores the anticodon of the tRNA (Wan *et al.*, 2014) and therefore potentially, the project could in future simply use these

already developed aaRSs to aminoacylate appropriately mutated tRNAs (mutated at the anticodon only) *in vitro*, as originally envisaged at the beginning of the project. Such a system could further allow the use of native *E. coli* aaRSs to re-charge native *E. coli* tRNAs if required since the mutant PylRSs and cognate tRNAs should be orthogonal to the *E. coli* system.

7.2.1.1. Pyrrolysyl-tRNA synthetase and $tRNA_{CUA}$ pair

Pyrrolysine (Pyl) is found in nature as 22nd amino acid (selenocysteine, Sec being 21st) and is encoded by UAG codon, normally known as amber termination codon (Krzycki, 2005). Methanogens are archaeal microorganisms which utilise carbon dioxide and convert it into methane during anaerobic respiration in the process called methanogenesis. During formation of methane, genes associated with the process encode essential Pyl by conserved in-frame amber termination codon (Soares *et al.*, 2005). These genes have been found in *M. barkeri*, *Methanosarcina mazei*, *Methanosarcina acetivorans* and *Methanococcoides burtonii*, and Pyl is present in *Methanosarcinacea* and Gram-positive *Desulfitobacterium hafniense* (Paul, Ferguson, and Krzycki, 2000; Krzycki, 2005). While Sel cannot be found as free metabolite and instead is formed directly on tRNA, Pyl resembles other 20 amino acids with cognate aaRS and tRNA. It was confirmed in *in vitro* and *in vivo* studies that Pyl has its cognate archaeal class II Pyrrolysyl-tRNA synthetase (PylRS) and is aminoacylated onto tRNA.

PylRS does not activate natural amino acids and does not aminoacylate tRNAs other than $tRNA_{CUA}$, making PylRS and $tRNA_{CUA}$ an orthogonal pair in *E. coli* and other tested eukaryotic cells. Therefore, if there is a requirement to expand genetic code for pyrrolysine, an expression of only two genes, *pylS* and *pylT* for PylRS and $tRNA_{CUA}$, respectively, is needed (Blight *et al.*, 2004). However, for further expansion of genetic code and incorporation of UAAs into the proteins, many studies investigated *M. barkeri* PylRS (MbPylRS) variants of the active site to use in different organisms, including common *E. coli* (Elliott *et al.*, 2014). While amber termination codon can be used to encode

UAA, the enzyme also has little specificity for anticodon of $tRNA_{CUA}$, giving the possibility for expansion.

7.2.2. *In vitro* transcription/translation

The *in vitro* translation or transcription/translation system supports the production of the protein in a cell-free environment. This is an alternative and convenient way of protein production when overexpression of the product becomes toxic to the organism, or when it goes into insoluble fractions making it difficult to purify, or more importantly when the protein function is affected due to degradation via proteolytic lysis ("The Basics: In Vitro Translation.," no date). The extension of this *in vitro* system is far beyond being only an alternative to protein synthesis. It can be employed for using cognate aaRSs and tRNAs pairs during *in vitro* translation when the pair cannot be used in the eukaryotic or prokaryotic system due to interaction with the native system. Hence, and this is parallel with the interest of this project in terms of a long term goal, *in vitro* translation or transcription/translation system can be used to incorporate UAAs in proteins.

7.2.3. *In vitro* translation in UAA incorporation

Cell-free protein translation technologies are often chosen to introduce UAA in proteins. This was firstly described by Schultz and his co-workers in 1989 (Noren, Anthony-Cahill, Griffith, and Schultz, 1989). They used *E. coli* extracts to successfully introduce UAAs in beta-lactamase by replacing Phe66 with an amber stop codon within the gene and supplying chemically aminoacylated tRNA suppressor with UAA. The suppressor tRNA used in this study was *S. cerevisiae* $tRNA^{Phe}$ with anticodon mutation to recognise amber stop codon in the beta-lactamase gene. From the earlier studies, it was confirmed that *S. cerevisiae* $tRNA^{Phe}$ will not interact with endogenous *E. coli* PheRS found in *E. coli* extracts. Kwok and Wong (1980) proved that less than 1 % of *S. cerevisiae* $tRNA^{Phe}$ can be acylated by *E. coli* PheRS, hence Schultz's choice. Incorporating UAA into proteins became even easier when the reconstituted cell-free technologies emerged. These systems enabled the users to pick and choose what components they require for *in vitro* translation. For example, the

reconstituted cell-free system allowed for the replacement of endogenous *E. coli* synthetases for engineered enzymes or enzymes from different species. It also enabled to manipulate with the set of tRNAs and to introduce orthogonal tRNA and aaRS pairs (Forster, Cornish, and Blacklow, 2004) or pre-charged tRNA (Forster *et al.*, 2003). Using *in vitro* translation via cell-free methods for incorporation of UAA into proteins have been developing for decades and thanks to the progression, proteins and peptides with UAAs can now be produced for various purposes (Chong, 2014).

In order to translate *in vitro*, extracts need to have all necessary components that participate in translation and these are: tRNAs, aaRSs, 70S or 80S ribosomes, initiation/elongation/termination factors. However, this component list is not exhausted and the user has to supply amino acids, ATP and/or GTP as the energy source, an energy regenerating system (different for prokaryotes – phosphoenol pyruvate and pyruvate kinase, and for eukaryotes – creatine phosphate and creatine phosphokinase), as well as other co-factors (for example K^+, Mg^{2+}). The difference between *in vitro* translation and transcription/translation system is used RNA and DNA template, respectively (Chong, 2014; "The Basics: In Vitro Translation.," no date). However, the system envisaged in this project would be much simplified. Because tRNAs will be pre-charged, neither aaRSs nor the energy system required to regenerate them are needed, leaving (in addition to the pre-charged tRNAs) just ribosomes and associated factors, GTP for ribosomal translocation and the required mRNA encoding the protein, or more likely, the peptide of interest.

In vitro systems based on extracts from *E. coli* are more suitable for coupled *in vitro* transcription/translation. This is because exogenous mRNA is targeted by endogenous nucleases causing its degradation. The translation machinery from *E. coli* is also less complicated and complex to eukaryotic. The *E. coli* cell-free system is available as reconstituted purified *E. coli* system PURExpress® *In vitro* Protein Synthesis Kit (NEB).

7.2.4. Comparison between the proposed system and current methods for incorporating multiple UAAs

Current methods that aim to incorporate two UAAs *in vivo* or *in vitro* involve either using two termination codons, one termination (amber) and one quadruplet codon, or finally, two quadruplet codons. Quadruplet codons and amber codons are read from orthogonal mRNA by specially engineered orthogonal ribosome ribo-Q1 (Neumann *et al.*, 2010). As an example, mutually orthogonal pairs were developed, MbPylRS-$tRNA_{CUA}$ and AzPheRS-$tRNA_{UCCU}$. AzPheRS was engineered based on well-known MjTyrRS and its cognate $tRNA_{CUA}$ which was changed to accommodate quadruplet anticodon as $tRNA_{UCCU}$. MbPylRS was found to be specific for its substrate, N6-[(2-propynyloxy)carbonyl]-L-lysine (CAK). These two pairs were designed to incorporate two different UAAs encoded by different codons and thanks to their orthogonality, the aminoacylation was specific. As a result, the protein GST-calmodulin included two UAAs, p-azido-L-phenylalanine (AzPhe) (by AzPheRS-$tRNA_{UCCU}$ pair) and CAK (by MbPlyRS-$tRNA_{CUA}$). The protein was produced in *E. coli* (Neumann *et al.*, 2010).

The development of technology using quadruplet codons and ribosome ribo-Q1 is complex and requires engineering of different components. The advantage of this method is that, in theory, there can be up to 256 quadruplet codons available to encode UAAs (Neumann *et al.*, 2010). Alternatively, a much simpler approach is to engineer variants of PylRS and $tRNA_{CUA}$ pairs and to assign codons unused in ProxiMAX to UAAs of interest, what could be an extension of this project. Since PylRS has little specificity in recognition of anticodon, the native anticodon recognising amber suppression can be changed to accommodate others. This method would avoid a requirement for the development of a synthetic ribosome (Neumann *et al.*, 2010). It would also focus on engineering only one of the aaRSs, preferably PylRS, rather than trying to find mutually orthogonal pairs (an example above). As a development of this project, multiple UAAs would be inserted in the synthetic peptide or protein via pre-charged tRNAs during *in vitro* translation. Aminoacylation of tRNAs with distinct anticodons would be isolated and independent of each other and therefore, mutually orthogonal aaRS-tRNA

pairs would not be needed. PylRS-$tRNA_{CUA}$ pair was already successfully coupled with various codons (Wan *et al.*, 2014) demonstrating a sought flexibility for future directions described.

Commonly used amber termination codon for insertion of UAAs into proteins permits incorporation of only one UAA at a time without competition from natural AAs. Using one codon is considered here as a limitation since the project aimed to create a system that will allow for incorporation of multiple UAAs simultaneously. Only one UAA insertion is common when, for example, MbPylRS-$tRNA_{CUA}$ and MjTyrRS-$tRNA^{Tyr}$ are used (Vargas-Rodriguez *et al.*, 2018). The advantage is that these tRNAs can have mutations in their anticodon, but when used in eukaryotes or prokaryotes, not much is possible in terms of codon reassignment unless the whole genome is synthetic. This is very challenging and time-consuming pointing out that cell-free protein synthesis and using pre-charged tRNAs is more realistic to achieve. Therefore, engineering MbPylRS or MjTyrRS for acceptance of D-AAs and then aminoacylating tRNAs (with mutated anticodons) is preferable in cell-free protein synthesis.

Nevertheless, *in vivo* methods have their advantages over cell-free systems for the production of proteins with UAAs. For example, if it is necessary to produce a protein with UAA in large quantities, *in vivo* production would achieve far higher yields than cell-free systems which are not suitable for large scale production. In addition, while cell-free systems permit to use more chemically complex amino acids, *in vivo* methods are more comprehensive because they enable, for example, functional assessments of engineered protein in the natural biological environment (Passioura and Suga, 2014). Moreover, using *in vivo* approaches for incorporation of even only one UAA via orthogonal engineered aaRS-tRNA pair allows for recycling of tRNAs whereas when pre-charged tRNAs are employed, they are discarded once AA is delivered.

In terms of methods engaging pre-charged tRNAs, chemical aminoacylation is challenging and others, more convenient methods are available. Flexizyme system offers high-throughput acylation

of tRNAs via engineered ribozymes. Prefix ‘flexi’ defines the flexible approach of the system towards various tRNAs and UAAs. This technology offers a lot of advantages including mutation of anticodon because flexizymes require only conserved CCA-3’ of tRNA in the reaction of aminoacylation (Passioura and Suga, 2014). There are three flexizymes available, with different specificity for UAAs. When comparing to the proposed system, Flexizyme technology has great flexibility towards different tRNAs and UAAs what is a substantial advantage, as such plasticity will never be achieved with a single engineered aaRS. Flexizyme is also capable of aminoacylating tRNAs with D-AAs (Katoh, Tajima, and Suga, 2017).

The aim of introducing D-AAs into proteins in the cell-free system via pre-acylation of tRNA by engineered aaRSs and tRNAs must consider the translation machinery in such system. D-AAs are stereochemically different from L-AAs, and since the latter are naturally occurring, they will be favoured by the translational machinery. While favouring is not an issue because there is no competition when amino acids are codon-assigned, however, if the components such as ribosomes or factors associated with translation discriminate against tRNAs charged with non-natural AAs, this can become problematic for the whole process. It was reported before that while in the effort of introducing D-AAs via amber termination codon *in vivo*, D-AA-tRNAs were poorly accepted (less than 10 % when compared with canonical L-AAs) by elongation factors of the translation machinery. Subsequently, the newer investigation revealed that the efficiency of D-AAs incorporation varies and results demonstrated that eight D-AAs achieve over 40 % efficiency, four D-AAs between 10 and 40 %, and the remaining seven D-AAs, indeed less than 10 % what is considered as poor efficiency (Passioura and Suga, 2014). Therefore, to create a technology for the incorporation of UAAs including D-AAs, it must be considered if they are compatible with the system and if they are not, perhaps, engineered ribosomes and/or associated factors have to be included into the cell-free system.

Nevertheless, while incorporating D-AAs into the proteins still remains challenging with currently available *in vivo*, *in vitro* or cell-free systems, there is still space for discovery and development of novel technology that would include improved incorporation of D-AAs at multiple positions.

7.3. Conclusion

Two very high-quality libraries were produced each encoding over 34 million unique *E. coli* AlaRS variants as determined by NGS. However, the development of a library screening protocol remained challenging and served to identify more straightforward (albeit less commercially viable) alternatives to achieve the main goal of the larger project.

Chapter 8: References

Altman, S. (1975). Biosynthesis of transfer RNA in Escherichia coli. *Cell, 4*(1), 21-29.

Ashraf, M., Frigotto, L., Smith, M. E., Patel, S., Hughes, M. D., Poole, A. J., Hebaishi, H. R., Ullman, C. G., and Hine, A. V. (2013). ProxiMAX randomization: a new technology for non-degenerate saturation mutagenesis of contiguous codons. *Biochemical Society Transactions, 41*(5), 1189-1194.

Barciszewska, M. Z., Perrigue, P. M., and Barciszewski, J. (2016). tRNA--the golden standard in molecular biology. *Molecular Biosystems, 12*(1), 12-17.

The Basics: In Vitro Translation. (no date). Retrieved from https://www.thermofisher.com/uk/en/home/references/ambion-tech-support/large-scale-transcription/general-articles/the-basics-in-vitro-translation.html#7

Beebe, K., Mock, M., Merriman, E., and Schimmel, P. (2008). Distinct domains of tRNA synthetase recognize the same base pair. *Nature, 451*(7174), 90-93.

Berg, P. (1956). Acyl adenylates; the interaction of adenosine triphosphate and L-methionine. *The Journal of Biological Chemistry, 222*(2), 1025-1034.

Beuning, P. J., Gulotta, M., and Musier-Forsyth, K. (1997). Atomic Group “Mutagenesis” Reveals Major Groove Fine Interactions of a tRNA Synthetase with an RNA Helix. *Journal of the American Chemical Society, 119*(36), 8397-8402.

Beuning, P. J., and Musier-Forsyth, K. (1999). Transfer RNA recognition by aminoacyl-tRNA synthetases. *Biopolymers, 52*(1), 1-28.

Biotech, GSL. SnapGene software, available at snapgene.com.

Blight, S. K., Larue, R. C., Mahapatra, A., Longstaff, D. G., Chang, E., Zhao, G., Kang, P. T., Green-Church, K. B., Chan, M. K., and Krzycki, J. A. (2004). Direct charging of tRNA(CUA) with pyrrolysine in vitro and in vivo. *Nature, 431*(7006), 333-335.

Cavarelli, J., and Moras, D. (1993). Recognition of tRNAs by aminoacyl-tRNA synthetases. *FASEB Journal, 7*(1), 79-86.

Chang, K. Y., Varani, G., Bhattacharya, S., Choi, H., and McClain, W. H. (1999). Correlation of deformability at a tRNA recognition site and aminoacylation specificity.

Proceedings of the National Academy of Sciences of the United States of America, 96(21), 11764-11769.

Chong, S. (2014). Overview of cell-free protein synthesis: historic landmarks, commercial systems, and expanding applications. *Current Protocols in Molecular Biology, 108*, 16 30 11-11.

Chou, C. C., Patel, M. T., and Gartenberg, M. R. (2015). A series of conditional shuttle vectors for targeted genomic integration in budding yeast. *FEMS Yeast Research, 15*(3).

Crick, F. H. (1958). On protein synthesis. *Symposia of the Society for Experimental Biology, 12*, 138-163.

Day, R. N., and Davidson, M. W. (2009). The fluorescent protein palette: tools for cellular imaging. *Chemical Society Reviews, 38*(10), 2887-2921.

Demeshkina, N., Jenner, L., Westhof, E., Yusupov, M., and Yusupova, G. (2012). A new understanding of the decoding principle on the ribosome. *Nature, 484*(7393), 256-259.

Dignam, J. D., Guo, J., Griffith, W. P., Garbett, N. C., Holloway, A., and Mueser, T. (2011). Allosteric interaction of nucleotides and tRNA(ala) with *E. coli* alanyl-tRNA synthetase. *Biochemistry, 50*(45), 9886-9900.

Elliott, T. S., Bianco, A., and Chin, J. W. (2014). Genetic code expansion and bioorthogonal labelling enables cell specific proteomics in an animal. *Current Opinion in Chemical Biology, 21*, 154-160.

Eun, H.-M. (1996). 1 - Enzymes and Nucleic Acids: General Principles. In H.-M. Eun (Ed.), *Enzymology Primer for Recombinant DNA Technology* (pp. 1-108). San Diego: Academic Press.

Ferreira Amaral, M. M., Frigotto, L., and Hine, A. V. (2017). Beyond the Natural Proteome: Nondegenerate Saturation Mutagenesis-Methodologies and Advantages. *Methods in Enzymology, 585*, 111-133.

Fischer, A. E., Beuning, P. J., and Musier-Forsyth, K. (1999). Identification of Discriminator Base Atomic Groups That Modulate the Alanine Aminoacylation Reaction. *The Journal of Biological Chemistry, 274*(52), 37093-37096.

Forster, A. C., Cornish, V. W., and Blacklow, S. C. (2004). Pure translation display. *Analytical Biochemistry, 333*(2), 358-364.

Forster, A. C., Tan, Z., Nalam, M. N., Lin, H., Qu, H., Cornish, V. W., and Blacklow, S. C. (2003). Programming peptidomimetic syntheses by translating genetic codes designed de novo. *Proceedings of the National Academy of Sciences of the United States of America, 100*(11), 6353-6357.

Francklyn, C., and Schimmel, P. (1989). Aminoacylation of RNA minihelices with alanine. *Nature, 337*(6206), 478-481.

Francklyn, C., Shi, J. P., and Schimmel, P. (1992). Overlapping nucleotide determinants for specific aminoacylation of RNA microhelices. *Science, 255*(5048), 1121-1125.

Frigotto, L., Smith, M., Brankin, C., Sedani, A., Cooper, S., Kanwar, N., Evans, D., Svobodova, S., Baar, C., Glanville, J., Ullman, C., and Hine, A. (2015). Codon-Precise, Synthetic, Antibody Fragment Libraries Built Using Automated Hexamer Codon Additions and Validated through Next Generation Sequencing. *Antibodies, 4*(2), 88-102.

Guffanti, E., Ferrari, R., Preti, M., Forloni, M., Harismendy, O., Lefebvre, O., and Dieci, G. (2006). A minimal promoter for TFIIIC-dependent in vitro transcription of snoRNA and tRNA genes by RNA polymerase III. *The Journal of Biological Chemistry, 281*(33), 23945-23957.

Guo, M., Chong, Y. E., Shapiro, R., Beebe, K., Yang, X. L., and Schimmel, P. (2009). Paradox of mistranslation of serine for alanine caused by AlaRS recognition dilemma. *Nature, 462*(7274), 808-812.

Guo, M., Shapiro, R., Schimmel, P., and Yang, X. L. (2010). Introduction of a leucine half-zipper engenders multiple high-quality crystals of a recalcitrant tRNA synthetase. *Acta Crystallographica Section D Biological Crystallography, 66*(Pt 3), 243-250.

Hanes, J., and Plückthun, A. (1997). In vitro selection and evolution of functional proteins by using ribosome display. *Proceedings of the National Academy of Sciences of the United States of America, 94*(10), 4937-4942.

Hecht, S. M., Alford, B. L., Kuroda, Y., and Kitano, S. (1978). "Chemical aminoacylation" of tRNA's. *The Journal of Biological Chemistry, 253*(13), 4517-4520.

Heim, R., Cubitt, A. B., and Tsien, R. Y. (1995). Improved green fluorescence. *Nature, 373*(6516), 663-664.

Heim, R., Prasher, D. C., and Tsien, R. Y. (1994). Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proceedings of the National Academy of Sciences of the United States of America, 91*(26), 12501-12504.

Hill, K., and Schimmel, P. (1989). Evidence that the 3' end of a tRNA binds to a site in the adenylate synthesis domain of an aminoacyl-tRNA synthetase. *Biochemistry, 28*(6), 2577-2586. doi:10.1021/bi00432a035

Ho, S. P., Britton, D. H., Stone, B. A., Behrens, D. L., Leffet, L. M., Hobbs, F. W., Miller, J. A., and Trainor, G. L. (1996). Potent antisense oligonucleotides to the human multidrug resistance-1 mRNA are rationally selected by mapping RNA-accessible sites with oligonucleotide libraries. *Nucleic Acids Research, 24*(10), 1901-1907.

Hoagland , M. B. (1955). An enzymic mechanism for amino acid activation in animal tissues. *Biochimica et Biophysica Acta*, *16*(2), 288-9.

Hoagland, M. B., Stephenson, M. L., Scott, J. F., Hecht, L. I., and Zamecnik, P. C. (1958). A soluble ribonucleic acid intermediate in protein synthesis. *The Journal of Biological Chemistry, 231*(1), 241-257.

Holley, R. W. (1963). Large-scale preparation of yeast "soluble" ribonucleic acid. *Biochemical and Biophysical Research Communications, 10*, 186-188.

Holley, R. W. (1968). Alanine transfer RNA (Nobel lecture). Retrieved from https://www.nobelprize.org/uploads/2018/06/holley-lecture.pdf

Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Marquisee, M., Merrill, S. H., Penswick, J. R., and Zamir, A. (1965). Structure of a ribonuclei acid. *Science, 147*(3664), 1462-1465.

Hou, Y. M., and Schimmel, P. (1988). A simple structural feature is a major determinant of the identity of a transfer RNA. *Nature, 333*(6169), 140-145.

Hou, Y. M., and Schimmel, P. (1989). Evidence that a major determinant for the identity of a transfer RNA is conserved in evolution. *Biochemistry, 28*(17), 6800-6804.

Hughes, M. D., Nagel, D. A., Santos, A. F., Sutherland, A. J., and Hine, A. V. (2003). Removing the Redundancy From Randomised Gene Libraries. *Journal of Molecular Biology, 331*(5), 973-979.

Kaishima, M., Ishii, J., Matsuno, T., Fukuda, N., and Kondo, A. (2016). Expression of varied GFPs in Saccharomyces cerevisiae: codon optimization yields stronger than expected expression and fluorescence intensity. *Scientific Reports, 6*, 35932.

Katoh, T., Tajima, K., and Suga, H. (2017). Consecutive Elongation of D-Amino Acids in Translation. *Cell Chem Biol, 24*(1), 46-54. doi:10.1016/j.chembiol.2016.11.012

Kille, S., Acevedo-Rocha, C. G., Parra, L. P., Zhang, Z. G., Opperman, D. J., Reetz, M. T., and Acevedo, J. P. (2013). Reducing codon redundancy and screening effort of

combinatorial protein libraries created by saturation mutagenesis. *ACS Synthetic Biology, 2*(2), 83-92.

Kim, S. H., Quigley, G. J., Suddath, F. L., McPherson, A., Sneden, D., Kim, J. J., Weinzierl, J., and Rich, A. (1973). Three-dimensional structure of yeast phenylalanine transfer RNA: folding of the polynucleotide chain. *Science, 179*(4070), 285-288.

Kirchner, S. and Ignatova, Z. (2014). Emerging roles of tRNA in adaptive translation, signalling dynamics and disease. *Nature Reviews Genetics, 16*(2), 98–112.

Krzycki, J. A. (2005). The direct genetic encoding of pyrrolysine. *Current Opinion in Microbiology, 8*(6), 706-712.

Kuhn, S. M., Rubini, M., Fuhrmann, M., Theobald, I., and Skerra, A. (2010). Engineering of an Orthogonal Aminoacyl-tRNA Synthetase for Efficient Incorporation of the Non-natural Amino Acid O-Methyl-L-tyrosine using Fluorescence-based Bacterial Cell Sorting. *Journal of Molecular Biology, 404*(1), 70-87.

Kwok, Y., and Wong, J. T. (1980). Evolutionary relationship between Halobacterium cutirubrum and eukaryotes determined by use of aminoacyl-tRNA synthetases as phylogenetic probes. *Canadian Journal of Biochemistry, 58*(3), 213-218.

Lang, Y., Zhang, Y., Zhan, L., Feng, Z., Zhou, X., Yu, M., and Mo, W. (2014). Expression, purification, and characterization of rhTyrRS. *BMC Biotechnology, 14*, 64.

Lee, N., Bessho, Y., Wei, K., Szostak, J. W., and Suga, H. (2000). Ribozyme-catalyzed tRNA aminoacylation. *Nature Structural Biology, 7*(1), 28-33.

Liu, D. R., and Schultz, P. G. (1999). Progress toward the evolution of an organism with an expanded genetic code. *Proceedings of the National Academy of Sciences of the United States of America, 96*(9), 4780-4785.

Lorenz, C., Lunse, C. E., and Morl, M. (2017). tRNA Modifications: Impact on Structure and Thermal Adaptation. *Biomolecules, 7*(2).

Maršavelski, A., Lesjak, S., Močibob, M., Weygand-Đurašević, I., and Tomić, S. (2014). A single amino acid substitution affects the substrate specificity of the seryl-tRNA synthetase homologue. *Molecular Biosystems, 10*(12), 3207-3216.

Melancon, C. E., 3rd, and Schultz, P. G. (2009). One plasmid selection system for the rapid evolution of aminoacyl-tRNA synthetases. *Bioorganic and Medicinal Chemistry Letters, 19*(14), 3845-3847.

Murakami, H., Ohta, A., Ashigai, H., and Suga, H. (2006). A highly flexible tRNA acylation method for non-natural polypeptide synthesis. *Nature Methods, 3*(5), 357-359.

Murphy, F. V. t., and Ramakrishnan, V. (2004). Structure of a purine-purine wobble base pair in the decoding center of the ribosome. *Nature Structural & Molecular Biology, 11*(12), 1251-1252.

Musier-Forsyth, K., and Schimmel, P. (1999). Atomic Determinants for Aminoacylation of RNA Minihelices and Relationship to Genetic Code. *Accounts of Chemical Research, 32*(4), 368-375.

Nagai, T., Ibata, K., Park, E. S., Kubota, M., Mikoshiba, K., and Miyawaki, A. (2002). A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nature Biotechnology, 20*(1), 87-90.

Naganuma, M., Sekine, S., Chong, Y. E., Guo, M., Yang, X. L., Gamper, H., Hou, Y. M., Schimmel, P., and Yokoyama, S. (2014). The selective tRNA aminoacylation mechanism based on a single G*U pair. *Nature, 510*(7506), 507-511.

Naganuma, M., Sekine, S., Fukunaga, R., and Yokoyama, S. (2009). Unique protein architecture of alanyl-tRNA synthetase for aminoacylation, editing, and dimerization. *Proceedings of the National Academy of Sciences of the United States of America, 106*(21), 8489-8494.

Neumann, H., Wang, K., Davis, L., Garcia-Alai, M., and Chin, J. W. (2010). Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature, 464*(7287), 441-444.

Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C., and Schultz, P. G. (1989). A general method for site-specific incorporation of unnatural amino acids into proteins. *Science, 244*(4901), 182-188.

Odegrip, R., Coomber, D., Eldridge, B., Hederer, R., Kuhlman, P. A., Ullman, C., FitzGerald, K., and McGregor, D. (2004). CIS display: In vitro selection of peptides from libraries of protein-DNA complexes. *Proceedings of the National Academy of Sciences of the United States of America, 101*(9), 2806-2810.

Ohuchi, M., Murakami, H., and Suga, H. (2007a). The flexizyme system: a highly flexible tRNA aminoacylation tool for the translation apparatus. *Current Opinion in Chemical Biology, 11*(5), 537-542.

Ohuchi, M., Murakami, H., and Suga, H. (2007b). The flexizyme system: a highly flexible tRNA aminoacylation tool for the translation apparatus. *Current Opinion in Chemical Biology, 11*(5), 537-542.

Park, E. J., Yoon, S., Oh, E. J., Hong, S. H., Kim, C. G., and Park, S. D. (1996). A simplified electrophoresis method for analyses of high molecular weight RNA. *Molecules and Cells, 6*(1), 23-26.

Passioura, T., and Suga, H. (2014). Reprogramming the genetic code in vitro. *Trends in Biochemical Sciences, 39*(9), 400-408.

Pastrnak, M., Magliery, T. J., and Schultz, P. G. (2000). A New Orthogonal Suppressor tRNA/Aminoacyl-tRNA Synthetase Pair for Evolving an Organism with an Expanded Genetic Code. *Helvetica Chimica Acta, 83*(9), 2277-2286.

Pastrnak, M., and Schultz, P. G. (2001). Phage selection for site-specific incorporation of unnatural amino acids into proteins in vivo. *Bioorganic and Medicinal Chemistry, 9*(9), 2373-2379.

Paul, L., Ferguson, D. J., Jr., and Krzycki, J. A. (2000). The trimethylamine methyltransferase gene and multiple dimethylamine methyltransferase genes of Methanosarcina barkeri contain in-frame and read-through amber codons. *Journal of Bacteriology, 182*(9), 2520-2529.

Perona, J. J., and Hadd, A. (2012). Structural diversity and protein engineering of the aminoacyl-tRNA synthetases. *Biochemistry, 51*(44), 8705-8729.

Poole, A. J. (2015). *The integration of ProxiMAX randomisation with CIS display for the production of novel peptides.* (PhD). Aston University, Birmingham, UK. Retrieved from https://research.aston.ac.uk/en/studentTheses/the-integration-of-proximax-randomisation-with-cis-display-for-th

RajBhandary, U. L., Stuart, A., Faulkner, R. D., Chang, S. H., and Khorana, H. G. (1966). Nucleotide sequence studies on yeast phenylalanine sRNA. *Cold Spring Harbor Symposia on Quantitative Biology, 31*, 425-434.

Ramaswamy, K., Saito, H., Murakami, H., Shiba, K., and Suga, H. (2004). Designer ribozymes: programming the tRNA specificity into flexizyme. *Journal of the American Chemical Society, 126*(37), 11454-11455.

Redden, H., and Alper, H. S. (2015). The development and characterization of synthetic minimal yeast promoters. *Nature Communications, 6*(1), 7810.

Robertson, S. A., Ellman, J. A., and Schultz, P. G. (1991). A general and efficient route for chemical aminoacylation of transfer RNAs. *Journal of the American Chemical Society, 113*(7), 2722-2729.

Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J. C., and Moras, D. (1991). Class II aminoacyl transfer RNA

synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA(Asp). *Science, 252*(5013), 1682-1689.

Santoro, S. W., Wang, L., Herberich, B., King, D. S., and Schultz, P. G. (2002). An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. *Nature Biotechnology, 20*(10), 1044-1048.

Schatz, M. C., Delcher, A. L., and Salzberg, S. L. (2010). Assembly of large genomes using second-generation sequencing. *Genome Research, 20*(9), 1165-1173.

Schiestl, R. H., and Gietz, R. D. (1989). High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. *Current Genetics, 16*(5-6), 339-346.

Shimomura, O., Johnson, F. H., and Saiga, Y. (1962). Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea. *Journal of Cellular and Comperative Physiology, 59*, 223-239.

Soares, J. A., Zhang, L., Pitsch, R. L., Kleinholz, N. M., Jones, R. B., Wolff, J. J., Amster, J., Green-Church, K. B., and Krzycki, J. A. (2005). The residue mass of L-pyrrolysine in three distinct methylamine methyltransferases. *The Journal of Biological Chemistry 280*(44), 36962-36969.

Sood, S. M., Hill, K. A., and Slattery, C. W. (1997). Stability of Escherichia coli alanyl-tRNA synthetase quaternary structure under increased pressure. *Archive of Biochemistry and Biophysics, 346*(2), 322-323.

Sood, S. M., Slattery, C. W., Filley, S. J., Wu, M. X., and Hill, K. A. (1996). Further characterization of Escherichia coli alanyl-tRNA synthetase. *Archive of Biochemistry and Biophysics, 328*(2), 295-301.

Strazewski, P., Biala, E., Gabriel, K., and McClain, W. H. (1999). The relationship of thermodynamic stability at a G x U recognition site to tRNA aminoacylation specificity. *RNA, 5*(11), 1490-1494.

Swairjo, M. A., Otero, F. J., Yang, X. L., Lovato, M. A., Skene, R. J., McRee, D. E., Ribas de Pouplana, L., and Schimmel, P. (2004). Alanyl-tRNA synthetase crystal structure and design for acceptor-stem recognition. *Molecular Cell, 13*(6), 829-841.

Tamura, K., and Hasegawa, T. (1999). Relationship of the CCA sequence of tRNA with the early evolutional aspect of aminoacyl-tRNA synthetases. *Nucleic Acids Symposium Series*(42), 211-212.

Tang, L., Gao, H., Zhu, X., Wang, X., Zhou, M., and Jiang, R. (2012). Construction of "small-intelligent" focused mutagenesis libraries using well-designed combinatorial degenerate primers. *BioTechniques, 52*(3), 149-158.

Toh, Y., Hori, H., Tomita, K., Ueda, T., and Watanabe, K. (2009). Transfer RNA Synthesis and Regulation.

Topal, M. D., and Fresco, J. R. (1976). Base pairing and fidelity in codon-anticodon interaction. *Nature, 263*(5575), 289-293.

Van den Brulle, J., Fischer, M., Langmann, T., Horn, G., Waldmann, T., Arnold, S., Fuhrmann, M., Schatz, O., O'Connell, T., O'Connell, D., Auckenthaler, A., and Schwer, H. (2008). A novel solid phase technology for high-throughput gene synthesis. *BioTechniques, 45*(3), 340-343.

Varani, G., and McClain, W. H. (2000). The G x 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.

Vargas-Rodriguez, O., Sevostyanova, A., Söll, D., and Crnković, A. (2018). Upgrading aminoacyl-tRNA synthetases for genetic code expansion. *Current Opinion in Chemical Biology, 46*, 115-122.

Virnekas, B., Ge, L., Pluckthun, A., Schneider, K. C., Wellnhofer, G., and Moroney, S. E. (1994). Trinucleotide phosphoramidites: ideal reagents for the synthesis of mixed oligonucleotides for random mutagenesis. *Nucleic Acids Research, 22*(25), 5600-5607.

Waldmann, T. (2006). Sloning announces the development of SlonoMax™ gene variant libraries of previously unmatched quality - Creating new possibilities for the directed evolution of proteins. . Retrieved from http://www.businesswire.com/news/home/20060928005357/en/Sloning-Announces-Development-SlonoMax-TM-Gene-Variant

Waldmann, T. (2013). Achieving a new quality level in tailoring genetic diversity. Retrieved from http://www.scientistlive.com/content/19548

Waldmann, T., and Fuhrmann, M. (2006). Slonomics: An Advanced Technology for Automated Gene Synthesis. *Innovations in Pharmaceutical Technology*. Retrieved from http://iptonline.com/articles/public/page59loresnonprint.pdf

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

Wang, Q., and Wang, L. (2008). New Methods Enabling Efficient Incorporation of Unnatural Amino Acids in Yeast. *Journal of the American Chemical Society, 130*(19), 6066-6067.

Watson, J. D., and Crick, F. H. C. (1953). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature, 171*(4356), 737-738.

Xiao, H., Bao, Z., and Zhao, H. (2015). High Throughput Screening and Selection Methods for Directed Enzyme Evolution. *Industrial & Engineering Chemistry Research, 54*(16), 4011-4020.

Yang, F., Moss, L. G., and Phillips, G. N., Jr. (1996). The molecular structure of green fluorescent protein. *Nature Biotechnology, 14*(10), 1246-1251.

Yang, G., and Withers, S. G. (2009). Ultrahigh-throughput FACS-based screening for directed enzyme evolution. *ChemBioChem, 10*(17), 2704-2715.

Young, D. D., and Schultz, P. G. (2018). Playing with the Molecules of Life. *ACS Chemical Biology, 13*(4), 854-870.

Zacharias, D. A., Violin, J. D., Newton, A. C., and Tsien, R. Y. (2002). Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science, 296*(5569), 913-916.