“Utilisation of amber suppression/non-natural amino acid technology for protein engineering and cellular control”

by

Josephine Lydia Morris

A Thesis submitted to Cardiff University for the degree of Doctor of Philosophy

Cardiff University

March 2013

DECLARATION

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or other award.

Signed …… (candidate) Date ……

STATEMENT 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of ……(insert MCh, MD, MPhil, PhD etc, as appropriate)

Signed …… (candidate) Date ……

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated.

Other sources are acknowledged by explicit references. The views expressed are my own.

Signed …… (candidate) Date ……

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed …… (candidate) Date ……

Acknowledgements

Firstly I'd like to thank my supervisors Dr Eric Tippmann and Professor Phil Stephens who have given me the opportunity to undertake the work presented here; provided crucial supervision throughout the research and preparation of this thesis.

I would like to acknowledge Dr Sohie Shim, who performed the ESI-MS analysis on my protein samples; Dr Damien Murphy who performed the EPR spectroscopy and data analysis; Dr Elisabetta Canetta who performed the Raman spectroscopy; Dr E. Joel Loveridge for greatly assisting me during the circular dichroism measurements; Dr Pierre Rizkallah who assisted with protein crystal formation and collection and who performed x-ray diffraction at Diamond Light Source, Didcot.

Without the support from the staff within the Schools of Chemistry and Dentistry the work presented would not have been possible and I appreciate their help. I am grateful to Professor Rudolf Allemann and Dr Dafydd Jones for the use of their laboratory equipment and helpful scientific comments.

I'd like to thank Dr Alicja Antonczak and Dr Zuzana Simova for showing me the way when I first started and throughout the first and second year of my research. They were both a source of intellectual knowledge and friendship. I'd also like to thank the other members of Physical Organic Chemistry group, past and present, particularly Jamie Johnson and Azzedine Bouleghlimat, for their friendship and scientific discussions. Also, the Wound Biology Group within the School of Dentistry and members of Dr Dafydd Jones' lab within the School of Bioscience, who have supported me and provided beneficial discussions.

I would like to acknowledge the support and love provided by my Mum and Dad, my sisters, Alana and Naomi and my partner Gareth. In addition, I'd like to thank my friends, particularly, Dr Matt Edmundson, Dr Mike Gamble, Andy Hartley, Sam Reddington, Dr Amy Baldwin, and Seni Chanapai who have also provided support.

Abstract

The amber suppression technology is an intracellular methodology that allows position specific incorporation of a specific non-natural amino acid (NAA) into proteins using imported NAA-specific machinery during protein translation. The method has been utilised to incorporate over 50 NAAs into proteins (e.g. those that confer unique reactivity (and allow subsequent conjugation of additional factors), installation of post-translational modification mimics, modulation of protein function and those that aid in structural determination). Therefore, the aim of the work presented within this Thesis was to explore the use of this technology in novel applications; namely the installation and study of a reactive moiety within a defined environment and the creation of a 'biological switch' to control the production of the protein and subsequently a cellular phenotype.

It was demonstrated that incorporation of the NAA, azidophenylalanine, within the hydrophobic pocket of T4 Lysozyme (T4LazF) could provide a protein scaffold to stabilise, shield and thereby allow exploration of the chemical reactivity of the photoreactive aryl azide moiety using various spectroscopic techniques. Specifically, electron paramagnetic spectroscopy of irradiated T4LazF demonstrated that the singlet phenyl nitrene species had been captured. To create a 'biological switch' to control the production of a protein and the subsequent cellular phenotype, the duality of the amber suppression method was the basis for the novel application rather than the chemical reactivity of the NAA. *E. coli* and mammalian cell motility readout systems were successfully created using flagellin (non-flagellate) and Rac1 (GTPase protein involved in lamellipodia production) knockout cell lines respectively in conjunction with specially created plasmid constructs. However, complexities regarding the implementation of amber suppression in order to control this motility via the presence/absence of NAA limited the functionality of these readout systems.

The work presented has advanced the field of amber suppression and NAA technology by demonstrating that generating reactive intermediates derived from NAAs within a defined chemical environment of a protein provides a novel technique to generate and study highly reactive intermediates. In addition, it has been shown that the amber suppression technology has potential to act as a biological switch to control cellular responses.

Table of contents

List of figures

List of tables

Abbreviations

AA – amino acid

aaRS – aminoacyl tRNA synthetase

aa-tRNA – aminoacyl tRNA

AMP – adenosine monophosphate

Amp – ampicillin

APS – ammonium persulphate

ATP – adenosine 5'-triphosphate

azF – Azidophenylalanine

azRS – aminoacyl tRNA synthetase specific for azF

Bs - *Bacillus stearothermophilus*

*Bs*tRNA$_{CUA}$ - *Bacillus stearothermophilus* tRNA recognising UAG codon

Bp – base pair

BGH - bovine growth hormone

Carb - carbenicillin

CAM – chloramphenicol

CD – circular dichroism

CHO – Chinese Hamster Ovary

CLLL (CL) - post-translational modification signal sequence of rac1 protein, i.e. the isoprenylation, proteolysis and methylation of the C-terminal CAAX motif

CMV - Cytomegalovirus

cDNA - complementary DNA

Da – Dalton

DNA – deoxyribonucleic acid

DN – Dominant negative

DMEM - Dulbecco's modified Eagle's medium

dNTP – deoxynucleoside triphosphate

E. coli – *Escherichia coli*

*Ec*TyrRS – *E. coli* tyrosyl tRNA synthetase

EDTA - Ethylenediaminetetraacetic acid

EF-Tu – elongation factor Tu

eGFP – enhanced green fluorescent protein

EDTA - Ethylenediaminetetraacetic acid

EPR – Electron paramagnetic resonance

ESI-MS – electrospray ionisation mass spectroscopy

FBS – foetal bovine serum

fMet - N-formylmethionine

fMet-tRNAf – initiator tRNA

FPLC – fast protein liquid chromatography

FRET – fluorescent resonance energy transfer

FTIR – Fourier transform infrared spectroscopy

GAP – GTPase-activating proteins

GDP – guanine diphosphate

GEF - Guanine nucleotide exchange factor

GFP – green fluorescent protein

GPCR – G protein-coupled receptors

GTP – guanine triphosphate

h– hour

HEK - Human Embryonic Kidney cells 293

HPLC – High performance liquid chromatography

IF – initiator factor

iodoF – iodophenylalanine

iodoRS – aminoacyl tRNA synthetase specific for iodoF

IPTG - Isopropyl β-D-1-thiogalactopyranoside

IR – infrared spectroscopy

IRES – internal ribosome entry site

Kan – kanamycin

Kb – kilobase pairs

kDa – kilodalton

KO - knockout

*lac*I – gene encoding the lactose repressor

LB – lysogeny broth

lpp – lipoprotein promoter

M – molar

M. jannaschii (Mj) – Methanococcus jannaschii

MEF – Mouse embryonic fibroblast

MCS – multicloning site

Min – minute

mRNA – messenger ribonucleic acid

MW – molecular weight marker

NaOH- sodium hydroxide

NAA – nonnatural amino acid

Ni-NTA - nickel-nitriloacetic acid

OD_{600} – optical density at λ = 600 nm

pA – polyadenosine

PAGE – polyacrylamide gel electrophoresis

PAKs - p21-activated serine/threonine kinases

pBpa - p-benzoyl-L-phenylalanine

PBS – phosphate buffered saline

PCR – polymerase chain reaction

PDB – protein data bank

PEI - Polyethylenimine

PLG – phase lock gel

PMSF - phenylmethylsulfonyl fluoride

PTM – post translational modification

PYLIS – pyrrolysine insertion sequence

Rac1 KO – Rac knockout fibroblast cells

RhoGDI – Rho guanosine nucleotide dissociation inhibitors protein

RBS – ribosome binding site

RF – release factor

RKCL –polybasic region and post-translational modification signal sequence of rac1 protein, i.e. the isoprenylation, proteolysis and methylation of the C-terminal CAAX motif

RNA - ribonucleic acid

RT – room temperature

RT-PCR – reverse transcription polymerase chain reaction

rpm – revolutions per minute

rRNA – ribosomal ribonucleic acid

RS – tRNA synthetase

S. cerevisiae - *Saccharomyces cerevisiae*

SD – standard deviation

SDS – sodium dodecyl sulphate

SeMet - selenomethionine

SEC – size exclusion chromatography

SECIS – selenocysteine insertion sequence

SOB – super optimal broth

SOC - super optimal broth with catabolite repression

SV40 - Simian vacuolating virus 40

T4L – T4 lysozyme

T4L153azF - T4 lysozyme with azidophenylalanine at position 153

T4L153iodoF - T4 lysozyme with iodophenylalanine at position 153

T4L153TAG - T4 lysozyme with TAG codon at position 153

Taq – thermostable DNA polymerase from *Thermus aquaticus*

TEMED – Tetramethylethylenediamine

Tet - tetracycline

Tm – melting temperature

Tris – tris(hydroxymethyl)aminomethane

tRNA - transfer ribonucleic acid

$tRNA_{CUA}$ – orthogonal transfer ribonucleic acid recognising UAG codon

UV-vis – ultraviolet-visible spectroscopy

UV – ultraviolet

WASP – Wiskott–Aldrich Syndrome protein

WAVE – WASP family verprolin homology domain-containing protein

WT – wild type

Zeo^r – zeocin resistance gene

Zeo – zeocin antibiotic

λ – wavelength

Amino acids

Alanine, Ala, A

Arginine, Arg, R

Asparagine, Asn, N

Aspartic acid, Asp, D

Azidophenylalanine, azidoPhe, azidoF

Cysteine, Cys, C

Glutamine, Gln, Q

Glutamic acid, Glu, E

Glycine, Gly, G

Histidine, His, H

Iodophenylalanine, iodoPhe, iodoF

Isoleucine, Ile, I

Leucine, Leu, L

Lysine, Lys, K

Methionine, Met, M

Phenylalanine, Phe, F

Proline, Pro, P

Pyrrolysine, Pyl, O

Selenocysteine, Se-cys, U

Serine, Ser, S

Threonine, Thr, T

Tryptophan, Trp, W

Tyrosine, Tyr, Y

Valine, Val, V

Nucleotides

Adenosine, A

Dihydrouridine, D

Cytosine, C

Guanine, G

Inosine, I

Lysidine

Thymine, T

Uracil, U

Pseudouracil

1 CHAPTER 1: INTRODUCTION

1.1 Introduction

The aim of this Chapter is to introduce the reader to the topics that relate to the experimental study that is covered within this thesis. Firstly, the process of translation will be described for both the prokaryotic system and eukaryotic system, as both systems are utilised within the experimental work reported in this thesis. This includes the nature of the genetic code, the relationship to translation and processes which ensure the fidelity of the translation of the genetic code to protein. Namely, the specific recognitions and aminoacylation reactions that occur between the aminoacyl tRNA synthetases and tRNA molecules. This will focus on the tyrosine aminoacyl tRNA synthetases and tRNA from all the kingdoms of life. These particular points are outlined due to their importance in the methods used within this thesis. This Chapter will then continue to introduce the variety of methods used to expand the chemical, biological and physical properties of proteins by the introduction of non-natural amino acids, focussing on the recently established method utilising genetic code expansion. This method utilises the *in vivo* process of translation along with an orthogonal aminoacyl tRNA synthetases and tRNA pair and essentially recodes the amber stop codon to instead code for a non-natural amino acid. The previous applications of this amber suppression method, in both prokaryotic and eukaryotic systems, will be outlined. Explanation and rational for the work conveyed within this thesis will be considered, in particular in relation to the two novel applications addressed. The first application involves the elucidation of reactive intermediates by creating a defined and stabilising environment, which was thought could be used to reduce the complications of bimolecular reactions which occur when studying highly reactive intermediates. The environment considered is the hydrophobic pocket within the well-studied protein, T4 lysozyme and the photoreactive azide group providing an example of a highly reactive species. The second application explores the use of the amber suppression method to control cellular phenotype by creating a switch controllable by the presence/absence of non-natural amino acid. Both the prokaryotic and eukaryotic systems were used in the investigations. In both cases the phenotype selected for control was motility. The Chapter will then conclude

with the hypothesis for the investigations performed and reported within this Thesis.

1.2 Translation

Translation is the process by which genetic information in the form of messenger RNA (mRNA) is converted into a sequence of corresponding amino acids to form a peptide chain. In the cell, this process is undertaken by the ribosome and involves a number of protein (e.g. aminoacyl tRNA synthetases) and RNA molecules (e.g. the amino acid - RNA adapter molecule, tRNA). Translation can be divided into three steps, *initiation* of protein synthesis, *elongation* of the peptide chain and *termination* and release of the peptide. During initiation the ribosome and required protein factors assemble on the mRNA, the start site of the gene is located and the initiator tRNA is bound. Throughout elongation, amino acids are delivered to the ribosome by aminoacylated-tRNAs, codon : anticodon recognition occurs and the amino acids are joined on to the growing chain via a peptide bond. At termination, a termination codon is read, which does not code for an amino acid because there is no tRNA with an anticodon that recognises it, causing the ribosome to dissociate and the peptide to be released from the complex. A number of components are involved within translation and a selection function to ensure fidelity of protein synthesis. The factors differ between prokaryotic and eukaryotic translation and will be considered below.

1.3 Prokaryotic translation

1.3.1 Ribosome

The ribosome is a large complex made up of two subunits comprising of protein and RNA (for review see Ramakrishnan 2002). The complex has three tRNA binding positions; the aminoacyl site (A-site), the peptidyl site (P-site) and the exit site (E-site; Figure 1). The small subunit (30S) comprises the 16S rRNA molecule and 21 different ribosomal proteins (r-proteins, S1-S21). The proteins are predominantly basic in order to stabilise the negatively charged RNA. The rRNA plays an active role

in translation by intercalating with mRNA and facilitating its proper alignment within the ribosome complex. The large subunit (50S) comprises the 23S, 5S rRNA and 34 proteins and confers peptidyltransferase and translocase activity.

1.3.2 Initiation

The process of initiation requires initiation factors, IF-1, IF-2 and IF-3 (for review see Laursen *et al.* 2005). IF-1 and IF-3 bind to the free 30S subunit preventing re-association with the 50S and blocks the A-site preventing tRNA from binding to the initiation complex. During translation initiation the ribosome binding site (RBS; comprising the purine-rich Shine-Dalgarno sequence and start codon, typically AUG) of the mRNA is recognised by the 16S component of the small subunit of the ribosome (Figure 1A). IF-2-GTP is responsible for transporting the first aminoacyl-tRNA (fMet-$tRNA_f$) to the P-site of the initiation complex. If the base pairing between initiator tRNA and start codon is correct, IF-3 is released from the complex. The 50S subunit then associates with the initiation complex concurrently with hydrolysis of IF-2-GTP and release of IF-1 and IF-2.

1.3.3 Elongation

During elongation aminoacyl-tRNA synthetases (aa-tRNAs) are transported to the A-site by elongation factor, EF-Tu-GTP (Figure 1B). If complementary base-pairing occurs within the A-site, GTP hydrolysis occurs and EF-Tu-GDP is released. This GTP hydrolysis in response to correct aa-tRNA association acts as the rate-limiting step within translation elongation, providing a time interval during which an incorrect aminoacyl tRNA, which would bind less strongly to the mRNA codon, can dissociate from the ribosome rather than being incorporated incorrectly.

The peptidyltransferase reaction, catalysed by the 23S rRNA (of the large subunit), occurs between the free carboxyl group of the amino acid connected to the tRNA present within the P-site (peptidyl-tRNA) and the free amino group of the amino acid connected to the tRNA within the A-site (aa-tRNA). Translocation of the ribosome is promoted by elongation factor, EF-G-GTP. GTP hydrolysis triggers a conformational change within the ribosome which promotes the translocation of

the ribosome. The process results in the next 3' codon to enter the A-site and the recently de-aminoacylated tRNA moves from the P-site to the E-site. The ribosome is then left with a peptidyl tRNA bound at the P site, and an empty A site. The process is repeated by the delivery of the next aa-tRNA to the A-site by EF-Tu and the de-aminoacylated tRNA is ejected from the E-site.

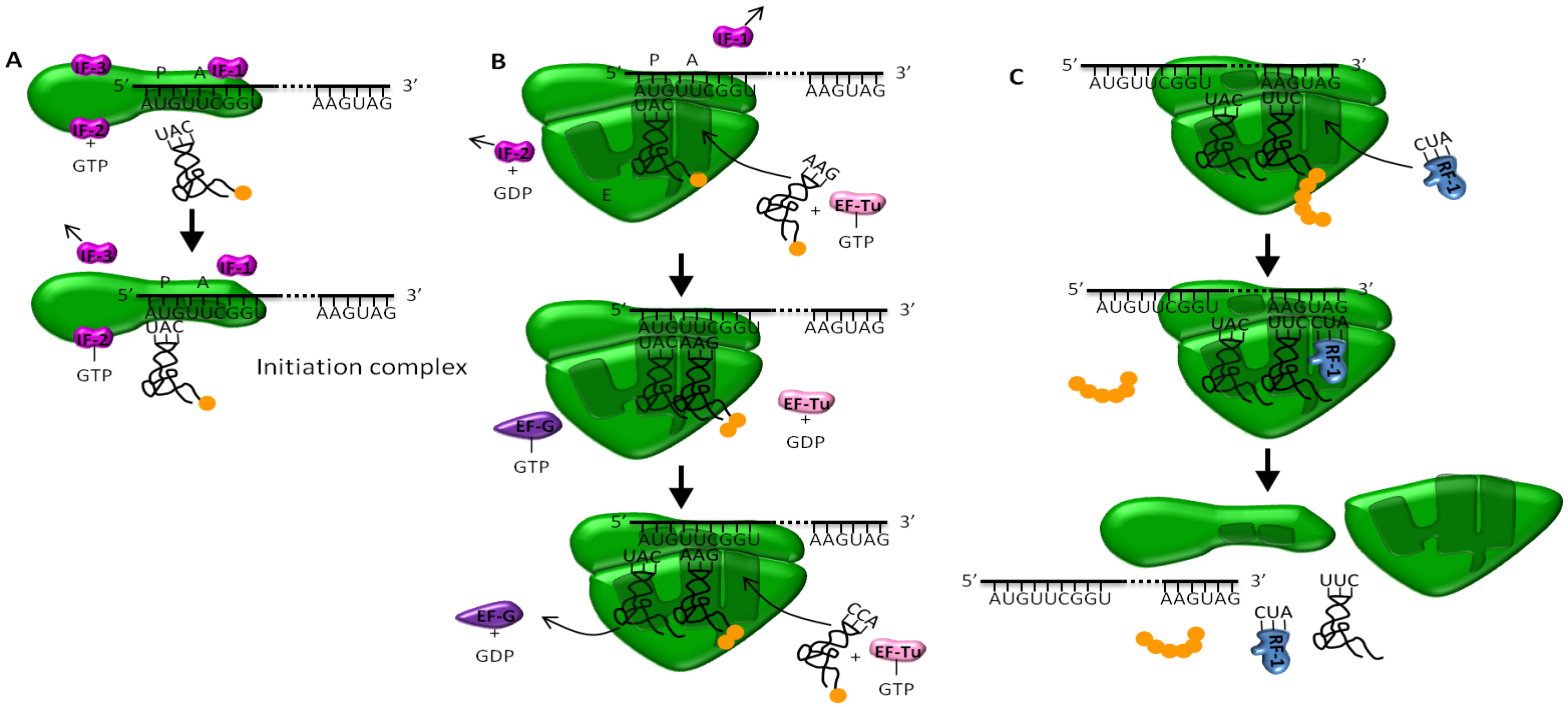


Figure 1. Prokaryotic translation. A) Initiation. RBS of the mRNA is recognised by the 16S component of the small subunit of the ribosome (green) along with initiation factors IF-1, IF-2 and IF-3. IF-2-GTP transports the first aminoacyl-tRNA (fMet-tRNA$_f$; yellow) to the P-site of the initiation complex, IF-3 is released from the complex. The 50S subunit then associates to the initiation complex concurrently with hydrolysis of IF-2-GTP and release of IF-1 andIF-2-GDP.

B) Elongation. aa-tRNAs are transported to the A-site by EF-Tu-GTP. A peptidyltransferase reaction, catalysed by the 23S rRNA, occurs between the two amino acids. Translocation of the ribosome is promoted by hydrolysis of EF-G-GTP.

C) Termination. Stop codons are recognised by release factors, which upon binding cause conformational changes that allow disassociation of the polypeptide, tRNA and ribosome subunits. Subsequent binding of IF-3 to the 30S subunit results in disassociation of the release factors, mRNA and tRNA.

The cycle of elongation repeats throughout the reading frame of the target gene, decoding each codon to its corresponding amino acid, which is added to the growing peptide chain. The recycling of EF-Tu-GDP to GTP requires the elongation factor EF-Ts, which facilitates the exchange of GDP to GTP.

1.3.4 Termination

Cycles of elongation are ended by the presence of a translational termination codon, namely UAA, UAG or UGA (ochre, amber, opal respectively), within the A-site. These codons are not recognised by any tRNA, but are recognised instead by release factors, RF (Figure 1C). RF-1 terminates gene translation at UAA and UAG stop codons, RF-2 at UAA and UGA stop codons, and RF-3 facilitates the function of RF1 and RF2 (Nakamura et al., 2000). The binding of RF-1 and RF-2 to the ribosomal complex causes conformational changes that allow hydrolysis of the ester bond between the polypeptide and tRNA. After release of the peptide, the release factors catalyse the disassociation of the ribosome subunits. Subsequent binding of IF-3 to the 30S subunit results in disassociation of the release factors, mRNA and tRNA.

1.4 Eukaryotic translation

1.4.1 Ribosome

Translation of eukaryotic genes is much more complex than that of prokaryotes with the involvement of numerous factors. As with the prokaryotic system, the eukaryotic ribosome is a large complex of comprising protein and RNA. The small subunit (40S) is composed of the 18S rRNA and approximately 30 proteins; and the large subunit (60S) is composed of 28S, 5.8S, and 5S rRNAs and about 45 proteins.

1.4.2 Initiation

At translation initiation, the eukaryotic initiation factors, eIF-1, eIF-1A and eIF-3 bind the 40S ribosomal subunit and eIF-2 (in complex with GTP) associates with the initiator methionyl tRNA (Figure 2). The eIF-4 group of factors recognise and bring eukaryotic mRNA to the ribosome. Namely, eIF-4E mediates recognition of the 5' terminal 7-methylguanosine cap, eIF-4G binds to both eIF-4E and to poly-A binding protein (PABP) that is associated with the poly-A tail at the 3' end of the mRNA. The initiation factors eIF-4E and eIF-4G, in association with eIF-4A and eIF-4B, then bring the mRNA to the 40S ribosomal subunit, with eIF-4G interacting with eIF-3. The 40S ribosomal subunit, in association with the bound methionyl tRNA and eIFs, then scan the mRNA 5' to 3' to locate the AUG initiation codon. Subsequent to the recognition of the initiation AUG codon, eIF-5 triggers the hydrolysis of GTP bound to eIF-2. The eIFs are then released, and the ribosomal 60S subunit binds to the 40S subunit to form the 80S initiation complex of eukaryotic cells.

1.4.3 Elongation

The process of eukaryotic elongation is similar to that of prokaryotes utilising the three sites within the ribosomal complex, however aa-tRNAs are escorted to the complex via the GTP-complexed elongation factor, eEF-1α (in place of EF-Tu; Figure 1B). Here, as for prokaryotic translation, is a point of proofreading (before GTP hydrolysis) of the codon-anticodon pairing before the peptide bond forms. Translocation in eukaryotes requires eEF-2 and is again coupled to GTP hydrolysis. The GTP exchange factor for eukaryotic translation is eEF-1β. The elongation cycle continues as for prokaryotic translation.

1.4.4 Termination

In eukaryotic cells a single release factor (eRF-1) recognises all three termination codons in conjunction with eRF-3 in a process much like prokaryotic termination (Figure 1C).

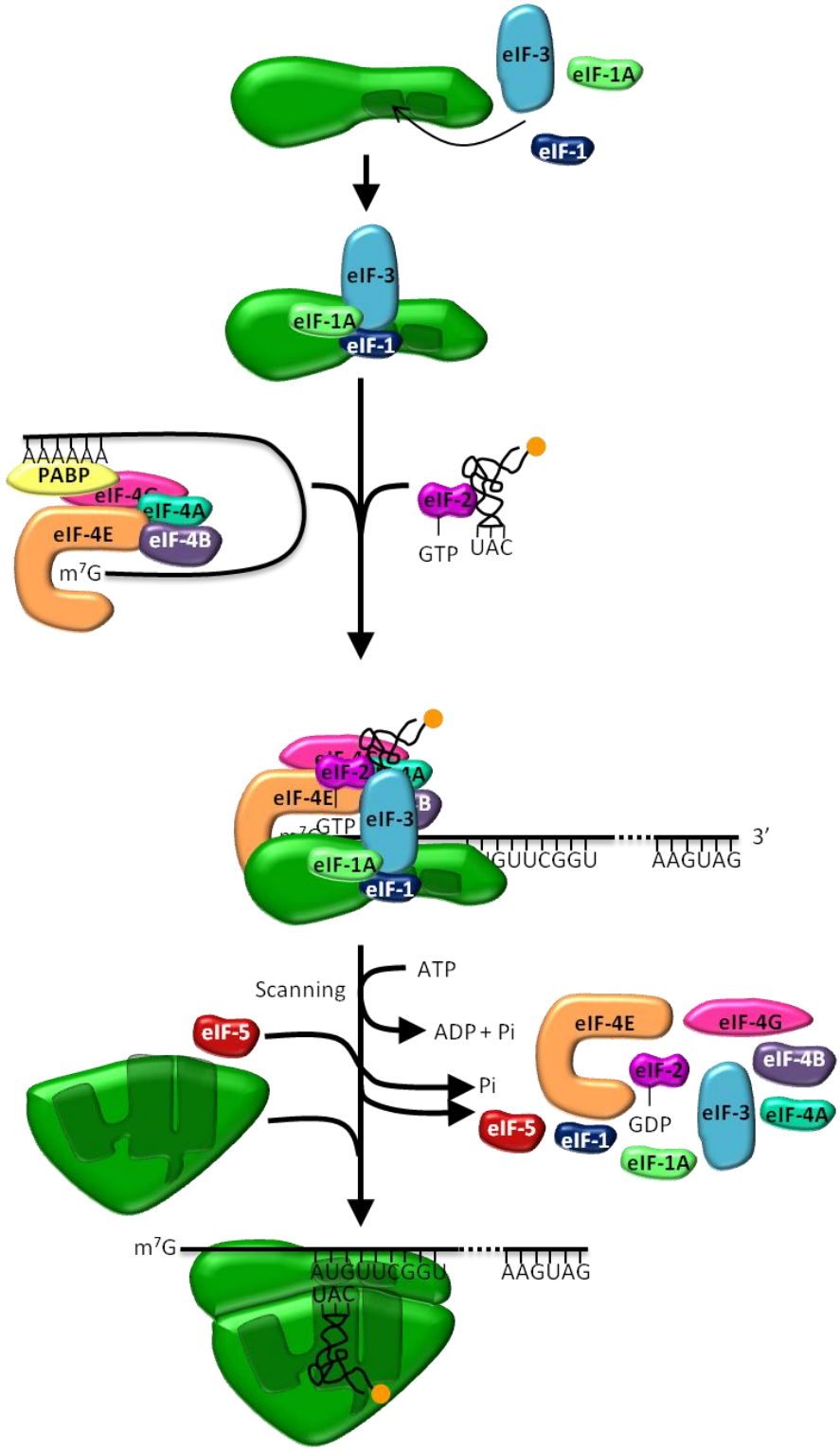


Figure 2. Eukaryotic translation initiation. eIF-1, eIF-1A and eIF-3 bind the 40S ribosomal subunit and eIF-2-GTP associates with the initiator methionyl tRNA. The eIF-4 group of factors recognise and bring eukaryotic mRNA to the 40S ribosome subunit with eIF-4G interacting with eIF. The 40S ribosomal subunit scans the mRNA to locate the AUG initiation codon. eIF-5 triggers the hydrolysis of GTP bound to eIF-2. The eIFs are then released, and the ribosomal 60S subunit binds to the 40S subunit to form the 80S initiation complex.

1.5 Genetic code

Messenger RNA (mRNA) comprise a combination of the four nucleotides, U, T, G, and C. These nucleotides are read by the ribosome by three-letter codons, which are interpreted to specify for one amino acid. These codons have a corresponding anticodon present on tRNA that carries a specific amino acid. This link between RNA and amino acid is known as the genetic code (Figure 3). The code is degenerate in that a particular amino acid can be encoded by more than one codon. Many organisms typically have less than 45 unique tRNAs, which is less than the total number of potential codons (64). Therefore some codon : anticodon interactions may display non-canonical Watson-Crick base pairing, which is predominantly due to the "wobble" effect. Namely, the third position in a tRNA codon is less crucial in decoding mRNA, because it can sometimes perform non-standard basepairing (for example, G-U) or be replaced by a modified base pair (Inosine, I), allowing it to recognise more than one codon (I-A, I-C, I-U).

Nature provides the genetic coding for the most common 20 amino acids, three codons that terminate translation (amber, ochre, and opal with codons UAG, UAA, and UGA respectively) plus the rarely incorporated amino acids, pyrrolysine (Hao *et al.* 2002) and selenocysteine (Boeck *et al.* 1991) for use in protein construction (Figure 3). The incorporation of the natural 20 amino acids occurs using the cognate process of translation, as described within section 1.2. The three stop codons are used with differing frequencies. The amber codon is of particular interest, as this is the least frequently used stop codon in the *E. coli* genome (9 % bacterial proteins, very rarely used in essential genes; 23 % in Homo sapiens (Nakamura 2007)). Interestingly, the incorporation of pyrrolysine and selenocysteine occurs at the UAG and UGA stop codons respectively. The process to incorporate the 21st and 22nd amino acids is complex and requires extra information on the mRNA.

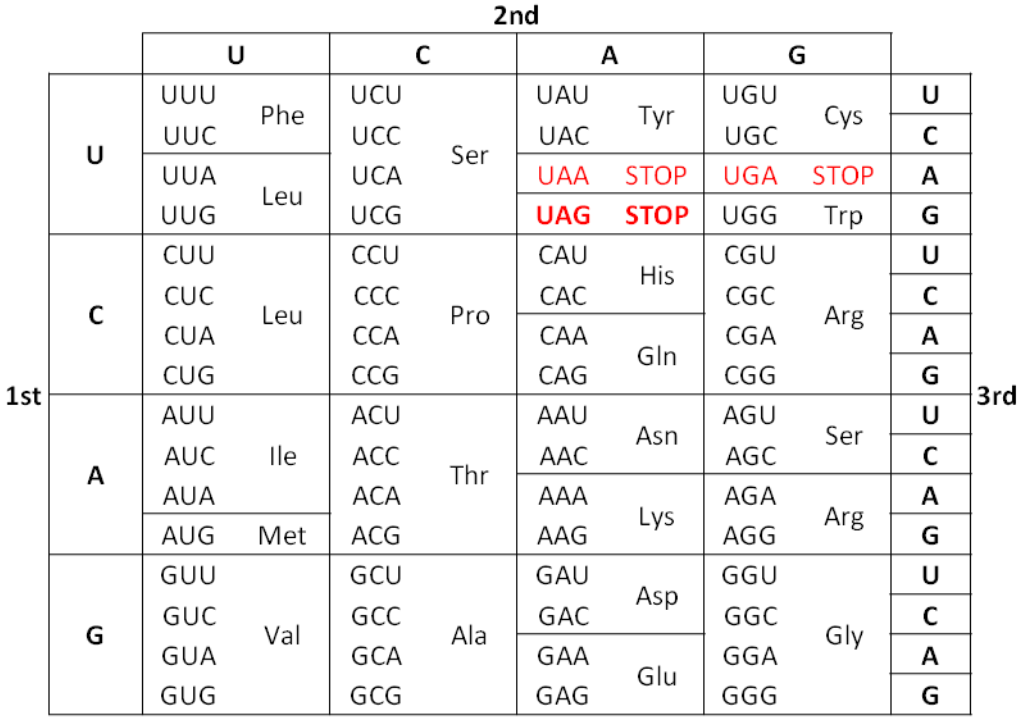


Figure 3. Genetic code. The 3-nucleotide codons and their corresponding amino acids. Translation termination codons are indicated in red. Amber codon, TAG, is indicated in bold.

1.5.1 Incorporation of pyrrolysine and selenocysteine

The incorporation of the 21st and 22nd amino acids, selenocysteine (Sec) (Boeck et al. 1991) and pyrrolysine (Pyl) (Ibba and Söll 2002) are of particular interest because their incorporation is encoded by termination codons. Selenocysteine is similar to cysteine, but carries selenium in place of sulphur. The selenocysteine insertion element, SECIS, found on the mRNA of *E. coli* genes specifies the incorporation of selenocysteine at the UGA codon. Firstly $tRNA^{Sec}$, is aminoacylated with cysteine, which is then converted to selenocysteine by selenocysteine synthase (SelA). Sec-$tRNA^{Sec}$, along with GTP and selenocysteine-specific elongation factor, SelB, binds to the SECIS element at the ribosome, which allows the incorporation of Sec into the protein. Unlike with Sec, Pyrrolysine has a specific synthetase (PylRS) which charges the $tRNA^{Pyl}$ with Pyl (Polycarpo *et al.* 2004). The PYLIS element, within the mRNA, is thought to enhance the incorporation of Pyl at the specific UAG codon (Théobald-Dietrich *et al.* 2005).

1.6 Aminoacylation

A number of components function to ensure fidelity of protein synthesis. Namely, the accuracy of codon : anticodon recognition and the aminoacylation of the tRNA (Figure 4). The sequence of amino acids that are used to make up the peptide are determined by the three base-pair codons on the mRNA. The mRNA codon is recognised, via complementary base-pairing, by the aminoacyl-tRNA anticodon region. The tRNA is charged with its cognate amino acid in two steps. Firstly, the α-phosphate of ATP reacts with the amino acid to form a mixed anhydride intermediate, aminoacyl-adenylate, and inorganic pyrophosphate. In the second step, an ester bond is formed between the 3'-terminal ribose of the cognate tRNA and the amino acid moiety of the aminoacyl-adenylate (Equation 1). This yields an aminoacyl-tRNA and AMP. These steps are catalysed by the specific aminoacyl tRNA synthetase (aaRS). Once the tRNA is correctly charged, it is able to participate in codon : anticodon interaction and the amino acid can be incorporated into the growing peptide chain during translational elongation (see section 1.3.3 and 1.4.3).

Equation 1. Two step aminoacylation of tRNA. Firstly, the α-phosphate of ATP reacts with the amino acid (AA) to form a mixed anhydride intermediate, aminoacyl-adenylate (aaRS·aa-AMP), and inorganic pyrophosphate (PPi). In the second step, an ester bond is formed between the 3'-terminal ribose of the cognate tRNA and the amino acid moiety of the aminoacyl-adenylate. This yields an aminoacyl-tRNA (aa-tRNA) and AMP. These steps are catalysed by the specific aminoacyl tRNA synthetase (aaRS).

$$AA + ATP + aaRS \leftrightarrow aaRS{\cdot}aa\text{-}AMP + PP_i$$

$$aaRS{\cdot}aa\text{-}AMP + tRNA \leftrightarrow aaRS + aa\text{-}tRNA + AMP$$

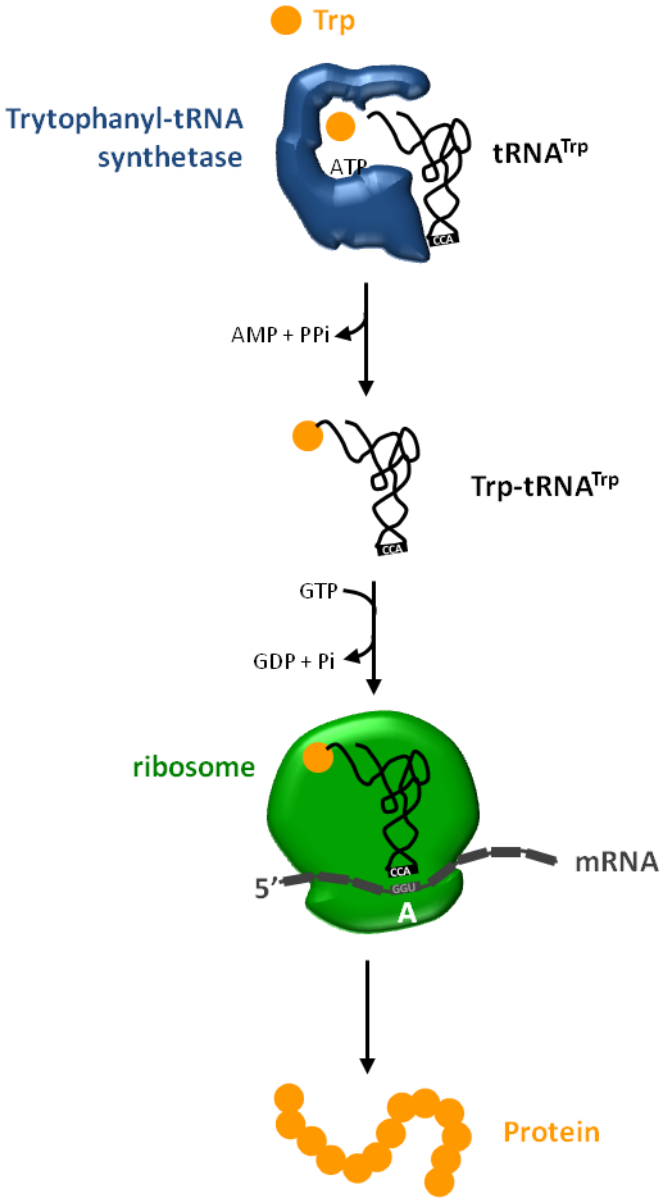


Figure 4. Role of aminoacyl-tRNA formation in the elongation phase of protein synthesis. An uncharged tRNA is first aminoacylated with the appropriate amino acid to generate an aminoacyl tRNA. The aminoacyl-tRNA is delivered to the ribosomal A site via a translation elongation factor. The aminoacyl-tRNA anticodon interacts with the corresponding codon in mRNA. The example shown illustrates how this leads to the translation of the codon GGU as tryptophan.

1.7 The aminoacyl tRNA synthetase

In most bacteria a family of 20 aminoacyl-tRNA synthetases (aaRS) are responsible for charging the correct tRNA with the correct amino acid. Each synthetase enzyme is specific for one amino acid, but capable of aminoacylating one or more isoaccepting tRNA (Sprinzl *et al.* 1987). In eukarya, multiple aaRSs for certain amino acids are present in the cytoplasm, while different aaRSs are present in the mitochondria. Each synthetase catalyses a similar reaction, requiring ATP and Mg^{2+}, and utilises similarly structured substrates, namely amino acid and tRNA. Therefore, the synthetases must share common features to enable them to catalyse the aminoacylation reaction, but enough variation in order to recognise slight differences in tRNA and amino acid. Physical, chemical, and mutagenic studies (For review see Beuning and Musier-Forsyth 1999; Ibba and Söll 2000) demonstrate that the binding of tRNA into the tRNA-binding cleft of aaRSs occurs along the inside of the tertiary L-shaped tRNA structure in a general shape-related manner (see section 1.8). However, the specific interactions the aaRSs makes with the tRNA tend to be spatially diverse with specificity being achieved by the specific interactions of crucial position-specific amino acid residues. There are two broad classes of synthetase. The members of each class are broadly characterised by their sequence motifs, active site structure, mode of tRNA binding and the OH group of the cognate tRNA that is aminoacylated.

Class I aaRSs, which includes the tyrosyl-aaRS, contain HIGH and KMSKS motifs with the active site containing a Rossmann fold (Rao and Rossmann 1973). The peptide backbone of the Gly-His region of the HIGH motif associates with the adenine of ATP which is bound in an extended conformation, while the two His residues interact with the triphosphate both directly, and via water molecules. The MSK, which is part of a loop, folds over the binding site, with the Lys interacting with the α- and ɣ-phosphates. For class I enzymes, the acceptor stem of the tRNA interacts with the aaRS via the minor groove side with the variable loop facing the solvent and therefore, the amino acid is attached to the 2'-OH of the terminal adenosine of the tRNA. However, for tyrosyl-tRNA synthetase the recognition of tRNA resembles

that of a class II synthetase, the tRNA associating by the major groove side of the acceptor stem (Yaremchuk *et al.* 2002). Also to note is that class I aaRSs tend to be monomeric, with the exception of TyrRS and TrpRS, which are obligate dimers.

Class II aaRSs contain conserved motifs, 1, 2 and 3. The active site of these enzymes are composed of an anti-parallel β-fold, containing parts of motif 2 and 3. Motif 1 consists of a long α-helix linked to a β-strand, and participates in the dimer interface. In Class II enzymes, the ATP binds in a bent conformation. Class II aaRSs approach the major groove side of the tRNA acceptor stem with the variable loop facing the synthetase. It is this binding which causes class II aaRS, with the exception of $tRNA^{Phe}$ (attaches Phe to 2'-OH), to attach amino acids to the 3'-OH of the terminal adenosine of the tRNA. Class II enzymes tend to be obligate dimers.

The class I and II aaRSs are then sub-divided into groups of RSs that have similar modules or type of amino acid they charge.

1.8 tRNA

Transfer RNA (tRNA) are small RNA molecules comprising 75 to 95 nucleotides. They contain standard RNA nucleotides, namely A, C, G, U, but also modified nucleotides, which are introduced post-transcriptionally (Figure 5A). tRNAs have an important role in the fidelity of translating genetic information in to protein by acting as an adapter between the mRNA, amino acid and the protein synthesis machinery (Crick 1955, 1958). All tRNAs have a similar primary (Sprinzl et al. 1987) secondary (Holley *et al.* 1965) and tertiary (Kim *et al.* 1973; Ladner *et al.* 1975) structure which allows them to be recognised by the translational machinery. Namely, the secondary cloverleaf and tertiary L-shaped structures respectively (Figure 5). This structure is found in both prokaryotic and eukaryotic tRNAs; however, some differences in basic primary structures, such as kingdom-specific identity elements, are noted below.

The cloverleaf structure is formed by a series of inverted repeats that allow the linear RNA molecule to form four double helical segments (Figure 5A). The 5' and 3' terminals of the tRNA associate to form the accepter stem. At this stem the -OH

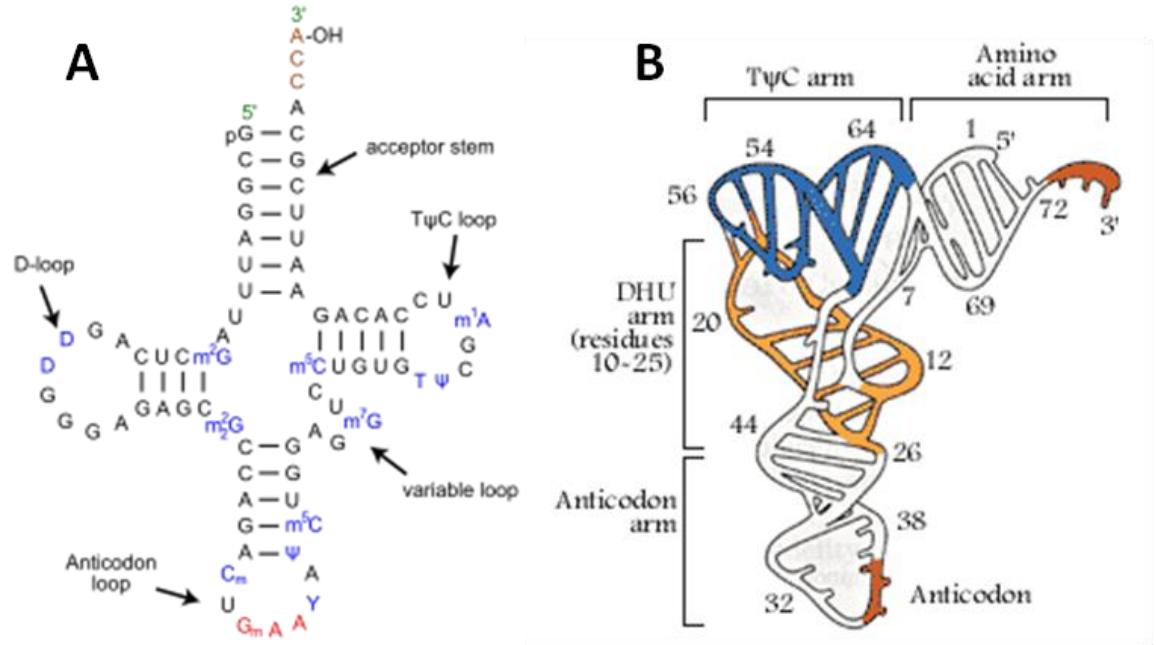


Figure 5. A) Secondary cloverleaf structure (adapted from Kim et al. 1973). Stem and loop structures are labelled. Modified residues are marked in blue and anticodon is marked in red.

B) Tertiary L-shape of prokaryotic tRNA (adapted from Kim et al. 1974). Arms are labelled and certain residue numbers are indicated.

group of A76 of the 3'-CCA single stranded extension is coupled to the amino acid during aminoacylation of the tRNA. The TψC arm, named because of the presence, in most tRNAs, of the modified uracil, pseudouridine, ψ, between a thymine and cytosine residue. This sequence in eukaryotic initiator tRNAs is substituted with AUC or AψC. The variable arm can be between four and twenty-one nucleotides in length. The anticodon loop contains the three nucleotide anticodon (nucleotides 34, 35, 36) which complementarily base-pairs with the corresponding mRNA codon during translation. The first position of the anticodon, the "wobble" position, often contains modified bases that allow the formation of non-Watson-Crick base-pairing (see section 1.5). The DHU arm or D-arm is named because of the presence of the modified nucleotide dihydrouridine, D. The number of residues in the stem and loop regions is so well conserved, that all tRNAs are referred to using a standard numbering system (Schimmel and Soll 1979). The variations in size occur within the variable and D-arms.

The tertiary, L-shape structure of tRNA is formed by the coaxial stacking of the double-stranded helices (Figure 5B). The CCA region of the accepter stem and anticodon loop form the tips of the L shape, with the DHU and TψC arms folding in to form the corner of the L-shape. A hydrogen-bond network creates associations throughout the molecule. However, the nucleotides of the CCA extension and the anticodon are exposed and do not form part of this hydrogen bond network (Robertus *et al.* 1974). This allows the anticodon loop to twist out, which enables interaction with the codon of the mRNA.

1.9 Identity elements

A major part in maintaining the fidelity of protein translation depends upon the strict recognition of the tRNA by its cognate aaRS and subsequent charging with the correct amino acid. The tRNA is relied upon to make the association with the correct aaRS, but also to refrain from associating with other incorrect aaRSs. As mentioned above, all tRNAs have similar primary, secondary and tertiary structures in order to function at the ribosome, but there are key differences in their structures that are important in discerning each tRNA's cognate synthetase. These

key differences are known as the identity elements of the tRNA (Giegé *et al.* 1998). The position and nature of these identity elements varies among different tRNAs. However, these identity elements are often found within the accepter stem (discriminator base), the anticodon and the variable arm.

The use of naturally occurring amber suppressors, chemical modification and the design of synthetic RNAs with desired single or multiple mutations, in silico calculations, and crystal structures have been used to provide valuable evidence regarding the elucidation of identity elements in tRNA-aaRS recognition (for review see Giegé et al. 1998). Originally the anticodon region was considered to confer tRNA identity and this is understandable as this is the obvious variable region of the tRNA and it is this region that provides the link between mRNA and protein. However, the first important observation was that many aaRSs aminoacylate more than one tRNA, i.e. tRNAs coding for the same AA, but with different anticodons. Therefore, it is more than just the anticodon region that determines the specificity of aminoacylation. There was further evidence that the anticodon region may not be the major identity element in some tRNAs, as mutation of *E. coli* $tRNA^{Gln}$, $tRNA^{Ser}$ and most importantly $tRNA^{Tyr}$ (Goodman *et al.* 1968) anticodons to CUA, creating amber suppressor tRNAs, did not perturb correct aminoacylation of these tRNAs (Capecchi and Gussin 1965; Inokuchi *et al.* 1979; Kaplan *et al.* 1965; Weigert and Garen 1965). Therefore, further investigation was required in order to elucidate the identity elements, and this research often concluded that the identity elements were predominantly specific to each tRNA-aaRS pair. The identity elements specific to $tRNA^{Tyr}$ will subsequently be considered further. The identity elements of $tRNA^{Tyr}$ between species is also important, therefore, these will also be considered.

1.9.1 TyrRS-$tRNA^{Tyr}$ identity elements

The discriminator base, A73 has been shown to be one of the major identity elements in $tRNA^{Tyr}$ from all kingdoms of life (Figure 6). The A73 base is stacked on the C72 base in the prokaryotic TyrRS-$tRNA^{Tyr}$ complex to extend the RNA helix, whereas the A73 base in the *Methanococcus jannaschii* archaean (*Mj*)TyrRS-$tRNA^{Tyr}$ projects away from the helical axis. The base pairs in region 1-72 are also important

in $tRNA^{Tyr}$ recognition, but displays a kingdom specific discrimination. The crucial identity element that distinguishes a bacterial $tRNA^{Tyr}$ from an archaeal or eukaryotic $tRNA^{Tyr}$ is the C1-G72 (G1-C72 in prokaryotic $tRNA^{Tyr}$) tRNA recognition element located at the top of the accepter stem (Fechter *et al.* 2001; Quinn *et al.* 1995; Steer and Schimmel 1999). Due to the similarities of the chemical groups presented to the minor groove, and differences presented in the major groove in tRNAs with either G-C or C-G pairs, it is thought that eukaryotic and archaeal TyrRS recognises its cognate tRNA via the major groove of the acceptor stem (Fechter *et al.* 2000; Fechter et al. 2001; Kobayashi *et al.* 2003).

Comparing the crystal structure of *Mj*TyrRS in complex with $tRNA^{Tyr}$ and tyrosine (Kobayashi et al. 2003; Zhang *et al.* 2005b), yeast (Fechter et al. 2000; Fechter et al. 2001), human (Yang *et al.* 2002) and prokaryotic TyrRSs (Brick *et al.* 1989; Yaremchuk et al. 2002), showed that there were similarities between human and *Mj*TyrRS, but significant differences between *Mj* and prokaryotic TyrRSs. The bacterial TyrRS has an extra C-terminal domain which recognises the long variable domain of its cognate $tRNA^{Tyr}$ (Yaremchuk et al. 2002). While the long variable arm is characteristic of prokaryotic $tRNA^{Tyr}$-TyrRS recognition, the short variable arm of eukaryotic and archaeal $tRNA^{Tyr}$ is not (Marck and Grosjean 2002).

The anticodon region has also been shown to be important in $tRNA^{Tyr}$-TyrRS recognition, but varies between kingdoms (Kobayashi et al. 2003). The portion that recognises the anticodon loop is similar between *Mj* and human, but varies for prokaryotes (Figure 7A). Prokaryotic TyrRS differ distinctly from archaeal and eukaryotic TyrRS in tRNA anticodon recognition. For prokaryotic systems, it has been shown that the central base of the anticodon, ψ35 (flipped out in comparison to G34 and A36, which are stacked), is the major element of recognition and is recognised by the extra C-terminal domain (Himeno *et al.* 1990; Hou and Schimmel 1989; Sherman *et al.* 1992; Yaremchuk et al. 2002). In contrast, in archaeal and eukaryotic $tRNA_{Tyr}$ all three anticodon bases are on the same side and G34 is stacked with hydrophobic residues of the TyrRS. However, G34 is recognised by the carboxyl group of Asp in all TyrRSs.

Residues Tyr32 and Asp158 within *Mj*TyrRS are responsible for tyrosine binding identity (Figure 7B). Y32 and D158 coordinate the p-hydroxyl group of the phenyl ring of tyrosine. Glu36, Tyr151, Gln155, Gln173 of *Mj*TyrRS are responsible for coordinating the amino and carboxylic acid portion of tyrosine (Zhang et al. 2005b). Equivalent residues within prokaryotic TyrRS are responsible for recognition of Tyr. Bedouelle et al (1993) have shown that Glu 152 of the prokaryotic TyrRS acts as a negative discriminant towards non-cognate tRNAs by electrostatic and steric repulsion.

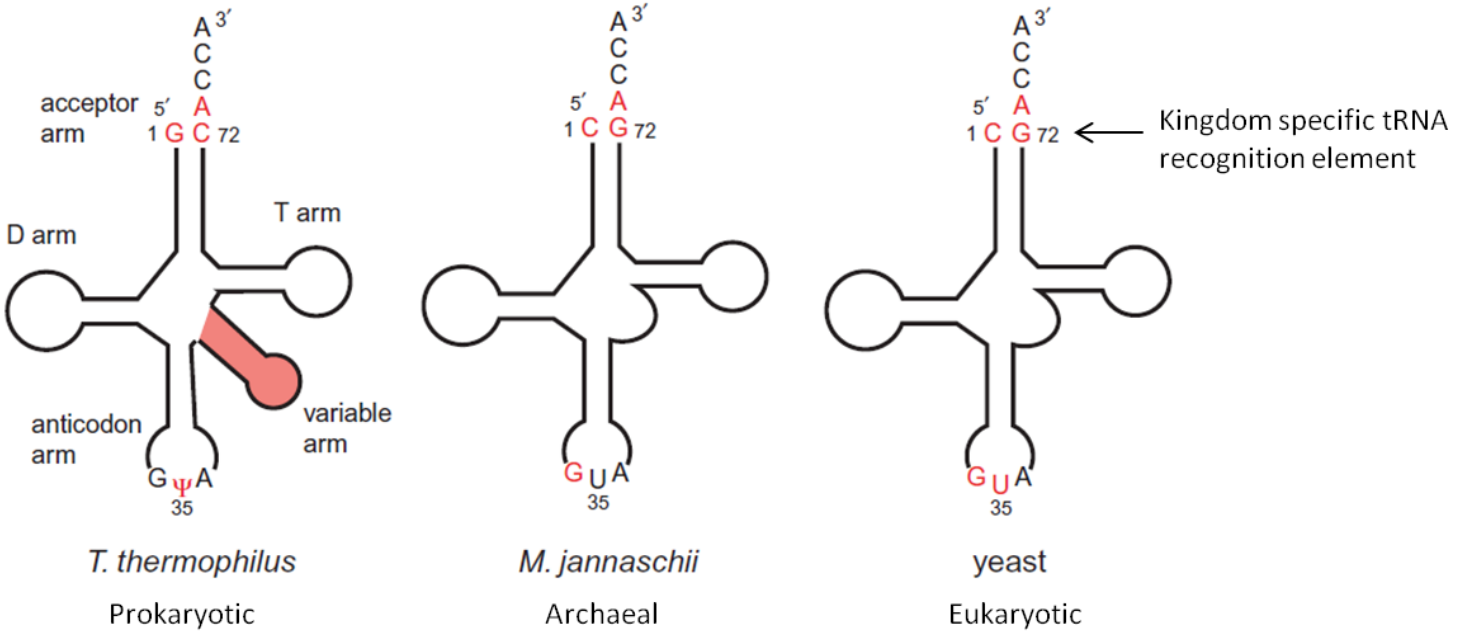


Figure 6. Cloverleaf structures of tRNA^{Tyr} showing key identity elements (in red) relevant to the three kingdoms of life. (Adapted from Kobayashi et al. 2003)

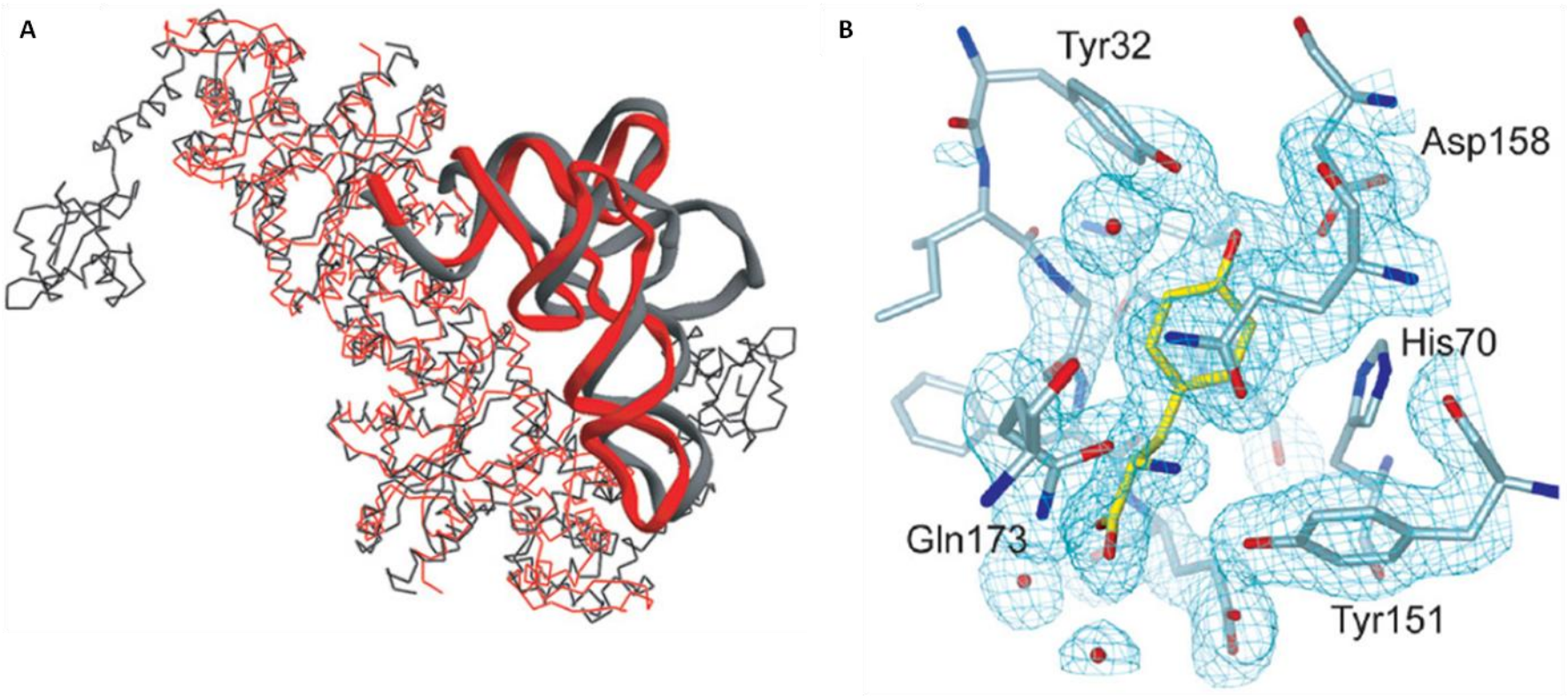


Figure 7. A) Comparison of overall structures of archaea (*M. jannaschii;* red; PDB 1J1U) and bacterial (*T. thermophilus;* black; PDB 1H3E) TyrRS-tRNA^{Tyr} (taken from Kobayashi et al. 2003). Only one tRNA is shown.

B) Coordination of Tyr within AA binding site of bacterial TyrRS-tRNA^{Tyr} showing residues important in discriminating Tyr (taken from Kobayashi et al. 2003).

1.10 Methods to incorporate NAAs

While genetically encoded amino acids can also be modified post-translationally to aid in increasing the variety in protein structure and function, this limitation in building material for protein construction constricts the biological, chemical and physical properties a protein can possess. Therefore, the incorporation of non-natural amino acids containing moieties with distinct biological, chemical or physical properties into proteins can expand the chemical functionality of a protein (for review see Dougherty 2000).

1.10.1 Chemical synthesis

There have been a number of methods developed to increase the range of amino acids that can be incorporated into proteins (for review see Wang and Schultz 2005). Chemical modification, for example via conjugation to cysteine residues, can be used to add non-natural moieties to proteins. For example, conjugation of a photoactivatable cross-linking agent to study protein-protein interactions (Chen *et al.* 1994). This can be a site-specific method as site-directed mutagenesis can be used to introduce cysteine residues into proteins, which are then modified. This technique has limited efficiency and uses a naturally occurring AA, which can create problems if the target protein contains multiple cysteines or requires cysteines to function.

Chemical synthesis of peptides can be used to incorporate NAA into proteins. The method of stepwise solid-phase peptide synthesis (SPPS) (Kent 1988; Merrifield 1963), has been exploited to incorporate NAAs into proteins. During this method, amino acids are chemically added to a growing chain immobilised on a bead. This method is particularly useful if toxic AAs or AAs not compatible with translational machinery are required. However, the method has a number of limitations, namely, the length of peptide is limited to <100 AAs, the peptide often has poor solubility and the method gives poor yields and is time-consuming.

1.10.2 Semisynthesis

To try to increase the length of peptide produced, native chemical ligation (NCL) was used to attach peptide fragments to one another (Dawson *et al.* 1994), for example protein expressed using recombinant systems (Offord 1987; Wallace 1995). NCL has been utilised to incorporate a variety of NAAs into cytochrome c (Dyckes, Creighton and Sheppard 1974, Wallace and Clark-Lewis 1992, Wallace 1993). NAAs were incorporated into cytochrome c using autocatalytic re-ligation of peptide fragments generated by cyanogen bromide (CNBr) cleavage. CNBr cleavage results in a C-terminal homoserine lactone, which reacts with a free N-terminus to reform the full-length protein.

The technique termed expressed protein ligation (EPL) utilises C-terminal thioesters generated from thiol-cleavage of an intein-mediated protein-splicing reaction in which a mutation in the intein blocks the final splicing step (Muir *et al.* 1998; Severinov and Muir 1998). The resulting α-thioester derivative can subsequently react with an N-terminal Cys, homocysteine (Roy *et al.* 1999) or selenocysteine peptide (Hondal *et al.* 2001). Numerous applications using EPL have been explored, for example, the incorporation of fluorescent moieties, spin-labels, FRET partners and ketone-containing constructs (Ayers *et al.* 1999); and incorporation of post-translationally modified AAs, for example, phosphotyrosine to study the signalling of a protein tyrosine kinase C-terminal, Src kinase (Muir et al. 1998). Semisynthetic techniques, however, do require cleavage and ligation sites and cannot be applied to *in vivo* study of protein.

1.10.3 *In vitro* biosynthesis

In vitro techniques utilising the existing protein biosynthetic machinery to incorporate NAAs in to protein have been established (Railey White *et al.* 2012). Chemical modification of enzymatically aminoacylated tRNAs were used to incorporate NAAs with synthetic reagents and probes. Modification of this technique, using truncated tRNAs (with the 3'-terminal mono- or dinucleotide removed) allowed attachment of mono- and dinucleotides chemically

aminoacylated with a variety of NAAs (Baldini *et al.* 1988). However, this technique gave poor yields, NAA incorporation was in competition with natural AA incorporation, and the NAA could be incorporated genome-wide. To try to improve the site-specificity of NAA incorporation, the *in vitro* method was augmented to incorporate chemically aminoacylated tRNAs in response to stop codons (Noren *et al.* 1989). An orthogonal amber suppressor tRNA derived from yeast phenylalanine-tRNA was constructed. Noren et al, (1989) used this technique to incorporate NAAs *p*-nitrophenylalanine, *p*-fluorophenylalanine and homophenylalanine in to Phe66 of β-lactamase. The yield was small, but the specificity and fidelity of NAA incorporation was shown to be high. Over 50 amino acids, with a variety of side chains for a variety of applications, have been incorporated site-specifically into proteins using this *in vitro* technique. Extensions of this method have been used to explore the incorporation of more than one NAA (Ozawa *et al.* 2012), use of frameshift suppression of four-base codons (Hohsaka and Sisido 2002; Taki *et al.* 2002); use of unnatural bases within the codon (Bain *et al.* 1992); microinjection of mutant RNA and chemically aminoacylated yeast $tRNA_{CUA}^{Phe}$ into *Xenopus* oocytes (Nowak *et al.* 1995). However, all the *in vitro* methods to incorporate NAAs into proteins are technically demanding, give poor yields (as the supply of chemically aminoacylated tRNAs is non-renewable), and the stability of the aminoacyl-tRNA linkage limits the NAAs that can be used.

1.10.4 *In vivo* biosynthesis

The possibility of utilising *in vivo* expression to incorporate NAAs into protein has been explored. Supplementing the growth media of an auxotrophic strain (cannot biosynthesise a specific amino acid) with an analogue of the amino acid allows incorporation of this analogue into proteins (Richmond 1962). For example, supplementing a Phe auxotrophic strain of *E. coli* with *p*-fluorophenylalanine allowed incorporation of this NAA into β-galactosidase (Cohen and Munier 1956); selenomethionine was incorporated using a methionine auxotroph (Cowie and Cohen 1957) and various replacements of hydrogen with fluorine, methylene with oxygen or sulphur, and ring substitutions have all been performed (Brawerman and

Yčas 1957; Cowie *et al.* 1959; Rennert and Anker 1963; Richmond 1959). This method relies on the close analogy between natural AA and NAA to be substituted. Therefore, to try to expand the variety of NAAs that can be used, the specificity of the aaRSs was relaxed using mutagenesis. The A294G mutation in the PheRS increased the substrate binding pocket size allowing it to accommodate and subsequently incorporate a variety of Phe analogues, including p-bromo-, p-iodo-, p-cyano-, p-ethynyl-,and p-azidoPhe (Kirshenbaum *et al.* 2002; Sharma *et al.* 2000). Another method to increase the number of NAAs that can be incorporated included the attenuation of the proofreading ability of RSs (Döring *et al.* 2001). Among a variety of applications, using aaRSs with relaxed substrate specificity has been particularly useful in incorporating NAAs that harbour heavy atoms, which aid in phase determination in X-ray crystallography. For example, incorporating selenomethionine in place of methionine (Yang *et al.* 1990a; Yang *et al.* 1990b). Once again, the method is not site-specific as incorporation of NAA occurs throughout the proteome, therefore NAA is required to be tolerated at all these positions limiting the variety of NAAs that can be used. Additionally, there is competition between the natural AA and it's NAA analogue, and yields are variable.

To overcome the limitations of the methods outlined above, the Schultz research team developed a method that allows genetic coding for NAAs resulting in the position specific incorporation of the specified NAA with high translational fidelity and efficiency (for review see Xie and Schultz 2006).

1.11 Expanding the genetic code

The genetic code allows incorporation of the 20 naturally occurring amino acids, and the rarely incorporated amino acids pyrrolysine and selenocysteine, with three translation termination codons (see section 1.5). Expansion of this genetic code is possible by introducing an aaRS and tRNA orthogonal to the endogenous machinery of the cell. The amber suppression method relies on the cellular process of translation (Figure 8). The basis of the method developed by Schultz's research group, is to recode the amber codon, UAG, to specify non-natural amino acids instead of terminating translation (Liu *et al.* 1997a; Liu *et al.* 1997b; Noren et al.

1989). The amber translational termination codon is a prime candidate for this recoding because, most notably, it is rarely used in *Escherichia coli* (9 % bacterial proteins, very rarely used in essential genes; 23% in Homo sapiens (Nakamura 2007)) and there are naturally occurring amber suppressors (Eggertsson and Soll 1988). For example, naturally occurring amino acids selenocysteine and pyrrolysine are incorporated into a growing peptide in response to stop codons when specific elements in the mRNA sequence and structure are present.

To utilise amber suppression, a TAG mutation is introduced site-specifically into the gene of choice; and a unique transfer-RNA (tRNA) accommodating a CUA anticodon within the anticodon loop and a corresponding aminoacyl-tRNA synthetase (aaRS) are provided to the cell (Wang *et al.* 2000). The tRNA must be constructed such that it is recognised by the orthogonal synthetase but not by endogenous synthetases. The orthogonality of the synthetase is required in order to specifically aminoacylate the orthogonal tRNA and not any endogenous tRNAs and furthermore, should exclusively aminoacylate the specified non-natural amino acid. The orthogonal tRNA and aaRS must also perform efficiently during translation, and the non-natural amino acid must be transported efficiently into the cytoplasm from the culture medium and not confer toxicity to the cell.

To achieve amber suppression in *E. coli* cells the tyrosyl tRNA ($tRNA_{Tyr}$)/synthetase pair from the archaea *Methanococcus jannaschii* were imported orthogonally in to the *E. coli* (Wang et al. 2000). The *M. jannaschii* tyrosyl aaRS (*Mj*TyrRS) and $tRNA_{Tyr}$ have properties that are favourable to produce an orthogonal pair for non-natural amino acid incorporation. In particular, *Mj*TyrRS has recognition elements for the C1:G72 tRNA eukaryotic recognition element as opposed to the G1:C72 tRNA prokaryotic recognition element (Figure 9), therefore it does not recognise endogenous *E. coli* tRNA (Fechter et al. 2001). In addition, the identity elements outside the anticodon region are mainly responsible for recognition of the *Mj*$tRNA_{Tyr}$ by *Mj*TyrRS (Fechter et al. 2001; Steer and Schimmel 1999), thereby allowing mutation of the anticodon loop to allow amber suppression without perturbing its recognition by *Mj*TyrRS (Wang et al. 2000). Also to note is that

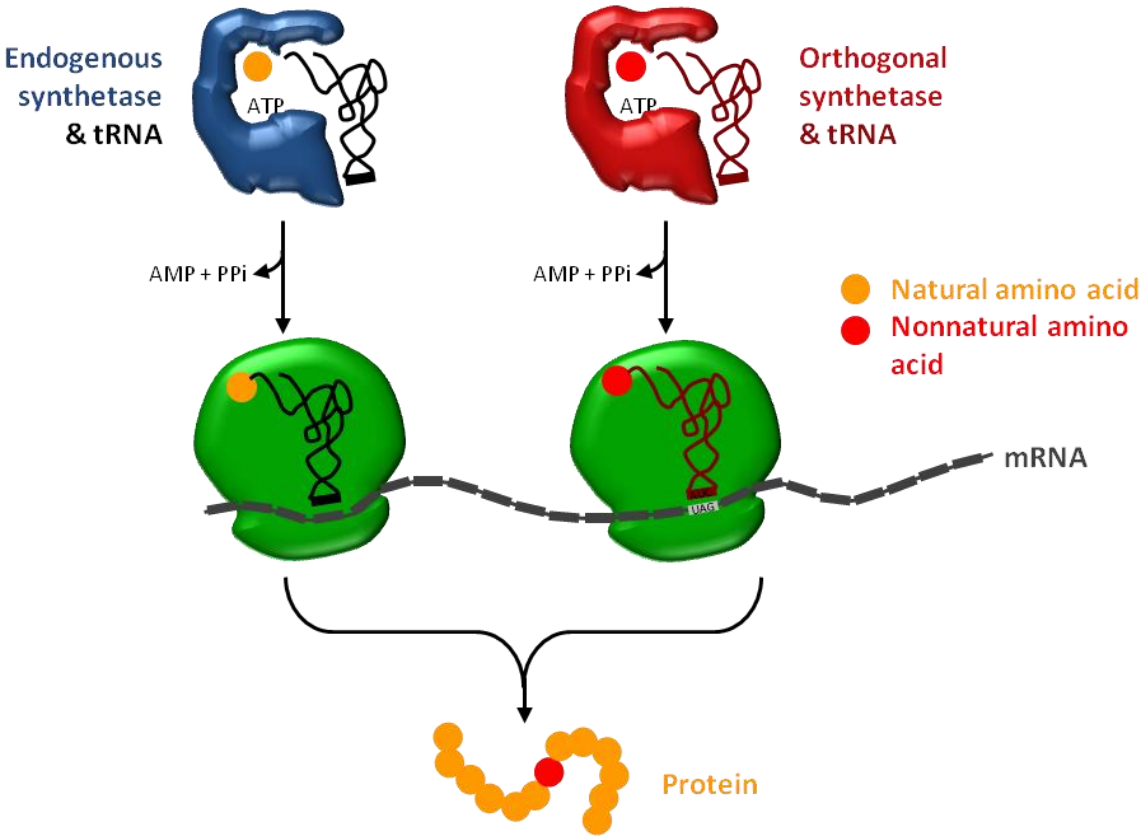


Figure 8. Endogenous translation compared to orthogonal system used for incorporation of non-natural amino acid during amber suppression. Process of translation showing the reading of mRNA (gray) by the ribosome (green), to code for amino acids (yellow - natural; red - nonnatural), using tRNA charged with amino acid by a corresponding aminoacyl tRNA synthetase.

*Mj*TyrRS has no editing mechanism and therefore aminoacylation is not subjected to proofreading (Jakubowski and Goldman 1992).

1.11.1 Production of tRNA$^{\mathbf{NAA}}_{\boldsymbol{CUA}}$

As noted above, the identity determinants of the anticodon region of tRNA^{Tyr} are not crucial in recognition, and the presence of naturally occurring amber suppressor tRNA^{Tyr} were important in the expansion of the genetic code using the amber suppression technology as developed by Schultz et al.

A fully orthogonal suppressor tRNA$_{CUA}$ for use in the expanded genetic code application developed by Schultz et al was required. Wang et al (2000) have shown that an orthogonal amber suppressor can be generated using the *Mj*tRNA$^{Tyr}_{GUA}$ by mutating the anticodon to CUA, creating *Mj*tRNA$^{Tyr}_{CUA}$, (Figure 9).

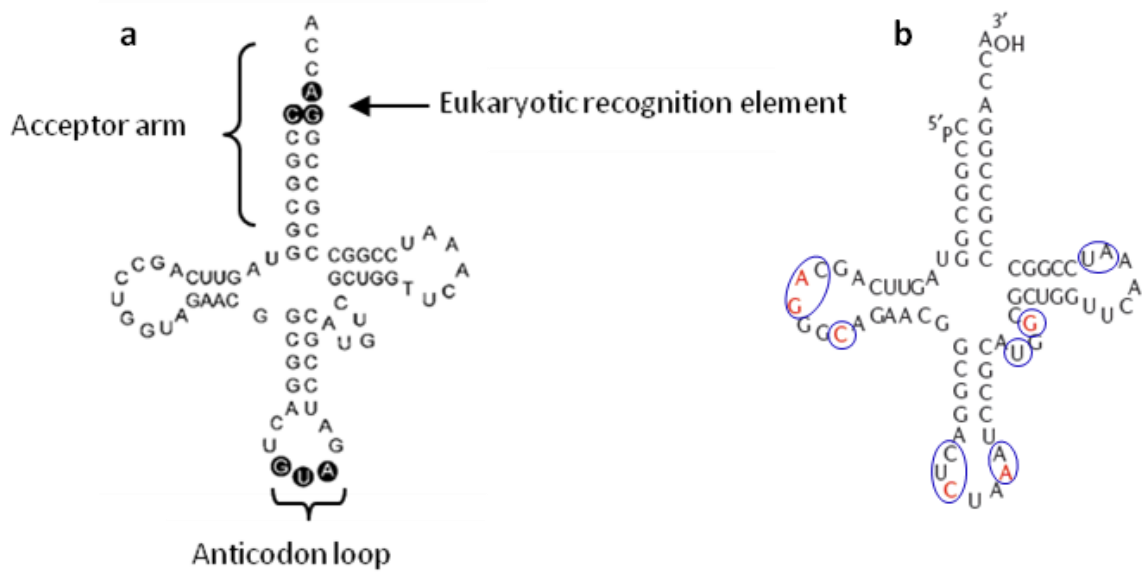


Figure 9. (a) *Methanococcus jannaschii* tyrosyl tRNA, (adapted from Wang and Schultz 2005).

(b) Orthogonal amber suppressor *Mj*tRNA$^{Tyr}_{CUA}$ (adapted from Xie and Schultz 2006). Mutated residues are indicated in red. Residues involved in tRNA library formation are indicated in blue circles. These residues were randomised to optimise orthogonality of *Mj*tRNA$^{Tyr}_{CUA}$ within *E. coli* (Wang and Schultz 2001).

However, *Mj*tRNA$_{CUA}^{Tyr}$ was shown to be recognised by endogenous RSs in *E. coli* (Wang et al. 2000), therefore further mutations were required (Wang and Schultz 2001). Two libraries were created, one where only residues in the anticodon loop were mutated and a second where non-conserved residues within the loops were mutated (Figure 9). A two-step selection was employed to select orthogonal tRNAs (Wang *et al.* 2001). A negative selection step, where a plasmid encoding the toxic ribonuclease barnase gene harbouring a TAG at codon 3 was used to select against *Mj*tRNA$_{CUA}^{Tyr}$s that were aminoacylated by endogenous *E. coli* aaRSs (Figure 10). The *Mj*tRNA$_{CUA}^{Tyr}$ would be aminoacylated by an endogenous aaRS, with a natural amino acid then subsequently incorporated, to produce full-length barnase, resulting in cell death. The surviving cells would harbour an orthogonal or non-functional *Mj*tRNA$_{CUA}^{Tyr}$. A subsequent positive selection step utilised the *Mj*tRNA$_{CUA}^{Tyr}$-containing plasmids along with a plasmid encoding β-lactamase harbouring a TAG mutation and an *Mj*TyrRS. The *E. coli* were grown on ampicillin containing plates and therefore only cells harbouring a functional *Mj*tRNA$_{CUA}^{Tyr}$ that can be recognised and aminoacylated by *Mj*TyrRS could produce full-length functional β-lactamase and survive. Using this method, an *Mj*tRNA$_{CUA}^{Tyr}$was selected for incorporation of NAAs (Figure 9B).

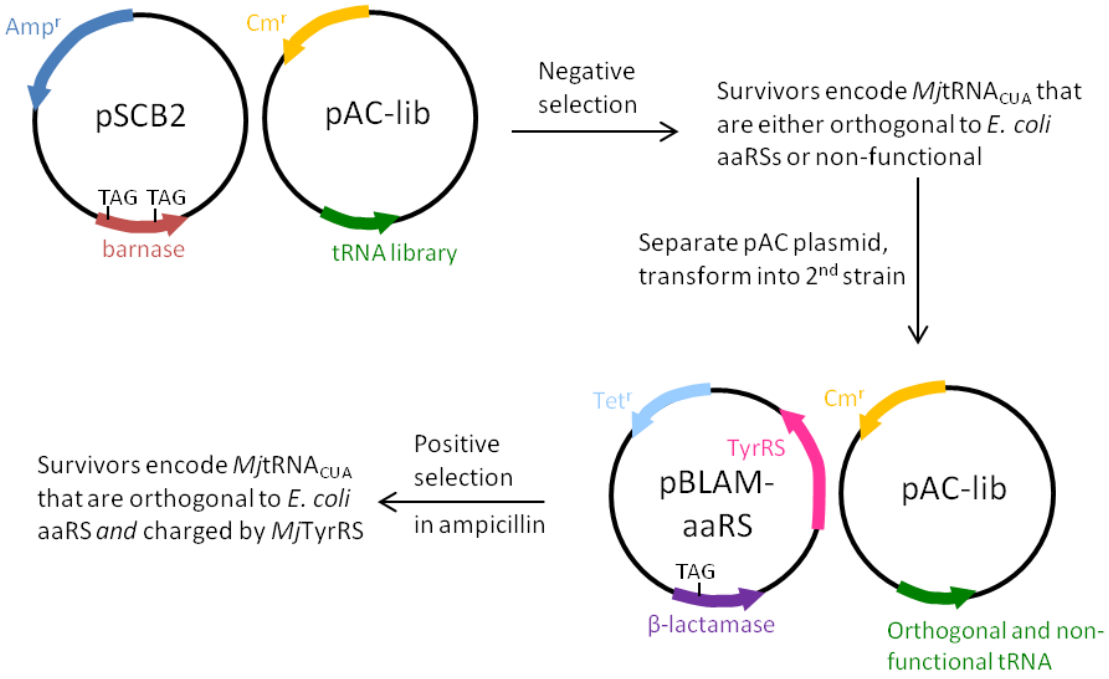


Figure 10. Two step method to select orthogonal *Mj*tRNA$_{CUA}^{Tyr}$ (Wang and Schultz 2001)

1.11.2 Production of aaRSs specific to NAAs

*Mj*TyrRSs that selectively charge mRNA with a specific non-natural amino acid (not tyrosine or any other natural amino acid), but maintain specificity to the tRNA, have been evolved in the process of directed evolution (Wang et al. 2000). The structure of the active site of *Mj*TyrRS was evaluated and residues that were responsible for coordinating Tyr were considered for mutation (Figure 11). Originally the structure of the analogous *Bacillus stearothermophilus* TyrRS (Brick et al. 1989) was used because the structure of *Mj*TyrRS had not been solved.

Residues Tyr32, Asn123, Asp158, Ile159, and Leu162 were subjected to random mutagenesis (via NNK mutagenesis). The library of mutant TyrRSs were subjected to a method of selection via iterative application of a two step process (Figure 12). A positive selection step utilised an Asp112TAG mutant of the chloramphenicol acetyltransferase (CAT) gene along with the plasmid encoded mutant *Mj*tRNA$_{CUA}^{Tyr}$ and library of mutant *Mj*TyrRS. Cells containing these plasmid constructs were grown in the presence of the chosen NAA. Surviving cells contained an *Mj*TyrRS mutant that was able to aminoacylate *Mj*tRNA$_{CUA}^{Tyr}$ with amino acid (either natural or NAA). These surviving colonies were replica plated, one grown in the absence of NAA, and the other in its presence. Those that did not survive in the absence of NAA were isolated from the replica plate as these colonies encoded mutant TyrRSs which aminoacylated the orthogonal *Mj*tRNA$_{CUA}^{Tyr}$ with NAA and not any natural amino acid. The subsequent negative step utilised a barnase gene with numerous TAG mutations. The plasmids containing the positively selected library of mutant *Mj*TyrRS and mutant *Mj*tRNA$_{CUA}^{Tyr}$were transformed into *E. coli* and grown in the absence of NAA. Mutant *Mj*TyrRSs not specific to NAA will charge *Mj*tRNA$_{CUA}^{Tyr}$ with natural amino acid and suppress the TAG mutations, producing toxic barnase that will result in cell death (Chin *et al.* 2002b; Pastrnak *et al.* 2000; Wang et al. 2001; Wang and Schultz 2005).

A number of aaRSs have been evolved to incorporate numerous NAAs. An aaRS specific for iodophenylalanine (iodoF) (Xie *et al.* 2004), azidophenylalanine (azF)

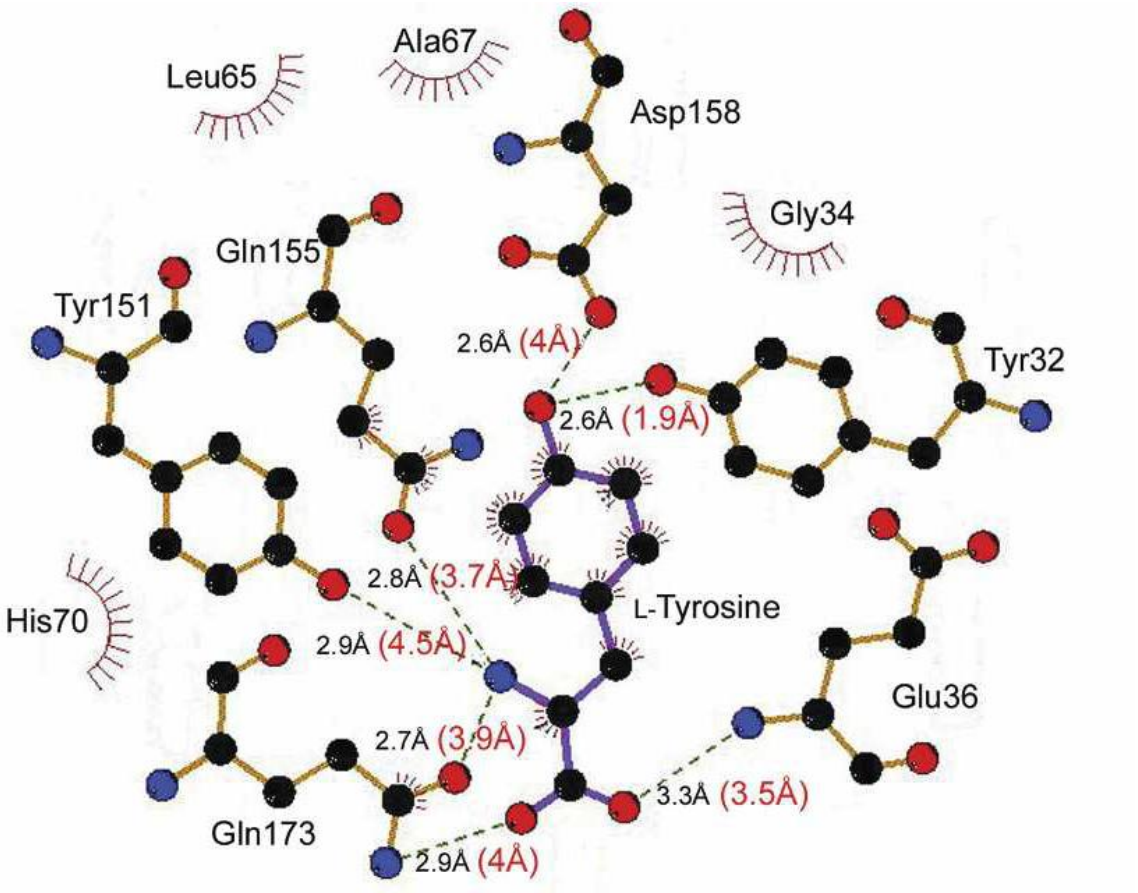


Figure 11. Schematic representation of a tyrosine bound to *Mj*TyrRS active site. The residues within the binding pocket are shown. It was these residues that were mutated in order to form aaRSs for specific NAAs. Hydrogen bonds are shown with dashed lines and the distances are shown (Taken from Zhang et al. 2005b).

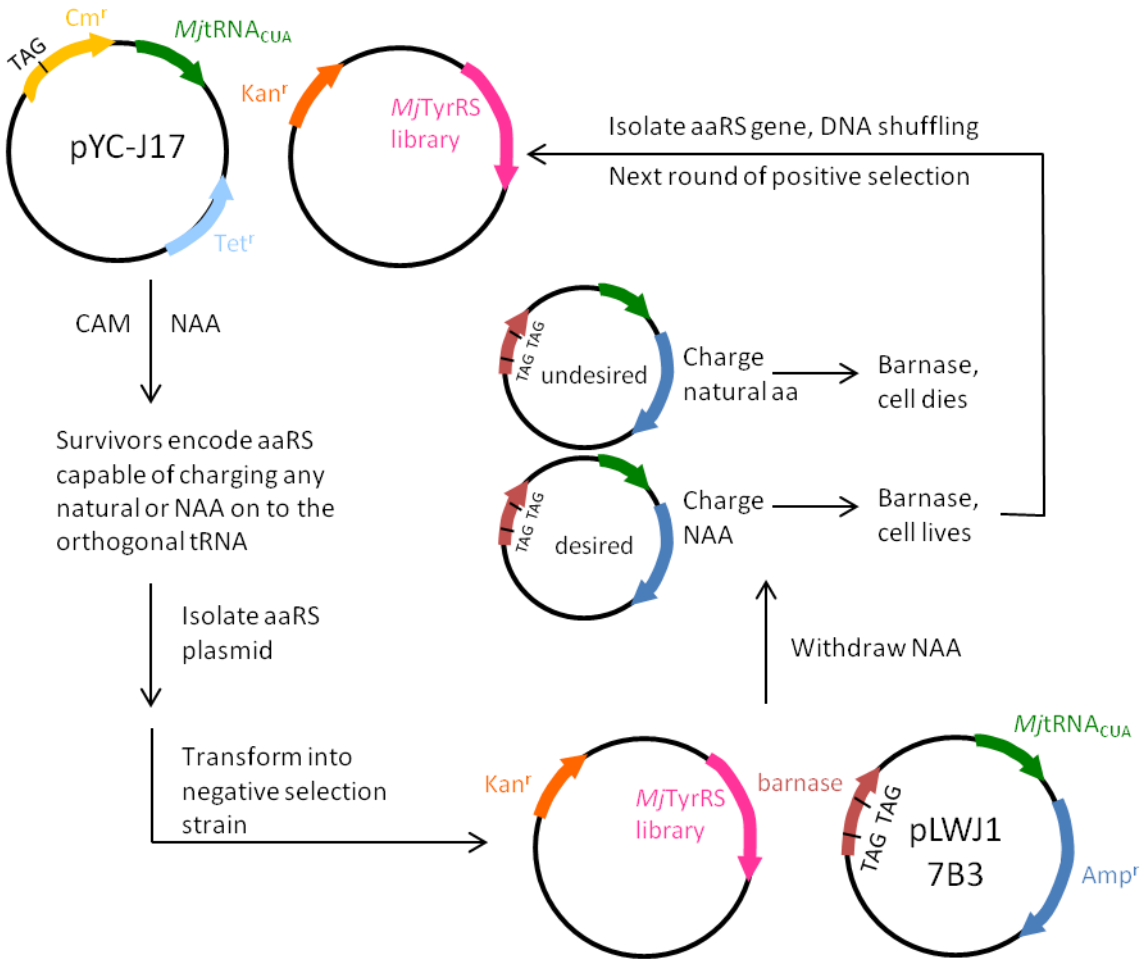


Figure 12. Two step system for selecting aaRSs specific to a NAA for use in amber suppression technology in *E. coli*.

(Chin *et al.* 2002a) have previously been engineered and are used in the experimental work discussed within this Thesis.

1.11.3 Applications of the amber suppression technology in *E. coli*

It has previously been shown that the amber suppression method makes it possible to incorporate a wide variety of non-natural amino acids into specified positions in a variety of proteins with high translational fidelity and efficiency (Wang et al. 2001; Xie and Schultz 2006). As a proof of principle, the close tyrosine analogue *O*-methyl-*L*-Tyrosine (Figure 13, **3**) was the first NAA to be incorporated using this method (Wang et al. 2001). Subsequently, novel chemical, physical and biological properties can be introduced in to a protein via the introduction of NAAs. This allows selective conjugation to peptic or nonpeptic molecules, for example dyes, drugs, polymers and carbohydrates; the study of protein-protein interactions and protein motions; the alteration of protein characteristics; study of protein localisation; the study or alteration of enzyme reactivities; and the introduction of therapeutic effect. Therefore, this technology has been applied to a variety of applications; a few of which are detailed below.

1.11.3.1 Post-translational modification (PTM) mimics

The NAA *p*-Carboxymethyl-l-phenylalanine (Figure 14, **13**) was incorporated into human signal transducer and activator of transcription-1 (STAT1) protein and used to mimic phosphotyrosine, which is important in signalling pathways. NAA *p*-Carboxymethyl-l-phenylalanine is resistant to hydrolysis, unlike phosphotyrosine and therefore eased functional analysis (Xie, Supekova and Schultz 2007). Another PTM of proteins, sulfation, was mimicked by incorporating sulfotyrosine (Figure 15, **34**) into human thrombin. The sulfation of proteins is not well understood and it was suggested that genetically encoding sulfotyrosine could elucidate its function (Liu and Schultz 2006). The glycosylation of proteins is a PTM that can have important therapeutic applications. Unlike protein expression using *E. coli*, expression using mammalian cells produces glycosylated proteins. It would be beneficial to be able to produce glycosylated protein using the more cost effective *E. coli* expression system. Despite the attempt to incorporate glycosylated NAAs

(GlcNAc-Ser and GalNAc-Thr) into proteins using amber suppression technology (Xu *et al.* 2004; Zhang *et al.* 2004), it was shown not to be possible (Antonczak *et al.* 2009).

1.11.3.2 Modulation of protein function

The modulation of protein function has also been studied by site-specifically incorporating NAAs. Tyrosine 66 within the chromophore of green fluorescent protein (GFP) has been modified with *p*-amino-L-phenylalanine (Figure 14, **2**), *p*-methoxy-L-phenylalanine (Figure 13, **3**), *p*-iodo-L-phenylalanine (Figure 13, **4**), *p*-bromo-L-phenylalanine (Figure 13, **5**), and L-3-(2-naphthyl)alanine (Figure 13, **6**). These NAAs altered the absorbance and emission spectra of the resulting GFP (Wang *et al.* 2002). The bioluminescent photoprotein aequorin was also subjected to modulation of spectral characteristics via incorporation of the same NAAs in the H-bonding residue position (Y82) within the chromophore-binding site (Rowe *et al.* 2010).

The mechanism and modulation of enzymes has also been attempted using site-specific NAA incorporation. For example, the radical reaction ribonucleotide reductase was studied by the incorporation of 3-nitro-tyrosine (Figure 14, **14**) (Yokoyama *et al.* 2010) and 3-amino-tyrosine (Figure 14, **18**) (Minnihan *et al.* 2009). The reaction of Aristolochene Synthase has also been probed using site-specific incorporation of NAAs with increasing ring complexity in place of Trp 334 (Faraldos et al. 2011). The redox-active amino acid 3,4-dihydroxy-l-phenylalanine (Figure 14, **9**) was incorporated into protein, and provided a means to monitor and control electron transfer in proteins, it was suggested that this could facilitate the study of electron transfer in proteins, as well as enable the engineering of redox proteins with novel properties (Alfonta *et al.* 2003). The photoisomerisation of genetically encoded NAA phenylalanine-4'-azobenzene (Figure 14, **12**) has been shown to regulate the binding affinity of catabolite activator protein to its promoter (Bose *et al.* 2005). The incorporation of an α-hydroxy acid, p-hydroxy-L-phenyllactic acid, allowed non-natural mutagenesis of the protein backbone that can be used in site-specific backbone hydrolysis and probing of backbone stability (Guo *et al.* 2008a).

1.11.3.3 Structural determination

Site-specific incorporation of radioactive labelled NAA into proteins can aid in elucidating protein structure via NMR. For example, the NAAs 2-amino-3-(4-(trifluoromethoxy) phenyl) propanoic acid (OCF_3Phe, Figure 15, **25**), $^{13}C/^{15}N$-labeled *p*-methoxyphenylalanine (OMePhe; Figure 13, **3**), and ^{15}N-labeled o-nitrobenzyl-tyrosine (oNBTyr; Figure 15, **26**) were incorporated individually into 11 positions around the active site of the 33 kDa thioesterase domain of human fatty acid synthase. It was shown that this site-specific incorporation of labelled NAAs aided in structural determination of a large protein (Cellitti *et al.* 2008; Deiters *et al.* 2005; Jackson *et al.* 2007). ^{19}F-NMR has also been shown to be possible via the incorporation of NAA trifluoromethyl-L-phenylalanine (tfm-Phe; Figure 14, **22**). Incorporation of this NAA has aided in the study of substrate binding, inhibitors and cofactors, as well as in the reactions of the enzymes nitroreductase and histidinol dehydrogenase (Jackson et al. 2007). The heavy iodine atom within the NAA p-iodophenylalanine was used in single-wavelength anomalous dispersion phasing to elucidate the structure of T4 Lysozyme (Xie et al. 2004).

1.11.3.4 Introduction of novel reactivity

The alkynyl group containing NAA *p*-propargyloxyphenylalanine (Figure 15, **30**) and the aryl azide containing NAA *p*-azidophenylalanine (azF; Figure 15, **31**) show similar reactivity and applications. The azide and alkyne groups are able to form triazoles by copper(I) catalysed cycloaddition reactions and therefore a variety of groups can be conjugated onto the proteins in a site-specific manner (Chin et al. 2002a). For example, azF has been incorporated into proteins to allow attachment, via click reaction, of a polyethylene glycol (PEG) molecule (Chin et al. 2002a; Deiters *et al.* 2004). The attachment of PEG is therapeutically important as its presence has been shown to increase protein stability and bioavailability (Veronese and Pasut 2005). Incorporation of other photoactivatable NAAs, namely, *p*-benzoylphenylalanine (Figure 14, **15**) (Chin et al. 2002b; Chin and Schultz 2002; Lee *et al.* 2009), 4'[3-(trifluoromethyl)-3H-diazirin-3yl]-phenylalanine (TFMD; Figure 14, **21**) (Tippmann *et al.* 2007) and p-(2-tetrazole)-phenylalanine (Figure 13, **8**) (Wang *et al.* 2010a) can also be used to introduce orthogonal chemical "handles" into proteins in a site-specific manner and allow mapping of peptide–protein interactions, allow

photocrosslinking of proteins, and the introduction of fluorescent probes. Fluorescent dyes have been conjugated to *p*-propargyloxyphenylalanine (Deiters and Schultz 2005). The NAA O-allyl-l-tyrosine (Figure 15, **33**) contains an allyl group which has been shown to be versatile in organic transformations and therefore allow conjugation of additional molecules (Zhang *et al.* 2002). The incorporation of NAA p-boronophenylalanine (Figure 15, **32**) also provides a method for selective chemical modification of proteins (Brustad *et al.* 2008a).

1.11.3.5 Probes

The reactive keto-containing NAA, p-acetylphenylalanine (Figure 14, **7**) and its derivatives, have been used to attach fluorophores to proteins, for example Alexa Fluor488 hydroxylamine derivatives (Wang *et al.* 2003; Zeng *et al.* 2006). The attachment of a fluorophore to a NAA and traditional incorporation at a cysteine residue could allow single molecule fluorescence resonance energy transfer (smFRET) and be used to study protein motions (Brustad *et al.* 2008b). The NAAs L-4-Cyanophenylalanine (pCNPhe; Figure 14, **16**) and 4-ethynylphenylalanine (pENPhe) have also been used in an attempt to monitor protein conformational changes using FRET. The two NAAs were incorporated in to the hydrophobic core (Phe153) of T4 Lysozyme and were positioned to allow FRET to occur between the NAAs and three local Tryptophans (Trps) (Miyake-Stoner *et al.* 2009). *p*-nitrophenylalanine (pNO2-Phe) was incorporated into GCN4 basic region leucine zipper (bZIP) protein and caused fluorescence quenching of a nearby Trp, allowing protein structure and function to be probed (Tsao *et al.* 2006). A another approach is to incorporate the fluorescent amino acid L-(7-hydroxycoumarin-4-yl)ethylglycine (Figure 14, **24**), which is sensitive to pH and polarity and can therefore be used to probe protein localisation and trafficking, protein conformation changes, and protein−protein interactions (Wang *et al.* 2006). NAAs have also been utilised to introduce novel moieties that can be investigated using Infrared (IR) spectroscopy, for example pCNPhe (Schultz *et al.* 2006), and azF (Ye *et al.* 2009).

1.11.3.6 Other applications

In *E. coli*, the orthogonal aaRS/tRNA pair developed for incorporation of these NAAs is derived from a *Methanococcus jannaschii* tyrosine aaRS/tRNA pair, therefore the

NAAs are all tyrosine derivatives, namely, are composed of a phenyl ring with varying substituents. However, other orthogonal pairs have been utilised to expand the chemically diversity of the incorporated NAAs, for example lysine (Guo *et al.* 2008b), leucyl (Wu *et al.* 2004) and pyrrolysine (Hancock *et al.* 2010; Mukai *et al.* 2008; Neumann *et al.* 2008) aaRS-tRNA pairs. Amber suppression technology has also been applied to yeast (Deiters *et al.* 2003), *C. elegans* (Greiss and Chin 2011; Parrish *et al.* 2012), Drosophila (Bianco *et al.* 2012; Mukai *et al.* 2010b) and mammalian cells. The use of the pyrrolysyl-tRNA synthetase (PylRS)–$tRNA_{CUA}$ pair from *Methanosarcina* species is of particular importance because it is orthogonal to synthetases and tRNAs in *E. coli*, yeast, mammalian cells and *C. elegans*.

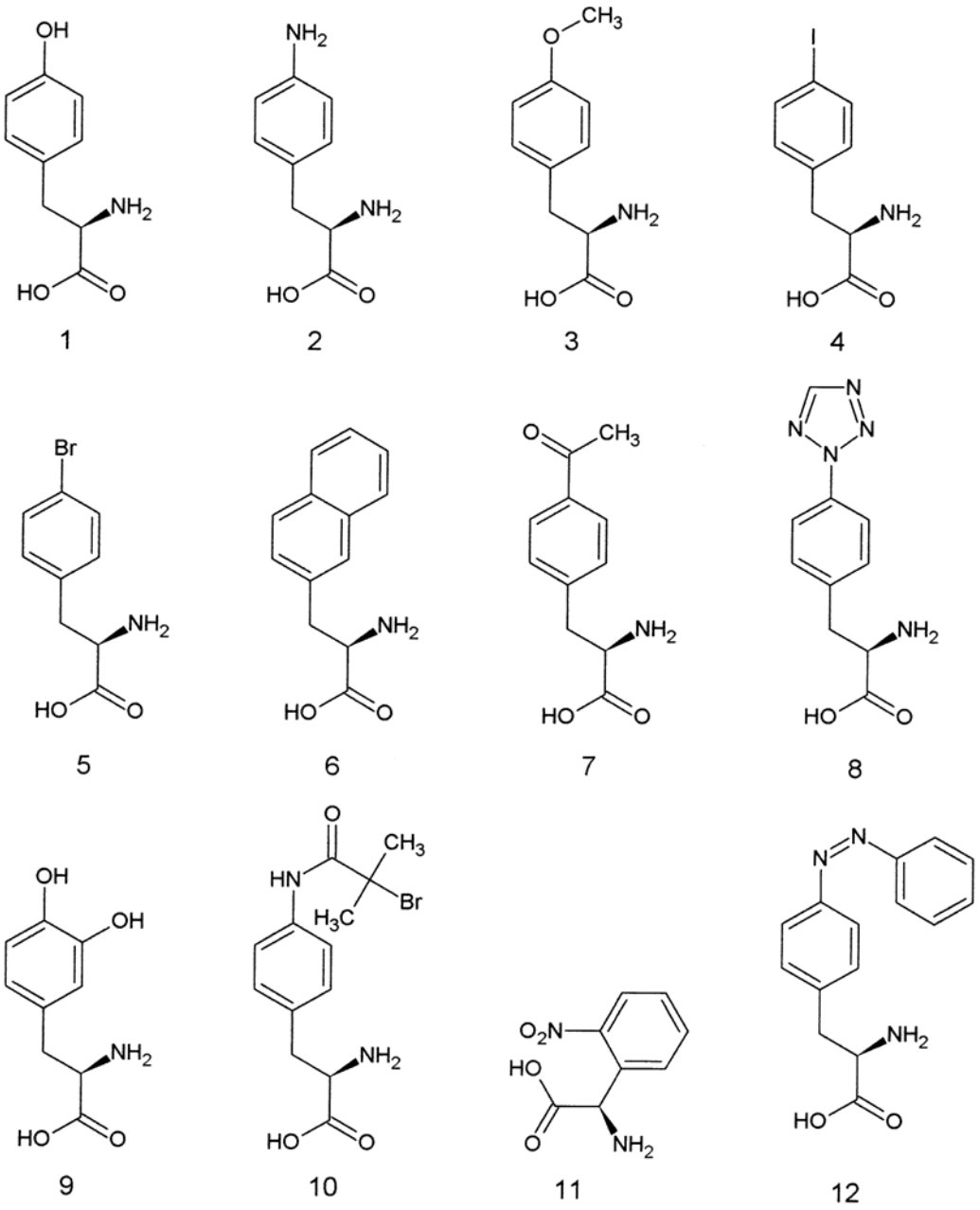


Figure 13. NAAs incorporated into proteins using the amber suppression method. NAA names are referred to within the main text. Drawn using Isis draw (MDL Information Systems).

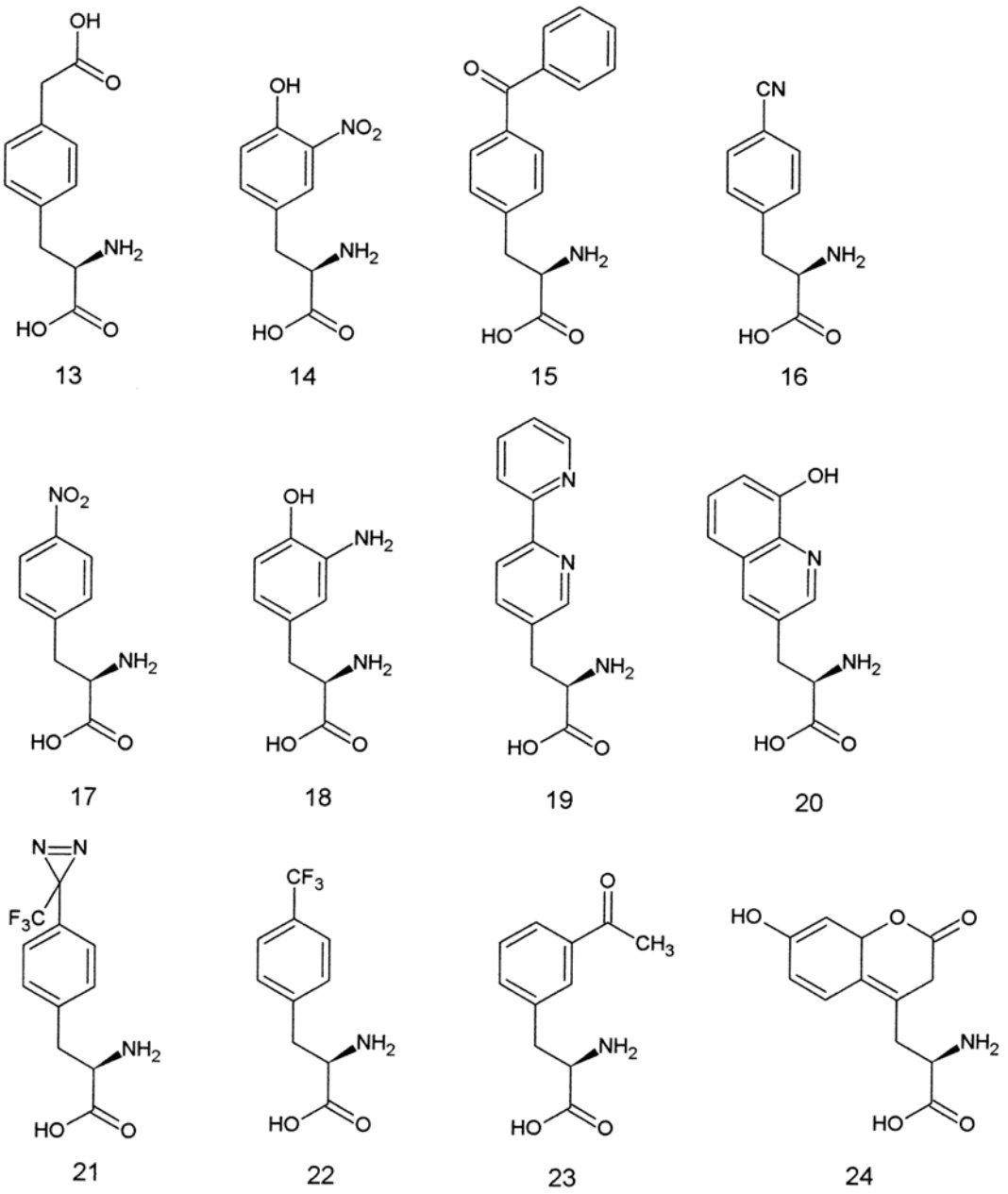


Figure 14. NAAs incorporated into proteins using the amber suppression method. NAA names are referred to within the main text. Drawn using Isis draw (MDL Information Systems).

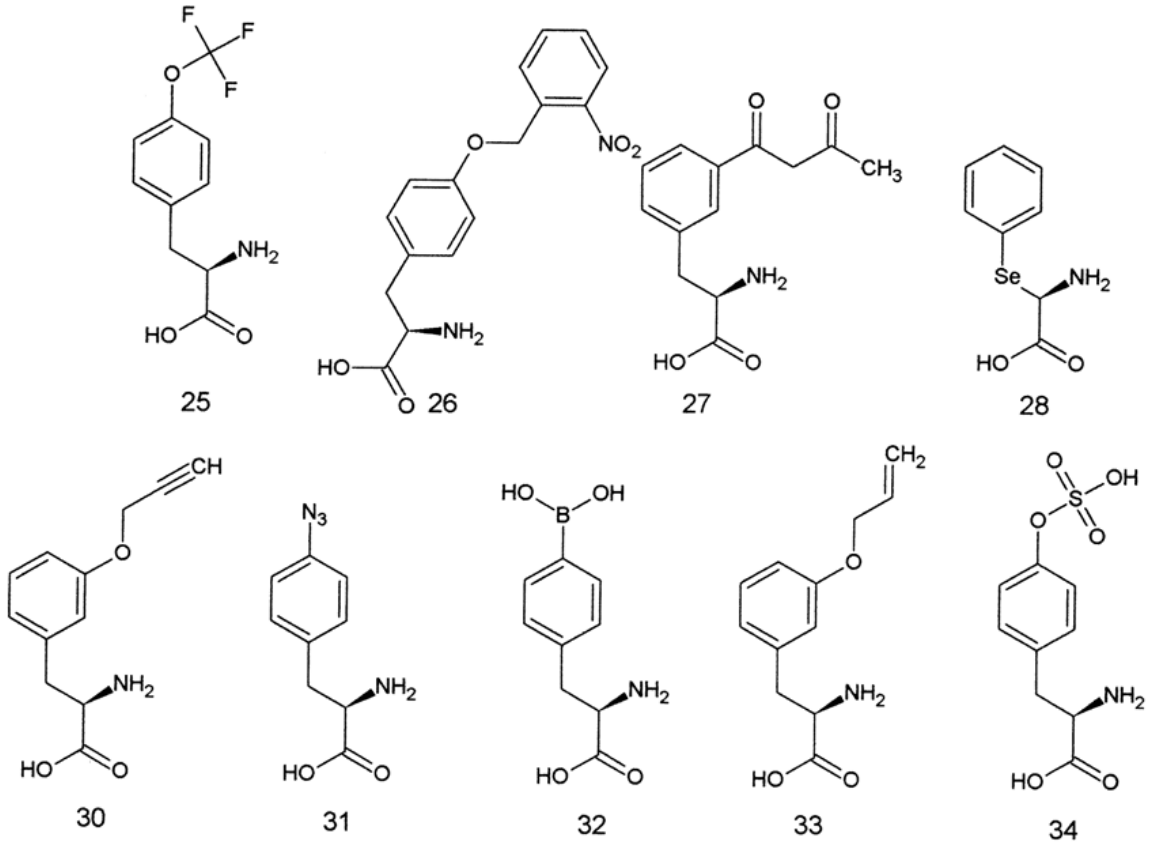


Figure 15. NAAs incorporated into proteins using the amber suppression method. NAA names are referred to within the main text. Drawn using Isis draw (MDL Information Systems).

1.11.4 Amber suppression technology in mammalian cells

As with the evolution of the amber suppression method in prokaryotes, the application of the method in eukaryotes started with the development of an aaRS/tRNA pair orthogonal to the endogenous eukaryotic aminoacylation system. As mentioned previously, the identity elements vary greatly between eukaryotes (and archaea) and prokaryotes. Therefore, importing tRNA/synthetase pairs from prokaryotes was initially used to develop the method.

1.11.4.1 Production of tRNA$^{NAA}_{CUA}$

Edwards and Schimmel (1990) demonstrated the orthogonality of the *E. coli* TyrRS/tRNA^{Tyr} pair within *Saccharomyces cerevisiae.* The work by Yokoyama's group made progress towards creating a method to incorporate NAA via amber suppression in eukaryotes by semi-rationally designing the *E. coli* TyrRS to recognise the NAA 3-iodo-L-tyrosine (Sakamoto *et al.* 2002). Initially a mutant *E. coli* TyrRS and tRNA$^{Tyr}_{CUA}$ orthogonal pair was used, as this system had previously been shown to be orthogonal and capable of NAA incorporation in a lower eukaryote (yeast) (Deiters et al. 2003). However, the mutant *E. coli* tRNA$^{Tyr}_{CUA}$ (*Ec*tRNA) does not contain the A box required for efficient transcription by RNA polymerase III in mammalian cells (Galli *et al.* 1981). Introduction of a pseudo-A box resulted in a non-functional tRNA that is not recognised by *Ec*TyrRS (Sakamoto et al. 2002; Wang *et al.* 2007b). *Bacillus. stearothermophilus* tRNA^{Tyr} (*Bs*tRNA^{Tyr}), however, does contain both the internal promoters (A and B box) in its native sequence, and the presence of these allows efficient expression within mammalian cells (Bedouelle 1990). Subsequently, mutant *E. coli* TyrRS and *B. stearothermophilus* tRNA$^{Tyr}_{CUA}$ were shown to incorporate 3-iodotyrosine within mammalian cells (Kiga *et al.* 2002; Sakamoto et al. 2002).

*Bs*tRNA$^{Tyr}_{CUA}$ was created by mutating the anticodon to CUA (Liu *et al.* 2007). Liu et al (2007) suggested this was possible because G34 of the anticodon is only a weak identity element for recognition by TyrRS (Hou and Schimmel 1989), therefore mutation to G34C did not perturb recognition by *Ec*TyrRS. Furthermore, it was

demonstrated that expression of *Bs*tRNA$^{Tyr}_{CUA}$ in mammalian cells could be enhanced by addition of 5' flanking sequence of human tRNA^{Tyr} (Liu *et al*., 2007). This mutant *Bs*tRNA in conjunction with the 5' flanking region of human tRNA^{Tyr} was therefore utilised within this Thesis.

1.11.4.2 Production of aaRSs specific to NAAs

Due to the complex nature of the mammalian system, aaRSs for NAAs were evolved in yeast. As for amber suppression in *E*. *coli*, a library of mutant aaRSs were created based on the five residues in the active site of the homologous TyrRS from *Bacillus stearothermophilus.* This library was introduced into the *Saccharomyces cerevisiae, (S. cerevisiae)* selection strain, MaV203:pGADGAL4 (2 TAG; uracil and histidine auxotroph) (Vidal *et al.* 1996) which were grown in the presence of NAA (and absence of uracil and histidine). In the positive selection step, NAA is incorporated by aaRS to produce full-length functional transcriptional activator GAL4 and subsequent activation of the GAL4- responsive *HIS3*, *URA3*, and *lacZ* reporter genes (Figure 16). Expression of these genes allows the cell to biosynthesise histidine and uracil and therefore survive. During the subsequent negative selection step, the cells are grown on media containing 5-fluorootic acid (no NAA). Those cells expressing URA3, as a result of suppression of the GAL4 amber mutations with natural amino acids, convert 5-FOA to a toxic product, killing the cell (Chin *et al.* 2003a; Chin *et al.* 2003b).

aaRSs for the NAAs, *p*-acetyl-L-phenylalanine, *p*-benzoyl-L-phenylalanine, *p*-azido-L-phenylalanine, O-methyl-L-tyrosine, and *p*-iodo-L-phenylalanine, have been evolved using this approach (Hino *et al.* 2005; Liu et al. 2007). The mutant *E*. *coli* azRS was utilised within this Thesis.

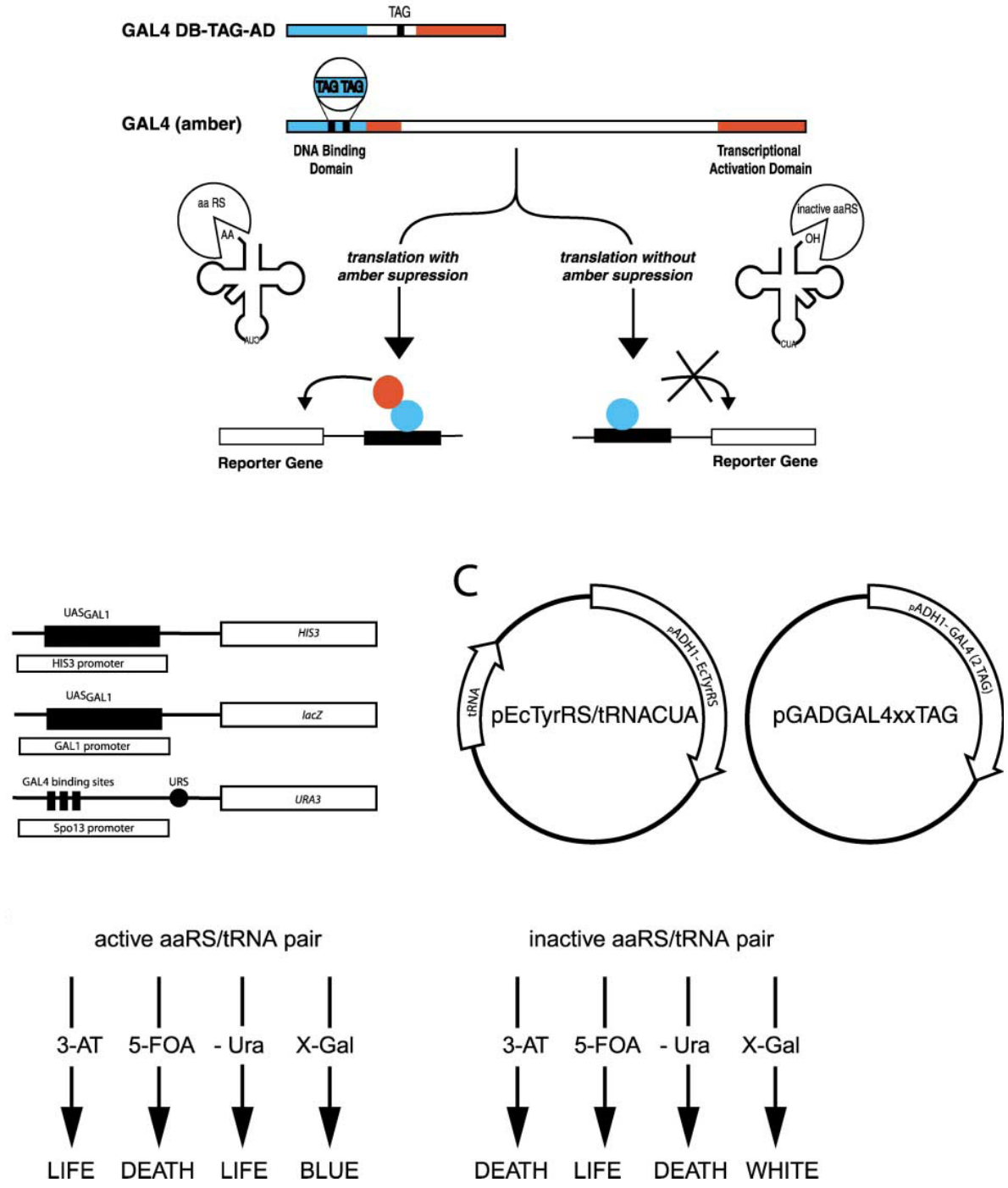


Figure 16. Two step system for selecting aaRSs specific to a NAA for use in amber suppression technology in mammalian cells (Taken from Chin et al. 2003b).

1.11.4.3 Applications of the amber suppression technology in mammalian cells

As with the production of NAA-containing proteins using *E*. *coli*, the proteins expressed using mammalian cells have been used for a variety of applications. Those applications utilising the evolved TyrRSs will be considered below. Typically proteins not expressed well in *E*. *coli* or the study of proteins in-situ are situations where amber suppression technology in mammalian cells are required.

Liu et al (2007), demonstrated that the newly evolved *Ec*TyrRSs could incorporate cognate NAAs into GFP using CHO and HEK293T cells. The incorporation of NAA was analysed by purifying the expressed protein and subjecting the sample to nanoscale reversed-phase liquid chromatography/ tandem mass spectrometry.

Photocrosslinkable NAAs have been incorporated into proteins using mammalian cells in order to study *in situ* protein-protein interactions. Hino et al incorporated the photoactive NAAs, p-benzoyl-L-phenylalanine (Hino et al. 2005) and p-Trifluoromethyl-diazirinyl-L-phenylalanine (Hino *et al.* 2011) into the Src homology 2 (SH2) domain of the adaptor protein Grb2 using CHO cells. Upon irradiation with 365 nm light, Grb2 variants containing NAA in the positions proximal to the ligand-binding pocket could be cross-linked with the transiently expressed epidermal growth factor.

Wang et al (2007b), applied NAA technology to investigate neuronal signalling. Although the group did demonstrate NAA incorporation within mouse hippocampal neurons, the study of NAA-potassium channels was performed using protein expression in HEK293 cells. The NAAs were incorporated into the inactivation peptide of the potassium channel and used to study the bulkiness of the side chain on the rate of inactivation of the channel.

The research performed by Ye et al (2009; 2008; 2010) has utilised amber suppression in mammalian cells (HEK293 cells) to incorporate a variety of NAAs into, and aid in the study of G protein-coupled receptors (GPCRs). To introduce novel chemical reactivity into the proteins the keto-containing NAAs, p-acetyl-L-phenylalanine and (Acp) and *p*-benzoyl-L-phenylalanine (Bzp), were used. Conjugation of keto-reactive hydrazide-fluorophore allowed study of rhodopsin

protein by UV-visible absorbance spectroscopy before and after photobleaching (Ye et al. 2008). Incorporation of the azide-containing NAA, azF, has allowed further study by IR spectroscopy (Ye et al. 2009; 2010).

Takimoto et al (2009) showed that further engineering of the anticodon-binding domain of the mutant TyrRS, via the mutation D265R, improved amber suppression in HEK293 cells. To demonstrate that this aaRS could facilitate study of mammalian protein-protein interactions it was used to incorporate azF into *E*. *coli* glutathione S-transferase (GST), in order to mediate photoactivated cross-linking between the two dimers of the GST homodimer.

Shen et al (2011), used a lentiviral based system to stably deliver *E*. *coli* aaRS and the $tRNA_{CUA}^{NAA}$ (plus H1 promoter driven $tRNA_{CUA}^{NAA}$) expression cassette into neural stem cells (HCN-A94). Typically, the cells used for amber suppression are robust cell lines, namely, HEK293, HeLa or CHO cells. However, Shen et al, has shown that amber suppression can also be applied to specialised cells. Firstly, to test amber suppression in the HCN-A94 cells, the group showed that the aaRS specific for benzoylphenylalanine (evolved *E*. *coli* TyrRS) could be used to incorporate benzoylphenylalanine into GFP. For a more functional application, the group incorporated the fluorescent NAA, 2-amino-3-(5-(dimethylamino)-naphthalene-1-sulfonamido)propanoic acid (DanAla) using the $tRNA_{CUA}^{Leu}$,/DanAlaRS pair, into the voltage sensing domain of *Ciona intestinalis* voltage-sensitive phosphatase, which was then used to optically monitor the membrane depolarization of neurons differentiated from HCN-A94 cells (Shen et al. 2011).

1.12 Studying reactive intermediates using a defined protein environment

The third chapter within this thesis will address a novel application of amber suppression technology. Namely, the site-specific incorporation of an NAA containing a photoreactive moiety into a defined protein environment. The NAA *p*-azido-L-phenylalanine was chosen as the reactive NAA and the well studied protein, T4 Lysozyme was chosen as the protein scaffold. The photoreactions of the aryl azide was subsequently studied using spectroscopy.

1.12.1 p-azido-L-phenylalanine

The amino acid *p*-azido-L-phenylalanine (azF; Figure 15, **31**) contains a photoreactive aryl (phenyl) azide moiety. Aryl azides are used in a variety of chemical reactions that have many useful applications, for example in organic synthesis, production of photoimaging devices and a biochemical method of photoaffinity labelling. Aryl azide chemistry has also been utilised to mediate crosslinking between biological molecules (Knowles 1972; Tanaka *et al.* 2008). This photocrosslinking has been widely used in biology, for example, mapping RNA structural neighbours, crosslinking proteins, studying ligand : enzyme interactions and site selectively modifying protein by reaction with azide (Buchmueller *et al.* 2003; Chen and Ebrigh 1993; Fleming 1995; Sugawara and Matsuda 1995; Zhang *et al.* 2005a). Despite widespread use, the basic features of aryl chemistry in the context of an aqueous, roughly biological environment has not been widely investigated.

1.12.2 Aryl azide photochemistry

Put simply, photolysis of phenyl azide, at wavelengths 254 – 400 nm, results in loss of molecular nitrogen, which results in a species with the formula C_6H_5N (Huisgen 1955; Huisgen *et al.* 1958). However, the photoreaction of phenyl azide is more complex because different reactions occur depending on the temperature and substituents on the phenyl ring. Previous work performed on phenyl azide, utilising

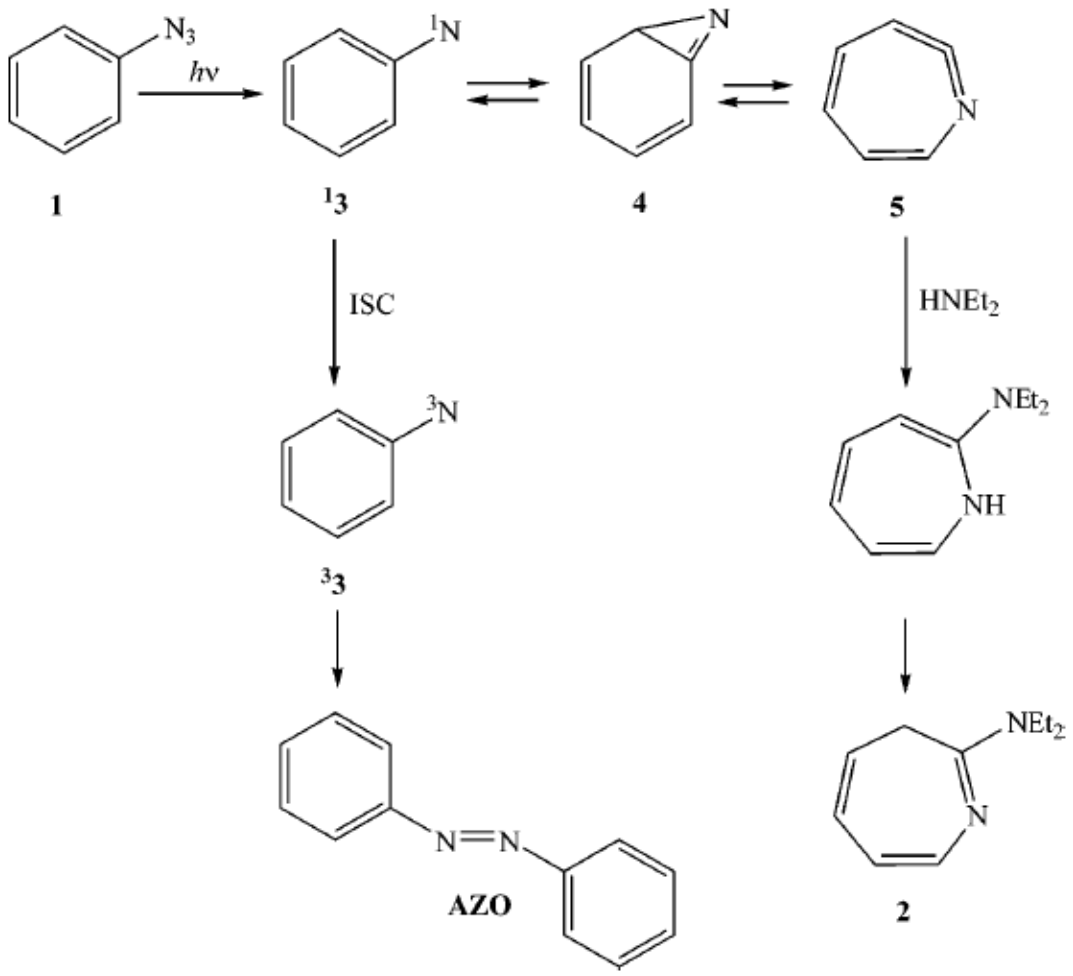


Figure 17. Modern view of the mechanism of phenyl azide photochemistry. UV photolysis of phenyl azide (1) produces singlet phenylnitrene $^{1}3$ and molecular nitrogen. In the liquid phase, $^{1}3$ isomerises over a small barrier to form didehydroazepine / cyclic ketenimine, 5. This species can be trapped to produce azepine 2. At cryogenic temperatures, (10 – 77 K) singlet phenyl nitrene ($^{1}3$) is produced. This species lacks the energy to undergo isomerisation and preferentially relaxes to triplet phenyl nitrene $^{3}3$ by intersystem crossing (ISC). This process most likely has no barrier to surmount and therefore is favoured at very low temperature. Decay of triplet arylnitrenes and/or formation of corresponding azo-compounds occurs.

a variety of methods, implicates the production of a variety of intermediates (for review see Gritsan and Platz 2006; Schuster and Platz 1992). Certain interesting points relating to the complexities of these reactions will be noted in the following text.

At low temperature (4 – 100 Kelvin), particularly in rigid media, reactive intermediates are stabilised because their unimolecular reactions are slowed, and bimolecular reactions are prevented by inhibition of diffusion. This increased stability enables the application of a variety of spectroscopic techniques that can aid in the determination of the structure of the intermediates, most notably EPR. Previously Wasserman photolysed a frozen crystalline solution of phenyl azide at 77 K. His observed spectrum showed a large new resonance absorption band at ~6700 G, which was assigned to triplet phenyl nitrene (Smolinsky *et al.* 1962; Wasserman 1971). Wasserman et al (1971), showed that it was a persistent species at this temperature. Furthermore, studies using UV-vis and IR spectra at low temperature also indicated formation of triplet nitrene as the primary photoproduct (Reiser *et al.* 1966). However, continued irradiation rapidly destroyed the triplet nitrene, and ketenimine was the product of secondary photolysis (Leyva *et al.* 1986). This secondary photolysis was the cause for much of the controversy surrounding the intermediates of photolysis of phenyl azide.

Previous work photolysing phenyl azide in solution at room temperature with a pulsed laser generated a reactive intermediate that absorbed strongly at 340 nm, formed rapidly (< 20 ns) and reacted slowly (over period of several hundred microseconds) to form tarry polymer and a small amount of azobenzene (Li *et al.* 1988). Li et al (1988), concluded that the reactive intermediate could be either triplet phenyl nitrene, dehydroazepine, benzazirine or some species not seen at low temperature.

Final resolution of the reaction process was determined utilising a combination of multiple techniques, for example, absorption and emission spectroscopy along with flash photolysis techniques (Leyva et al. 1986), matrix IR spectrum (Chapman and Le Roux 1978), time-resolved IR and UV-vis spectroscopy (Li et al. 1988; Schrock and Schuster 1984). In 1992, three research groups all proposed similar schemes to

explain the complex phenyl azide photochemistry (Budyka *et al.* 1992; Gritsan and Pritchina 1992; Schuster and Platz 1992). Photolysis of phenyl azide in the condensed phase forms singlet phenyl nitrene and molecular nitrogen (Figure 17). The singlet species rapidly sheds excess vibrational energy by collisions with solvent, overcomes a small barrier leading to isomerisation to form dehydroazepine (cyclic ketenimine), a high energy species that can be chemically trapped and characterised by TRUV and TRIR spectroscopy. Later computational work showed this to be a two-step process involving benzazirine (Karney and Borden 1997). At cryogenic temperatures, (10 – 77 K) singlet phenyl nitrene is produced. This species lacks the energy to undergo isomerisation and preferentially relaxes to triplet phenyl nitrene by intersystem crossing. This process most likely has no barrier to surmount and therefore is favoured at very low temperature.

1.12.3 Spectroscopy of NAAs

The incorporation of NAAs containing spectroscopically active moieties makes it possible to track the incorporation of the non-natural amino acid by the presence of this spectroscopic signal. The incorporation of the azide group containing amino acid, azF is an ideal candidate as the spectral properties of azide are distinct from those naturally found in proteins. The azide group is of particular interest because the photoreaction of this moiety alters its spectroscopic signals, particularly the EPR spectra that arise due to the creation of unpaired electrons during photolysis of the azide group. This technique allows the valuable study of the complex reactions of the azide group within a biological relevant setting, which is lacking from the literature. AzF, and it's photoreaction products, should also confer orthogonal IR and UV-vis spectra. This study is relevant because the reactivity of aryl azides in a biological context is particularly important for the quality of structural information obtained from those experiments mentioned above.

1.12.4 T4 Lysozyme

T4 lysozyme (T4L) is a 164 amino acid endoacetylmuramidase enzyme that is expressed during the lytic phase when the T4 bacteriophage infects *E. coli* cells. T4L

catalyses the hydrolysis of 1,4-β-linkages between N-acetylmuramic acid and N-acetyl-D-glucosamine residues in peptidoglycan layer of the bacterial cell wall. This cleavage ultimately causes lysis of the host cell and release of progeny phage particles. Both wild type and numerous mutant T4 lysozymes from recombinant sources have been extensively studied (Dixon *et al.* 1992; Eriksson *et al.* 1993; Eriksson *et al.* 1992; Hurley *et al.* 1992; Matsumura *et al.* 1988). The primary structure has been determined (Tsugita and Inouye 1968), the protein crystallised (Matthews *et al.* 1973) and the tertiary structure solved to 1.7 Å (Remington *et al.* 1978; Weaver and Matthews 1987) (Figure 18). In addition, the amber suppression technique has been applied to T4L. Previous work utilising incorporation of NAAs in T4L include elucidation of protein structure using the heavy atom Iodine (p-iodophenylalanine) in single-wavelength anomalous dispersion phasing (Xie et al. 2004). Also, incorporation of 4-cyano-phenylalanine and 4-ethynyl-phenylalanine at the same site to act as FRET donors with tryptophan has been utilised to probe the disruption of the hydrophobic core upon protein unfolding (Miyake-Stoner et al. 2009).

The wealth of knowledge regarding T4 lysozyme lends this protein to be used as a scaffold for utilisation within this study. Substitution at position 153 places the NAA within the hydrophobic core of the large lobe of T4L (Figure 18). It was hypothesised that the presence of the NAA, p-azido-l-phenylalanine, within this hydrophobic pocket can be exploited to study the reactive chemistry of the azide group without the complication of bimolecular reactions occurring. In addition, the hydrophobic pocket may stabilise the reactive intermediates to allow their study.

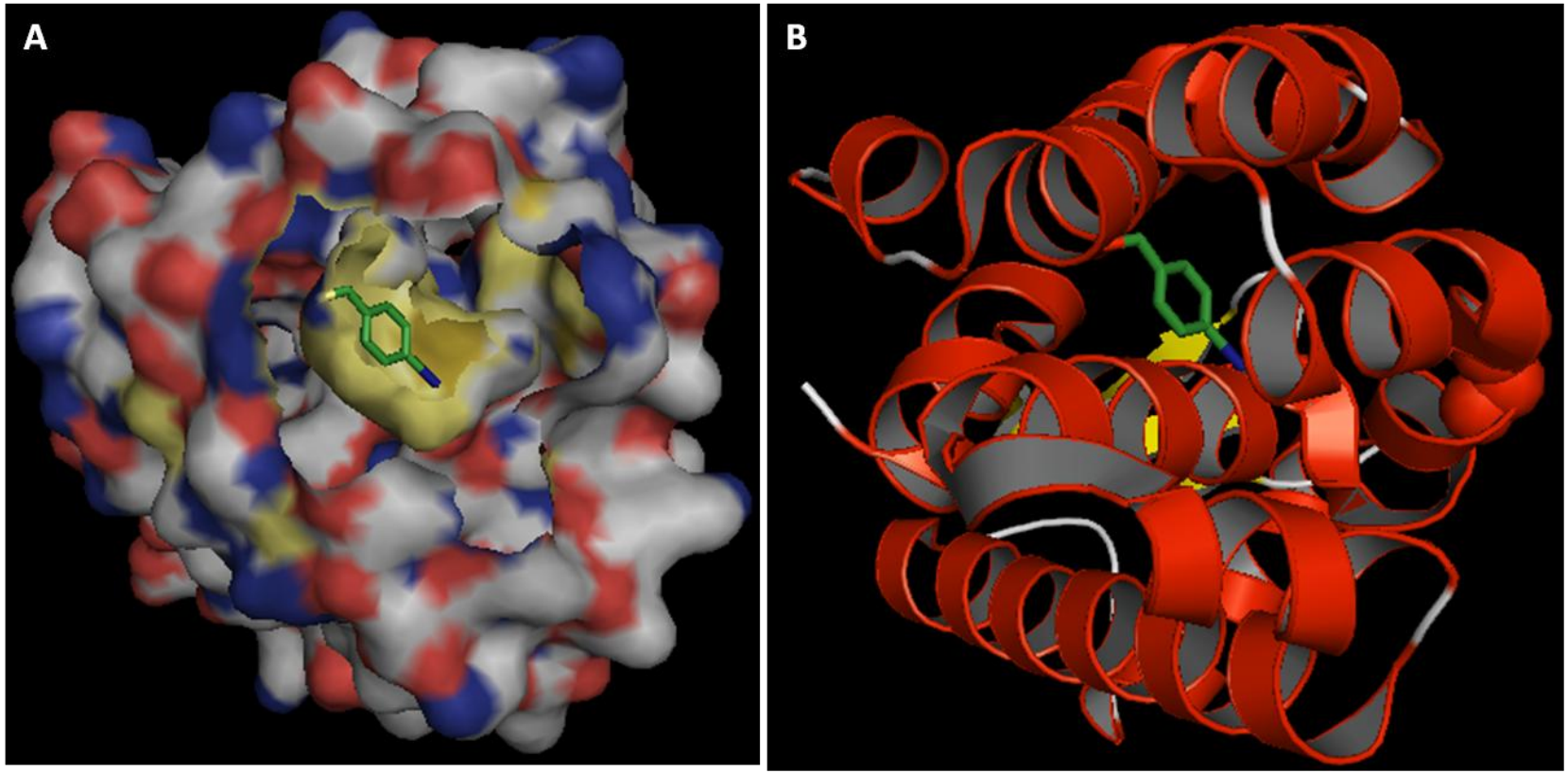


Figure 18. Structures of T4 Lysozyme (PDB ref: 1T6H).

A) Surface structure of T4L. The outer surface of T4L is coloured by element (C, grey; N, blue; O, red), the hydrophobic pocket within T4L is coloured yellow for clarity. P-iodoPhe is coloured green. Created using Pymol (DeLano and Lam 2006).

B) Ribbon structure of T4L. Helices (red) and beta sheet (yellow). Position of azidoPhe shown as green stick structure, nitrogen, N is blue. Figure created using Pymol (DeLano and Lam 2006)

1.13 Controlling phenotype by amber suppression

The fourth, fifth and sixth chapters within this thesis will address another novel application of amber suppression technology. In previous work using the amber suppression in mammalian cells, the cells are harvested and protein extracted and the protein studied *in vitro* or the non-natural amino acid is incorporated into GFP, luciferase and fluorescence or another visualisation technique is used to monitor protein after incorporation of non-natural amino acid. Therefore, the main aim of the work presented within Chapters 4, 5 and 6 was to produce a system where the protein was expressed and studied intracellularly, linking the incorporation of NAA to a novel, easily read output. Essentially, the hypothesis was to create a switch that was controllable by the presence or absence of NAA. Therefore, allowing control over protein expression and subsequently control over the phenotype of the cell via the supplementation of the media with a NAA. Modulation of motility was chosen for phenotypic control as it produces an easily read output and is applicable to both prokaryotes and eukaryotes. It has been postulated that cellular motility is an ancient process with its origin 1 billion years ago (Pollard and Borisy 2003) and that the default status for a cell is to be mobile and that it is the lack of space, the components of the extracellular matrix and cellular signalling that are responsible for inhibiting migration (Binamé *et al.* 2010).

1.13.1 Bacterial motility

The motility of *E. coli* is conferred by the rotating helical flagella filaments. The bacterial flagellum consists of three distinct parts. The basal body is embedded in the cell surface and forms part of the motor that produces the flagellar rotational force that propels the bacterium forward. The hook, is a flexible linker that joins the basal body to the filament and transmits torque to the filament. The filament, which protrudes into solution, converts this torque to thrust by means of its corkscrew shape (Macnab 2003) (Figure 19). The filament is constructed entirely using flagellin proteins (<30,000 subunits), which self assemble to form a filament 5 - 15 µm in length, with a diameter of about 200 Å.

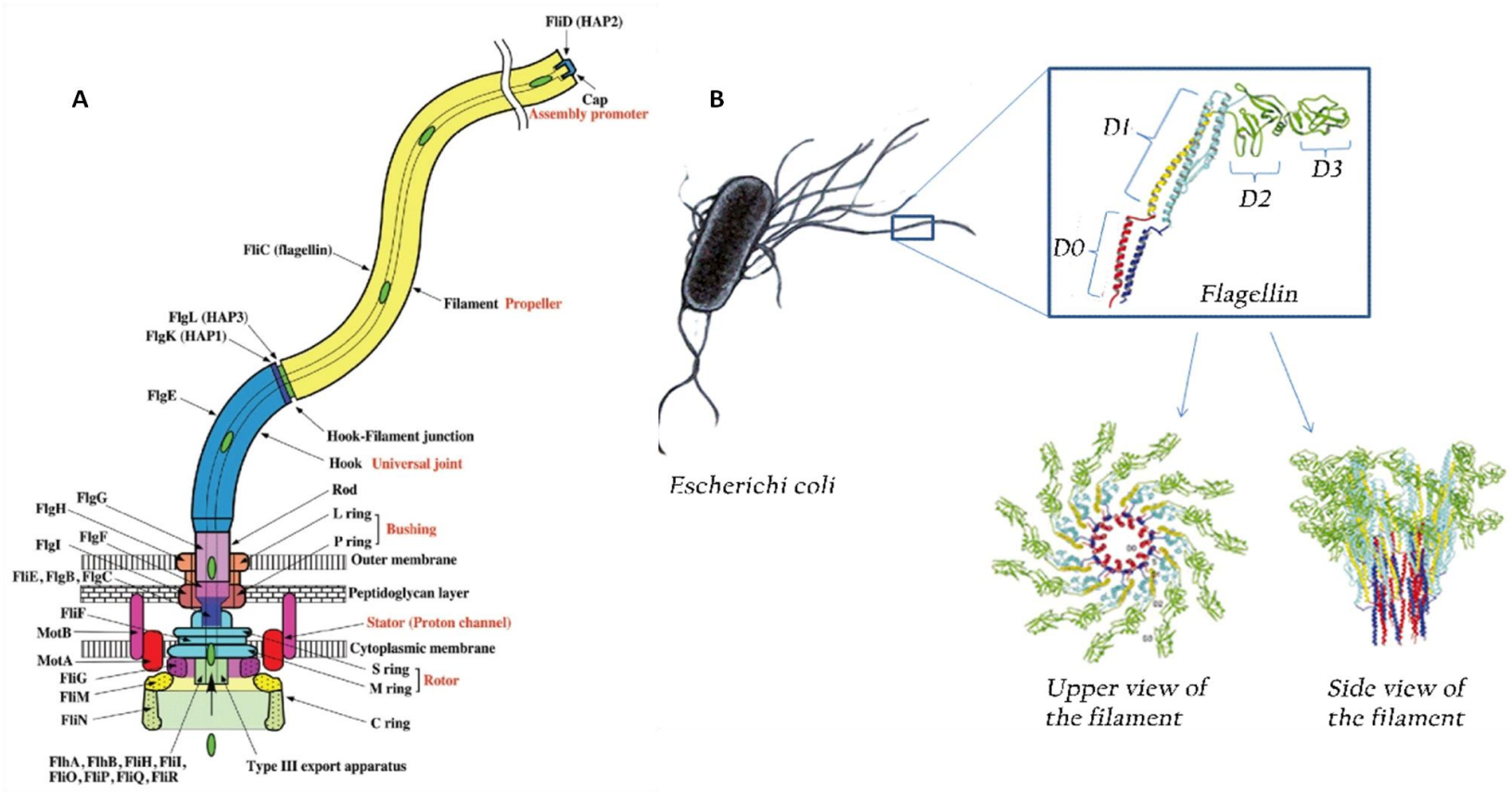


Figure 19. Structure of bacterial flagella

A) The basal body comprising rotor and stator; the hook; the flagellin filament and the cap are shown (taken from Yonekura *et al.* 2002; Yonekura *et al.* 2003).

B) Domains 0 (D0) and 1 (D1) promote polymerisation and form the inner core, domains 2 and 3 protrude into solution and form the outer ring (taken from Yonekura et al. 2003).

A flagellin protein consists of domains which have distinct properties (Yonekura et al. 2003). The domains 0 and 1 are formed from the N- and C- terminal of the peptide and are responsible for initiating polymerisation of flagellin by forming stable coiled-coils upon polymerisation (Vonderviszt *et al.* 1991). Domain 2 and 3 protrude from the filament into solution. Domain 3 is variable among strains and is responsible for antigenic properties (Kuwajima 1988b). It has been previously shown that insertion of foreign proteins can be tolerated within domain 3 (Le Moigne *et al.* 2006; Lu *et al.* 1995).

1.13.1.1 Flagellar export

Typically proteins are exported by *E. coli* via type I or type II bacterial export pathway in which the signal for export is found within an N- or C- terminal peptide (Homma *et al.* 1987a; Homma *et al.* 1987b; Jones *et al.* 1989). However, it is thought that flagellin protein and other proteins required for flagella assembly, namely the hook protein, the hook-associated proteins, and the component proteins of the basal body are exported via a flagellum-specific pathway (Type III). Kuwajima et al (1988a; 1989) performed a series of investigations utilising truncated flagella protein to ascertain which part of the protein encoded for export via this pathway. The work showed that the middle portion could be deleted or the C-terminal severely truncated without affecting transport. It has been suggested that the signal occurs in the N-terminal region of the protein itself and not present outside the coding region of flagellin gene (Kuwajima et al. 1989; Végh *et al.* 2006).

1.13.2 Mammalian cell motility

The ability for a metazoan cell to move is important for a variety of biological processes, whether it is, for example, embryonic axis patterning during development, wound healing, leukocyte migration from circulation into surrounding tissue during immune surveillance, axonal growth or metastasis and invasion of tumour cells (For review see Rørth 2009). The molecular processes that occur during cell migration are also common to other cellular process, for example, adhesion, cytokinesis and endocytosis and this paints a complex picture within the cell.

1.13.2.1 Amoeboid motion

The motility of a single cell depends on a variety of factors, such as, the strength of their adhesion; the type of substratum; external migratory signals; mechanical pliability, and the organization of the cellular cytoskeleton. Differences in these factors results in different modes of migration. For example, amoeboid motion describes the movement of a cell when the actin cytoskeleton is tending towards a disordered state and the adhesion to the substratum is weak (does not involve integrins). This amoeboid motion is observed in leukocytes cells of the immune system (Friedl and Weigelin 2008), haematopoietic stem cells (Giebel *et al.* 2004) and some tumour cells. This type of migration is studied by the model motility of *Dictyostelium discoideum* (Firtel and Meili 2000).

1.13.2.2 Metazoan motion

On the other hand, cells like fibroblasts display typical metazoan or mesenchymal motion (cell crawling) due to the elaborate cytoskeletal structures and adhesions. This motion is typically slow. In simple terms, movement of a polarised cell is now widely accepted to be driven by the extension of flat sheet-like layer of cytoplasmic protrusion (~0.2 µm thick) at the leading edge of the cell (by actin polymerisation, see section 1.13.2.4), the lamellipodium, as described in the seminal work by Abercrombie in the 1970's (Abercrombie 1980; Abercrombie *et al.* 1970, 1971). After protrusion of this lamellipodium, new adhesion sites to the substratum at the front are established (Figure 20). This is followed by cell body contraction (and detachment of cell adhesions to the rear of the cell), which acts to translocate the cell body forward and retraction of the cell rear, and is mediated by actin- and myosin- based contractions. As a result of this cyclical process the polarised cell tends to move in the same direction, until a random membrane protrusion makes stable contacts to substratum and can define a new direction of movement. These changes in direction can allow long time scale random movements, which enable the cell to explore its environment (Petrie *et al.* 2009).

This mesenchymal mode of migration is important for wound healing, and depends upon the activity of Rac GTPase (section 1.13.2.6) and therefore it is this type of motion that is considered in the following study.

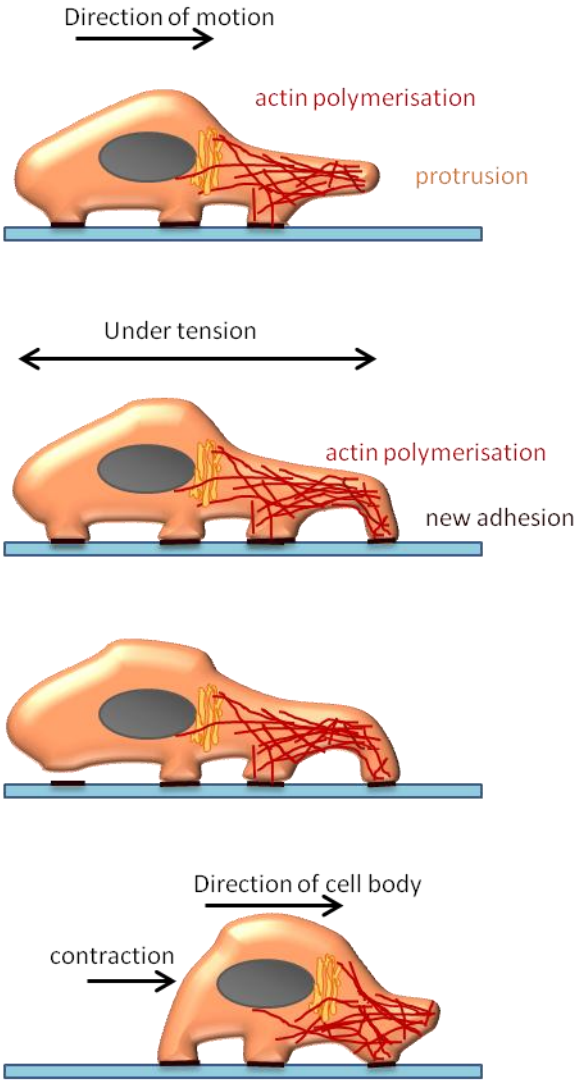


Figure 20. Schematic of the three stages of metazoan cell movement. Initially the polarised cell produces a lamellipodium protrusion. New adhesion sites to the substratum at the front, via focal adhesions, are established. This is followed by cell body contraction and detachment of adhesions and subsequent retraction of the cell rear.

1.13.2.3 Scratch wound model

The use of a 2D scratch model of a wound is a well established tool for use in the study of cell migration (Figure 21) (Vaughan and Trinkaus 1966). A scratch wound is created by mechanically scratching away a broad section of cells from a confluent cell monolayer. The response to the scratch involves both reaction to the cell injury and to the creation of a space within the scratched area (Nikolić *et al.* 2006; Poujade *et al.* 2007). The cells bordering the wound sense the free space, polarise (by a process controlled by cdc42 (Etienne-Manneville and Hall 2003; Fairn *et al.* 2011; Nobes and Hall 1995)), extend lamellipodia into this space to promote cell migration in this direction. Typically, the movement of cells during the simple scratch wound assay move as a collective sheet to close the wound created. The movement here is a collective migration based on the mesenchymal migratory mode. However, there has been evidence produced that suggests that confluent fibroblastic cells can also respond to a scratch wound as independent cells (Matsubayashi *et al.* 2004).

There is remaining discussions as to whether the cells behind the wound edge move forward as an act of passive dragging by the leading cells or whether they also orchestrate lamellipodia formation themselves (Fenteany *et al.* 2000; Ridley *et al.* 1992). It has previously been proposed that the leading cells recognise the scratch and produce the migratory force which pulls the sheet of cells in order to close the wound. The passive dragging of cells has been proposed to be orchestrated by the link between adherens junctions of adjacent cells formed by homotropic interactions of cadherins, and the linking of actin filaments at these positions to create a continuum of cytoskeleton (Hartsock and Nelson 2008; Tsukita *et al.* 1992). More recent work demonstrates that cells at the leading edge of the wound and submarginal cells behind it, are both actively motile and are involved in the closing of the wound (Farooqui and Fenteany 2005; Fenteany et al. 2000; Poujade et al. 2007; Vitorino and Meyer 2008). However, those cells at the leading edge confer more protrusive and motile activity (Omelchenko *et al.* 2003).

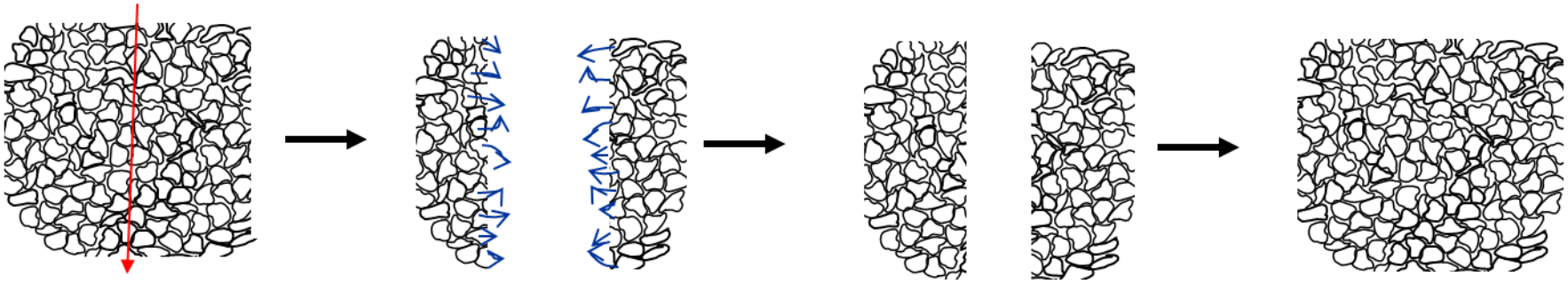


Figure 21. Scratch wound assay. A monolayer of confluent cells is scratched using a p200 tip, the cells sense the space created, polarise and migrate in to the wound space. The wound then starts to close, until a monolayer of confluent cells is apparent once more.

1.13.2.4 Motility and the role of actin

Actin, microtubules and intermediate filaments make up the cytoskeleton of a metazoan cell. Within the cell, actin filaments provide mechanical support for maintaining cellular shape; provide a path for the movement of intracellular cargo and polymerisation provides the driving force for cellular motility. It has been shown that actin comprises approximately 10 % of total cellular protein in fibroblasts (Bray and Thomas 1975).

The morphology and physical behaviour of migrating cells are predominantly orchestrated by the dynamics of the filamentous actin (F-actin) cytoskeleton which produce lamellipodia and are coupled to the substrate via focal adhesions. The most extended region of the lamellipodia is composed of a network of highly branched actin filaments that are actively polymerising (Glacy 1983). It is this polymerisation of actin beneath the plasma membrane that produces the physical force to push the membrane forward (For review see Gardel *et al.* 2010). The current view, the dentritic-nucleation model (Figure 22), is that actin filaments form a branched network with actin filaments lying at an angle of 70° with branching mediated by Arp2/3 complex (Pollard 2007; Svitkina and Borisy 1999). During polymerisation, globular G-actin is arranged head-to-tail to form F-actin in a process catalysed by ATP hydrolysis. F-actin is created in such a way that the filament displays polarity, with one end termed “barbed” or “plus” end, where actin polymerisation predominantly occurs; and the other “pointed” or “minus” end. The process of polymerisation can be stimulated by extracellular stimuli, for example, the creation of a scratch wound, which culminates in signalling cascades including Rac1 GTPase signalling.

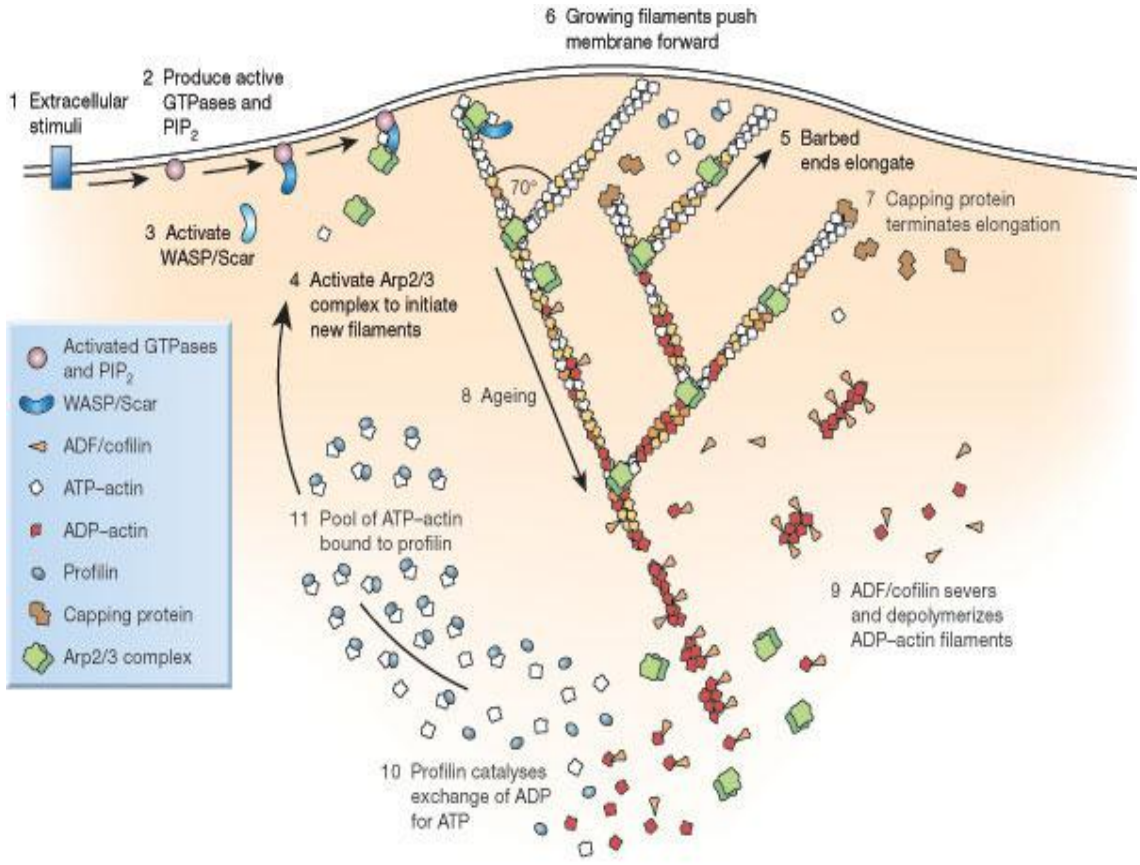


Figure 22. The dentritic-nucleation model for protrusion of lamellipodia. Extracellular signals (1) activate signalling pathways that lead to activation of GTPases (2), for example Rac1. These then activate Wiskott–Aldrich syndrome protein (WASP) and related proteins (3), which in turn activate Arp2/3 complex. Arp2/3 complex initiates filament branching on the side of an existing filament (4). Each new filament grows (5), fed by a high concentration of profilin-bound actin stored in the cytoplasm, and this pushes the plasma membrane forward (6). Capping protein binds to the growing ends, terminating elongation (7). Actin-depolymerising factor (ADF)/cofilin then severs and depolymerises the ADP filaments, mainly in the 'older' regions of the filaments (8, 9). Profilin re-enters the cycle at this point, promoting dissociation of ADP and binding of ATP to dissociated subunits (10). ATP–actin binds to profilin, refilling the pool of subunits available for assembly (11). (Taken from Pollard 2003).

1.13.2.5 GTPases and their role in cell migration

The spatiotemporal regulation of the physical structures associated with cell migration is regulated by a variety of signalling molecules and their downstream effectors, namely, the Rho family GTPases (For review see Raftopoulou and Hall 2004). GTPase activity of Rho is required to maintain cell adhesion during movement and Ras orchestrates the production of contractile actomyosin structures (stress fibres) and focal adhesions (Nobes and Hall 1995; Ridley and Hall 1992). Cdc42 is required to maintain cell polarity, which includes the localisation of lamellipodial activity to the leading edge and the reorientation of the Golgi apparatus in the direction of movement, and orchestrates production of filopodia (Nobes and Hall 1995, 1999). Rac is essential for protrusion of lamellipodia which is important in forward cellular migration (Nobes and Hall 1995; Ridley et al. 1992). When functional Rac1 is withheld from fibroblast cells, either by the presence of dominant negative (N17Rac1) Rac1 protein (Nobes and Hall 1999; Ridley et al. 1992), siRNA knockdown (Monypenny *et al.* 2009), or Rac1 knockout (Guo *et al.* 2006; Vidali *et al.* 2006), the cells display reduced ability to produce membrane ruffles and lamellipodia; thus limiting the migratory abilities of the cells. However, injection of constitutively active V12Rac1 in to fibroblasts was sufficient to induce lamellipodia and membrane ruffles (Ridley et al. 1992).

1.13.2.6 Ras-related C3 botulinum toxin substrate 1 (Rac1)

Ras-related C3 botulinum toxin substrate 1 (Rac1) is a small (~21 kDa) signalling GTPase protein, and is a member of the Rac subfamily of the Rho family and Ras superfamily of small GTP-binding proteins (Raftopoulou and Hall 2004). The GTPase functions by cycling between active plasma membrane associated GTP-bound form and inactive GDP-bound state. When activated, Rac1-GTP can bind to a variety of effector proteins to illicit numerous and varied responses, for example, secretory processes, phagocytosis of apoptotic cells, epithelial cell polarisation and most importantly its role in mediating actin-induced lamellipodial formation and subsequent cell migration. Rac1 is activated by tyrosine kinases and G-protein coupled receptors and is often dependent on phosphoinositide 3-kinase (PI 3-kinase) activity.

1.13.2.6.1 Overall structure

The Rac1 protein tertiary structure consists of a six strand (five parallel, one anti-parallel) central β-sheet and with six α-helices and two 3_{10} helices (Hirshberg *et al.* 1997). This overall structure is similar to that of members of the Ras superfamily, however, Rac1 contains an additional 13 residue insertion, residues 123 – 135, between β5 and α4. The nucleotide binding pocket of all Ras GTPases consists of highly conserved loops, namely, the two guanine recognition loops ^{115}TKLD and ^{158}SAL and two phosphate binding loops ^{10}GDGAVGKT and ^{57}DTAG (Hirshberg et al. 1997). Two regions, switch I (residues 28-40) and switch II (residues 59-74), of the Rac1 GTPase have been shown to be important in GTP binding and undergo conformational changes during the GDP-GTP exchange reaction (see section 1.13.2.6.2) (Rapley *et al.* 2008; Worthylake *et al.* 2000). Also involved in GTP binding is a hydrogen bond between the side chain of Cys 18 and the O2 atom of the α-phosphate. Phe 28 also interacts with the guanine base of the bound GTP. The Mg^{2+} ion is bound by the hydroxyl of Thr 35, within switch I (Grizot *et al.* 2001). The C-terminal region of Rac1, including the polybasic region and prenylation domain, is important in its association with the plasma membrane (see section 1.13.2.6.3).

Rac1 makes interactions with numerous protein factors, such as guanine nucleotide exchange factors (GEFs) and GTPase-activating proteins (GAPs), which are involved in GDP-GTP exchange. The region between amino acids 53 – 72 of Rac1 is required for specific recognition and activation by the GEFs with residue Trp 56 within β3 important in specific recognition of Rac1 by Rac1-specific GEFs (Gao *et al.* 2001). Rac1 also interacts with Rho guanosine nucleotide dissociation inhibitors (RhoGDIs) through hydrogen bonds which involve a number of residues of Rac1 mainly within the switch II region, namely, Thr 35, Tyr 64, Arg 66, His 103, and His 104, but also via hydrophobic interactions involving Leu 67 and Leu 70 (Grizot et al. 2001). The Rac1-RhoGDI complex is also involved in the activation of the NADPH oxidase complex (Grizot et al. 2001; Lapouge *et al.* 2000).

Figure 23. Ribbon structure of Rac1, PDB structure 2RMK (deposited by Modha *et al.* 2008). Switch 1 - light blue, Switch 2 – blue, Rac1 important determinant Trp56 – yellow, Polybasic region – pink, prenylation sequence (CLLL) – light pink, Mg2+ -grey sphere, Phosphomethylphosphonic acid guanylate ester (GTP analogue) – CPK colouring. Figure was produced using Pymol (DeLano and Lam 2006).

1.13.2.6.2 GTP exchange of Rac1

As the name suggests, Rac1 GTPases can exchange and hydrolyse GTP in reactions catalysed by guanine nucleotide exchange factors (GEFs) and GTPase-activating proteins (GAPs), respectively (Figure 24) (For review see Van Aelst and D'Souza-Schorey 1997). Many GEFs activate Rac1, for example, α/β-PIX, Tiam1, DOCK180 and some members of the Vav subfamily (Rapley et al. 2008; Worthylake et al. 2000). It has been demonstrated that Dbl-homology (DH) and pleckstrin-homology (PH) domains within GEFs, such as Tiam1, are responsible for catalysis of nucleotide exchange and the activation of G proteins by interacting with the two switch regions and causing their displacement and remodelling (Worthylake et al. 2000). More specifically, the interactions are mediated via hydrogen bonds to the amide nitrogens of Thr 35 and Val 36 and the hydroxyl of Tyr 32 of Rac1. The induced conformation changes disrupt binding of the nucleotide and Mg^{2+} ion, and thereby promote their release (Rapley et al. 2008; Worthylake et al. 2000).

1.13.2.6.3 Post-translational modification of C-terminal of Rac1

As for the majority of the Ras superfamily of proteins, Rac1 is post-translationally modified by attachment of a lipid moiety (more specifically, a geranylgeranyl group) to its C-terminal cysteine in response to the consensus C-terminal CAAX (C = Cys, A = aliphatic, X = any amino acid; CLLL for Rac1) tetrapeptide sequence. The C-terminal tetrapeptide undergoes a process of prenylation, proteolysis, and carboxyl methylation in association with a cytosolic chaperone, Rho guanosine nucleotide dissociation inhibitors (RhoGDIs) (Michaelson *et al.* 2005; Michaelson *et al.* 2001) (Figure 25). In addition, RhoGDIs bind Rac1 to prevent its interaction with the plasma membrane by shielding the geranylgeranyl group (Fukumoto *et al.* 1990; Michaelson et al. 2001). It is this proteolysed/prenylated/carboxylated cysteine residue, which, when coupled with the polybasic residues immediately upstream, comprise the membrane-targeting sequence (Clarke 1992). Once correctly modified, bound to GTP and associated with the plasma membrane, Rac1 can elicit signalling cascades, for example, those involved in the formation of actin protrusions and orchestrate directed migration (Figure 26) (For review see Raftopoulou and Hall 2004).

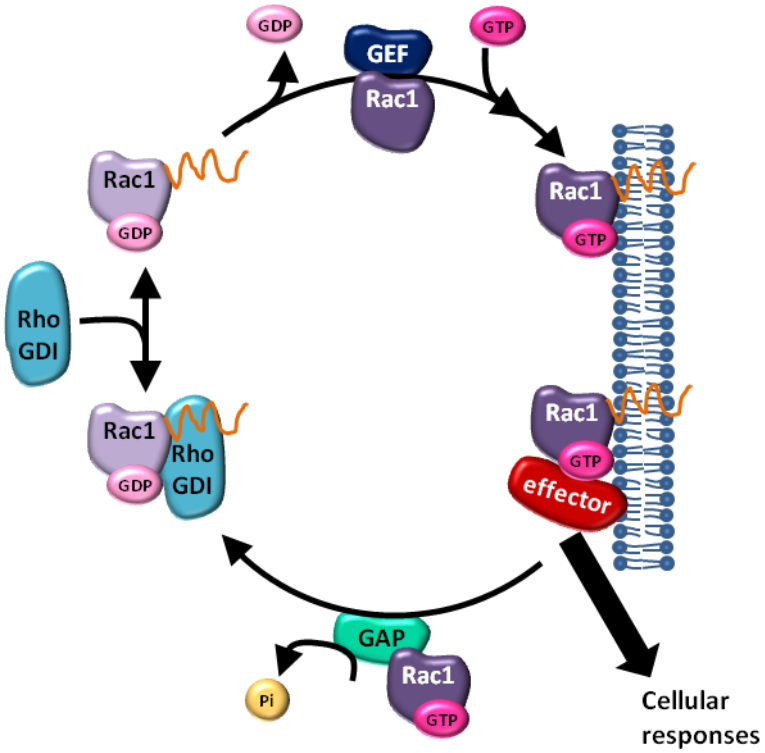


Figure 24. The GTPase cycle of Rac1. The GTPase functions by cycling between active plasma membrane associated GTP-bound form and inactive GDP-bound state. When activated, Rac1-GTP can bind to a variety of effector proteins to illicit numerous and varied response. The intrinsic GTPase activity of the Rac1 can hydrolyse GTP to GDP, with catalysis from GTPase activating proteins (GAPs), GTP exchange factors (GEFs) facilitate exchange of GDP with GTP and thus favour the active, GTP-bound state.

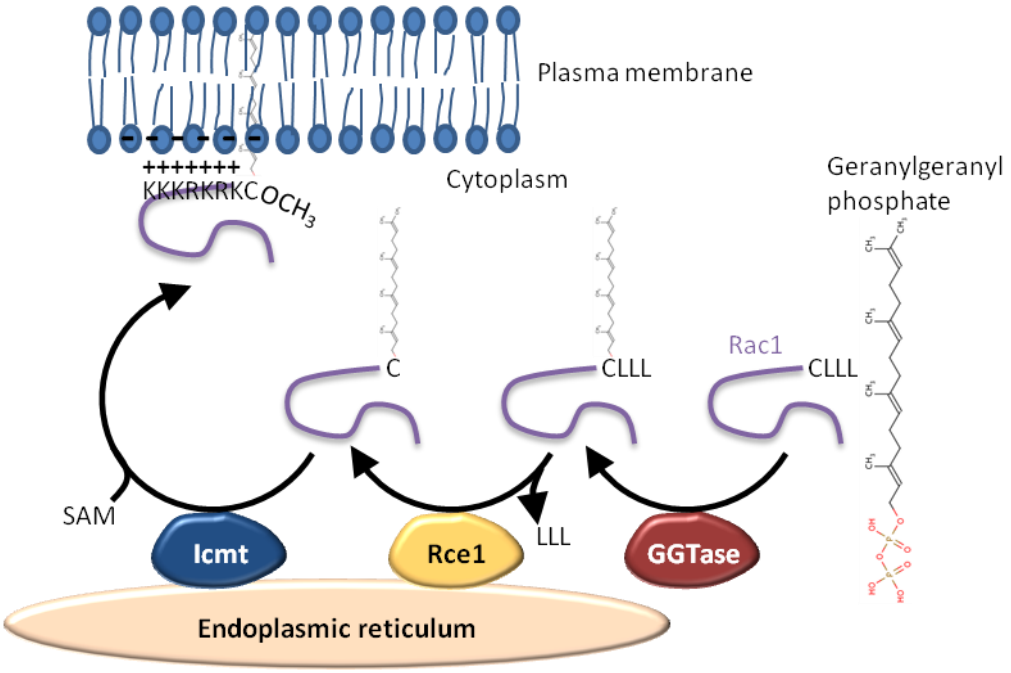


Figure 25. Post translational modification of the C-terminal of Rac1. In the cytoplasm, Geranylgeranyltransferase (GGTase) adds a geranylgeranyl group to the cysteine residue of CLLL motif. At the cytosolic leaflet of the endoplasmic reticulum Ras and a-factor converting enzyme (Rce1) removes the LLL tripeptide and isoprenylcysteine carboxyl methyltransferase (Icmt) methylates the cysteine in a process requiring *S*-adenosylmethionine (SAM). Rac1 then travels to the plasma membrane (PM), where the geranylgeranyl group inserts into, and the polybasic region associates with, the PM.

1.13.2.6.4 Signalling cascades elicited by Rac1

Most notably, Rac1 activates the scaffold protein WAVE/Scar complex, which includes the components WAVE, Abi, Nap125, Sra-1 and HSPC-300 (Figure 26). It has been suggested that Rac1 activates the WAVE/Scar complex by inducing the dissociation of the Abi-Nap125-Sra-1 trimer from WAVE (Vartiainen and Machesky 2004). Phosphorylation has also been implicated in the activation of WAVE/Scar (Ardern *et al.* 2006; Cory *et al.* 2003). Activated WAVE/Scar in turn activates the Arp2/3 complex (Machesky *et al.* 1999) (see section 1.13.2.4; Figure 22). Both the Arp2 and Arp3 subunits can bind ATP and this ability has been suggested to be involved in the activation of Arp2/3-initiated actin polymerisation (Dayel and Mullins 2004; Kiselar *et al.* 2007). Recent work has suggested that activation of the Arp2/3 complex is brought about by the conformational changes within Arp2 stimulated by its phosphorylation (Narayanan *et al.* 2011). Activated Arp2/3 then functions to create branched actin by associating with an existing actin filament and creating a nucleation point for polymerisation of actin by forming an actin-like heterodimer that mimics the barbed end of an actin filament (Robinson *et al.* 2001).

Rac1 signalling via a phosphatidylinositide 4-phosphate 5-kinase (PIP 5-kinase) to induce the formation of PIs(4,5)P_2 has also been shown. PI(4,5)P_2 can act to promote actin polymerisation by removing capping proteins from the barbed ends of existing actin filaments at the plasma membrane (Carpenter *et al.* 1999).

PI 3-kinases and their lipid product PI(3,4,5)P_3 have also been implicated in Rac1-induced control of cell migration (Reif *et al.* 1996). The production of PI(3,4,5)P_3 leads to an increase in GTP-Rac in many cell types and is thought to act via direct interaction with Rac GEFs, for example Tiam-1. It has been suggested that a major role of PI(3,4,5)P_3 is the induction of membrane translocation via GEFs PI(3,4,5)P_3–binding pleckstrin homology (PH) domain (Raftopoulou and Hall 2004). Rac1-GTP is also capable of activating PI 3-kinase, resulting in production of PI(3,4,5)P_3, which establishes a positive feedback loop.

In addition, Rac1 can activate a variety of downstream targets, for example p21-activated serine/threonine kinases (PAKs) that phosphorylate a plethora of

substrates implicated in cell migration (For review see Bokoch 2003), including MLCK (inhibitory) (Sanders *et al.* 1999), RLC (activating) (Chew *et al.* 1998), LIMK (activating, which inhibits cofilin, regulating its actin-severing activity) (Yang *et al.* 1998), and stathmin (which controls microtubule stability) (Wittmann *et al.* 2004).

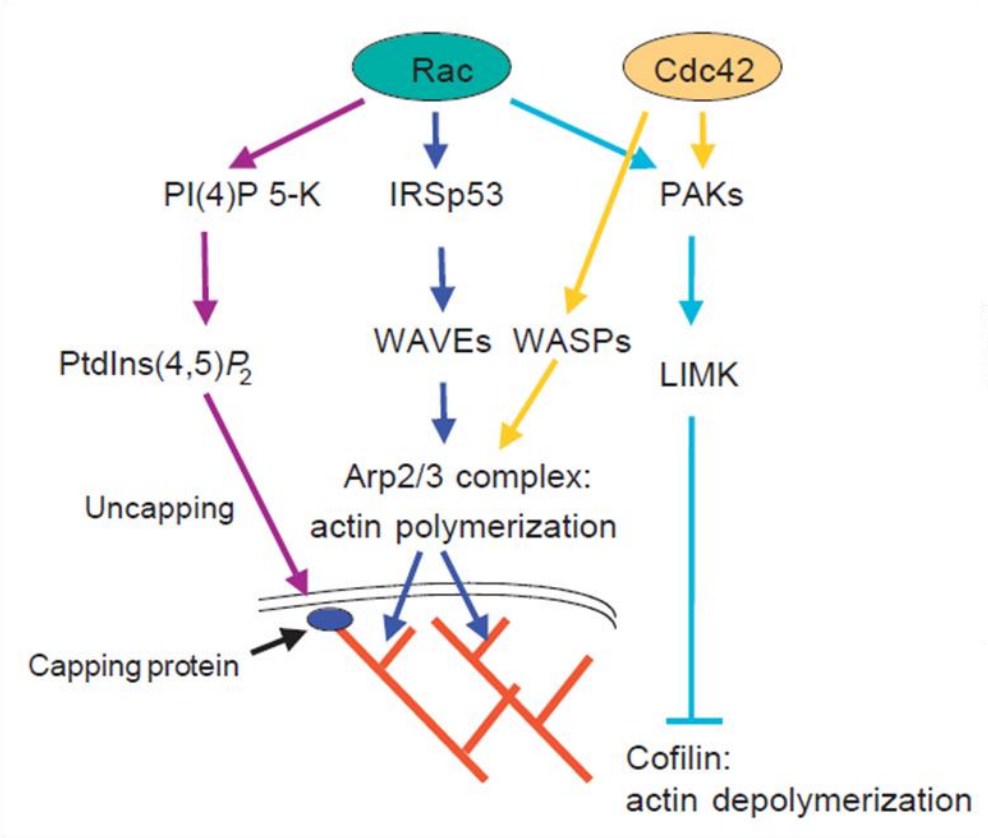


Figure 26. A diagram displaying a model for Rac-induced lamellipodium extension. Rac signalling stimulates Arp2/3 complex-induced actin polymerisation by interacting with a complex of IRSp53 and WAVE proteins. Arp2/3 complex acts as a nucleation point for the formation of new actin filament branches. Rac can also induce actin polymerisation via filament uncapping, which generates phosphatidylinositol 4,5-bisphosphate and inhibit cofilin-induced depolymerisation (via PAKs-LIMK signalling). Cross-talk of Rac with Cdc42 pathways via IRSp53 and/or PAKs may act to regulate Rac signalling. (Taken from Ridley 2001).

1.14 Aims

The aims of the work presented within this Thesis were to expand upon the current applications of amber suppression technology and incorporation of NAAs into proteins, as are discussed above. Despite having a vast range of applications in inserting chemical, physical or biological reactivities into proteins, there is potential to apply this method to disparate applications.

Firstly, the aim was to use amber suppression technology to site-specifically incorporate a specific chemical moiety into a defined environment within a protein. The hypothesis was that this protein scaffold could be used to stabilise, shield and therefore allow exploration the chemical reactivity of photoreactive moieties. The well studied protein, T4 lysozyme was used to provide the defined chemical environment, and the photoreactive, aryl azide containing NAA azidophenylalanine was used as an example of a highly reactive species.

In the second application, the amber suppression method itself was the basis for the novel application rather than the chemical reactivity of the NAA. It was explored whether the amber suppression method could be used to create a biological switch controlled by the presence or absence of NAA. This would provide a novel method to control protein expression and could be linked, therefore, to modulation of a specific phenotype. The flagellin protein was chosen as the target protein within *E*. *coli* for modulation of motility. The Rac1 GTPase was chosen as a target protein in mammalian cells because it is involved in the production of lamellipodia, which has the potential to control cell migration.

2 CHAPTER 2: GENERAL MATERIALS AND METHODS

2.1 Instruments

Electroporation of bacterial cells was performed using ECM 399 Electroporator system (BTX Harvard Apparatus, 45-0000). Protein purification was performed using AKTApurifier UPC10 (28-4062-68, GE Healthcare) using a Superdex 75 10/300 GL size exclusion column (17-5174-01, GE Healthcare). Polymerase chain reactions were performed using a TC-300 Thermal cycler (Techne, VWR 731-1501). DNA and protein concentration measurements were obtained using a NanoDrop ND-1000 UV-vis spectrophotometer (Thermo Scientific). ddH_2O was obtained using a PURELab Option S-R-7-15 (ELGA) water purifier, which produced water up to 15 MΩ-cm. Gel images and motility assays were visualised using a UV-transilluminator (UVP) and images were captured using a Canon PowerShot G10 camera with UV-filter (Canon). UV-Vis spectra were recorded using a Jasco 630 UV-visible spectrophotometer equipped with a Peltier temperature controller (Jasco).

2.2 Materials

The chemicals, plastic and glassware were purchased from Sigma Aldrich, VWR, Melford and Fisher unless otherwise stated.

2.2.1 Non-natural amino acids

The non natural amino acids (NAAs) p–azido-L-phenylalanine and p-iodo-L-phenylalanine were obtained from BACHEM. NAAs were dissolved in sterile ddH_2O with 5 M NaOH added until fully dissolved.

2.2.2 Bacterial antibiotics

Bacterial antibiotics were made up as indicated in Table 1, aliquoted and stored at -20 °C until required. The desired concentration of antibiotic was added to microbiological media for immediate use.

Table 1. Bacterial antibiotics

Antibiotic	Manufacturer	Stock solution concentration	Solute	Working concentration
Carbenicillin disodium	Melford	100 mg / ml	ddH_2O	100 µg / ml
Chloramphenicol	Fluka	34 mg / ml	ethanol	34 µg / ml
Kanamycin Sulphate	Applichem	50 mg / ml	ddH_2O	50 µg / ml
Tetracycline hydrochloride	Applichem	5 mg / ml	ethanol	25 µg / ml

2.2.3 Buffers

2.2.3.1 Buffers for preparation of chemically competent cells

TFB I buffer contained 30 mM KOAc, 50 mM $MnCl_2$, 100 mM RbCl, 10 mM $CaCl_2$, 15 % Glycerol, was adjusted to pH 5.8 with acetic acid; sterile filtered and stored at 4 °C.

TFB II contained 10 mM MOPS, 75 mM $CaCl_2$, 10 mM RbCl, 15 % Glycerol, adjusted to pH 7.0, sterile filtered, and stored at 4 °C.

2.2.3.2 Buffers for DNA preparation

Buffer P1 (resuspension buffer) contained 50 mM Tris-Cl, 10 mM EDTA, 100 µg/ml RNase A, pH 8.0 and was sterile filtered and stored at 4 °C.

Buffer P2 (lysis buffer) contained 200 mM NaOH, 1 % SDS (w/v). The solution was sterile filtered and stored at room temperature.

Buffer P3 (neutralisation buffer) contained 3 M Potassium acetate, pH 5.5. The solution was sterile filtered and stored at room temperature. Prior to use buffer P3 was pre-chilled to 4 °C.

The composition of Buffer N3 (neutralisation buffer), Buffer ERC, and the wash buffers, Buffer PB and Buffer PE are confidential, proprietary components of the QIAprep spin miniprep kit (Qiagen).

Buffer QBT (equilibration buffer; Qiagen) contained 750 mM NaCl, 50 mM MOPS, pH 7.0, 15 % isopropanol (v/v) and 0.15 % Triton® X-100 (v/v). Buffer was stored at room temperature.

Buffer QC (wash buffer; Qiagen) contained 1.0 M NaCl, 50 mM MOPS, pH 7.0 and 15 % isopropanol (v/v). Buffer was stored at room temperature.

Buffer QF (elution buffer; Qiagen) contained 1.25 M NaCl, 50 mM Tris-Cl, pH 8.5 and 15 % isopropanol (v/v). Buffer was stored at room temperature.

Buffer EB (Elution buffer, Qiagen) contained 10 mM Tris-Cl, pH 8.5.

TE buffer contained 10 mM Tris·Cl, 1 mM EDTA, pH 8.0. Buffer was stored at room temperature.

2.2.3.3 Buffers for enzymatic reactions

Antarctic phosphatase buffer (for dephosphorylation of DNA) contained 50 mM Bis-Tris-Propane-HCl, 1 mM $MgCl_2$, 0.1 mM $ZnCl_2$, pH 6.0.

Quick Ligation reaction buffer contained 132 mM Tris-HCL, 20 mM $MgCl_2$, 2 mM dithiothreitol, 2 mM ATP, 15 % Polyethylene glycol (PEG 6000) , pH 7.6.

NEBuffer 1 contained 10 mM Bis-Tris-Propane-HCl, 10 mM $MgCl_2$, 1 mM Dithiothreitol, pH 7.0. NEBuffer 2 contained 50 mM NaCl, 10 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM Dithiothreitol, pH 7.9. NEBuffer 3 contained 100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM Dithiothreitol, pH 7.9. NEBuffer 4 contained 50 mM Potassium Acetate, 20 mM Tris-acetate, 10 mM Magnesium Acetate, 1 mM Dithiothreitol, pH 7.9.

2.2.3.4 Buffers for DNA electrophoresis

A 50 x stock of Tris-Acetate-EDTA (TAE) buffer contained 2 M Tris-acetate, 50 mM EDTA, 57.1 ml acetic acid, pH 8. A 1 x working solution was prepared in ddH_2O and used for DNA gel electrophoresis (section 2.3.2.7)

A 6 x stock of DNA electrophoresis loading buffer was made using 0.25 % bromophenol blue, 0.25 % xylene cyanol, 1 mM EDTA, 30 % glycerol, pH 8.0. DNA samples were diluted to an appropriate concentration using ddH_2O and the stock DNA loading buffer.

2.2.3.5 Buffers for microbiological media

M9 salts stock solution contained 42 mM Na_2HPO_4, 24 mM KH_2PO_4, 9 mM NaCl, 19 mM NH_4Cl.

A 1000 x stock of trace metal solution contained 50 mM $FeCl_3$, 20 mM $CaCl_2$, 10 mM $MnCl_2$ 10 mM $ZnSO_4$, 2 mM $CoCl_2$, 2 mM $CuCl_2$, 2 mM $NiCl_2$, 2 mM Na_2MoO_4, 2 mM Na_2SeO_3, 2 mM H_3BO_3 in ddH_2O. The stock was sterilised by syringe filtration through a 0.2 µm membrane.

2.2.3.6 Buffers for sodium dodecyl sulphate polyacrylamide gel electrophoresis

Buffers for making and running SDS-PAGE gels are indicated in Table 2. Buffers were made using ddH_2O and stored at room temperature.

Table 2. SDS-PAGE buffers

Buffer	Components	Concentration	Use
SDS-PAGE gel buffer	Tris-HCl	3 M	Making Tricine gels
	SDS	0. 3 %	
	pH	8.45	
Acrylamide/bis-acrylamide 37.5 : 1, 30 % solution	Acrylamide/bis-acrylamide 37.5 : 1	30 %	Making all protein gels
Cathode buffer 10 x	Tris-HCl	1 M	Cathode buffer when running tricine protein gels
	Tricine	1 M	
	SDS	1 %	
	pH	8.25	
Anode buffer 10 x	Tris-HCl	2.1 M	Anode buffer when running tricine protein gels
	pH	8.9	
Laemmli buffer 1 x	Tris	3 %	Running buffer for glycine protein gels
	Glycine	14.4 %	
	SDS	1 %	
Sample buffer 4 x	Tris-HCl	150 mM	Dilute protein samples for SDS-PAGE
	Coomassie blue G-250	0.05 %	
	Glycerol	30 %	
	β-mercaptoethanol	6 %	
	SDS	12 %	
SDS-PAGE Stain solution	Coomassie blue G-250	1 g / L	Staining SDS-PAGE gels for visualisation
	Acetic acid	10 %	
	Ethanol	40 %	
	ddH_2O	50 %	
SDS-PAGE destain solution	Acetic acid	10 %	Destaining SDS-PAGE gels for visualisation
	Ethanol	40 %	
	ddH_2O	50 %	

2.2.4 Restriction enzymes

All restriction enzymes were supplied by New England Biolabs (Table 3). Double digests were designed using the *Double Digest* tool on NEB website (New England Biolabs).

Table 3. Restriction enzymes

Enzyme	Recognition sequence	Buffer	Digestion temperature	Inactivation
*Age*I-HF	ACCGGT	4 + BSA	37 °C	65°C, 20 min
*Apa*I	GGGCCC	4 + BSA	25 °C	65°C, 20 min
*Bam*HI-HF	GGATTC	4	37 °C	-
*Bgl*II	AGATCT	3	37 °C	-
*Dpn*I	GA(Me)TC	4	37 °C	65°C, 20 min
*Eco*RI-HF	GAATTC	4	37 °C	65°C, 20 min
*Hind*III	AAGCTT	2	37 °C	65°C, 20 min
*Kpn*I-HF	GGTACC	4	37 °C	-
*Nco*I-HF	CCATGG	4	37 °C	65°C, 20 min
*Nde*I	CATATG	4	37 °C	65°C, 20 min
*Nhe*I-HF	GCTAGC	4 + BSA	37 °C	65°C, 20 min
*Not*I-HF	GCGGCCGC	4 + BSA	37 °C	65°C, 20 min
*Ssp*I-HF	AATATT	4	37 °C	65°C, 20 min
*Xba*I	TCTAGA	4 + BSA	37 °C	65°C, 20 min
*Xho*I	CTCGAG	4 + BSA	37 °C	65°C, 20 min

2.2.5 Other enzymes

T4 Polynucleotide Kinase (PNK), Antarctic Phosphatase, and Quick T4 DNA ligase were purchased from New England Biolabs.

2.2.6 *Escherichia Coli* Cells

Top10 (Invitrogen) genotype F^- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) ϕ80*lac*ZΔM15 ΔlacX74 *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L *end*A1 *nup*G were used for cloning,

plasmid preparation and protein expression using pBAD plasmids. NEB 10-beta Competent *E. coli* (High Efficiency) (New England Biolabs) genotype *araD139* Δ(*ara-leu*)7697 *fhuA lacX74 galK* (φ80 Δ(*lacZ*)*M15*) *mcrA galU recA1 endA1 nupG rpsL* (Str^R) Δ(*mrr-hsdRMS-mcrBC*) were used for cloning of mammalian plasmids. BL21 Star™(DE3) (Invitrogen) genotype F^- *omp*T *hsdS*B (r_B^- m_B^-) *gal dcm rne*131 (DE3) and BL21 Gold(DE3) (Stratagene) genotype E. coli B F^- *omp*T hsdS(r_B^- m_B^-) dcm^+ Te^r *gal* λ(DE3) *end*A were used for protein expression using pET plasmids and other plasmids requiring T7 polymerase expression. KAF95 is a K12 derivative, AW405 (Armstrong *et al.* 1967) constructed by Karen Fahrner (Berg and Turner 1993). The cell line carries the fliC726 allele, with a non-flagellate phenotype (Silverman and Simon 1973) and a cheY deletion so is smooth swimming and non-chemotactic. Further details can be found in Appendix A5. KAF95 was used in experiments to try to restore motility using a plasmid encoded FliC gene. The JW1908 strains of the Keio collection were created by Baba et al (2006). The strain has the genotype, AG1[*recA1 endA1 gyrA96 thi-1 hsdR17(rK-mK+) supE44 relA1*]. The group systematically made a set of precisely defined, single-gene deletions of all nonessential genes in *Escherichia coli* K-12 BW25113. The clone used in this study is a FliC knock-out strain and therefore has a non-flagellate phenotype and are therefore non-motile. Keio cells were used in experiments to try to restore motility using a plasmid encoded FliC gene.

2.2.7 Media

Lysogeny broth (LB) media contained 1 % tryptone, 1 % NaCl, 0.5 % Yeast extract. LB agar contained components as listed for LB, but with addition of 1 % agar. Agar was autoclaved, stored at room temperature and melted and plated as required, with addition of specified antibiotics when agar was hand-hot. LB media and LB agar was used for routine bacterial culture. LB medium containing 0.3 % agar was used for *E. coli* motility assays.

Super Optimal broth (SOB) with Catabolite repression (SOC) media contained 0.5 % Yeast Extract, 2 % Tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$

with addition of 20 mM filter sterilised glucose after autoclaving. SOC was used to rescue *E*. *coli* transformations.

Terrific broth (TB) autoinduction media contained 1.2 % Tryptone, 2.4 % yeast extract, 0.4 % glycerol which was then autoclaved. Sterile filtered 17 mM KH_2PO_4, 72 mM K_2HPO_4, 0.05 % glucose and 0.2 % α-lactose was added when cool. TB was used for protein expression media.

ZYM-5052 autoinduction medium contained 1 % Tryptone, 0.5 % yeast extract, 25 mM Na_2HPO_4, 25 mM KH_2PO_4, 50 mM NH_4Cl, 5 mM Na_2SO_4, 2 mM $MgSO_4$, 1× trace elements (section 2.2.3.5), 0.5 % glycerol, 0.05 % glucose and 0.2 % α-lactose. ZYM-5052 was used for protein expression. For induction from *ara*BAD promoter, 0.2 % arabinose was added directly to the growing culture.

PSI broth contained 5 % yeast extract, 20 % tryptone, 5 % $MgSO_4$ and was adjusted to pH 7.6. Psi broth was used in bacterial culture to make chemically competent cells.

M9 minimal media contained 1 x M9 salts (section 2.2.3.5), 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 2.0 % glucose, 0.5 µg/ml thiamine and 0.4 % glycerol.

All media was made up in ddH_2O. Autoclaving was performed at 121 °C for 15 minutes. Media was stored at room temperature until required.

2.2.8 Plasmids

Specific plasmid details are detailed within the text at the relevant sections and listed in the Appendix A1.

2.2.9 Primers

Primers were designed using the DNAstar program (DNAStar Inc. 1999). Melting temperatures, length and GC content were calculated using the online Finnzymes Tm calculator. Primers were produced by Integrated DNA technologies (IDT). The primers were designed optimally to have a melting temperature of between 60 and 70 °C, length of 20 – 30 nucleotides and GC content between 40 and 60 %. Primers are listed in Appendix A2.

2.3 Methods

2.3.1 Microbiology

2.3.1.1 Preparation of electrocompetant *E. coli* cells

A single colony from an agar plate containing required *E. coli* cells was used to inoculate a 5 ml LB culture, which was grown overnight (37 °C, with shaking). Fresh LB medium was inoculated with 1 ml of this overnight culture per 250 ml fresh LB medium, appropriate antibiotics were added where necessary and the culture grown until an OD_{600} of 0.5 - 0.8 was reached. The culture was placed on ice for 15 minutes and cells harvested by centrifugation (4000 x *g* for 15 minutes). All subsequent steps were carried out at 4 °C using pre-cooled materials. The pellet was then washed three times with 15 % glycerol (autoclaved); first wash the cells were resuspended in 1/1 culture volume, second with 1/2 volume and for the final wash with 1/5 volume. After the final wash, the pellet was resuspended in 3/250 volume 15 % glycerol, aliquoted in to 100 µl aliquots in pre-cooled tubes, frozen in liquid nitrogen and stored at - 80 °C.

2.3.1.2 Preparation of chemically competent *E. coli* cells

A single colony from an agar plate containing required *E. coli* cells was used to inoculate a 5 ml LB culture, which was grown overnight (37 °C, with shaking). PSI broth was inoculated with 1 ml of this overnight culture per 250 ml fresh medium, appropriate antibiotics were added where necessary and the culture was grown until an OD_{600} of 0.5 - 0.8 was reached. The culture was placed on ice for 15 minutes and cells harvested by centrifugation (4000 x *g*, for 15 minutes). All subsequent steps were carried out at 4 °C with pre-cooled materials. The pellet was resuspended in 0.4 x volume TFB I buffer (section 2.2.3.1), incubated for 15 min (ice) and pelleted by centrifugation (4000 x *g*, for 15 minutes). The pellet was resuspended in 0.04 x volume TFB II (section 2.2.3.1); aliquoted in to 100 µl aliquots in pre-cooled tubes, frozen in liquid nitrogen and stored at -80 °C.

2.3.1.3 Transformation of DNA in to electrocompetant *E. coli*

An aliquot of electrocompetant *E. coli* cells was thawed on ice. Electroporation cuvettes (0.2 µm, Bio-Rad) were pre-cooled on ice. Plasmid DNA (section 2.3.1.6), typically 1 µl, was added to the cells, and the mixture transferred to the cuvette and electorporated at 2400 mV. Immediately post electroporation, 400 µl SOC medium was added and the cell suspension was transferred to a 2 ml Eppendorf tube. After a recovery time of 1 hour (37 °C, with shaking), the cells were plated on to LB agar plates containing the required antibiotic and incubated overnight at 37 °C.

2.3.1.4 Transformation of DNA in to chemically competent *E. coli*

An aliquot of chemically competent Top10 cells was defrosted on ice. 1 – 5 µl of ligation reaction (section 2.3.2.6) or plasmid DNA (section 2.3.1.6) was added to this aliquot of cells and incubated on ice for 30 minutes. The aliquots were then heat shocked at 42 °C for 1 – 2 minutes (digital heatblock, VWR). 400 µl SOC media was added and the suspension pipetted into a clean 2 ml or 1.5 ml Eppendorf tube. The transformed cells were then incubated at 37 °C, with shaking, for 1 hour. After incubation, all cells were plated on to a 14 cm diameter LB agar plate containing the required antibiotic, spread to dry and incubated overnight at 37 °C.

2.3.1.5 Preparation of bacterial stocks

An 810 µl aliquot of *E. coli* growing in exponential growth phase (OD_{600} = >1) was added to a cryovial containing 190 µl sterile 80 % glycerol (final concentration 15 % glycerol). Aliquots were snap-frozen in liquid nitrogen and stored at -80 °C.

2.3.1.6 Isolation of plasmid DNA

Plasmid DNA was used to transform competent cells as described above, and plated on to agar plates containing the specific antibiotic required to select for the specific plasmid. Individual colonies were picked from the plate using a sterile p200 tip and dropped into a 25 ml universal tube containing 5 – 10 ml LB media containing the specific antibiotic. The starter culture was incubated overnight in an incubator at 37 °C with shaking.

For plasmid used in microbiological molecular biology experiments the cells were pelleted by centrifugation at 4000 x *g* for 10 – 15 minutes. The plasmid DNA was then isolated from the cell pellet using a QIAprep spin miniprep kit (Qiagen), following the manufacturers protocol. More specifically, the cells were resuspended in Buffer P1 (section 2.2.3.2), subjected to alkaline lysis (Buffer P2; section 2.2.3.2) and neutralised (Buffer N3; section 2.2.3.2). The solution was subjected to centrifugation (10 minutes at 17,900 x *g*) and the supernatant applied to a QIAGEN silica-gel–membrane by centrifugation at 17,900 *x g* for 1 minute. Nucleic acids were adsorbed on to the silica-gel membrane in the presence of chaotropic salts, which removed water from hydrated molecules in solution. Polysaccharides and proteins were not adsorbed and were removed via wash steps with Buffer PB followed by Buffer PE (section 2.2.3.2). Pure nucleic acids were eluted under low-salt conditions using Buffer EB (section 2.2.3.2).

Plasmid DNA for use in mammalian cell transfection was isolated using a QIAGEN Plasmid Midi Kit according to manufacturers' protocol. Briefly, the starter culture was used to inoculate 50 – 200 ml LB containing the appropriate antibiotic, which was grown for another 8 - 16 hours. The cells were harvested, resuspended in Buffer P1 (section 2.2.3.2) and subjected to alkaline lysis using Buffer P2 (section 2.2.3.2) for 5 minutes at room temperature. The lysis reaction was quenched by addition of chilled Buffer P3 (section 2.2.3.2) and incubated on ice for 15 minutes. The resulting suspension was subjected to centrifugation at ≥ 20,000 x *g* for 30 minutes at 4 °C. The supernatant was removed and subjected to further centrifugation at ≥ 20,000 x *g* for 15 minutes at 4 °C. The supernatant was applied to a Qiagen Anion-Exchange Resin column previously equilibrated using Buffer QBT (section 2.2.3.2), where the highly negatively charged phosphates of the nucleic acid backbone interact with the positively charged DEAE groups on the surface of the resin. The resin was washed using Buffer QC (section 2.2.3.2). The nucleic acid was eluted from the column using high-salt buffer, Buffer QF (section 2.2.3.2). The DNA was concentrated and desalted by isopropanol precipitation, collected by centrifugation (≥15,000 x *g* for 30 minutes at 4 °C), washed with 70 % ethanol, dried and finally resuspended in TE buffer (section 2.2.3.2).

2.3.1.7 Motility assay

A 1 µl drop of culture growing in exponential growth phase (OD_{600} = >1) was added to the centre of a plate containing LB plus 0.3 % agar. The plate was sealed using Parafilm to prevent the agar from drying out and then incubated at 37 °C. Images were acquired periodically to monitor growth of the halo of *E. coli* cells that spread from the drop in the centre.

2.3.2 Molecular Biology

2.3.2.1 Polymerase Chain Reaction (PCR)

2.3.2.1.1 Colony PCR

Colony PCR was used to screen colonies for the required inserted DNA plasmid of interest during DNA cloning. PCR was performed using EconoTaq® PLUS GREEN 2X Master Mix (0.1 units/µl of EconoTaq DNA Polymerase, Reaction Buffer (pH 9.0), 400 µM dATP, 400 µM dGTP, 400 µM dCTP, 400 µM dATP, 3 mM $MgCl_2$, and a proprietary mix of PCR Enhancer/Stabilizer and blue and yellow tracking dyes; Lucigen). PCR reactions were set up as described in manufacturer's manual. Briefly, an aliquot of mastermix was diluted 1/2 with ddH_2O, the required primers were added (1 µM) mixed, and the solution aliquoted into 0.2 ml PCR tubes. Colonies from an agar plate were picked and transferred into the aliquots. Typically, thermocycling was performed as indicated in Table 4. The products were analysed by agarose gel electrophoresis (section 2.3.2.7).

Table 4. Thermocycler parameters for EconoTaq PCR

Cycling step	Temperature	Time	# of cycles
Initial denaturation	94 °C	2 min	1
Denaturation	94 °C	30 s	25 - 35
Annealing	50 – 65 °C	30 s	
Extension	72 °C	1 min / Kb	
Final extension	72 °C	10 min	1
Hold	10 °C	Indefinitely	1

2.3.2.1.2 General and Mutagenic PCR

TAG and other codon mutagenesis, deletion and addition mutagenic PCR were performed using Phusion DNA polymerase (Finnzymes™), as suggested by manufacturer's protocol. For a TAG mutagenesis, typically, a reaction was set-up containing 1 x Phusion HF buffer, 200 µM each dNTP, 0.5 µM forward primer, 0.5 µM reverse primer, 1 ng template DNA (in ddH_2O) and 0.02 U/µl Phusion DNA polymerase. For complex templates, or when initial mutagenic reaction failed to produce correctly sized product, a variety of conditions were screened. These included the use of the GC buffer, additional primer (2 µM), additional template (<5 ng), additional $MgCl_2$ or addition of DMSO. Typically, thermocycling was performed as indicated in (Table 5). The products were analysed by agarose gel electrophoresis (2.3.2.7), purified (2.3.2.2) and utilised in future ligation reactions (2.3.2.6). Consensus Kozak sequences were maintained at initiation sites of translation for all target genes, or mutated post-cloning using mutagenic Phusion polymerase PCR, using primers as indicated in Appendix A2.

Table 5. Thermocycler parameters for Phusion mutagenic PCR

Cycling step	Temperature	Time	# of cycles
Initial denaturation	98 °C	30 s	1
Denaturation	98 °C	5 - 10 s	25 – 35
Annealing	65 – 72 °C	20 s	
Extension	72 °C	15 - 30 s / Kb	
Final extension	72 °C	10 min	1
Hold	10 °C	Indefinitely	1

2.3.2.2 Purification of DNA products

DNA from PCR reactions or digestions was purified using either a Qiagen MinElute PCR Purification Kit or MinElute Reaction Cleanup Kit (if the product was <4 Kb) and QIAquick PCR Purification Kit (if products were > 4 Kb) following manufacturer's protocols. Briefly, DNA was adsorbed onto a silica membrane, in the presence of high concentrations of salt (Buffer PB or Buffer ERC; section 2.2.3.2). The membrane was washed with ethanol-containing Buffer PE (section 2.2.3.2) which removed impurities and salts, the ethanol was removed using a spin step and the DNA eluted with Buffer EB (section 2.2.3.2). Purified DNA was analysed by agarose gel electrophoresis (section 2.3.2.7), stored at -20 °C until used in ligation reactions (section 2.3.2.6).

In the circumstances where digestion of the DNA yielded more than 2 products (for example, restriction digest of plasmid DNA during the cloning process, see section 2.3.2.4), agarose gel electrophoresis was performed (section 2.3.2.7) and the required band excised from the gel under UV-light. The DNA was purified from the agarose gel slice using Gel Extraction Kit using the provided proprietary buffers (E.Z.N.A Omega Bio-tek). Briefly, 1:1 w/v XP2 Binding buffer was added to the gel slice and the gel melted by incubating the mixture at 65 °C for 10 minutes. The solution was monitored for pH, and 5 µl 5 M sodium acetate was added to bring down the pH (pH ≤7.5). DNA was adsorbed onto a HiBind® DNA mini column,

washed using 300 µl of XP2 binding buffer, further washed with 2 x 700 µl of SPW wash buffer to remove impurities and the DNA eluted using Buffer EB (section 2.2.3.2).

2.3.2.3 Phosphorylation of PCR products

After mutagenic PCR, when the primers were not bought ready-phosphorylated, PCR products were phosphorylated using T4 Polynucleotide Kinase (PNK). The reaction was performed by adding 1 µl PNK to the ligation reaction mixture, set up as described in section 2.3.2.6, and incubated at 37 °C for 30 mins. The mixture was then subjected to ligation reaction without any prior enzyme inactivation or DNA purification.

2.3.2.4 Digestion of plasmid DNA by restriction enzymes

Restriction digests were performed to prepare DNA for ligation during DNA cloning. Typically, 0.5 - 5 µg plasmid or purified PCR DNA was digested in the appropriate buffer (NEBuffer1, 2, 3 or 4; section 2.2.3.3) using 1 - 5 U enzyme/µg DNA for 1 – 6 h at 37 °C (or 25 °C; see Table 3). Digestion was deemed complete when a single band (or two, for when an insert was being cut out) was visible when a sample was analysed by gel electrophoresis. The band was also compared to that of uncut DNA. Enzyme inactivation was performed after digestion steps by heating as indicated in Table 3 or DNA was purified (2.3.2.2) before subsequent use.

*Dpn*I digestion was used to digest methylated DNA that was present in purified PCR products, namely original template plasmid DNA generated in *E. coli*, before ligation into target plasmid, when the target plasmid carried the same resistance maker as the original plasmid. This was performed as for restriction enzymatic digests and typically performed after preparative restriction digestion.

2.3.2.5 Dephosphorylation of DNA

Re-ligation of digested plasmid DNA was prevented by treating the DNA with Antarctic Phosphatase in 1 x Antarctic phosphatase buffer (section 2.2.3.3). The DNA was not purified following restriction enzymatic digest inactivation. The reaction was allowed to proceed for 15 minutes at 37 °C for 5´ extensions or blunt-

ends and 60 minutes for 3´ extensions. The enzyme was inactivated by incubating for 5 minutes at 65 °C. Linearised and dephosphorylated vectors were applied to preparative agarose gels and the band corresponding to required DNA purified (2.3.2.2). The purified DNA was used in ligation reactions (2.3.2.6).

2.3.2.6 Ligation

Ligation reactions were carried out using Quick T4 DNA ligase from the Quick Ligation™ kit (NEB). For PCR products from Phusion mutagenic reactions, 25 ng purified DNA was diluted in ddH_2O to a final volume of 5 µl, 5 µl Quick Ligation reaction buffer was added (section 2.2.3.3), followed by 0.5 µl Quick T4 DNA ligase enzyme. For cloning reactions where DNA was to be inserted into a complimentarily digested plasmid, 50 ng plasmid was used and 1:3 ratio plasmid : insert was utilised. The volume was increased to 10 µl using ddH_2O, and 10 µl Quick T4 DNA ligase added. For both, the reaction was mixed by flicking the tube, it was then subjected to pulse centrifugation to settle the contents and incubated for 5 minutes at 25 °C. The reaction was then stored on ice until required for transformation into chemically competent cells (see section 2.3.1.4).

2.3.2.7 Agarose Gel electrophoresis

Agarose gel electrophoresis was performed for either analytical or preparative purposes by preparing 0.8 - 2.0 % agarose gels in 1 x TAE containing ethidium bromide at final concentration 0.5 µg/ml (see section 2.2.3.4). The agarose was added to the TAE buffer, melted (microwave) and allowed to cool to hand-hot temperature before the ethidium bromide was added. The gel was then poured into the cast and allowed to set. The gels were run in 1 x TAE at 50 – 100 V using a horizontal VWR tank and visualised using a UV transilluminator (Syngene) and camera (Canon) or a GelDoc System (Bio-Rad). Samples and the standards were prepared prior to loading by the addition of 6 x DNA loading buffer.

2.3.2.7.1 DNA electrophoresis standard

The 1 Kb ladder was purchased from New England Biolabs (N3232). The ladder was diluted in DNA loading buffer and loaded onto the agarose gel.

2.3.2.8 DNA quantification

A NanoDrop ND-1000 UV-vis spectrophotometer was used to determine the concentration of all DNA. A 1 – 2 µl sample or buffer was pipetted directly onto the optical measurement surface connected to one end of a fibre optic cable. During operation, the sample liquid bridges the gap. A pulsed xenon flash lamp was used as a light source and a spectrometer utilising a linear CCD array was used to analyse the light after passing through the sample at either a 1 mm or 0.2 mm pathlength. A computer running DNA NanoDrop software was used to analyse the output and calculate concentration. The 260/280 ratio was used as a guide to quantify purity, along with visualisation on an agarose gel. Only plasmid DNA with 260/280 ratio > 1.90 was used in downstream applications.

2.3.2.9 DNA sequencing

DNA sequencing was carried out by the in-house DNA sequencing Core, Molecular Biology Unit (Cardiff University) or by Eurofins MWG Operon.

2.3.2.10 Non-natural amino acids

For incorporation of non-natural amino acids into proteins expressed using microbiological organisms, non-natural amino acids were prepared by dissolving the required amount in ddH_2O and enough concentrated NaOH (<5 M) to dissolve the powder. For large expressions (> 1 L) the NAA powder could be added straight to the inoculated culture. Non-natural amino acid was added either at inoculation, when OD_{600} = 0.3 or approximately 1 hour – 30 minutes prior to induction.

2.3.3 Protein Analysis

2.3.3.1 Sodium dodecyl sulphate polyacrylamide gel electrophoresis

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) protein tricene (Table 6) or glycine (Table 7) gels were used to analyse protein after the purification steps. Typically, mini gels (7 cm x 10 cm x 0.75 mm) were prepared. Electrophoresis was performed using Mini-PROTEAN tetra cell system (Bio-rad 165-8003). Depending on the size of the protein for analysis, a suitable separating gel was assembled between ethanol-cleaned minigel plates. Ethanol was overlaid to

enable formation of straight edge by eliminating bubbles. After polymerisation of the separating gel, the stacking gel was prepared (on top of the separating gel), a comb was inserted and the gel allowed to polymerise. Gels were stored at 4 °C until required. The samples were run at 60 – 120 V in either cathode and anode buffer for tricine gels or Laemmli buffer for glycine gels (Table 2). The gels were subjected to electrophoresis until the tracking dye of the loading buffer came out the bottom of the gel.

Table 6. Tricine protein gel solutions

Solutions	**Separating gel, 10 %**	**Separating gel, 16 %**	**Stacking gel, 4 %**
Acrylamide/bis-acrylamide 37.5 : 1, 30 % solution	10 ml	16 ml	1.6 ml
Gel buffer (see Table 2)	10 ml	10 ml	3 ml
70 % glycerol	3 ml	4 ml	-
ddH_2O	7 ml	0.2 ml	7.4 ml
10 % APS	150 µl	133 µl	90 µl
TEMED	15 µl	13.2 µl	9 µl

Table 7. Laemmli protein gel solutions

Solutions	Separating gel, 10 %	Separating gel, 12 %	Stacking gel, 6 %
Acrylamide/bis-acrylamide 37.5 : 1, 30 % solution	5 ml	6 ml	1 ml
ddH_2O	3.8 ml	2.6 ml	2.65 ml
1 M Tris-HCl pH 8.8 **1 M Tris-HCl pH6.8**	6 ml -	6 ml -	- 1.25 ml
10 % SDS	150 µl	150 µl	50 µl
10 % APS	150 µl	150 µl	50 µl
TEMED	15 µl	15 µl	5 µl

2.3.3.1.1 SDS-PAGE electrophoresis standard

The Protein Marker, Broad Range (2-212 kDa; New England Biolabs (P7702S)) was diluted in loading buffer, prepared as outlined in section 2.3.3.1.2 and loaded on to the gel. It was used to estimate the molecular weight and purity of protein samples also run on the gel.

2.3.3.1.2 Sample preparation for SDS-PAGE

Samples (typically a 15 µl sample) and 15 µl molecular weight marker for SDS-PAGE were boiled with 1 x SDS-containing sample buffer (Table 2; 10 minutes) cooled to room temperature, collected by brief centrifugation and loaded in to wells of the assembled gel.

2.3.3.1.3 Visualisation of SDS-PAGE gels

After electrophoresis, the gels were removed from the plates and added to Coomassie stain solution (Table 2) for 0.5 – 1 hour, with rocking. The gels were washed with water and then destained using destain solution (Table 2) by repeatedly adding fresh destain solution and rocking until protein bands were visible.

2.3.3.2 Concentration of protein

Protein in large volumes was concentrated using solvent absorption technique utilising a Vivapore 10/20, MWCO 7500 (Satorius Stedim) or Amicon Ultra-15 ml centrifugal filter unit in conjunction with centrifugation at 4000 x *g*. For ultra-concentration or for small volumes (500 µL down to 15 µL), for example for crystallisation, an Amicon Ultra-0.5 ml device in conjunction with centrifugation at 14,000 x *g* was utilised.

2.3.3.3 Determination of protein concentration

Protein concentration was determined using the NanoDrop ND1000 as described in section 2.3.2.8, utilising the protein analysis NanoDrop software. The concentration was calculated using the absorption at 280 nm along with the extinction coefficient specific to the protein being analysed. This extinction coefficient was calculated using the ExPASy Protparam online tool (Gasteiger *et al.* 2005), which utilises the presence of aromatic amino acids present within the protein sequence.

3 CHAPTER 3: PHOTOCHEMISTRY OF ARYL AZIDES IN DEFINED PROTEIN ENVIRONMENTS

3.1 Introduction

This chapter explores whether the incorporation of a reactive moiety into a protein, in the form of a non-natural amino acid via amber suppression technology, can aid in the elucidation of reactive intermediates by creating a defined and stabilising environment and reducing the complications of bimolecular reactions that occur when studying highly reactive intermediates. More detailed examination of the photochemistry of aryl azides within a biological environment would also aid in expanding the utility of aryl azide chemistry in biology.

The amber suppression method can incorporate non-natural amino acids that contain reactive moieties not naturally found in proteins or biological situations. The non-natural, azide group containing amino acid, azidophenylalanine (azF; Figure 27), is of particular importance due to the wide use of aryl azide based photocrosslinking in biological applications, namely, in photoaffinity labelling, mapping RNA structural neighbours, crosslinking proteins, studying ligand : enzyme interactions and site selectively modifying protein by reaction with azide (Buchmueller et al. 2003; Chen and Ebrigh 1993; Fleming 1995; Sugawara and Matsuda 1995; Zhang et al. 2005a). Despite widespread use, there is still considerable scope for applying the unique chemistry of the aryl azide to biological systems. For example, aryl azide photochemistry involves much more than just crosslinking.

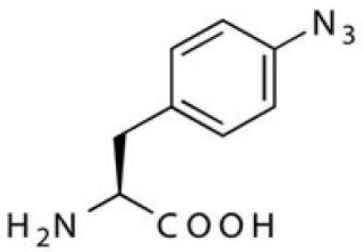


Figure 27. Azidophenylalanine

The incorporation of non-natural amino acids possessing spectroscopically active moieties makes it possible to track the incorporation of the non-natural amino acid by the presence of this spectroscopic signal. The incorporation of the azF is an ideal

candidate as the spectral properties of azide are distinct from those naturally found in proteins. The azide group is of particular interest because the photoreaction of this moiety alters its spectroscopic signals, particularly the electron paramagnetic spectroscopy (EPR) spectrum that arises due to the creation of unpaired electrons during photolysis of the azide group. EPR is not only sensitive but also highly informative with regards to the structural information that is elucidated for the transient radicals formed from the azide.

The use of amber suppression allows the position specific incorporation of the azide-containing non-natural amino acid into the target protein. The use of a protein scaffold allows the environment surrounding the azide to be determined. It was thought that surrounding the azide with hydrophobic residues, for example, within the hydrophobic pocket of T4 Lysozyme, would trap the reactive species and protect against the bimolecular reactions that complicate the study of the azido group's photochemical reaction intermediates.

The specific aims of this part of the project were to;

- Incorporate the non-natural amino acid, azF, into a protein scaffold using the amber suppression technology
- Use spectroscopic techniques to assess the incorporation of the azido group.
- Irradiate the azido-substituted protein to produce the triplet nitrene species, using conditions known to favour this product, namely the photolysis of the mutant protein at 77 K. The production of this triplet nitrene was then monitored using EPR.
- Further investigate whether other spectroscopic techniques could be used to study the photo-reactive intermediates, namely ultraviolet-visible (UV-vis), infrared (IR) and Raman spectroscopy.
- To investigate whether x-ray crystallography could be used to visualise the azido group and possibly the reactive intermediates or reacted product after photolysis.

Firstly, a reproducible, high-yielding protein expression system was required to express protein. Expression of wild type (WT) protein, iodoF-substituted and azido-

substituted protein was investigated. A variety of conditions were investigated, which included, the plasmid system utilised for expression, the media and the growth temperature. Secondly, the purification of the protein was investigated. Subsequently, the purified protein was used in crystallisation screens and the spectroscopic investigations.

3.2 Materials and methods

3.2.1 T4 lysozyme Expression

Electrocompetant BL21(DE3) gold *E. coli* cells were cotransformed with the previously constructed plasmids; pET101T4L153 / pSUPiodoRS, pETT4L153TAG / pAAazRS, pBADT4L153TAG / pDULEazRS or pBADT4LWT (details of which found in Appendix A1). A dual resistant colony was picked and used to inoculate a seed culture of LB containing the appropriate antibiotics. Seed cultures were grown overnight in LB at 37 °C and 1 : 250 dilution used to inoculate expression cultures. Cultures were either grown in LB media, TB autoinduction media or ZYM-5052 media containing appropriate antibiotics. 1 mM final concentration non-natural amino acid was added at an appropriate time, depending on expression conditions. Induction of expression was performed either by adding 0.5 mM final concentration isopropyl-β-D-thiogalactopyranoside (IPTG) at OD_{600} = 0.5, or via glucose/lactose auto-induction (lacO regulation) or via addition of arabinose (araC regulation) at OD_{600} = 4 - 7. Protein expression was carried out at 37 °C for 4 hours, 30 °C for 8 hours or at 25 °C overnight. Cells were pelleted by centrifugation at 4000 *x g* for 20 minutes at 4 °C, resuspended in lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, pH 8.0) and frozen at – 80 °C.

3.2.2 T4 lysozyme extraction

Pellets were defrosted, 1 mM phenylmethanesulfonylfluoride (PMSF) and 2 µl per 250 ml culture benzonase nuclease added and incubated for 30 minutes on ice. The expressed lysozyme lysed the cells and this provided evidence of production of functional lysozyme. Cells were further lysed by sonication (Vibra cell, Sonics) and the insoluble fraction was pelleted by centrifugation at 10,000 rpm, 40 minutes, 4 °C (rotor 5615, Rotanta 460R, Hettich).

3.2.3 T4 lysozyme purification

3.2.3.1 Nickel affinity chromatography

Full-length T4 lysozyme was purified from the supernatant collected from centrifugation step outlined in "T4 lysozyme extraction" using affinity purification between the histag on the C-terminal of the expressed protein and Ni-NTA beads. The supernatant was incubated with 1 ml / 1 L culture Ni-NTA (Novagen) for 1 hour (4 °C), with rotation. The mixture was applied to a filter column, washed with buffers containing increasing concentrations of imidazole (50 mM NaH_2PO_4, 300 mM NaCl, 20 / 30 / 50 / 75 mM imidazole, pH 8.0) to remove impurities and weakly bound peptide. Protein was eluted using 50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, pH 8.0. Wash steps and elution were analysed by SDS-PAGE, as described in general materials and methods.

3.2.3.2 Size exclusion chromatography (SEC)

T4 Lysozyme was further purified using either Superdex 75 size exclusion column (Amersham Biosciences) on AKTApurifier, previously equilibrated with 100 mM NaH_2PO_4, 0.55 M NaCl, pH 6.7. All buffers were filtered and de-gassed to ensure no precipitation or bubbles were introduced into the Superdex column. Protein was concentrated and loaded, in 250 – 500 µl aliquots onto the SEC column, using a 1 ml injection loop. The flow rate was 0.7 ml / minute and 1 ml fractions were collected until 1 column volume had passed through the column.

3.2.4 T4 Lysozyme crystallisation

Hanging drop and sitting drop plates were used. Possible buffers for crystallisation included 2.0 - 2.2 M aqueous sodium/potassium phosphate buffer, pH 6.7 - 7.1 with 15 mM hydroxyethyl disulfide (Xie et al. 2004) and 0.55 M NaCl, 14 mM mercaptoethanol, 1 mM $MgCl_2$, 0.01 M sodium phosphate, pH 6.7, using varying 2.2 M NaH_2PO_4, 1.8 M K_2HPO_4, pH 6.7 (Remington et al. 1978).

The crystallisation screen is summarised in Table 8.

Table 8. Crystallisation screen parameters

buffer	component	variable
crystallisation	Sodium/potassium phosphate	1.0 – 2.5 M
	pH	6.3 – 7.3
	hydroxyethyl disulphide	15 mM 4 – 8 µl/ml
	NaCl	0 – 0.35 M
protein	concentration	27 – 10 mg/ml
	: buffer ratio	5 : 5 – 5 : 1; 1 : 1 µl
temperature		4, 11, 25 °C

3.2.5 Spectroscopic analysis

3.2.5.1 Protein photolysis for EPR

Purified T4L153azF protein (8 mg/ml) was snap frozen in liquid nitrogen at 77 K in 50 % glycerol solution to maintain a glassy nature, so the light during photolysis could penetrate the sample and to reduce noise in the EPR spectrum. Photolysis in quartz – suprasil 4 mm EPR tubes with a 500W Oriel Instruments UV lamp, incorporating a Hg/Xe arc lamp (250 nm to >2500 nm) was used for all irradiations in the presence of a water filter. The UV output below 280 nm accounts for only 4 – 5 % of the total lamp output. The samples were irradiated for 2 min at 77 K in an EPR liquid N_2 finger dewar (Wilmad Glass) and rapidly transferred to the pre-cooled cryostat (10 K) for EPR measurements.

3.2.5.2 Electroparamagnetic resonance spectroscopy (EPR)

The continuous wave (CW)-EPR spectra were recorded on an a Bruker ESP300e series spectrometer operating at X-band frequencies (9.75 GHz), 100 kHz field modulation, 10 G modulation depth, mW microwave power and equipped with a high sensitivity cavity (ER 4119HS). The measurements were obtained by Dr Damien Murphy (Cardiff University). Measurements were taken at 10 K in an Oxford liquid helium cryostat. Twenty-five measurements were taken to increase

signal to noise ratio, scanning 1000 – 8000 G. EPR simulations were performed by Dr Jamie Platts (Cardiff University) using the Easyspin software (Stoll and Schweiger 2006), and the commercially available Bruker SIMFONIA software.

3.2.5.3 Ultraviolet-visible spectroscopy

A 625 µM solution of azF in water plus NaOH or 100 µM protein was irradiated at 302 nm for 30 minutes total with spectra taken at 30 second intervals to 5 minutes, then after 10 and 30 minutes. Spectra were taken using Jasco 630 UV-visible spectrophotometer.

3.2.5.4 Fourier transform infrared spectroscopy

Fourier transform infrared spectroscopy (FTIR) spectra were obtained using a Bruker IFS-66 FTIR Spectrometer. Solutions of Phe and azF NAA were prepared at 20 mM in NaOH and sandwiched between CaF_2 plates with a pathlength of 25 µm.

3.2.5.5 Circular dichroism

Solutions of 50 µM WT and T4L153azF protein were dialysed in to degassed 100 mM NaH_2PO_4, pH 6.7. A 2 µm pathlength quartz cuvette (Helma) was used to take measurements on an Applied Photophysics Chirascan Circular Dichroism Spectrometer. Measurements were taken at 20 °C under N_2.

3.2.5.6 Raman spectroscopy

Various concentrations of azF and iodoF were prepared in aqueous NaOH (1, 2, 5, 10, 20, 50, 100 mM). Samples were sandwiched between Quartz plates with a 1 µm spacer. Raman spectroscopy was performed by Dr Elisabetta Canetta (School of Bioscience, Cardiff University).

3.3 Results

3.3.1 Expression optimisation

A number of plasmid systems were available for the expression of mutant T4L, and the complementary synthetase and $tRNA_{CUA}^{Tyr}$ containing plasmid, to incorporate the non-natural amino acids by the amber suppression technology. These plasmid systems were investigated to find the optimal expression system.

3.3.1.1 Plasmids pT4L153TAG/pAAT4L153TAG and pBKiodoRS

Plasmid pT4L153TAG encodes a Phe153 → TAG mutant cysteine-free T4 lysozyme under the control of a bacteriophage *T5* promoter and t_0 terminator, and the $tRNA_{CUA}^{Tyr}$ gene under the control of the *lpp* promoter and *rrnC* terminator. Plasmid pBK-iodoRS encoded the iodoF-specific tRNA-aminoacyl synthetase (iodoRS) under the control of the constitutive *E. coli* GlnRS promoter and terminator (Figure 28). It had been shown that this system of plasmids could successfully incorporate iodoF into codon 153 of T4L and therefore the expression of T4L153iodoF was attempted, using the conditions as described by Xie et al (2004). Protein expressions were performed using *E. coli* BL21(DE3) co-transformed with pT4L153TAG and pBK-iodoRS. A variety of media were used, namely, LB, TB, 2YT, M9 minimal media with IPTG induction and autoinduction, however, the yield of mutant protein was low (data not shown). In addition, it was thought that for ease of purification of the truncated protein from the non-natural amino acid-containing full-length protein, a C-terminal histag would be useful. Therefore other plasmid systems were investigated.

The pAAT4L153TAG plasmid had been previously created (Dr A. Antonczak), that encoded a codon optimised Phe153 → TAG mutant cysteine-free T4 lysozyme, with a C-terminal histag sequence, under the control of a bacteriophage *T5* promoter and t_0 terminator, with *lacO* repression, and the $tRNA_{CUA}^{Tyr}$ gene under the control of the *lpp* promoter and *rrnC* terminator (Figure 28). This plasmid was similar to the pT4L153TAG, however during the construction by Epoch Life Science the DNA

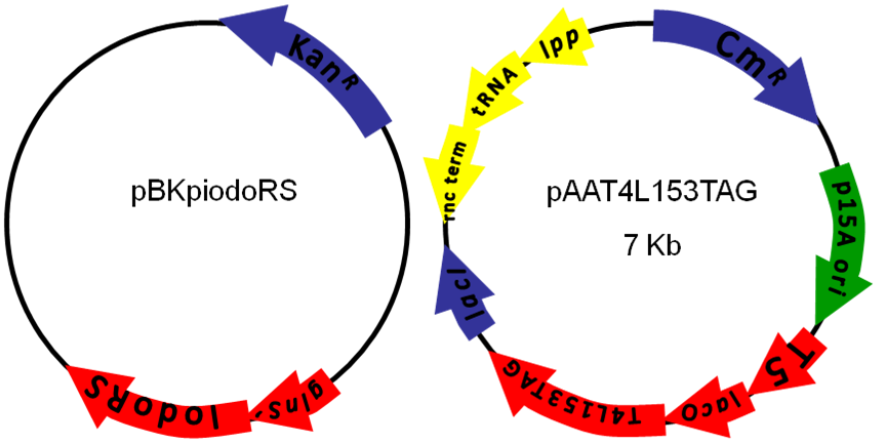


Figure 28. Plasmid maps of pBKiodoRS and pAAT4L153TAG/pT4L153TAG.

sequence used was codon optimised for use in *E. coli*. It was thought that the codon optimisation would allow increased expression of the mutant protein by reducing the depletion of amino acid charged tRNA for specific codons.

A variety of expressions were performed using plasmids, pAAT4L153TAG and pBKiodoRS, co-transformed into BL21(DE3) *E. coli.* However, the plasmid was not tolerated by the *E. coli* as there was limited growth of the cultures and cell death after approximately 1 hour in culture (Figure 29).

The T4L gene was re-sequenced and it was determined that during gene construction, the codon Thr152 of the T4L gene was omitted (Figure 31). There was also R154K mutation, and a GGSFG linker region between the T4L gene and hexahistidine tag, which may have affected the expression of the protein. Potential issues are discussed below. Due to the missing Thr152, mutagenic PCR was performed to try to re-introduce the missing codon. Three different primer combinations were tried, pAAT4L153TAG+ACG FP/RP, JLMt4l003-F/004-R and JLMt4l005-F/006-R (Appendix A2). However, the plasmid appeared to be refractory to the insertion of the codon, as no correctly-sized PCR could be obtained. Therefore other plasmid systems were investigated.

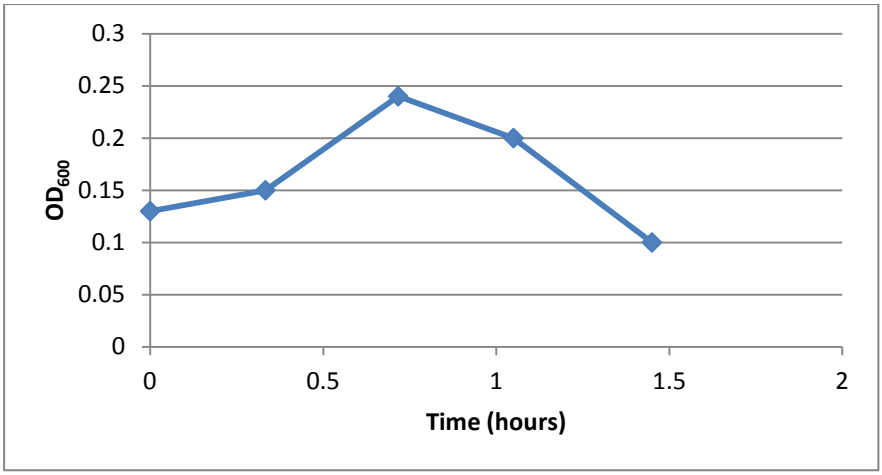


Figure 29. Graph showing growth curve of BL21(DE3) *E. coli* co-transformed with plasmids pAAT4L153TAG and pBKiodoRS. Seed cultures were grown in LB plus appropriate antibiotics overnight using a colony picked from a plate containing BL21(DE3) *E. coli* co-transformed with the plasmids. This seed culture was then used to inoculate LB containing the appropriate antibiotics, a sample was taken from the culture and an OD_{600} reading taken at the specified timepoints. The cell densities were plotted.

3.3.1.2 Plasmids pETT4L153TAG and pSUPiodoRS

The previously constructed ampicillin resistant pETT4L153TAG plasmid (Tippmann, E, 2008) encodes mutant cysteine-free bacteriophage T4 lysozyme, harbouring Phe153→TAG mutation, under the control of an IPTG inducible T7 promoter and terminator. The chloramphenicol resistant pSUPiodoRS (Figure 30) plasmid encodes a p-iodophenylalanine-specific tRNA aminoacyl synthetase (iodoRS) under the control of *glnS*' promoter and terminator, and 2 x 3 $tRNA_{CUA}^{Tyr}$ gene sequences under the control of *Pro*K promoter and terminator (Ryu and Schultz 2006). The presence of a C-terminal hexahistidine tag would, as mentioned previously, simplify purification of NAA-containing protein. The plasmid harboured a L99F, a M102A mutation and a GGSG linker region between the T4L and the hexahistidine tag (Figure 31). This plasmid system was more promising and effectively expressed mutant protein when co-transformed into *E. coli* BL21(DE3). Therefore these plasmids were used for subsequent expression optimisations.

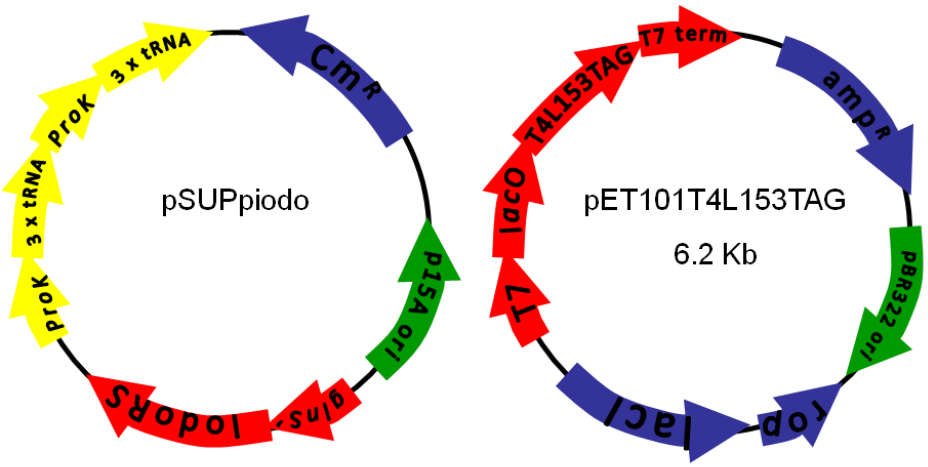


Figure 30. Plasmid maps of pSUPiodoRS and pET101T4L153TAG

```
pBADT4LWT          MDIFEMLRIDEGLRLKIYKDTEGYYTIGIGHLLTKSPSLNAAKSELDKAIGRNTNGVITK 60
pBADT4L153TAG      MDIFEMLRIDEGLRLKIYKDTEGYYTIGIGHLLTKSPSLNAAKSELDKAIGRNTNGVITK 60
pT4L153TAG         MNIFEMLRIDEGLRLKIYKDTEGYYTIGIGHLLTKSPSLNAAKSELDKAIGRNTNGVITK 60
pAAT4L153TAG       MNIFEMLRIDEGLRLKIYKDTEGYYTIGIGHLLTKSPSLNAAKSELDKAIGRNTNGVITK 60
pETT4L153TAG       MNIFEMLRIDEGLRLKIYKDTEGYYTIGIGHLLTKSPSLNAAKSELDKAIGRNTNGVITK 60
                   *:************************************************************

pBADT4LWT          DEAEKLFNQDVDAAVRGILRNAKLKPVYDSLDAVRRAALINMVFQMGETGVAGFTNSLRM 120
pBADT4L153TAG      DEAEKLFNQDVDAAVRGILRNAKLKPVYDSLDAVRRAALINMVFQMGETGVAGFTNSLRM 120
pT4L153TAG         DEAEKLFNQDVDAAVRGILRNAKLKPVYDSLDAVRRAALINMVFQMGETGVAGFTNSLRM 120
pAAT4L153TAG       DEAEKLFNQDVDAAVRGILRNAKLKPVYDSLDAVRRAALINMVFQMGETGVAGFTNSLRM 120
pETT4L153TAG       DEAEKLFNQDVDAAVRGILRNAKLKPVYDSLDAVRRAAFINAVFQMGETGVAGFTNSLRM 120
                   ************************************:** ******************

pBADT4LWT          LQQKRWDEAAVNLAKSRWYNQTPNRAKRVITTFRTGTWDAYKNLS---HHHHHH-- 171
pBADT4L153TAG      LQQKRWDEAAVNLAKSRWYNQTPNRAKRVITT-RTGTWDAYKNLS---HHHHHH-- 170
pT4L153TAG         LQQKRWDEAAVNLAKSRWYNQTPNRAKRVITT-RTGTWDAYKNL------------ 163
pAAT4L153TAG       LQQKRWDEAAVNLAKSRWYNQTPNRAKRVIT--KTGTWDAYKNLGGSGHHHHHH-- 172
pETT4L153TAG       LQQKRWDEAAVNLAKSRWYNQTPNRAKRVITT-RTGTWDAYKNLGGSGHHHHHH-- 173
                   ******************************  :**********
```

Figure 31. Comparison of T4L protein encoded by plasmids. The DNA sequence obtained was translated into protein sequence using ExPASy translate tool (Artimo P *et al.* 2012), these sequences were then compared using the ClustalW2 tool (Larkin *et al.* 2007).

BL21(DE3) gold *E. coli* cells were cotransformed with the plasmids pETT4L153TAG and pSUPiodoRS and used to express T4L153iodoF. The media was either supplemented with, or without, 1 mM iodoF. Cellular lysate from the two cultures were isolated and samples applied to an SDS-PAGE gel. Comparing the intensity of the band corresponding to the size of full-length T4L, it was demonstrated that when iodoF was supplemented in the media, full-length T4L was present (Figure 32). When iodoF was withheld from the media the band, corresponding to the size of full-length T4L, was dramatically lower in intensity. The band just above the size of full-length T4L was presumed to be truncated T4L (at codon 153). The intensity of this band in the culture not supplemented with iodoF was greater than that seen in the culture supplemented with iodoF.

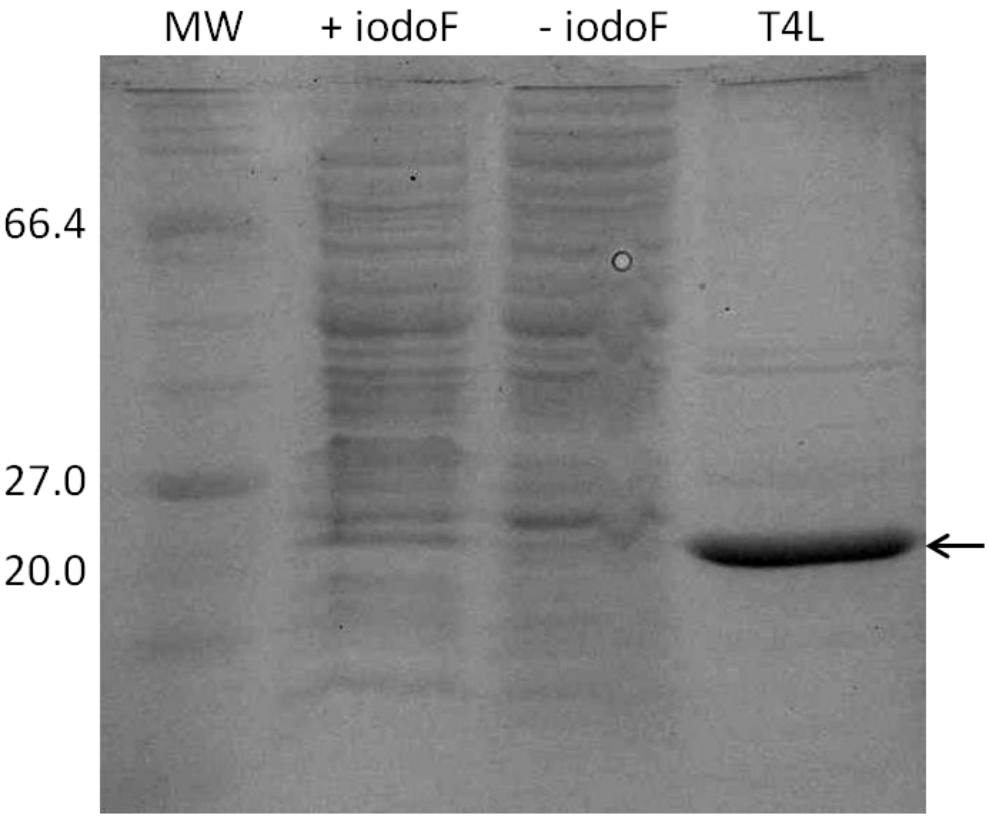


Figure 32. SDS-PAGE gel comparing cellular lysate of cultures harbouring plasmids pETT4L153TAG and pSUPiodoRS. Cultures were either grown with or without addition of iodoF. Arrow indicated size of full-length T4L protein.

Small scale expression cultures (200 ml) were used to find a robust method for expression of T4L153iodoF protein. Growth media LB with IPTG induction, TB autoinduction and ZYM-5052 autoinduction media; a variety of expression temperatures (25, 30, 37 °C) and time period of expression were tested. The soluble protein, from these expressions, was extracted, purified by Ni-NTA and analysed by SDS-PAGE, as described in materials and methods. The yields, ease of expression protocol and time of expression were evaluated and it was found that using ZYM-5052 autoinduction media, growing the cultures at 30 °C for 20 hours gave optimal expression. Therefore these plasmids were used to express protein to perform purification optimisation. The results of the investigation into the plasmids systems in their ability to express NAA-substituted T4 Lysozyme are summarised in Table 9

Table 9. Summary of plasmid systems and their ability to express NAA-substituted T4Lysozyme

Plasmid – T4L (provider)	**Complementary plasmid – aaRS**	**Expression *E. coli***	**Notes**	**Issues**	**Yield**	**Action**
pT4L153TAG (Dr J. Xie)	pBKiodoRS	BL21(DE3)	T5 promoter Not optimised sequence tRNA	Low expression No histag	0.3 mg/L pure	Different medias investigated
pAAT4L153TAG (Dr A. Antonczak)	pBKiodoRS	BL21(DE3)	T5 promoter *lac*O regulation Optimised sequence – T4L 75% nt sequence similarity to pT4L or pET plasmids Histag 2 x tRNA	Missing ACG codon at 152, thereby producing protein with NAA exposed on exterior of protein No growth	-	Could not re-introduce missing codon by site-directed mutagenesis No remedy
pET101T4L153TAG (Dr Eric Tippmann)	pSUPiodoRS (+ 6 x tRNA)	BL21(DE3)	T7 promoter *Lac*O regulation Histag	Only iodo RS available Low yield expression Lots of truncated protein produced	<3 mg/L pure	Optimisation of expression conditions Optimisation of purification
pBADT4L153TAG (Dr S. Miyake-Stoner)	pDULEazRS (tRNA)	Top10	*ara*BAD promoter *ara*C regulation histag	High yield Efficient incorporation of azF	<8 mg/L pure	Expression of WT T4L and T4L153azF for spectroscopic evaluations

3.3.2 Purification optimisation

3.3.2.1 Nickel affinity chromatography

T4L153iodoF protein expressed using BL21(DE3) gold *E. coli* cells cotransformed with the pETT4L153TAG and pSUPiodoRS plasmids was used during optimisation of purification. The use of nickel affinity chromatography was utilised to purify full-length NAA-containing T4L (19 kDa) away from whole cell lysate and in particular, truncated T4L protein (17.1 kDa). Only full-length protein should harbour a hexahistidine tag and therefore bind to the nickel affinity chromatography column. Analysis using SDS-PAGE demonstrated that binding of T4L153iodoF to the column was efficient as there was no band corresponding to full-length T4L (Lane 1, Figure 33). T4L153iodoF protein was eluted from the column in a purer form, however one major contamination remained (Lane 5, Figure 33).

This result demonstrated that NAA-containing T4L could be purified from whole cell lysate using Nickel affinity chromatography, subsequently this technique was used routinely as a first step purification after protein expression.

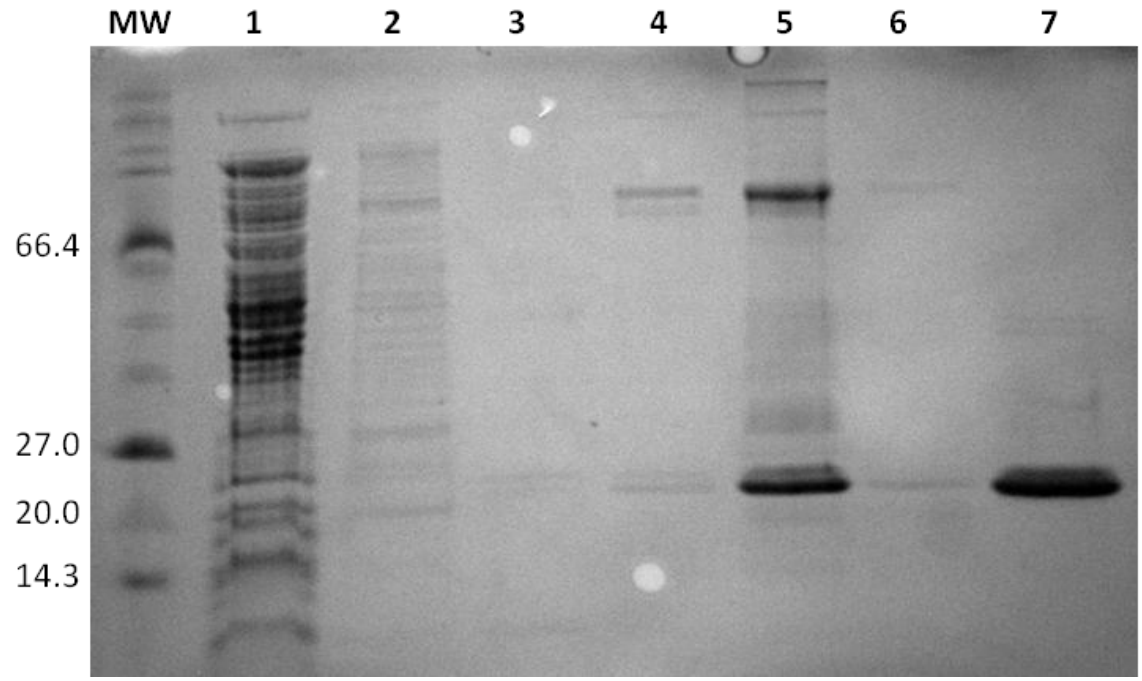


Figure 33. SDS-PAGE gel showing Ni-NTA purification of T4L153IodoF. Expression of T4L153IodoF was performed using LB, with IPTG induced expression for 3.5 hours at 37 °C. Cell culture lysate was incubated with Ni-NTA for 1 hr at 4°C, loaded onto filter column (flow-through lane 1), washed with 20 mM (lane 2), 30 mM (lane 3), 50 mM (lane 4) imidazole wash buffer, protein was eluted in 6 ml 250 mM imidazole buffer (lane 5). Lane 7 represents T4L44IodoF previously purified. A 15 µl sample of each wash was loaded onto 12 % tricene gel.

3.3.2.2 Size exclusion chromatography

Protein eluted from nickel column was considerably purer, but other bands were still present, as determined by SDS-PAGE gel. Therefore protein was further subjected to using size exclusion chromatography (SEC), using a Superdex75 column on an AKTApurifier, as described in materials and methods. The major impurity was of a considerably larger mass, therefore it was thought that SEC would be the most effective method to further purify NAA-containing T4L. Ni-NTA purified protein was concentrated and loaded onto the SEC column. The UV trace was monitored and fractions with peak UV, corresponding to the correct size protein, were analysed by SDS-PAGE. A typical SDS-PAGE gel (A) and AKTA trace (B) are shown in Figure 34. The fractions containing pure T4L, as identified by SDS-PAGE, were pooled and the yield equated to 0.29 mg purified protein / L culture.

This result demonstrated that using Nickel affinity chromatography followed by SEC purification step, NAA-containing protein could be purified effectively.

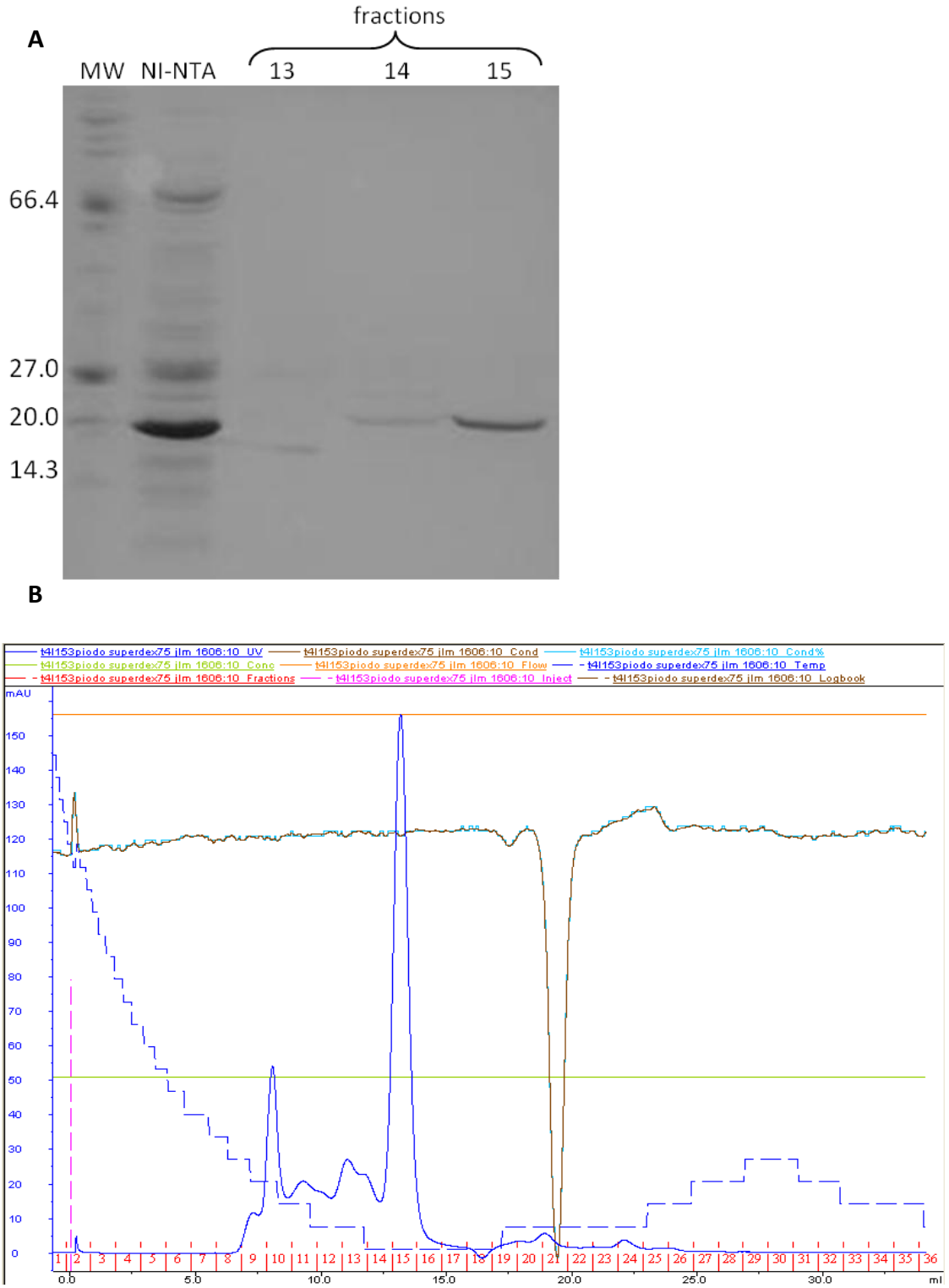


Figure 34. Purification of T4L153iodoF by size exclusion chromatography. A) A representative SDS-PAGE gel of nickel affinity purified T4L153iodoF (lane 1), sample of size exclusion 1ml fractions (Superdex75) 13, 14, 15.

B) Corresponding size exclusion chromatography trace. 2L ZYM-5052 media was used to express T4L153iodoF, T4L, protein was subjected to Ni-NTA purification, concentrated to 600 µl then applied to Superdex75 column in 2 x 300 µl aliquots. Peak UV (solid blue line) fractions, 13, 14, 15 were run on SDS-PAGE gel.

3.3.3 Investigation in to the expression of T4L153azF

Using the pET101T4L153/pSUPiodoRS system, an expression and purification system has been optimised to produce purified T4L153iodoF. However, the project aim was to produce T4L harbouring the azF amino acid within the protein's hydrophobic pocket, therefore, a system to incorporate the azF was required. There were a few possible methods to perform expression of T4L153azF, and these were investigated.

3.3.3.1 Using pSUPiodoRS to incorporate azF

It has been shown previously, that the synthetases engineered for a specific amino acid display cross-reactivity to other non-natural amino acids (Antonczak 2012 Thesis). This promiscuity was investigated by expressing T4L, using the plasmids pET101T4L153TAG and pSUPiodoRS, but supplementing the media with azF in place of iodoF. T4l153azF was successfully expressed, however the expression levels were low and other expression systems became available.

3.3.3.2 Using pAAazRS to incorporate azF

The plasmid, pAAazRS had previously been created and shown to successfully incorporate azF into proteins using amber suppression (Personal communication, Dr Matt Edmundson 2010). The ability of this plasmid to express azRS to allow incorporation of azF into T4L at codon 153 was also investigated. Analysis by SDS-PAGE demonstrated that T4l153azF was successfully expressed, the expression levels were compared to that expressed using the pDULEazF / pBADT4L153TAG system described in section 3.3.3.3 (Figure 36).

3.3.3.3 Using pDULEazF to incorporate azF

The plasmids pBADT4LWT, pBADT4L153TAG and pDULEazRS were kindly donated by Ryan Mehl (Oregon State University). The tetracycline resistant pDULE plasmid encoded an azF-specific synthetase, under an *lpp* promoter and *rrn*B terminator and tRNA under an *lpp* promoter and *rrn*C terminator (Figure 35). pDULEazF plasmid along with pBADT4L153TAG plasmid has been shown previously to produce

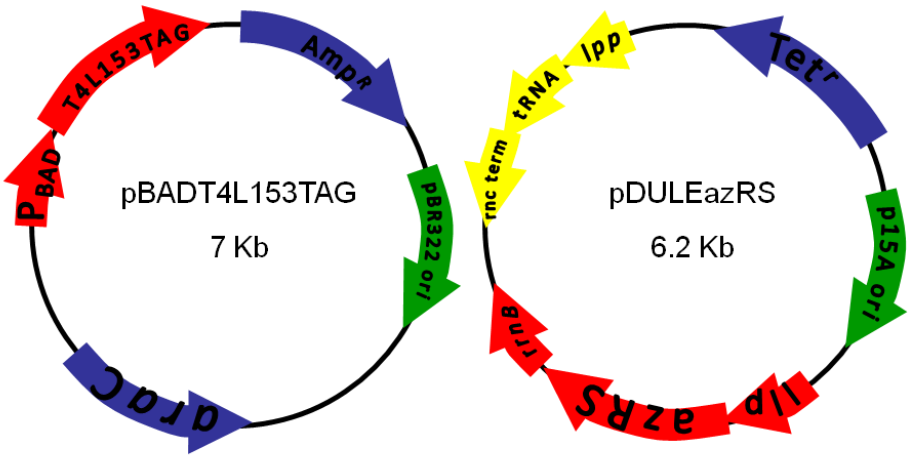


Figure 35. Plasmid maps of pBADT4L153TAG and pDULEazRS.

high yields of NAA-containing protein (Farrell *et al.* 2005; Tian *et al.* 2004), including T4L153NAA (Miyake-Stoner et al. 2009). The plasmids were sequenced and the protein encoded by these plasmids was compared to the previously used pET101T4L153TAG. The pBADT4LWT / pBADT4L153TAG plasmids encoded the cysteine-free T4LWT / 153TAG, under the control of *ara*BAD promoter with *ara*C regulation.

Comparing the protein sequences (determined by translation of DNA sequence data) it was noted that the pBAD T4L genes harboured a number of mutations that may have affected the expression levels. Namely, a N2D mutation, F99L and no linker region joining the protein to the hexahistidine-tag (Figure 31). In addition, it was thought that using the tightly regulated *ara*BAD promoter, by the *ara*C gene product, would permit more control over expression and therefore allow higher expression levels of T4L.

3.3.3.4 Comparison of plasmid systems, pET vs. pBAD

Due to the low-yielding expression using the plasmids system pETT4L153TAG and pSUPiodoRS or pAAazRS, the use of the plasmid system, pBADT4L153 and pDULEazF was investigated. A protein expression using the two plasmids along with the corresponding aaRS expressing plasmid (pETT4L153TAG and pAAazRS; pBADT4L153 and pDULEazF) were performed side-by-side. Miyake-Stoner et al, (2009) argues that because expression of T4L causes lysis of the host cell, the use of autoinduction media is contraindicated. Therefore, the method described by the group suggests that growing the cultures to high density followed by manual induction, recapitulated the conditions present during autoinduction, without the presence of the inducing component that could cause leaky production of T4L.

An expression using 200 ml ZYM-5052 media was used, the culture was grown at 37 °C, azF was added at OD_{600} = 0.9, expression was induced by addition of 0.2 % arabinose or lactose for the pBAD and pET cultures respectively at OD_{600} = ~5. Protein expression was allowed to proceed for 4 hours. The protein was extracted and protein subjected to Ni-NTA affinity purification. An SDS-PAGE gel was used to

compare the expression levels (Figure 36). It was demonstrated that using the expression conditions, as suggested by Miyake-Stoner (2009), using the pBAD system gives greater yields when compared to the pET system, when using the same expression method. It was therefore decided that the pBAD system should be utilised for expressing protein for use in spectroscopic evaluations.

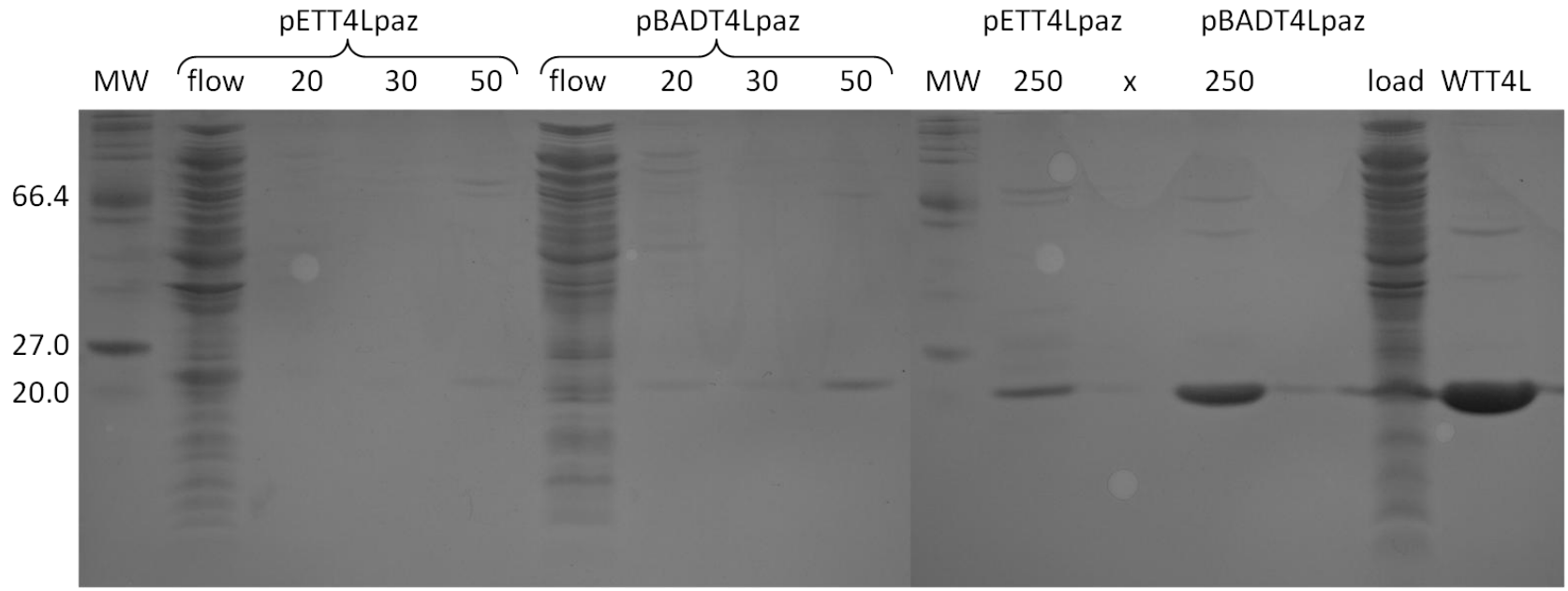


Figure 36. SDS-PAGE gel comparing the expression levels of T4L153azF using pETT4L153TAG (+pAAazRS) vs. pBADT4L153TAG (+pDULEazRS). 200 ml ZYM-5052 media was used, the culture was grown at 37 °C, azF was added at OD_{600} = 0.9, expression was induced by addition of 0.2 % arabinose or lactose for the pBAD and pET cultures respectively at OD_{600} = ~5, expression was allowed to proceed for 4 hours. The protein was extracted and protein subjected to Ni-NTA affinity purification, 20 30 and 50 mM imidazole concentrations were used in wash-steps, elution (250 mM imidazole) is representative of 15 µl sample from 2 ml.

3.3.4 Expression and purification of WT T4L

Plasmid pBADT4LWT, expressing WT T4L was used to produce a robust and high yielding protein expression method. The conditions that gave a high yield of protein was found to use the method developed by Miyake-Stoner (2009), in ZYM5052 media with manual induction by addition of arabinose (Studier 2005), as described in materials and methods. Protein was purified by nickel affinity chromatography and size exclusion chromatography to purity according to SDS-PAGE (Figure 37). Mass determined by ESI-MS (19511.8 ± 8.57), demonstrated that WT T4L had been isolated (Table 10; appendix A4). This purified protein was used in crystallisation optimisation, and spectroscopic analysis.

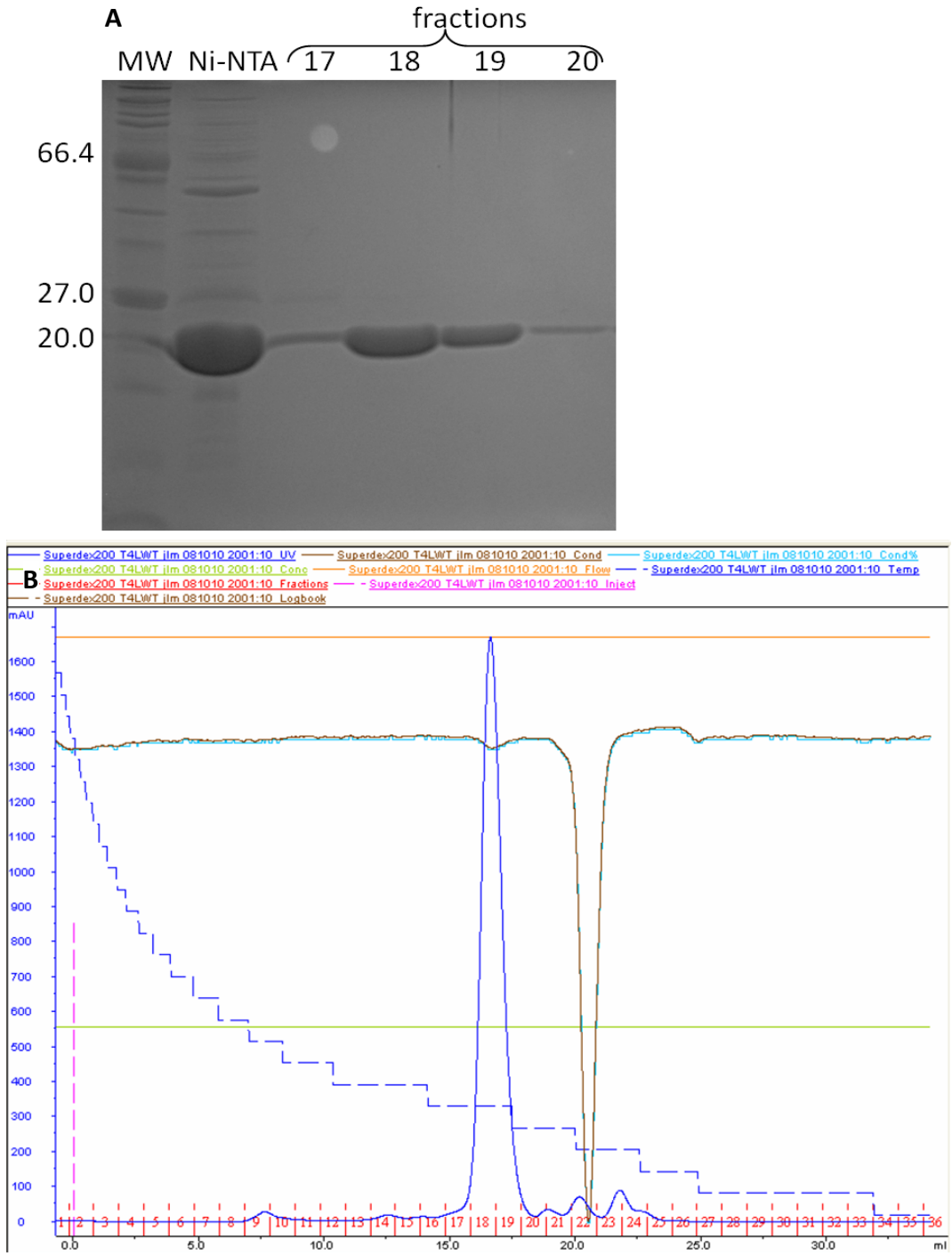


Figure 37. Purification of T4L WT by size exclusion chromatography. A) Representative SDS-PAGE gel of nickel affinity purified WT T4L (lane 1), sample of size exclusion 1ml fractions (Superdex75) 17, 18, 19, 20.

B) Corresponding size exclusion chromatography trace. ZYM-5052 media was inoculated with a 1 : 250 volume of saturated culture containing *E. coli* Top10 transformed with the plasmid pBADT4LWT. The culture was grown at 37 °C with expression induced by addition of 0.2 % arabinose at OD_{600} = 5 - 7, expression was allowed to proceed for 4 hours. The protein was extracted and subjected to Ni-NTA affinity purification. The elution fraction was further purified by size exclusion chromatography. Peak fractions (UV, blue line) were loaded onto 12 % SDS-PAGE gel.

3.3.5 Fidelity of aaRS

To establish the fidelity of the synthetase (expressed by the pDULEazRS plasmid) and tRNA towards azF, parallel expressions plus and minus amino acid were performed. ZYM5052 media containing Top10 *E. coli* cells cotransformed with pBADT4L153TAG and pDULEazRS were grown as indicated in materials and methods. One culture contained 1 mM azF, while the NAA supplement was withheld from the other. Protein expression and Ni-NTA purification was performed as described in materials and methods. Analysis by SDS-PAGE gel indicated the fidelity of pDULEazRS for incorporating azF into T4L at position 153 was > 99 % (Figure 38).

3.3.6 Expression and purification of T4L153azF

Using the expression conditions optimised using expression of the wild type protein (pBADT4LWT), azF substituted T4L was expressed, purified by Ni-NTA and SEC (Figure 39). A typical yield of T4L153azF was 7.5 mg purified protein / L culture. An accurate mass spec could not be obtained (Table 10), most probably due to the presence and reactivity of the azide group.

For spectroscopic analysis, protein expression, purification and handling was performed in the dark or under red light. This was done to try to limit the amount of UV-light that could allow the photoreaction of the azido group.

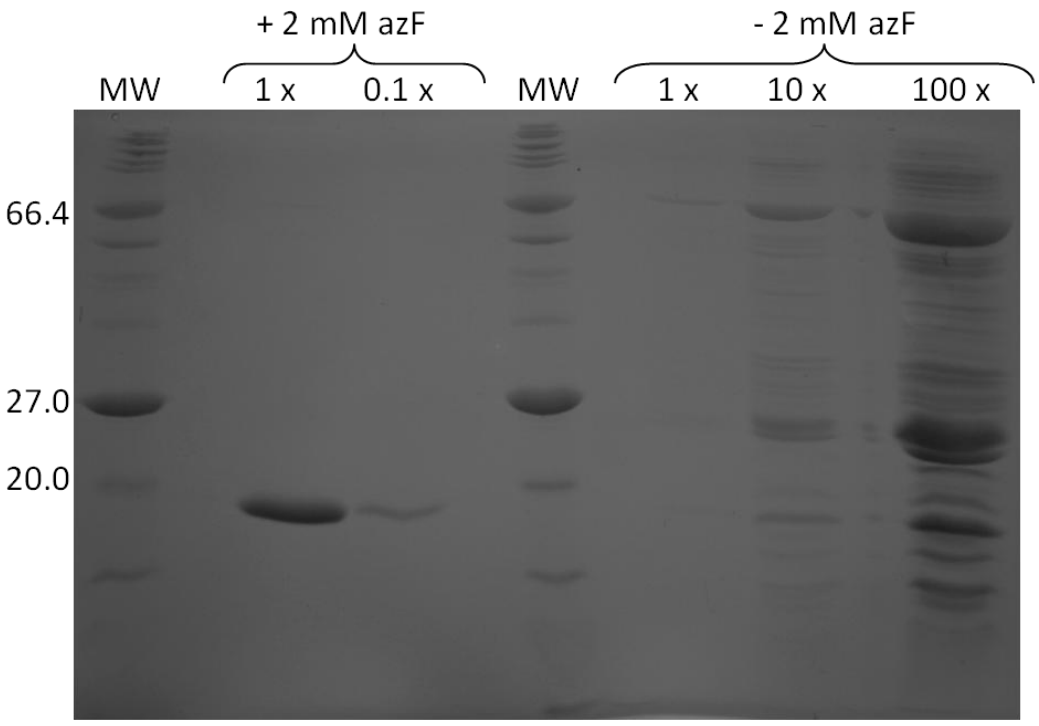


Figure 38. SDS-PAGE gel exploring the fidelity of the expressed synthetase for the NAA, azF. Parallel expressions plus and minus amino acid were performed. ZYM5052 media containing Top10 *E. coli* cells cotransformed with pBADT4L153TAG and pDULEcyanoRS were grown as indicated in materials and methods. One culture was supplemented with 1 mM azF, while the other was grown in the absence of NAA supplement. Expression in both cultures was induced by addition of 0.2 % arabinose. Protein was extracted and subjected to nickel affinity purification as described in materials and methods. Protein was concentration to 0.1 x, 10 x or 100 x.

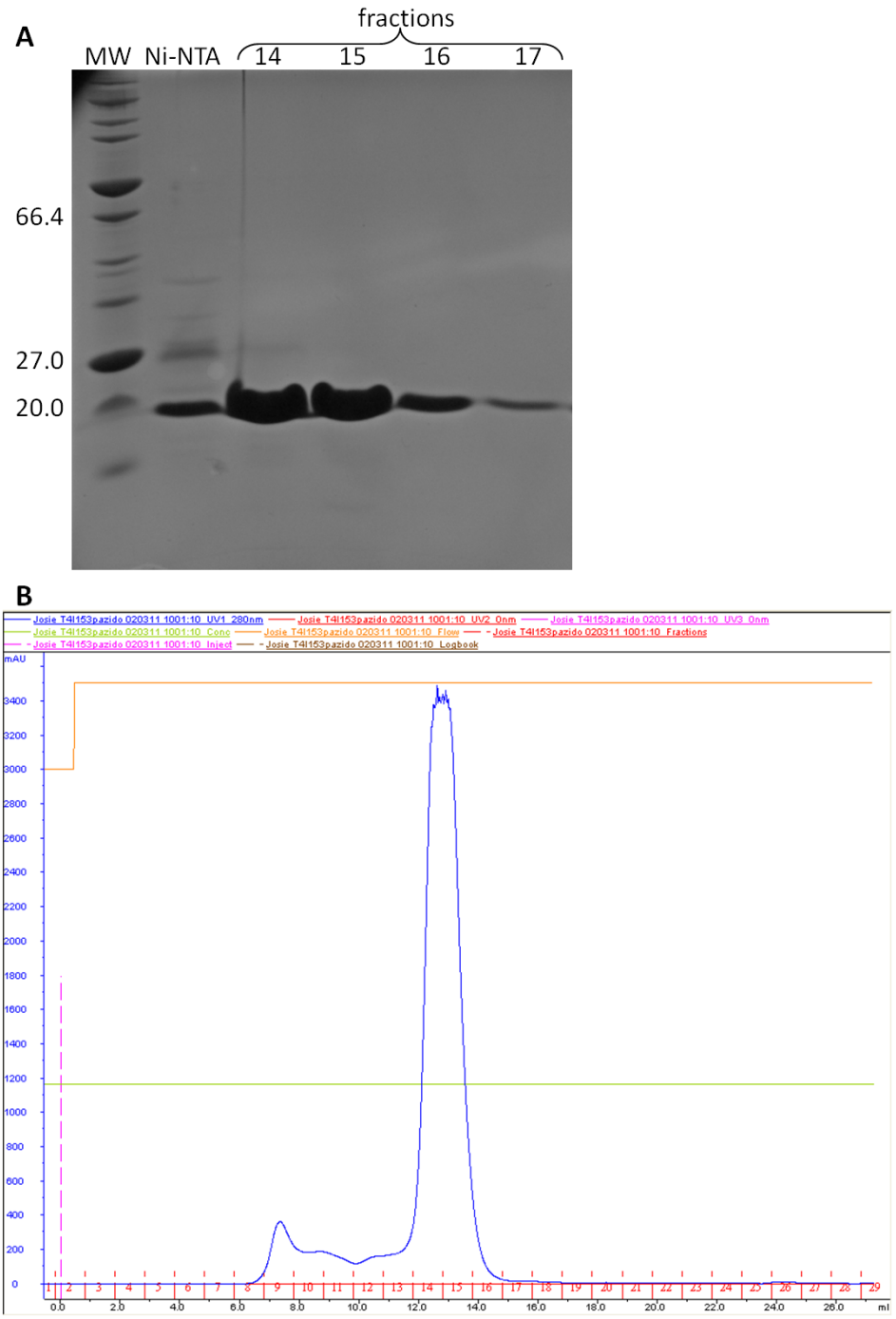


Figure 39. **Purification of T4L153azF by size exclusion chromatography. A)** Representative SDS-PAGE gel of nickel affinity purified T4L153azF (lane 1), sample of size exclusion 1ml fractions (Superdex75) 14, 15, 16, 17.

B) Corresponding size exclusion chromatography trace. ZYM-5052 media was inoculated with a 1 : 250 volume of saturated culture containing *E. coli* Top10 co-transformed with the plasmids pBADT4L153TAG and pDULEazRS. The culture was grown at 37 °C with expression induced by addition of 0.2 % arabinose at OD_{600} = 5 - 7, expression was allowed to proceed for 4 hours. The protein was extracted and subjected to Ni-NTA affinity purification. The elution fraction was further purified by SEC. Peak fractions (UV, blue line) were loaded onto 12 % SDS-PAGE gel.

3.3.7 Investigating the promiscuity of azRS

As mentioned previously, it has been shown that some of the synthetases are promiscuous towards the non-natural amino acids that they incorporate in response to TAG-containing target genes. The ability of pDULEazRS to incorporate iodoF into T4L153TAG was investigated. It was thought that if, using the same expression conditions, an accurate mass for iodoF-substituted T4L could be obtained the result would provide some evidence that the incorporation of NAAs into T4L using this plasmid and expression system was successful. The expression and purification of T4L153iodoF was performed using pBADT4L153TAG and pDULEazF supplementing the media with iodoF, as described in materials and methods. The mass obtained was 19638.95 ± 4.88 Da, demonstrating that incorporation of iodoF had occurred (Table 10; Appendix A4).

Table 10. Expected and observed mass of NAA-substituted T4L

Protein	Expected mass	Observed mass
T4L153iodoF	19638.1	19638.95 ± 4.88
T4L153azF	19553.2	19510 ± 3 19535 ± 6
WTT4L	19512.2	19511.8 ± 8.57

3.3.8 Crystallisation of WT and mutant T4L

Crystal screens were set up based on conditions used by Xi et al (2004) or other previous work crystallising WT T4L (Matthews et al. 1973; Remington et al. 1978; Weaver and Matthews 1987). Specifically, a vapour diffusion method, using sodium/potassium phosphate buffer containing hydroxyethyl disulphide. Hydroxyethyl disulphide has been shown to be crucial in T4L crystallisation as it required at interface of monomers. Purified WT, T4L153azF and T4L153iodoF were all used in crystal screens.

Crystals of WT protein were obtained in protein buffer: 100 mM NaH_2PO_4, 0.55 M NaCl, pH 6.7, reservoir buffer: 2 M NaH_2PO_4, 2 M K_2HPO_4, 15 mM hydroxyethyl disulphide, pH 6.7. Crystals of two shapes were obtained. These crystals diffracted to approximately 2.3 Å, but because the structure of WT T4L is so well studied, the structure was not solved (Appendix A10).

Very small crystals of T4L153azF were obtained in buffer 5:5 protein : reservoir, 1.5 M K/$NaPO_4$, 0.25 M NaCl pH 6.4, 6 µl / ml hydroxyethyl disulphide. However, these crystals did not diffract well as they were considered disordered.

3.3.9 Circular Dichroism of T4L

To ensure that the structure of T4 lysozyme was not perturbed by the incorporation of the NAA, azF and because of the difference in crystallisation efficiency, CD measurements were performed to check that the T4L153azF protein was correctly folded and comparable to that of wild type protein. The CD spectra were similar and indicated that both proteins contain α-helix, due to the presence of a peak at 208 and 222 nm, and β-sheet, due to the presence of a peak at 216 nm (Figure 40).

This result demonstrated that T4L153azF was folded similarly to WT.

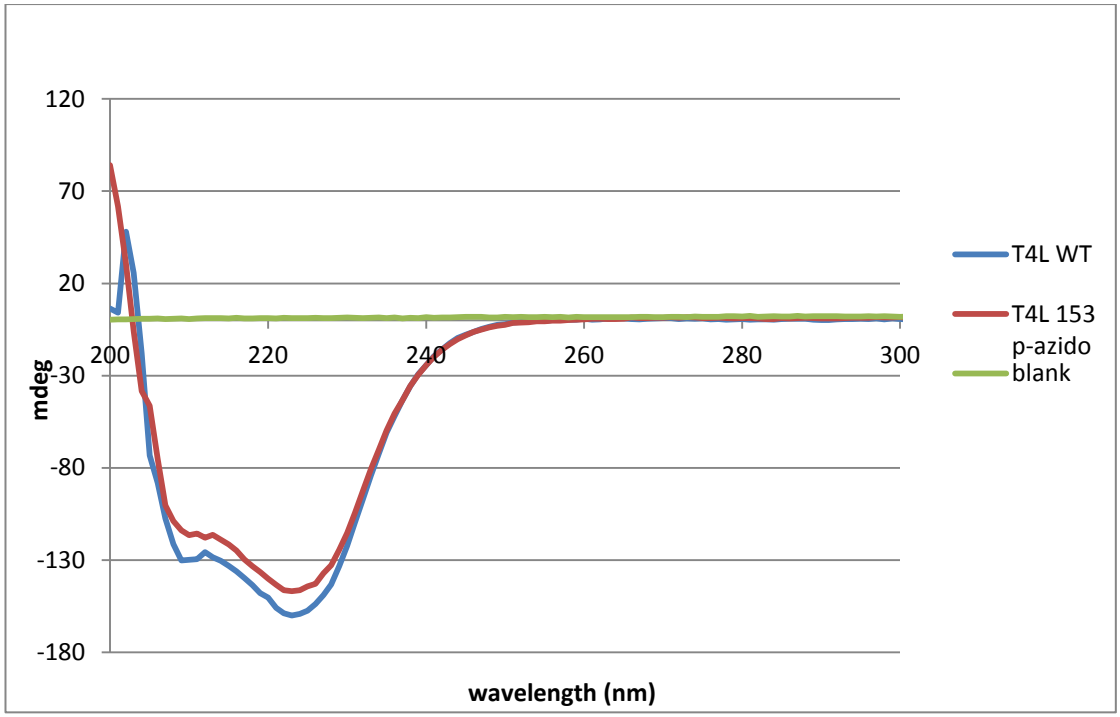


Figure 40. CD spectra of T4L and T4L153azF. Purified WT T4L and T4L153azF were concentrated to 50 µM and dialysed into degassed 100 mM NaH_2PO_4, pH 6.7. The samples were subjected to circular dichroism within a 0.2 um quartz cuvette. Buffer was used as blank.

3.3.10 Spectroscopic evaluation of azido substituted T4L

The aims of this chapter were to explore whether using a protein scaffold is an effective method to shield and therefore simplify the study of the photoreaction of the azido group. After a robust method to express and purify T4L153azF was determined, it was used to produce protein for spectroscopic investigations.

3.3.10.1 EPR of T4L153azF

As mentioned above, the reaction pathway taken by irradiation of phenyl azide is dependent on the temperature at which the reaction occurs (Figure 17). At 77 K, the major pathway taken is thought to lead to the formation of triplet nitrene. Therefore, the formation of triplet nitrene was probed for using continuous wave (CW) EPR, which was performed by Dr Damien Murphy (Cardiff University).

A sample of T4L153azF (8 mg/ml in 50 % glycerol to maintain glassy at low temperature) held within a quartz EPR tube, was frozen in liquid nitrogen and irradiated for 1 minute using a 400 W xenon arc lamp. The EPR measurements were taken at 10 K, with the system cooled by liquid helium. Twenty-five measurements were taken to increase signal to noise ratio, scanning 1000 – 8000 G.

The EPR spectra of T4L153azF showed a peak at 6869, which was representative of triplet nitrene formation (Figure 41). The inequality in the peak height and trough at 6869 were also classical of triplet nitrene. A spectrum was simulated by Dr James A. Platts (Cardiff University; Figure 41) and compared to the experimental spectrum. The spectrum was simulated using a g value of 2.0023 and D' = 1.0007 cm^{-1}. The experimental signal at 6869 G aligns to that of the simulated spectrum (Figure 41). However, the complexity of the experimental spectrum is most probably due to the presence of the protein.

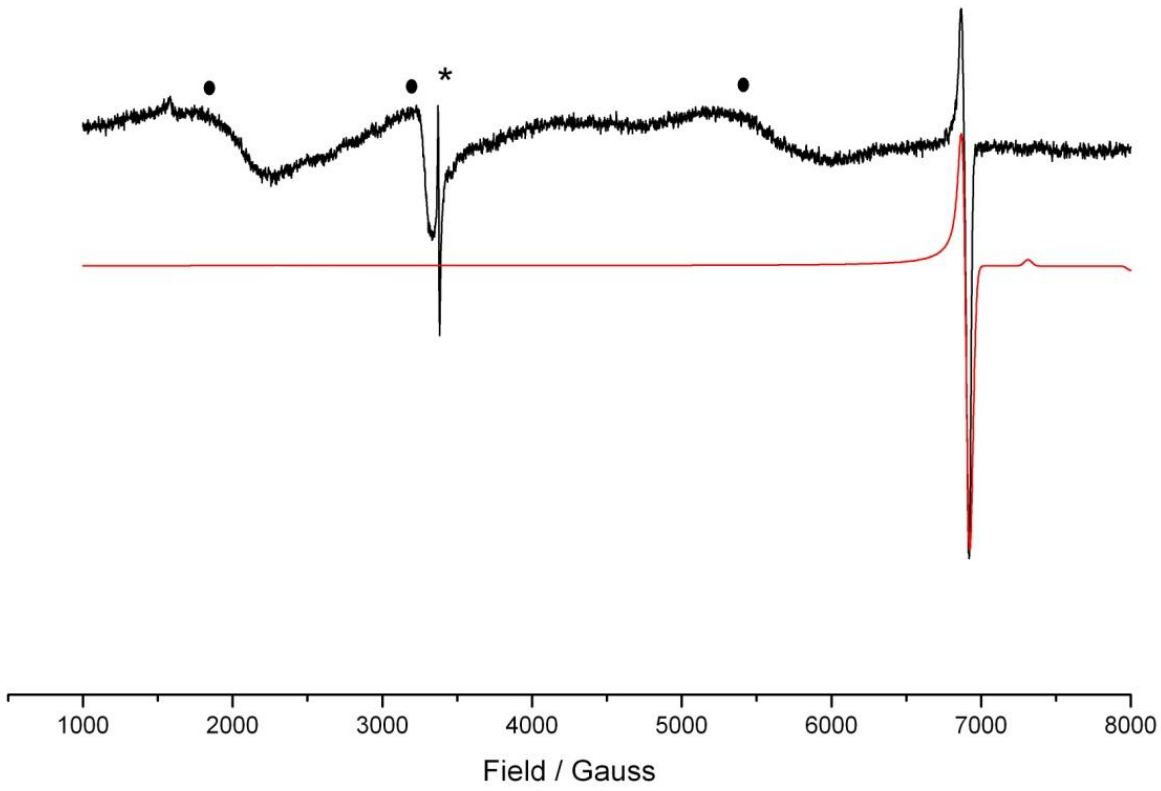


Figure 41. Experimental (black) and simulated X-band (red) CW EPR spectra. * = cavity artefact in the free spin region, • = background signal from sample. EPR spectra were obtained using azF substituted T4L (8 mg/ml in 50% glycerol to maintain glassy at low temperature). The sample was frozen in liquid nitrogen and irradiated for 1 minute using a 400 W xenon arc lamp. The EPR measurements were taken at 4 Kelvin, with the system cooled by liquid helium. Twenty-five measurements were taken to increase signal to noise ratio, scanning 1000 - 8000 G.

3.3.10.2 Ultraviolet-visible spectroscopy

To investigate the reactions of the azide group that could not be monitored using EPR some further spectroscopy was performed.

3.3.10.2.1 UV-vis of amino acid, azF

UV-vis spectral measurements were obtained using a solution of azF amino acid (625 µM; in NaOH). The sample was irradiated and spectra obtained at various timepoints, as described as in materials and methods. It was noted that there were a number of changes during photolysis (Figure 42). The initial absorption spectrum of azF showed a large absorption at approximately 250 nm, with a shoulder at approximately 280 nm. Upon irradiation the absorption at 250 nm decreased along with general increase at wavelengths 290 – 600 nm. The absorbance at 280 nm remained similar.

3.3.10.2.2 UV-vis of T4L153azF

UV-vis spectra of T4L153azF and WT T4L protein were also obtained (Figure 43). The UV spectra demonstrated similar absorbance, except T4L153azF showed higher absorbance at wavelengths approximately 250 – 270 nm.

The protein solutions were irradiated and spectra measured as described in materials and methods and UV-vis spectrum obtained throughout irradiation at specified timepoints. The spectra showed there was a change in absorption as the T4L153azF was irradiated (Figure 45) in a similar manner to the experiments with amino acid. There was a general increase in absorption over the entire range of measurements (260 – 460 nm), but the main increase in signal was between 300 and 340 nm. Whereas there was no change in the spectrum measured when WT T4L was irradiated, as the spectrum overlay similarly (Figure 44). The difference between spectra measured using WT and T4L153azF indicated that some changes occurred throughout irradiation and this was due to the presence of azide moiety. This data suggests that the azide group, present in the T4L153azF and not in WT, may have undergone some photoreaction during photolysis

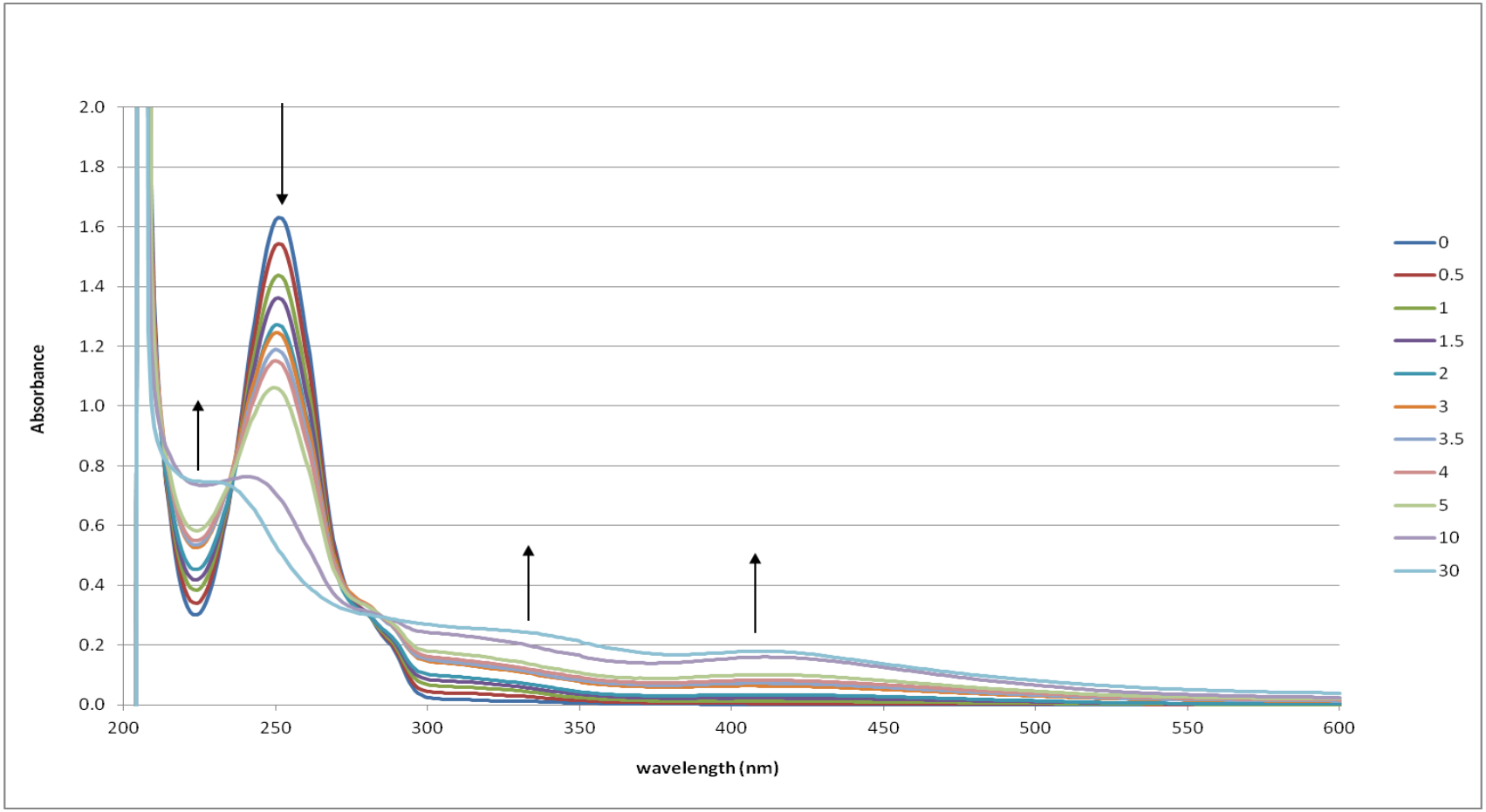


Figure 42. UV-vis spectra of azF. A solution of azF (625 µM; in NaOH) was irradiated using a UV lamp with wavelength 302nm for a total of 30 minutes and recording spectra every 30 seconds for the first 5 minutes then 10 minutes and 30 minutes.

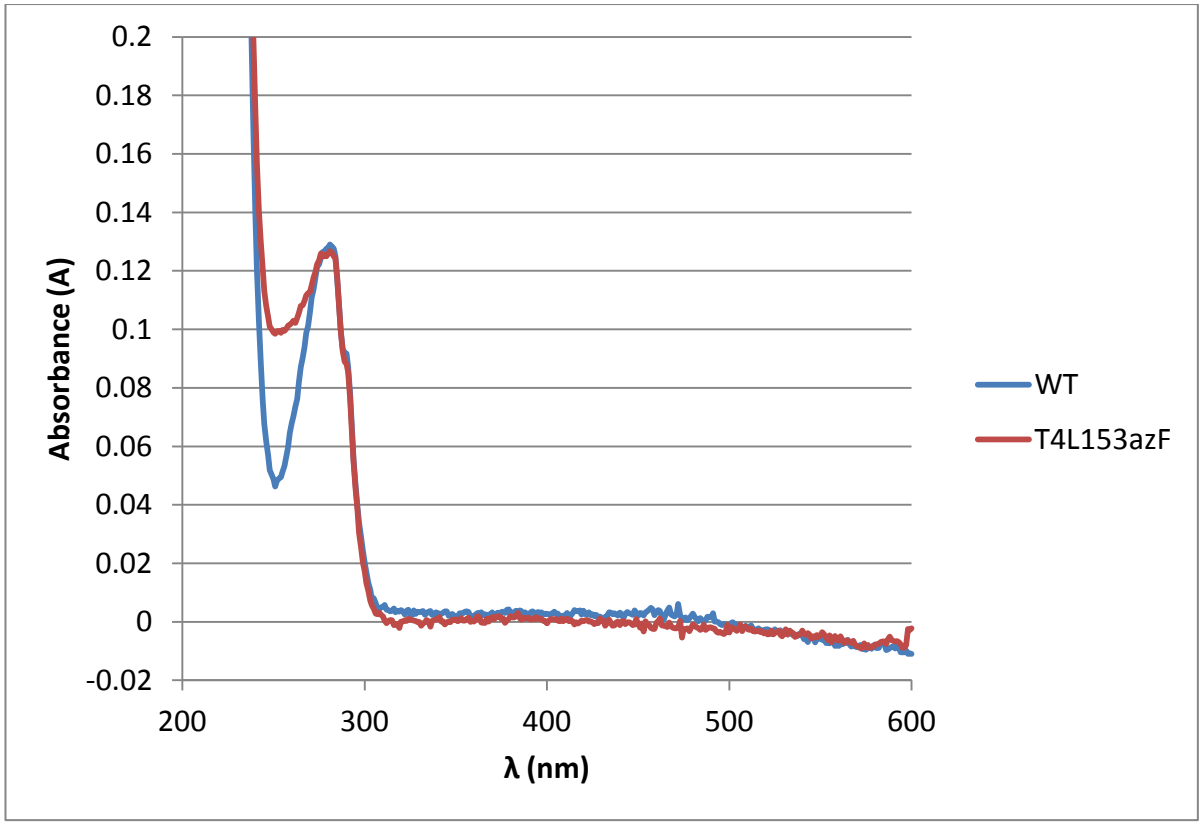


Figure 43. UV-vis spectrum WT T4L and T4L153azF. Protein at 50 µM in 100 mM NaH_2PO_4, pH 6.7 (degassed) was used to obtain spectra using a 0.2 µm quartz cuvette.

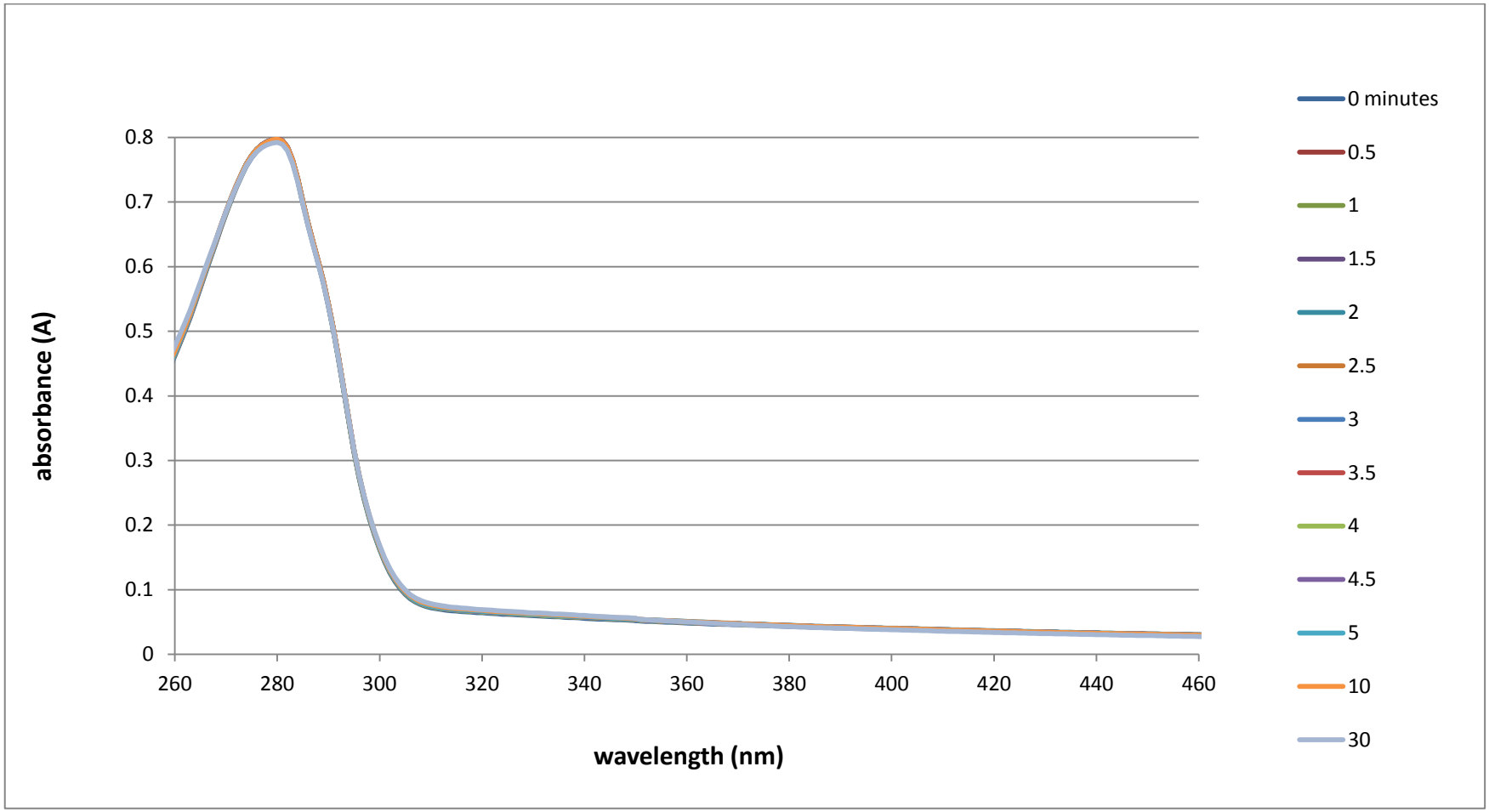


Figure 44. UV-vis spectra of irradiated WT T4L protein. Spectra of T4L153pazF (100 µM) were obtained, the solution was then irradiated (302nm) for a total of 30 minutes and spectra recorded every 30 seconds for the first 5 minutes then 10 minutes and 30 minutes.

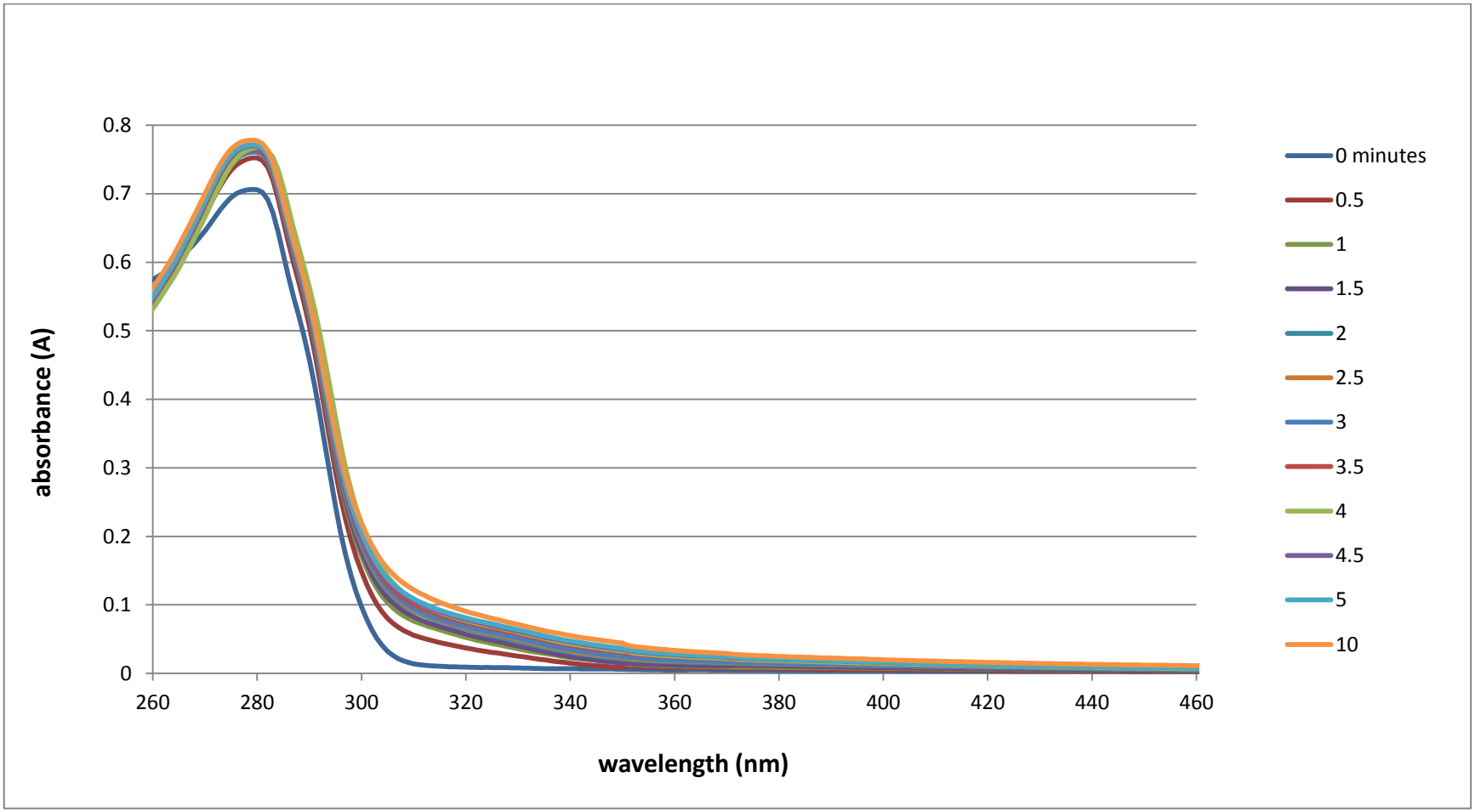


Figure 45. UV-vis spectra of irradiated T4L153azF protein. Spectra of T4L153pazF (100 µM) were obtained, the solution was then irradiated (302nm) for a total of 30 minutes and spectra recorded every 30 seconds for the first 5 minutes then 10 minutes and 30 minutes.

3.3.10.3 Infrared spectroscopy

Infrared spectroscopy is another powerful technique used to study reactions of azides. To test whether this technique could be applied to the study of the azide moiety present within the protein backbone, T4L, solutions of amino acid were used for IR measurements. A weak signal at approximately 2100 cm^{-1}, corresponding to azide, was obtained using 20 mM solutions of azF, which was not present in the spectrum for phenylalanine (Figure 46). The difference spectrum shows a difference of about 2 % transmittance at 2100 cm^{-1}. Although a peak corresponding to azide was obtained using a solution of azF, the concentration was too high (20 mM) to be obtained using protein as T4L forms precipitate at high concentrations.

3.3.10.4 Raman Spectroscopy

It was investigated whether Raman spectroscopy could be used to visualise the azide group. Spectra was measured and data analysed by Dr Elisabetta Canetta (School of Bioscience, Cardiff University). Spectra using NAA (in ddH_20 + NaOH) was obtained initially to ascertain the threshold of detection. It was encouraging that the symmetric stretching of the azide group, at 1289 cm^{-1}, could be visualised (Figure 47). However, this signal reduced as the concentration decreased and was no longer detectable at 5 mM. To be able to detect this peak in a protein sample, the concentration required was once again unobtainable.

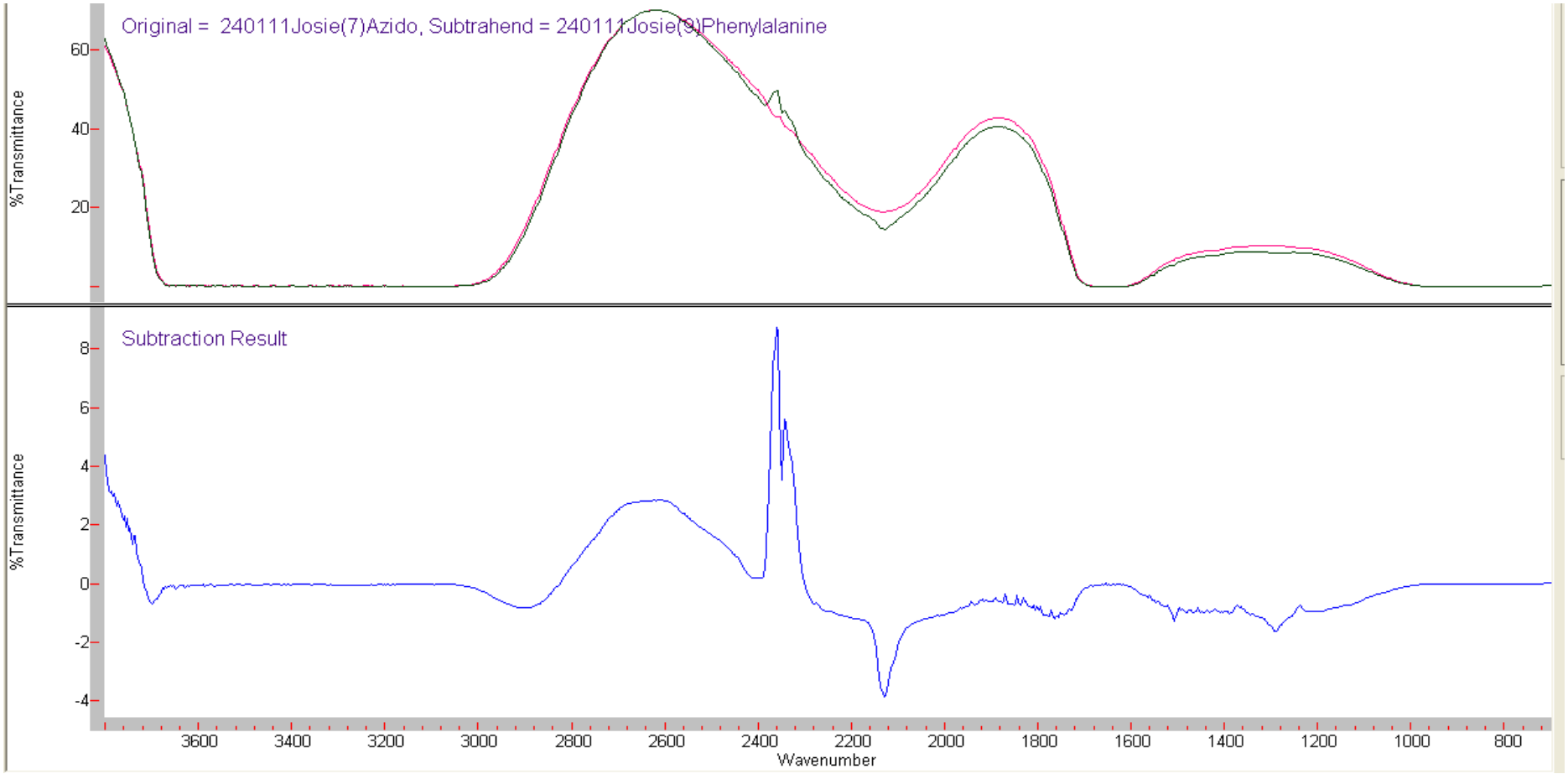


Figure 46. IR spectra of amino acids. Solutions of azF (Pink) and Phe (Blue; 20mM dissolved in H_2O + NaOH) were used with CaF_2 plates with a pathlength 25 µm to measure IR spectra. The difference spectrum was obtained to try to identify the peak corresponding to the azido group.

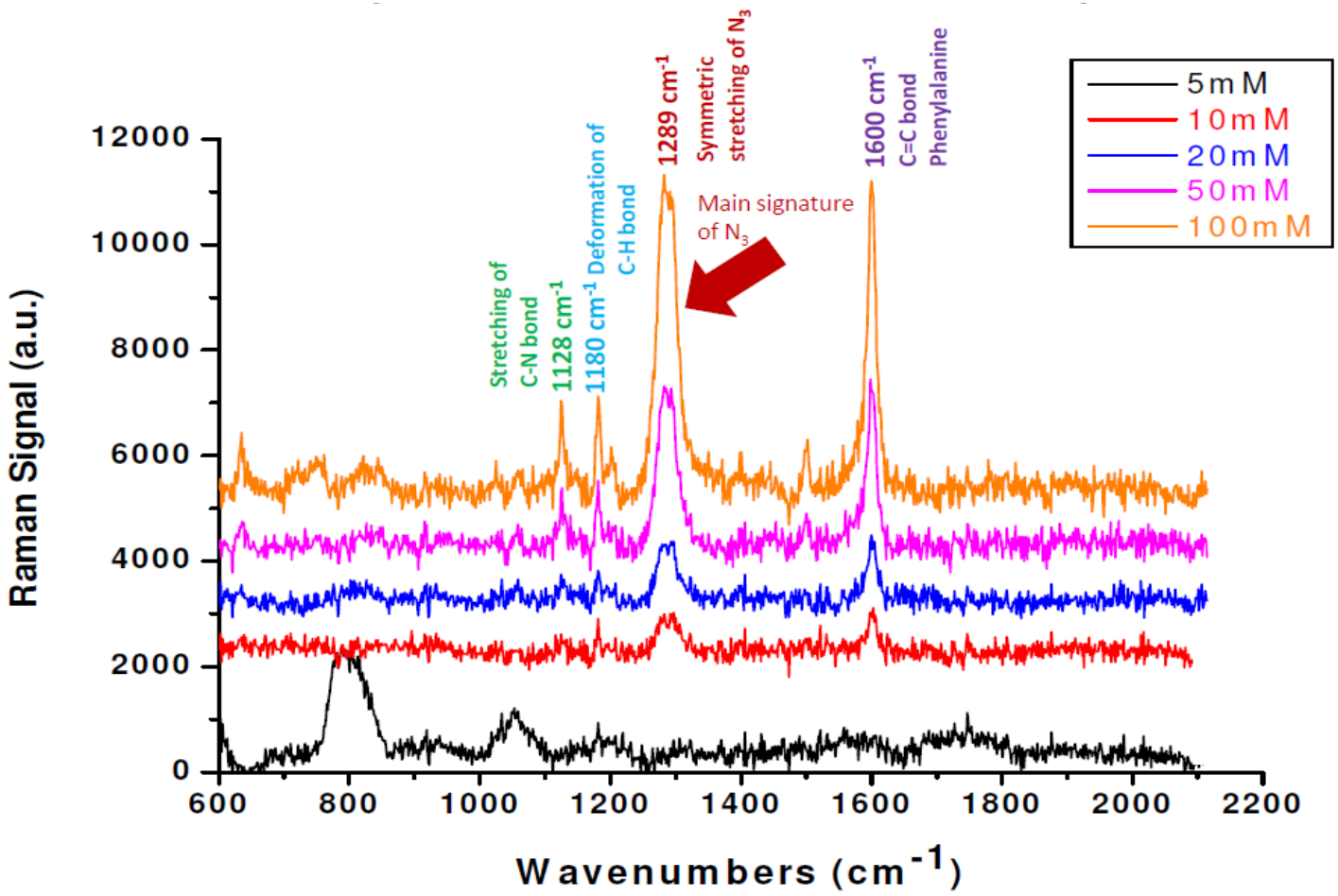


Figure 47. Raman spectroscopy of aqueous azF. Raman spectra were obtained for samples of azF (in ddH_20 + NaOH) at the specified concentrations. The samples were sandwiched between quartz slides, using a 1 mm spacer.

3.4 Discussion

3.4.1 T4L expression

3.4.1.1 Plasmid systems

The expression of WT and NAA-containing T4L was performed using numerous plasmid constructs. Each of these expressed protein with varying efficiency. Slight or major differences in plasmid construction could explain these differences. Namely, the differences in promoters used to drive expression of the target gene; mutations within the T4L gene; the presence/absence of a purification tag (for example, a hexahistidine tag); and differences in linker region between T4L and purification tag (Table 11).

Table 11. Differences between T4L-expressing plasmids

Plasmid	T4L promoter	Mutations	Linker	Purification tag
pT4L153TAG	T5		-	-
pAAT4L153TAG	T5	R154K ΔThr152 Optimised sequence	GGSG	histag
pET101T4L153TAG	T7	L99F M102A	GGSG	histag
pBADT4L153TAG	araBAD	N2D	S	histag

Before expression of T4L substituting azF at position 153, optimisation of expression and purification techniques were optimised using the cheaper and more photostable amino acid p-iodophenylalanine. Also, incorporation of iodoF was undertaken first as it had been shown previously that T4L153iodoF could be expressed, purified and crystallised (Xie et al. 2004).

The use of a T5 or a T7 promoter appeared to make little difference to the efficiency of protein expression. These are both exogenous promoters and this could explain the similarities in protein expression levels. The use of the T7 promoter limited the use of the *E*. *coli*(DE3) strains as these strains express the T7 polymerase. It was thought that the leaky expression of T4L before induction could have been detrimental to cell viability and protein production. Any expression of T4L could have caused lysis of the *E*. *coli* host cell and loss of the expressed protein to the media. Also, this would have limited the overall culture growth and subsequent protein production. It was thought that using the autoinduction medias, leaky expression from lac operon of pET plasmid would be limited due to the presence of glucose in the media (Studier 2005). This would reduce any problems caused by leaky expression of T4L and the subsequent lysis of the host cell. However, T4L production was still low.

Leaky expression from the T5 promoter controlling T4L expression on the pAAT4L153TAG plasmid could have occurred due to small amounts of lactose in the growth media. This leaky expression of Δ152 mutant protein could have been toxic to the host cell, causing cell death. This is especially important if any natural amino acid was incorporated at the TAG codon, (which is always a possibility, especially without any competition from the TAG-aaRS specific NAA) (Nehring *et al.* 2012), as this would have produced a full-length Δ152 mutant T4L which could have had a number of outcomes: retained T4L functionality and caused lysis of the host cell; conferred altered T4L functionality and caused lysis of the host cell; produced a non-functional toxic protein capable of aggregation, altered tertiary structure or other function conferring toxicity. Other explanations could include the reduced tolerance of the codon optimised sequence, the formation of an inaccessible plasmid due to the change in sequence, or the non-functional resistance gene. Even if growth occurred, the product of Δ152 mutant T4L would have placed the azido group of the azF on the surface of the protein, not in the hydrophobic region as anticipated. There was also R154K mutation, and a GGSFG linker region between the T4L gene and hexahistidine tag, which may have affected the expression of the protein. The issues relating to the attempted insertion of the missing codon into

the non-functioning of the plasmid, pAAT4L153TAG, strengthens the argument that the plasmid had formed an in-accessible form.

It had been shown previously that the pT4L153TAG and pBKpIodo plasmid system could successfully incorporate iodoF into codon 153 of T4L (Xie et al. 2004). Expression of T4L153iodoF protein was possible using these plasmids, however the presence of the histag on the other constructs made these more favourable. In addition, the expression of T4LiodoF was low.

The pBAD system appeared to express NAA-containing protein with the best efficiency. The notable changes in these T4L genes compared to the others included the araBAD promoter, the N2D, F99L mutations and the short S linker to the hexahistidine-tag. The araBAD promoter was likely to have had an impact on expression, because it is a tightly regulated promoter, via the doubly regulated *ara*C (Guzman *et al.* 1995). Put simply, when arabinose is absent but glucose is present, there is very little transcription from the promoter due to the doubly regulated regulon (Lee *et al.* 1987). Presence of glucose represses expression, while addition of arabinose causes transcriptional activation and expression is induced quickly (Miyada *et al.* 1984). This tight regulation would have limited the leaky expression of T4L prior to induction, may have increased culture growth resulting in higher cell density that can then express protein when the arabinose is added and limited the issues mentioned above.

The N2D mutation may have increased expression due to the change in N-terminal methionine cleavage. A process known to occur when expressing proteins using *E. coli* (Frottin *et al.* 2006; Hirel *et al.* 1989). The cleavage efficiency of the N-methionine is determined by the length of the side chain of the second amino acid. Although the mutation of an asparagine residue to aspartic acid within T4L does not increase the length of the side chain, it does alter the chemical properties of this residue, which might, in turn, alter the N-terminal methionine cleavage efficiency. The presence/absence of the N-terminal Met might alter the stability of the protein. This, and the other mutations, could have acted to increase the stability, and therefore allow accumulation of the protein

Expression from pBAD plasmids requires the use of the Top10 *E*. *coli* strain as opposed to the traditional protein expression strain BL21(DE3) because Top10 carries a *ara*D139 mutation which alters arabinose metabolism and allows control over protein over-expression (Gendron and Sheppard 1974). The use of Top10 may have also increased amber suppression as it has been demonstrated previously that incorporation of natural amino acid in response to TAG codon is lower than that observed in BL21 strains (O'Donoghue *et al.* 2012).

3.4.1.2 Plasmid encoding orthogonal aaRS and tRNA$_{CUA}^{NAA}$

The plasmid expressing the synthetase was also varied due to a number of reasons. Firstly, whether the tRNA$_{CUA}^{NAA}$, was present on the same plasmid as the synthetase or the target gene. Both aaRS and tRNA$_{CUA}^{NAA}$, were required for amber suppression. Also to note was that the number of copies of tRNA$_{CUA}^{NAA}$, varied among the plasmids. It appeared that the number of tRNAs encoded on a plasmid had little impact on efficiency of production of NAA-containing protein, because the pDULE plasmid encodes only 1 tRNA$_{CUA}^{NAA}$, and this was the plasmid used in the most efficient system. The number of copies of tRNA$_{CUA}^{NAA}$, or which plasmid harboured the tRNA could have influenced expression. In fact, the presence of more than 1 tRNA$_{CUA}^{NAA}$, on a plasmid may have been detrimental to protein expression, possibly due to toxicity (Young *et al.* 2010) It was also required that the plasmid pairs had compatible origins of replication, to ensure both plasmids were maintained in daughter plasmids after cell division; and this hindered the use of certain plasmid pairings.

Both the pBK and pSUP RS-containing plasmids used a *glnS*' promoter to express the synthetase. However the pDULE plasmid used and *lpp* promoter and this could explain the increase in expression when this plasmid was used to express azRS to incorporate azF into T4L. The aaRS in pDULE was originally engineered to incorporate cyanophenylalanine (Figure 48), not azF as for the other RSs used in expression of T4L153azF (see section 3.4.1.3). This also could have increased the expression of target protein by more efficient recognition, aminoacylation or incorporation of azF. The pAAazRS plasmid used a T5 promoter. The inducible

expression of azRS via this T5 promoter was not beneficial as the most efficient pDULE plasmid used a constitutive promoter.

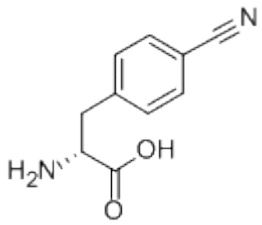


Figure 48. Cyanophenylalanine

3.4.1.3 Promiscuity of NAA-RSs

It was considered that the promiscuity of the orthogonal synthetases could be utilised to incorporate azF into T4L153TAG using the synthetase engineered for iodoF. The premise for considering this was based on previous work performed by group members (Antonczak 2012 Thesis) and other groups (Miyake-Stoner *et al.* 2010; Stokes *et al.* 2009). The work demonstrated that it was possible to use a variety of synthetases to incorporate a particular tyrosine-based non-natural amino acid. The differences in size and charge of the NAA were accommodated by the removal of the tyrosine-specific residue, Y32, within the AA binding pocket of aaRS and replacement with a smaller, non-charged amino acid (Appendix A3). This promiscuity was particularly important because the synthetase used to incorporate the azF in the pDULE plasmid was originally engineered for aminoacylating cyanophenylalanine, another nitrogen containing phenylalanine derivative (Figure 48).

3.4.1.4 Fidelity of amber suppression

Full-length protein was still apparent in cultures grown in the absence of NAA. This could be due to incorporation of a naturally occurring amino acid, typically tyrosine, as this was the original cognate amino acid for the engineered orthogonal RS and tRNA (Nehring et al. 2012). Although the synthetases have been engineered to recognise the specific NAA, and during the engineering process synthetases that

aminoacylate natural amino acids were selected against, it is possible that tyr has been incorporated at position 153 when NAA was withheld from the media. Tyrosine was the cognate amino acid for these synthetases therefore the probable choice for aminoacylation with a natural amino acid. Nehring et al (2012) have shown by in vitro catalysis experiments that these NAA-specific engineered synthetases show preference for tyrosine over the NAAs they were engineered for. The group suggested that accumulation of metabolically inert NAAs in the cell accounted for why the engineered aaRSs were charged with NAA and NAA incorporation occurred, despite the NAAs being weak substrates (Nehring et al. 2012). When NAA was withheld from the media, the engineered aaRSs preferentially charged tyrosine and incorporated tyr at the TAG position. Following this argument, it could be assumed that although Tyr incorporation occurred at a certain level when NAA was absent for the cells, this may not be representative of the Tyr incorporation seen in experiments when NAA was present. Therefore, T4L153Tyr should be at lower levels in the population of protein produced from experiments where NAA was present. The work presented within this chapter demonstrated that fidelity of NAA incorporation was approximately 99 %.

3.4.1.5 Functionality of mutant T4L

The functionality of the T4L protein was assayed by the ability of the protein to lyse the cells in which the protein was overexpressed. The output from this assay was the production of a thick cellular lysate, similar to that typically seen after hen egg lysozyme lysis. The initial freeze-thaw step after harvesting the culture would cause lysis of some of the cells, this would release the cellular contents, including T4L, and further aid in cell lysis. This simplified the protein isolation procedure, and ensured no exogenous lysozyme was used in culture lysis which would require removal via purification. This simple assay indicated that the presence of NAA had not altered the ability of the protein to fold, and T4L153azF/iodoF was still able to confer the typical function of T4L, namely, the lysis of *E. coli* cells.

3.4.2 T4L purification

The purification of expressed protein was crucial for the downstream applications, namely, to ensure the spectroscopic data obtained could be attributed to the T4L153azF protein and not any other contaminating protein. In addition, purity was also important during crystallisation as contaminating protein can prohibit the formation of well ordered crystals of pure protein.

Nickel affinity purification was an ideal first purification step as, in theory, this method should purify full-length, NAA-containing protein away from truncated protein. Truncated protein was produced where NAA incorporation had failed and the TAG codon had been "read" as a translation termination codon. This truncated protein did not contain a hexahistadine tag at the C-terminal of the protein. As the TAG codon was at codon 153 of 164 i.e. near the C-terminal, the truncated protein (17.1 kDa) was only 2.54 kDa smaller than full-length protein (19. kDa) and this method provided an easy way to separate these two populations of protein. Other purification methods might not distinguish between the slightly smaller protein, which has the same primary structure codon 1 to 152 and therefore similar properties. This purification step rendered the protein quite pure. The major contaminating protein was of a much larger mass which could easily be purified away using a SEC purification step.

The expression and purification of T4L153azF was performed, as much as possible, in the dark or under red light. The protein was expressed in a incubator with the glass covered, with the light off. Also, expression was typically performed in the evening or at night to ensure a light-free environment. A specially created dark room, with a red light, was used to purify and handle the protein; and red eppendorfs or aluminium foil coverings were used for intermediate steps. Although care was taken to only handle the protein in the dark, inevitably some protein handling could not be performed in the dark, for example, the centrifugation of culture or protein, therefore it was possible some photoreaction of the azide could have occurred prior to the experiments.

3.4.3 Crystallisation of T4L

It was thought that visualising NAA-containing T4L by x-ray crystallography would provide further evidence that NAA had been incorporated. It was also thought that the reactions that occur during photolysis could be probed using this method. Well diffracting crystals of WT T4L were obtained, but T4L containing NAA was more complex. Only small crystals of T4L153azF were isolated and these did not generate good diffraction patterns. It was considered that the crystals were disordered.

It has been shown previously that iodoF-containing T4L has been crystallised and the structure solved (Xie et al. 2004). In fact the presence of the heavy atom-containing NAA, iodoF was utilised in solving the structure of the protein. Xie et al, demonstrated that when using iodoF considerably less data was required to solve the protein structure when compared to an equivalent experiment utilising cysteine or methionine.

Purity of protein was particularly important in the crystallisation experiments, where any small amount of protein with a different structure or chemical make-up could reduce crystallisation efficiency. It is possible that even the smallest traces of contaminating protein prevented formation of ordered crystals. In addition, alterations in structure could have arisen from photoreaction of the azide and subsequent reaction with any part of the lysozyme protein. This would result in a new bond causing the alteration of protein structure to accommodate this. Any misfolded protein in the sample could interrupt crystal formation and this could explain the disorder seen in the collected protein crystals that were used for diffraction. Also an issue was that, due to the long crystal formation time, any light exposure could have caused re-arrangement in the protein subunits already incorporated into the protein crystal, this once again, could explain the disorder in the crystal. Even if the crystals were not exposed to light, any reacted protein monomers could have propagated structural rearrangements in the remaining subunits of the crystal. Although CD measurements indicated that the majority of the T4L153azF was folded similarly to WT any subtle differences may not have been seen or it is possible that the solution of T4L153azF was heterogeneous, but at a

low level as not to be detected by the CD. Evidence that the azido group was perturbing protein crystallisation was that WT crystals were easily isolated and were able to diffract well. In addition, Xie et al (2004) have previously obtained well-diffracting crystals of 153-iodo substituted T4L. This hypothesis could be investigated by performing CD measurements on the protein throughout the crystallisation process; or after irradiation of the protein.

3.4.4 EPR of T4L153azF

The EPR spectrum obtained using T4L153azF was compared to those obtained previously using other aryl nitrene species, specifically, the azido-functionalised cryptand investigated by Bucher et al (Bucher *et al.* 2005). It was apparent that the formation of triplet nitrene within T4L153azF had been captured by comparing the characteristic signal at 6869 G to the signal at 6680 G recorded for Bucher's cryptand. Further evidence came from comparing the observed spectra to the simulated spectrum, where the signals at 6869 G align well. However, owing to the large zero-field splitting term, a reliable value of G was difficult to extract from the simulated data using just the X-band frequency. Therefore an exact comparison cannot be made. Both of the spectra were also similar to that found by Wasserman in the paper originally describing the EPR signal of phenyl triplet nitrene at low temperatures (Wasserman 1971). The differences in spectra were due to differences in the chemical environment surrounding the azide. The complexity of the spectra was no doubt due to the presence of the protein that harboured the nitrene.

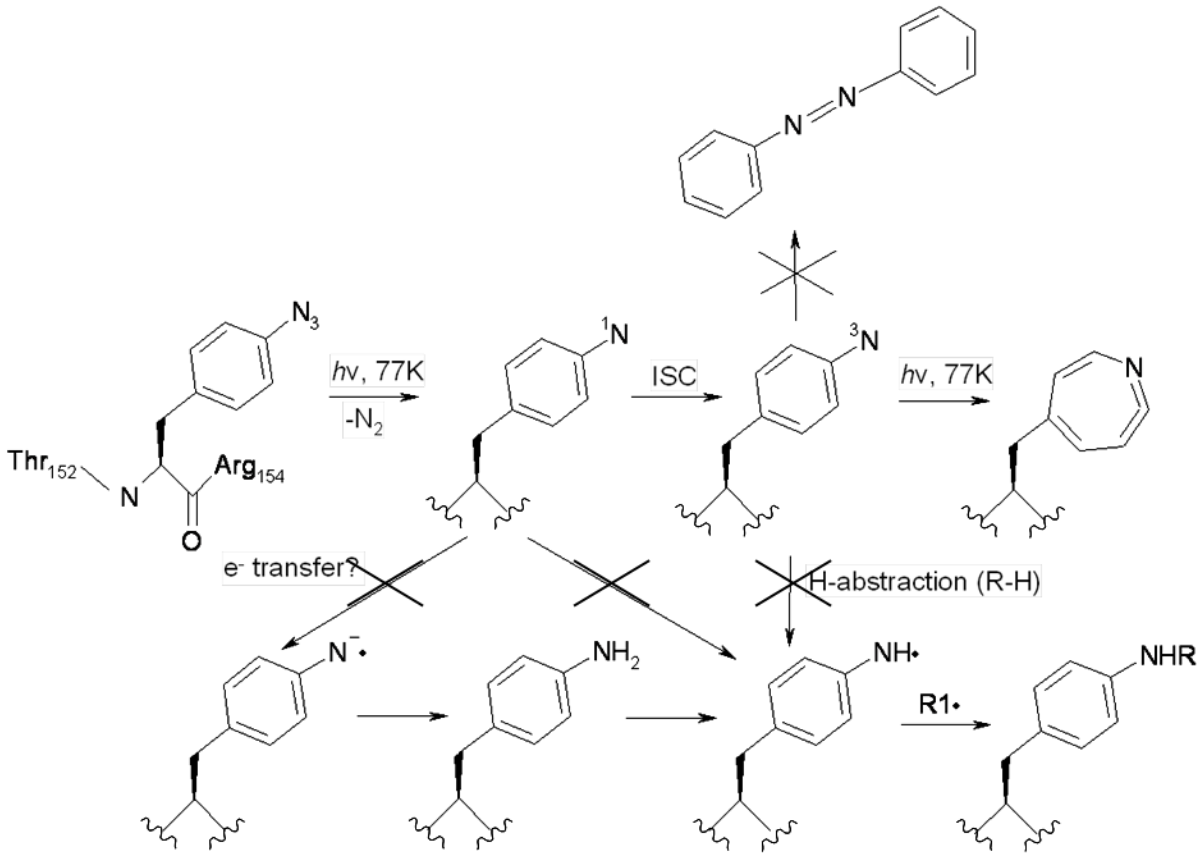
The azide was buried within the hydrophobic environment of T4L to shield the nitrene group from bimolecular reactions (i.e. dimerisation with azo formation) that has typically plagued previous attempts to study the photoreaction of this group (Gritsan and Platz 2006; Schuster and Platz 1992; Li et al. 1988). It was thought that trapping the highly reactive species in this pocket would aid the study of the intermediates. It was also thought that the hydrophobic pocket would result in stabilisation of the reaction intermediates. It was noted that the EPR measurements took approximately 30 minutes, therefore it was apparent that the

triplet nitrene was stable for at least 30 minutes. Further EPR studies could be used to study the stability of the triplet species by following the intensity of the peak at 6689 G throughout a length of time or possibly as the temperature was increased. Smolinsky et al, showed that an EPR spectrum was detectable for numerous azides at 77 K and the signal for phenylazide remained at a similar intensity for at least 18 hours after irradiation (Smolinsky et al. 1962).

It was thought that the hydrophobic environment of T4L could be used to stabilise the reactive intermediates produced during the irradiation of azide and thereby facilitate their study. This was explored by comparing the EPR spectrum of T4L153azF with spectra observed using irradiated protein with the azido group presented on the surface, and within the chromophore cavity of GFP (Reddington 2013 Thesis). As expected, the azide group, when present on the surface of GFP, did not show production of triplet nitrene after irradiation; or any other species detectable by EPR (Morris *et al* 2012). If any EPR-detectable species was produced, it would quickly react with components of the solution surrounding the protein. Interestingly, the GFP protein displaying the azido group within the chromophore (Phe66azF) produced an EPR profile with a predominant five-line pattern which was attributed to production of an anilino radical. This anilino radical was supposed to be produced as a result of hydrogen abstraction (Scheme 1) from the imidazolinone ring of His148. This assignment was supported by simulation data, which also revealed not only spin delocalization over the entire chromophore, but more importantly, a slight electron donating influence of the GFP chromophore group. This electron donating influence increased the spin density on the terminal N atom and resulted in hydrogen abstraction and formation of a complex system in conjugation with the imidazolinone ring. It was hypothesised that this intramolecular rearrangement seen was due to the close proximity of the other residues (< 3 Å) and this resulted in hydrogen abstraction that could have been promoted by the initial formation of either the singlet or triplet states. This was possible as both states are open shelled (Borden *et al.* 2000).

EPR was used exclusively to detect the formation of triplet phenyl nitrene which is proposed to form following intersystem crossing after the formation of singlet

nitrene; therefore the formation of other reactive intermediates cannot be excluded. The possible reaction pathways are proposed in Scheme 1, and are considered.



Scheme 1. Proposed reaction scheme of azF within T4L153azF. Drawn using ISIS draw.

Singlet nitrene was immediately produced following photolysis, but at this low temperature, this species lacks the energy to undergo isomerisation and preferentially relaxes to triplet phenyl nitrene by intersystem crossing, however, some singlet nitrene may have persisted. The production of the ketenimine, known to be produced from prolonged photolysis, was limited by photolysing the sample for a short time (2 minutes). This was performed in accordance with Platz and Wirz who observed a maximum relative intensity of a triplet EPR signal from phenyl azide after approximately 3 minutes of photolysis (Leyva *et al.* 1987). However some ketenimine species may have persisted. The production of the azobenezene bimolecule was reduced by the shielding of the reactive nitrenes and by the prevention of diffusion at this low temperature, therefore was unlikely to be present. Calculations by Sankaranarayanan et al (2010) suggested that there is a significant barrier for triplet phenyl nitrene to abstract a hydrogen (Sankaranarayanan *et al.* 2010). This in conjunction with the analysis of the distances between the putative nitrene and the nearest residues, from the T4L protein, provide evidence that hydrogen abstraction is unlikely to occur in this system. By analysing the T4L153iodoF crystallographic structure (PDB 1T6H), the internuclear distances between the putative aryl nitrene, N, and the proximal residues were calculated. It was shown that the nearest six residues were all non-polar and therefore not likely to provide an easily abstracted hydrogen. Also to note was that the approximate distances between the nitrene N and proximal C–H bonds range from ~ 3.3 – 5.5 Å (Figure 49). At this distance hydrogen abstraction was also shown to be unlikely.

Although the study of the photoreaction of azide at low temperatures has provided information regarding the possible reactions of the azide group and explored how a protein environment can be used to study azide reactions, the reactions discussed above are not those that would typically predominate at a biologically relevant temperature. Therefore, no conclusions can be made regarding the reactions of the azide that would occur within a biological setting. Exploration of the reaction at room temperature would help elucidate these reactions.

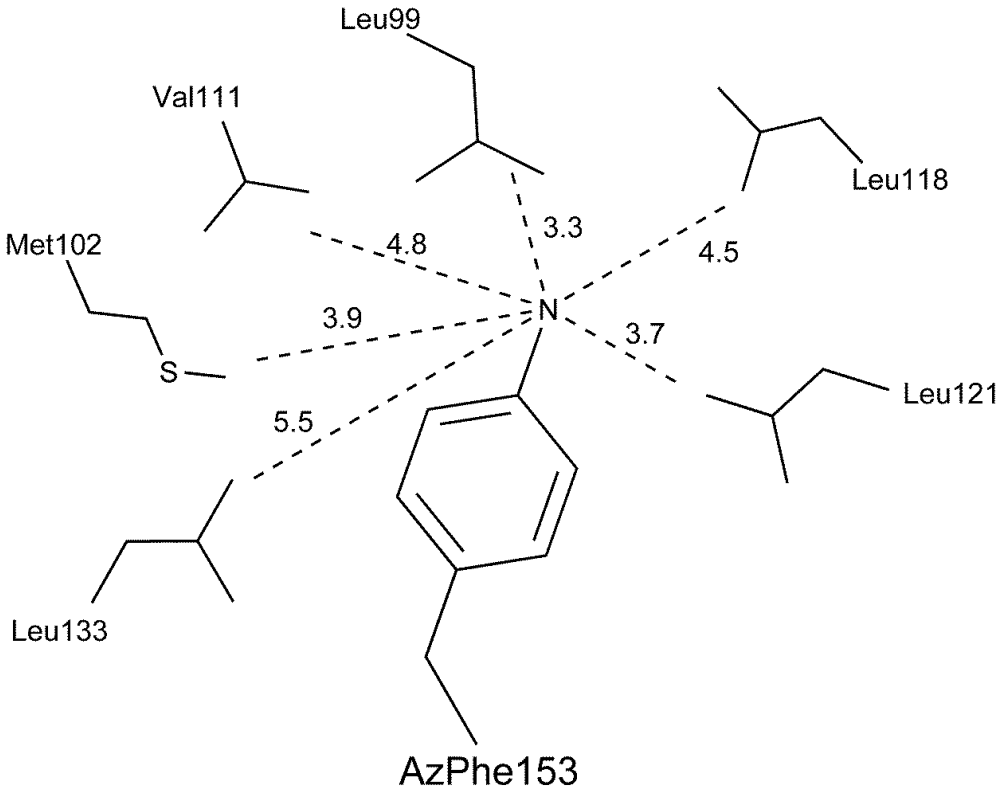


Figure 49. Hydrophobic pocket of T4L with the putative triplet nitrene shown in relation to closest residues. Distances to residues are shown. Drawn using Isis draw.

3.4.5 Further spectroscopic investigations

Other spectroscopic techniques were used to investigate the incorporation of the photoactive moiety into the protein. It was hypothesised that by performing these measurements of unreacted and reacted species at room temperature, this would represent those reactions that occur within the typical biological setting. However, due to the low abundance of azide within the protein, and the low levels of detection of the spectroscopic techniques used, it was possible to see only general differences in spectroscopic properties, or only detection of the azide group when solutions of amino acid were used.

3.4.5.1 CD spectroscopy

The CD spectra obtained during this study was representative of the PDB structure (Chapter 1, Figure 18) and similar to CD spectra previously obtained (Billsten *et al.* 1995; Zhang *et al.* 1993). Therefore the structure of the T4Ls utilised within this study were comparable to those previously studied. Furthermore, the presence of the NAA did not perturb structure and folding.

3.4.5.2 Mass spectrometry

An accurate mass was obtained for WT T4L and T4L153iodoF, using ESI, but not for T4L153azF. Thermoreaction of the azide group was most likely occurring and could explain why an accurate mass could not be obtained. This was concurrent with previous work (Chin et al. 2003a; Liu et al. 2007; Nehring et al. 2012). In contrast, ESI has been used to isolate mass of azF-containing protein (Chin et al. 2002a; Deiters et al. 2003). This difference could be explained by differences in the protein preparation and ESI measurement protocols. The production of T4L153iodoF and mass analysis using the same expression conditions as used in the production of T4L153azF, gave some evidence that NAA incorporation was occurring. Using iodoF for mass analysis had the advantage that its high molecular weight would give T4L153iodoF a mass much different to that of WT T4L.

3.4.5.3 UV vis spectroscopy

When the UV-vis spectrum obtained for azidophenylalanine was compared to that published for phenyl azide, the shape was similar but the slight difference could be attributed to the difference in substituents on the phenyl ring (Tomioka *et al.* 1993). After irradiation of the solution, the increase in absorption at wavelengths 290 – 600 nm, suggested the increase in double bond, the formation of bimolecular product or "tarry polymer". The change in colour from pale to dark yellow solution also indicated this. The reduction in signal at 250 nm and increase in signal at wavelengths 320 – 550 nm after irradiation was also seen by Tomioka et al (1993). Comparing the spectrum of azF to that of phenylalanine suggested that the photoreactivity demonstrated by the azido group is responsible for the change in absorption at wavelengths of 300 nm and above (Linstrom and Mallard 2005).

The UV-vis spectrum obtained using WT T4L shows typical protein UV-vis absorbance. Namely, the absorption in the far-UV range by peptide groups and aromatic side chains; absorption in the 240 – 300 nm region due to aromatic side chains (Table 12). There was very little absorption at wavelengths greater than 310 nm. The UV-vis spectrum obtained using T4L153azF also showed typical absorption of a protein, but when compared to that of WT protein, there appeared to be increased absorption at 240 – 260 nm. This could be attributed to the azF moiety. When irradiated, the absorbance at approximately 250 nm was drastically reduced over the period of irradiation. This could correspond to the photoreaction of the azide.

Table 12. Absorbance of protein moieties. As determined in water at neutral pH (data from Eftink 1991)

Moiety	Wavelength (nm)
Peptide groups	180 – 230
Disulfide bonds	~ 260
Aromatic sidechains	180 – 230 240 – 300
	λ_{max}
Trp Tyr Phe	280 275 258

The UV-vis data showed evidence that incorporation of azF in to the protein was successful and that the moiety displayed some photoreactivity. The changed in spectrum observed during irradiation suggested an increase in double bond content, due to the increase in absorption at wavelengths above 300 nm. This may be due to the formation of the dehydroazepine, which has a broad peak absorption at 320 to 380 nm (Leyva et al. 1986; Schrock and Schuster 1984). Of course, there will be limited formation of azobenzene-type products or tar because the azido was trapped within the hydrophobic pocket and reactive intermediates were shielded from each other. Therefore, this may explain why there was no increase in absorption at the higher wavelengths.

It was possible that the observed UV-spectrum did not represent the reaction of the azido group, but changes in the protein structure due to the reaction of the azido group. The formation of a new bond from the abstraction of a hydrogen could result in major protein structural rearrangements. Although the protein was expressed and purified in the dark as mentioned earlier, this could have caused this re-arrangement in the protein structure.

3.4.5.4 IR spectroscopy

A weak IR signal corresponding to azide was observed when a solution of 20 mM azF amino acid was used. The difference between Phe and azF at this wavelength was only 2 % transmittance. The background signals were also large. In order to visualise the spectrum of T4L153azF, an unobtainable concentration would be required, therefore these measurements were not undertaken.

Typically, a very strong characteristic IR peak at approximately 2100 cm^{-1} has been recorded for organic azides (Buchmueller et al. 2003) or protein-iron-azide complexes (Bogumil *et al.* 1994). The weak signal obtained within this study could be due to the aqueous environment of azF. However, this aqueous environment is required for the measurements of protein as T4L was not very soluble in organic solvents. The IR spectrum of H_2O within this stretch is not great and it was thought that visualisation of the azide would be possible.

Previous work has demonstrated that measurement of the IR spectrum of azide-containing protein was possible (Ye et al. 2009). Ye et al (2009) also incorporated an azide moiety (via azF) using amber suppression into the G protein–coupled receptor, rhodopsin. The group observed a peak at approximately 2100 cm^{-1} which corresponded to the antisymmetric stretch vibration of azide and demonstrated that the azide was sensitive to its electrostatic environment. These changes were used to monitor the change in electrostatic environments of specific interhelical networks during receptor activation. The sensitivity of the spectrometer used and the specific use of a photoactivatable protein may have influenced the ability to measure IR spectrum of the azide group.

Infrared measurements have been extensively used to study the binding of azide to myoglobin and haemoglobin (Alben and Fager 1972; Lim *et al.* 1998; McCoy and Caughey 1970). Once again the signal obtained from the azide group was strong. This could have been due to the use of sodium azide, which has a very strong signal. This combined with the ionic coordination to iron-containing haem groups could explain the difference in the ability to obtain good spectra. The haemoglobin protein used was isolated in large quantities from blood, for example from horse blood, therefore high concentrations could be easily reached.

3.4.5.5 Raman Spectroscopy

The use of Raman spectroscopy to investigate reactivity of the azide moiety within T4L153azF was also attempted. The symmetric stretching of the azide group, at 1289 cm^{-1}, could be visualised at concentrations above 10 mM. The signal was not visible at 5 mM. As seen for IR measurements, the concentration required to visualise the azide group within the T4L153azF protein was unobtainable. Therefore using Raman spectroscopy to investigate the photoreaction of azide within T4L153azF was not investigated further.

3.5 Conclusion

Using the amber suppression method, as developed by Schultz et al, it has been shown that it is possible to produce T4 Lysozyme protein harbouring an azide group within the protein's hydrophobic pocket. The use of this technique provided a method to incorporate an orthogonal, highly reactive moiety into a defined position within the protein. It was shown that spectroscopic techniques could be used to investigate whether the incorporation of NAA had been achieved. It was also shown that EPR is a technique that could be used to investigate the reaction of the azide group and more importantly that the use of a hydrophobic pocket within a protein could be used to investigate the reactive intermediates of azides. Namely, by shielding the reactive intermediates from one another to remove the possibility of production of azobenzene-type products that have typically plagued investigations into the reactions of azides. By comparing the data obtained using the hydrophobic pocket of a protein, to the use of surface exposed positions and the incorporation of the azide within the chromophore of GFP it appeared that the protein "matrix" surrounding the azide could be defined in such a way to influence the reaction.

Despite the widespread use of aryl azides in biological experiments, for example, photoaffinity labelling, no detailed examination of photochemical intermediates has been reported. This study has shown that generating reactive intermediates derived from non-natural amino acids within the hydrophobic pocket of a protein interior provides a novel, but also practical technique to generate and study highly reactive intermediates.

The observed EPR spectra appear to be strongly correlated to the protein "matrix" surrounding the nitrene. Thus, it was proposed that generating reactive intermediates derived from noncanonical amino acids within the tuneable "matrix" of a protein interior appears to be a novel, but also practical approach to generate and study highly reactive intermediates.

The work presented within this chapter, especially the EPR spectra generated using T4L153azF and azF present within other defined protein environments (see section 3.4.4) has been accepted for publication in the ACS journal, Organic Letters.

4 CHAPTER 4: UTILISATION OF THE AMBER SUPPESSION TECHNOLOGY TO CONTROL A FUNCTIONAL BACTERIAL READOUT SYSTEM

4.1 Introduction

The second novel application of amber suppression explored within this thesis is the creation of a biological switch controllable by the presence of NAA. The creation of a switch would provide a novel method to control protein expression and could be linked, therefore, to modulation of a specific phenotype. Modulation of motility was chosen as a simple, visual output. In this chapter, the application of this technology to a prokaryotic system was explored; in Chapters 5 and 6 the application to a eukaryotic system was explored.

The amber suppression technology has the benefit that two scenarios occur depending on the presence/absence of the non-natural amino acid. Ideally, when a non-natural amino acid is supplied to cells capable of performing amber suppression, the non-natural amino acid is incorporated in response to the UAG codon, read-through of the entire gene occurs, full-length, functional protein is produced, and the cell displays the specific phenotype (Figure 50). Conversely, when non-natural amino acid is withheld from the cells, the UAG codon is read as a translation termination signal, protein is truncated and therefore no functional protein is produced, and the cell does not display the specific phenotype. Thus, the aim of this chapter is to explore the duality of the amber suppression method to produce a biological switch, which can control the phenotype of an *E*. *coli* cell. In this case the presence/absence of the non-natural amino acid is the switch which is linked to an easily read output, namely, the modulation of the motility of *E*. *coli*.

The production of a biological switch is interesting in itself, however, the creation of read-out system has other consequences that can be exploited. Namely, the NAA concentration and the efficiency of the aaRS could be used as "handles" that can be used to fine-tune the expression efficiency, which would, in theory, correlate to the efficiency of swimming. It would be possible to use this read-out system to select new efficient aaRSs specific for NAAs (Link *et al.* 2006). The selection procedure would be similar to the positive selection method as outlined in Chapter 1 (Figure 12) but the selection of efficient aaRSs could be achieved by picking *E*. *coli*

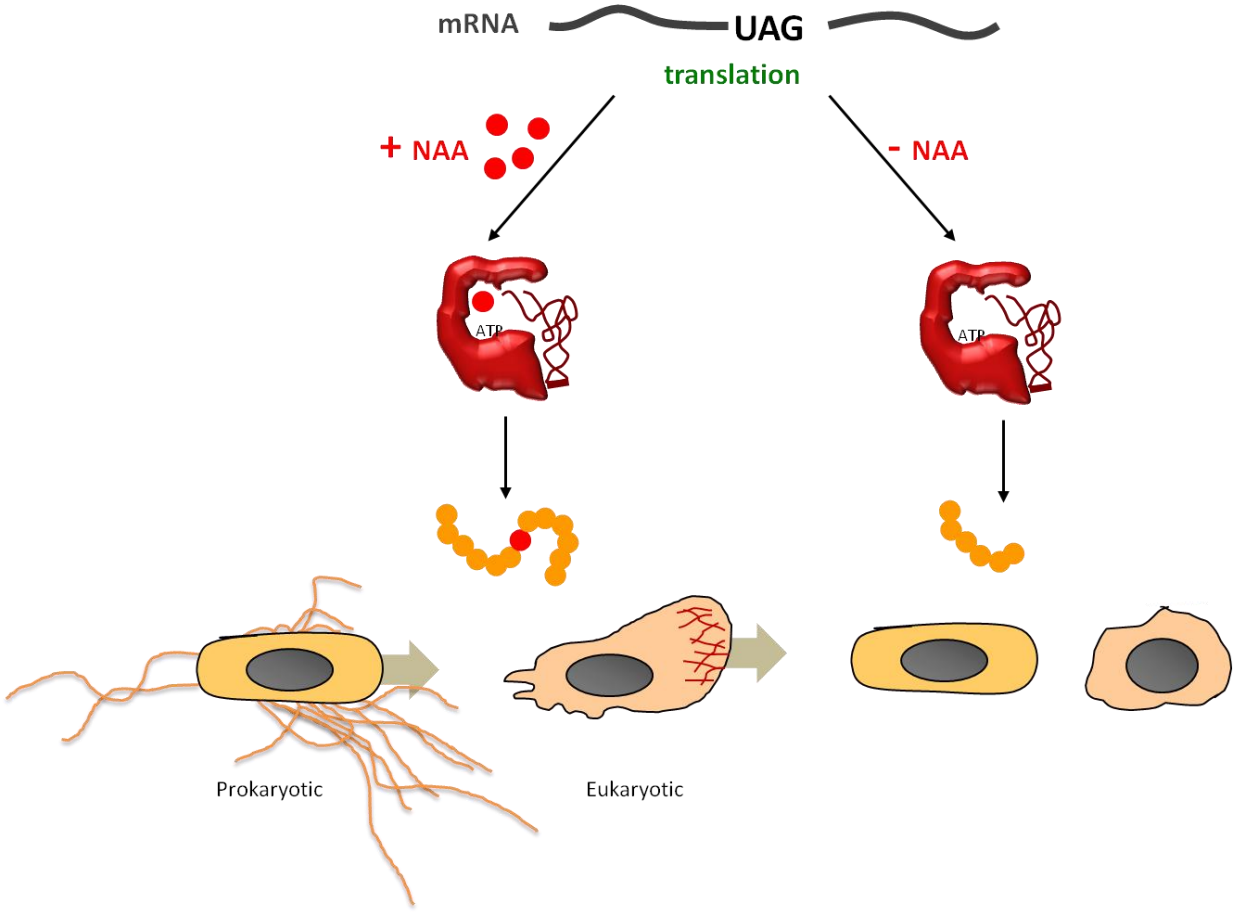


Figure 50. The two scenarios that occur during amber suppression technology depending on the presence/absence of a non-natural amino acid; and application in controlling cellular phenotype.

cells that display efficient swimming ability, namely, those from the outside of the swimming halo on a motility assay plate.

The addition of a NAA at an exposed position along the entirety of the flagella has interesting consequences, for example the ability to alter serotype and therefore antigen binding and immunological responses. In addition, the incorporation of azF at this position may allow the attachment of other factors, for example fluorogenic dyes, via click chemistry to the azF moiety (Beatty *et al.* 2005). Thus allowing site-specific labelling and visualisation of flagella, which can aid in investigations into flagella function (Turner *et al.* 2000).

The motility of *E*. *coli* is dependent on the formation of functional flagella. Without flagella the *E*. *coli* cell is non-motile. The multiple flagella, which are found on the cell surface, are composed of flagellin subunits that assemble to form flagella. The basal complexes that hold, and create rotational force to turn flagella, are located in the plasma membrane and associate with the flagella filament. The FliC gene encodes the flagellin protein. Strains of *E*. *coli* lacking functional FliC are non-motile. These non-motile strains provide a phenotypic background suitable for introducing plasmid-encoded FliC and observing the motility phenotype of the cell which will be contingent on the presence of nonnatural amino acid.

Specifically, the investigations within this chapter sought to:

- Establish that there were differences in motility between *E. coli* strains that exhibit wild type motility and those that are FliC-deficient
- Investigate whether plasmid-encoded FliC could restore motility to non-motile *E. coli* strains
- Investigate whether it was possible to control a bacterial phenotype (i.e. migration) as a function of the presence/absence of non-natural amino acid utilising the amber suppression technology.
- Investigate the applications of NAA display on an *E. coli* surface. Namely, the selection of new NAA-aaRSs, reaction of reactive moieties within NAAs displayed on *E*. *coli* cell surface.

4.2 Results

4.2.1 Analysis of *E. coli* motility

4.2.1.1 Functional FliC gene is required for motility

As the cell-based readout system required was centred on modulation of cell migration, firstly it was important to confirm that without a functional FliC gene, cells were nonmotile.

The *E. coli* strain BL21(DE3) were shown to be motile, as assayed by an agar plate motility assay, using a 1 µl drop of saturated culture onto a 0.3 % agar plate (Figure 51). The motile cells swam away from the inoculation point and spread over the agar to create a halo of cells surrounding the inoculation drop in the centre of the plate.

The KAF95 strain, previously constructed by Karen Fahrner (Berg and Turner 1993), was an *E. coli* K12 derivative, AW405 (Armstrong et al. 1967) which carried the fliC726 allele, with a nonflagellate phenotype (Silverman and Simon 1973). The strain also carried a cheY deletion, so would display smooth swimming motility and no chemotaxis. The KAF95 strain was shown to be non-motile, due to the lack of halo radiating from the inoculation drop in the centre of an agar plate motility assay (Figure 51). Instead, the growth of the colony occurred within a few millimetres of the inoculation point.

The JW1908 strain of the previously created Keio collection was an *E. coli* K-12 BW25113 derivative. The Keio collection was created by systematic construction of single gene knock-out mutants of all genes/ORFs (Baba et al. 2006) using Wanner's method (Datsenko and Wanner 2000). The clone, JW1908, used in this study was a FliC knock-out strain and therefore had a non-flagellate phenotype and was therefore non-motile. The strain will be subsequently referred to as “Keio”. The strain was shown to be non-motile (Figure 51).

These results demonstrated the non-motility of the two cell strains, KAF95 and Keio. These cell lines therefore, provided the basis for the investigations into

whether the motility of a bacterial cell could be controlled via the amber suppression method. The lack of the FliC gene and therefore lack of flagella gave the required background for the experiments introducing FliC gene back into these cells.

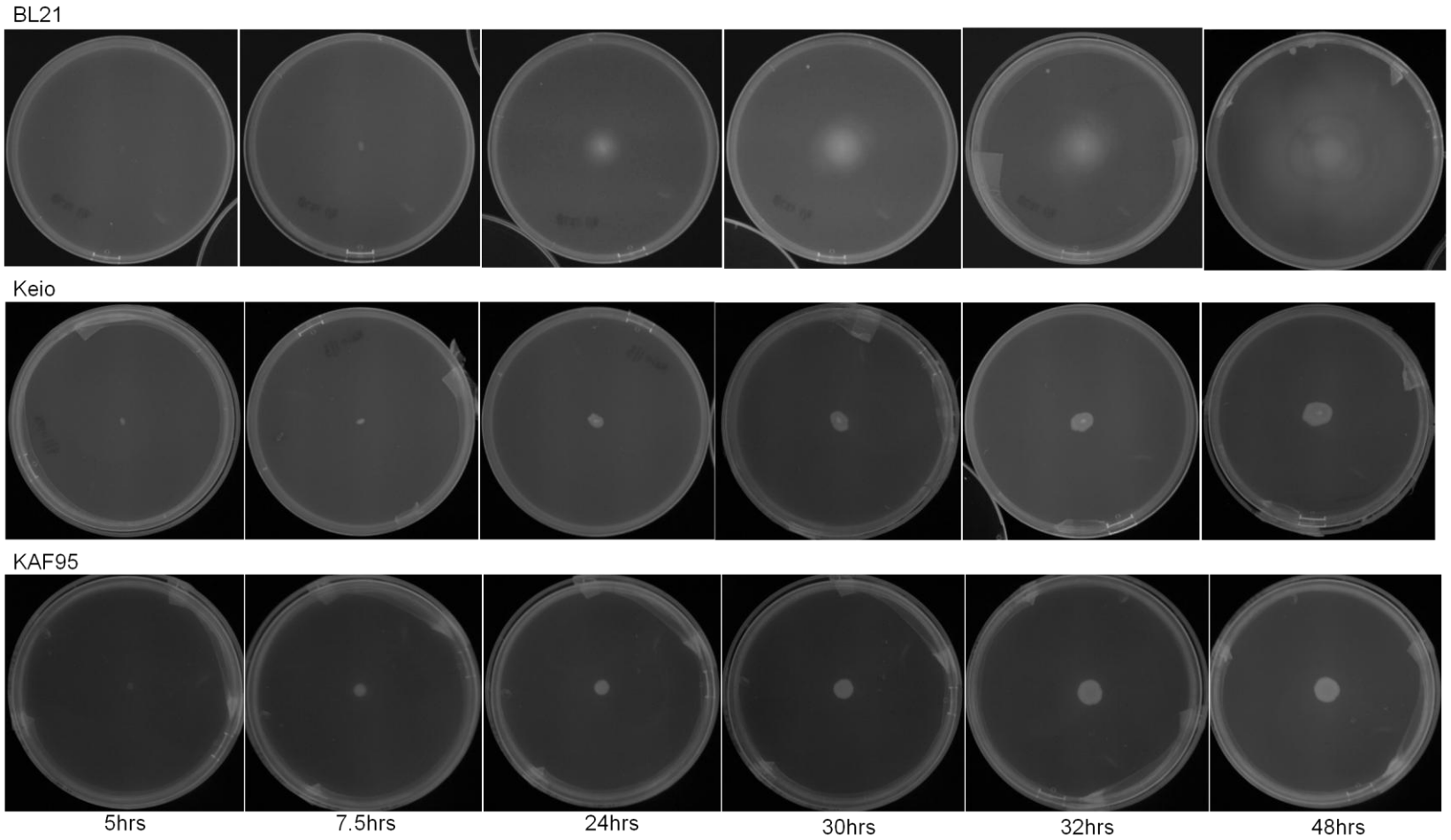


Figure 51. Comparison of motility; BL21(DE3), Keio, KAF95. A 1 µl drop of saturated culture containing either BL21(DE3), Keio or KAF95 was added to the centre of a 0.3 % agar plate (9 cm diameter). Ability of culture to swim was determined by taking photos at time points specified and monitoring halo of cells radiating from inoculation point.

4.2.2 Restoration of motility by plasmids-encoded FliC

To express plasmid encoded flagellin in the Keio and KAF95 strains, it was necessary to clone the FliC under the control of a promoter that could recognise by the endogenous *E*. *coli* machinery. This was because, unlike the BL21(DE3) strain, these strains do not harbour the T7 polymerase gene. The viral T5 promoter was therefore suitable, as it is recognised by the endogenous *E*. *coli* machinery. The plasmid pAA was chosen as this plasmid contains a T5 promoter driving expression of cloned genes and also, has previously been used in the application of the amber suppression method and in fact harboured the tRNAs required. Therefore this plasmid made an ideal vector for carrying the FliC gene for the subsequent implementation of amber suppression technology.

4.2.2.1 Cloning FliC into pAA plasmid

The FliC gene was amplified from the plasmid pCA24N-FliC by PCR using EconoTaq polymerase. The primers used were JLMFliC008 and JLMFliC010 which flanked the FliC gene, and added an *Xma*I and *Not*I to the 5' and 3' ends respectively (Appendix A2). Digested PCR product was ligated into the *Xma*I and *Not*I sites of complementarily digested and purified pAA plasmid. Colony PCR, using the primers rb_pAA_for and rb_pAA_rev, was used to screen the colonies for incorporation of the FliC gene (Figure 52). Putative positive colonies gave a PCR product of 1.5 Kb. Sequencing confirmed the insertion of FliC into plasmid pAA (Appendix A7). This resulted in the plasmid, pAAFliC (Figure 53).

4.2.2.2 Introduction of pAAFliC into non-motile strains

pAAFliC was used to transform Keio cells and used to perform a motility assay. However, motility was not restored to the non-motile Keio cells (Figure 54).

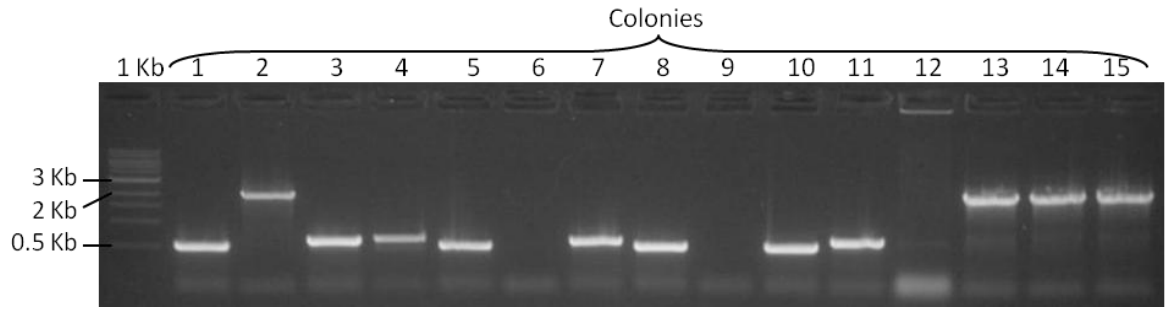


Figure 52. Agarose gel of colony PCR screening colonies for insertion of FliC gene into pAA using primers rb_pAA_for and rb_pAA_rev. Putative positive colonies gave a PCR product of 1.5 Kb. 1 Kb ladder (NEB) was used to approximate size of DNA fragments.

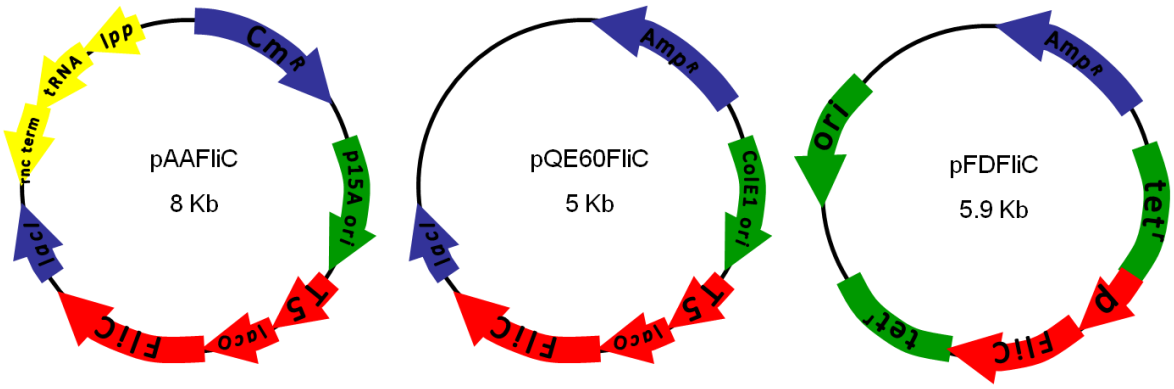


Figure 53. Plasmid maps of FliC-containing constructs used in investigations, namely, pAAFliC, pQE60FliC and pFDFliC.

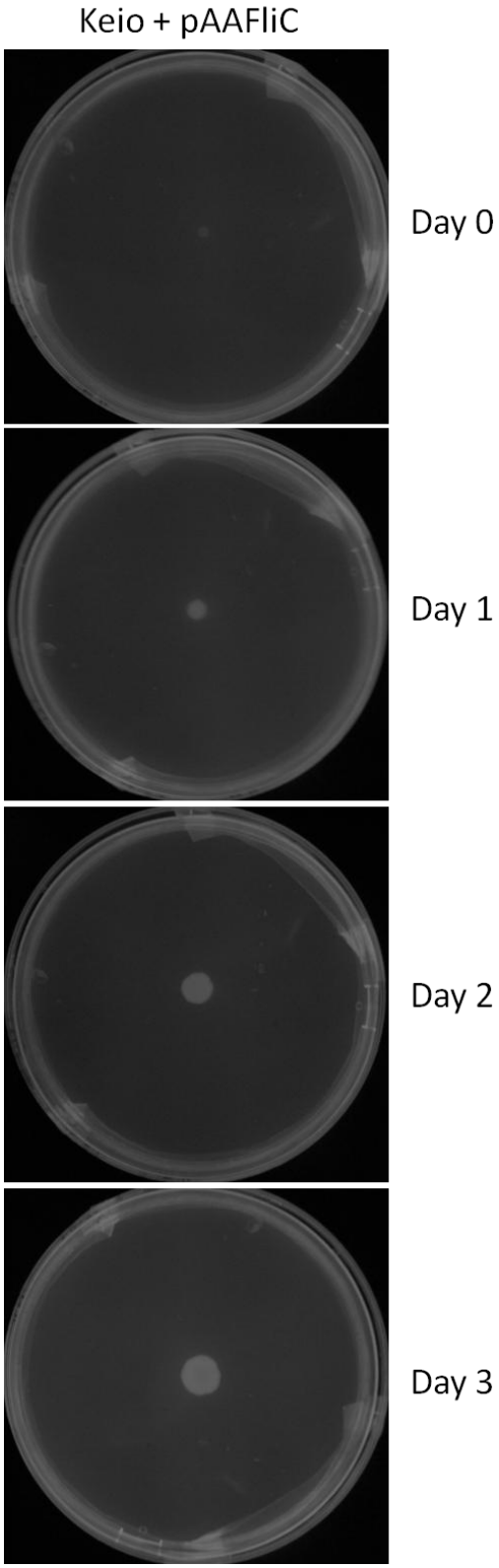


Figure 54. Swim assay of Keio + pAAFliC. Keio cells were transformed with plasmid pAAFliC. A colony was picked , grown to saturation and used to inoculate a 0.3 % agar motility assay plate (9 cm diameter). Images were captured at the specified timepoints.

4.2.2.3 Cloning FliC into pQE60 plasmid

Because pAA was an in-house produced plasmid, it was thought that either some complexity of the plasmid interrupted transcription or translation of the target gene. Therefore, FliC was cloned into the commercially available plasmid, pQE60 (Qiagen). This plasmid also controls expression of the target gene under the T5 promoter. The FliC gene was amplified from the plasmid pCA24N-FliC using PCR and primers JLMFliC014 and JLMFliC015 which flanked the FliC gene and added an *Nco*I and *Bgl*II to the 5' and 3' ends respectively. Digested PCR product was ligated into the *Nco*I and *Bgl*II sites of complementarily digested, de-phosphorylated and purified pQE60 plasmid. Colony PCR was used to screen the colonies for incorporation of the FliC gene, using primers pQE60 seq for and pQE60 seq rev, which were specific to the plasmid and therefore did not amplify endogenous FliC (Figure 55). A PCR product of 1.5 Kb indicated insertion of FliC into pQE60. DNA sequencing confirmed the incorporation of FliC gene into pQE60 plasmid. This created the plasmid pQE60FliC (Figure 53). The FliC gene contained the following mutations; Q76R, D121V, T400A (Appendix A7).

4.2.2.4 Introduction of pQE60FliC into non-motile strains

Keio cultures harbouring pQE60FliC were used in motility assays. Once again plasmid encoded FliC was not able to restore motility to the non-motile strains (Figure 56). Various inducing media was used to try to modulate the expression of flagellin, as it was thought that over expression could be detrimental to production of functional flagella. Cultures were grown in cultures containing either no added inducer, 0.2 % lactose, 0.5 mM or 25 µM IPTG (de Boer *et al.* 1989). In each case, the motility of the Keio cells could not be restored (Figure 56).

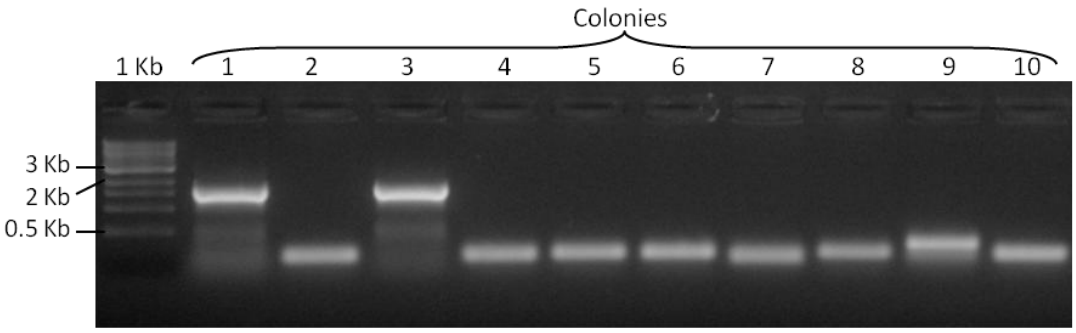


Figure 55. Agarose gel of colony PCR to screen for insertion of FliC into pQE60 (lanes 1-9), negative control colony (lane 10). PCR was performed using primers pQE60 seq for and pQE60 seq rev. Product of 1.5 Kb indicated insertion of FliC into MCS. 1 Kb ladder (NEB) was used to approximate size of DNA fragments.

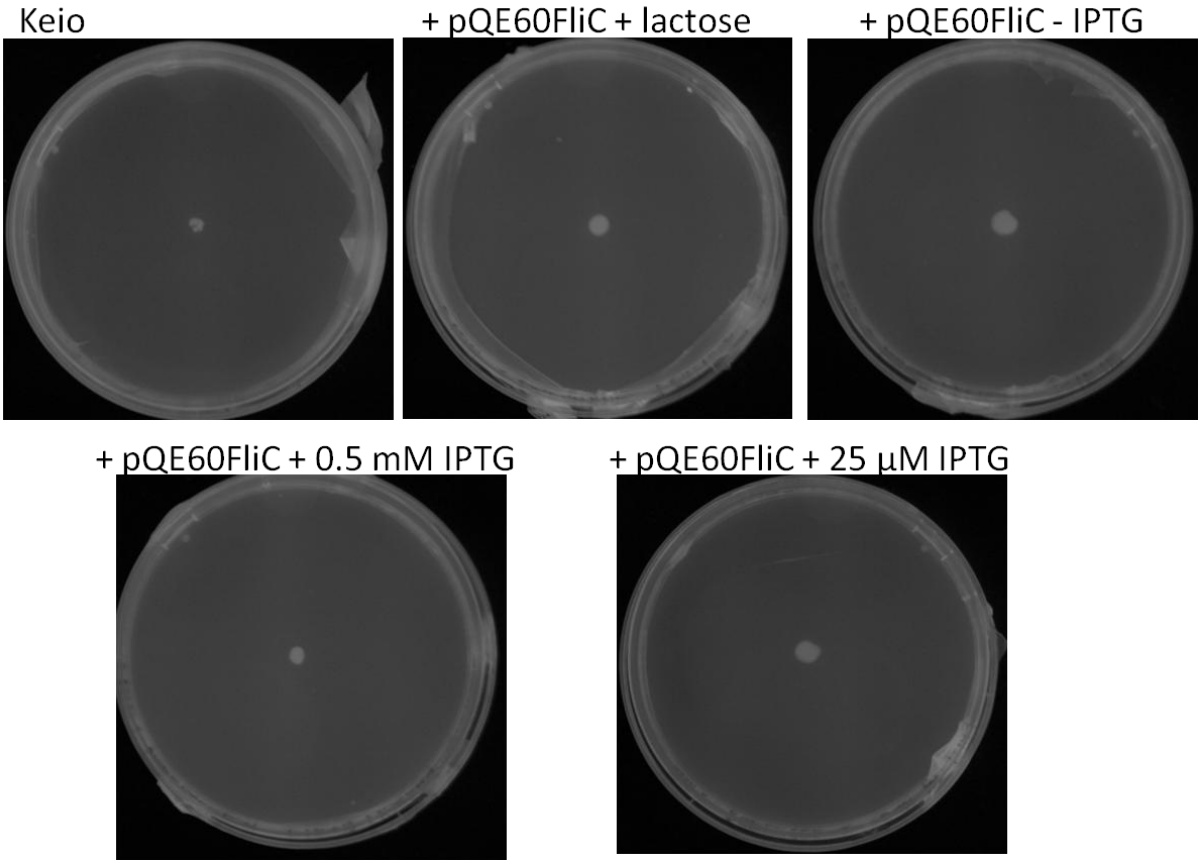


Figure 56. Swim assays of Keio plus pQE60FliC, investigating induction substrate and concentration. Keio were transformed with pQE60FliC, colonies picked and used to inoculate LB cultures with either lactose, 0.5 mM or 25 µM IPTG. After growth, saturated liquid cultures were used to drop a 1 µl drop into the centre of a swim assay plate (9 cm diameter) containing the same concentration of inducer. Photos were taken 1 day after inoculation.

4.2.3 Plasmid pFD313 and pFDFliC

To ascertain whether the issues with the plasmids encoding cloned FliC were related to the use of the T5 promoter, the use of the previously constructed pFD313-derived plasmid was investigated. The pFD313 plasmid was kindly donated by Dr Richard Berry (Rotary molecular motors group, The Oxford Biological Physics Group, Oxford University) and had previously been used in conjunction with the non-motile KAF95 strain of *E. coli* to produce motile *E. coli* with sticky flagella. These sticky flagella were encoded by the FliC gene with a deletion spanning codons 245-301. The plasmid is a pBR322 based cloning plasmid, harbouring the genomic fragment containing the FliC gene (Kuwajima 1988a). This fragment therefore includes the natural promoter and terminator of the FliC gene. The pFD313 plasmid was shown to restore motility to the two non-motile strains, KAF95 and Keio (Appendix A6).

4.2.3.1 Creation of the plasmid pFDFliC

Firstly, the section missing from the FliC gene, namely codons 245-301, was re-inserted into plasmid pFD313. During the deletion mutagenesis to create pFD313, a *Bgl*II and *Hind*III restriction site was inserted into the regions of the FliC gene flanking the deletion; therefore, these sites were used to re-insert the missing section. The primers, pFD313FliC FP and pFD313FliC RP (Appendix A2), were used to amplify the missing section of the FliC gene using PCR (template pCA24N) which added a *Bgl*II and *Hind*III site to the respective 5' and 3' ends of the PCR product. Digested PCR product was ligated into complementarily digested plasmid pFD313. Because of the small size of the insert, a difference could not be seen by colony PCR or by agarose gel electrophoresis of the purified plasmid, therefore sequencing was used to identify the presence of the inserted fragment. Sequencing confirmed that the plasmid pFDFliC contained full-length FliC. This created plasmid pFDFliC (Figure 53). The sequence contained a few mutations, namely Y246R, A247S, A301K, V302L, due to the presence of the *Bgl*II and *Hind*III sites at the ligation position (Appendix A7). The mutations were tolerated as they were present within the variable region of the flagellin and were therefore unlikely to alter the structure or

function of the flagellin molecule or the flagella. Also, these mutations were present in the FliC gene within the pFD313 plasmid that had previously been shown to produce functional flagella.

4.2.3.2 Restoration of motility using pFDFliC

The ability of the plasmid, pFDFliC, to restore motility to the non-motile strains KAF95 and Keio was investigated. The motility of both strains was restored using the pFDFliC plasmid (Figure 57). This was evident by the halo of migrated cells radiating from the inoculation point on the motility agar plate inoculated with cells transformed with pFDFliC. Whereas, no halo was observed on the motility assay plate containing cells not transformed with pFDFliC.

This result showed that plasmid encoded FliC (pFDFliC) could restore motility to the non-motile *E*. *coli* strains. Therefore the plasmid pFDFliC was further investigated for use in the application of the amber suppression technology.

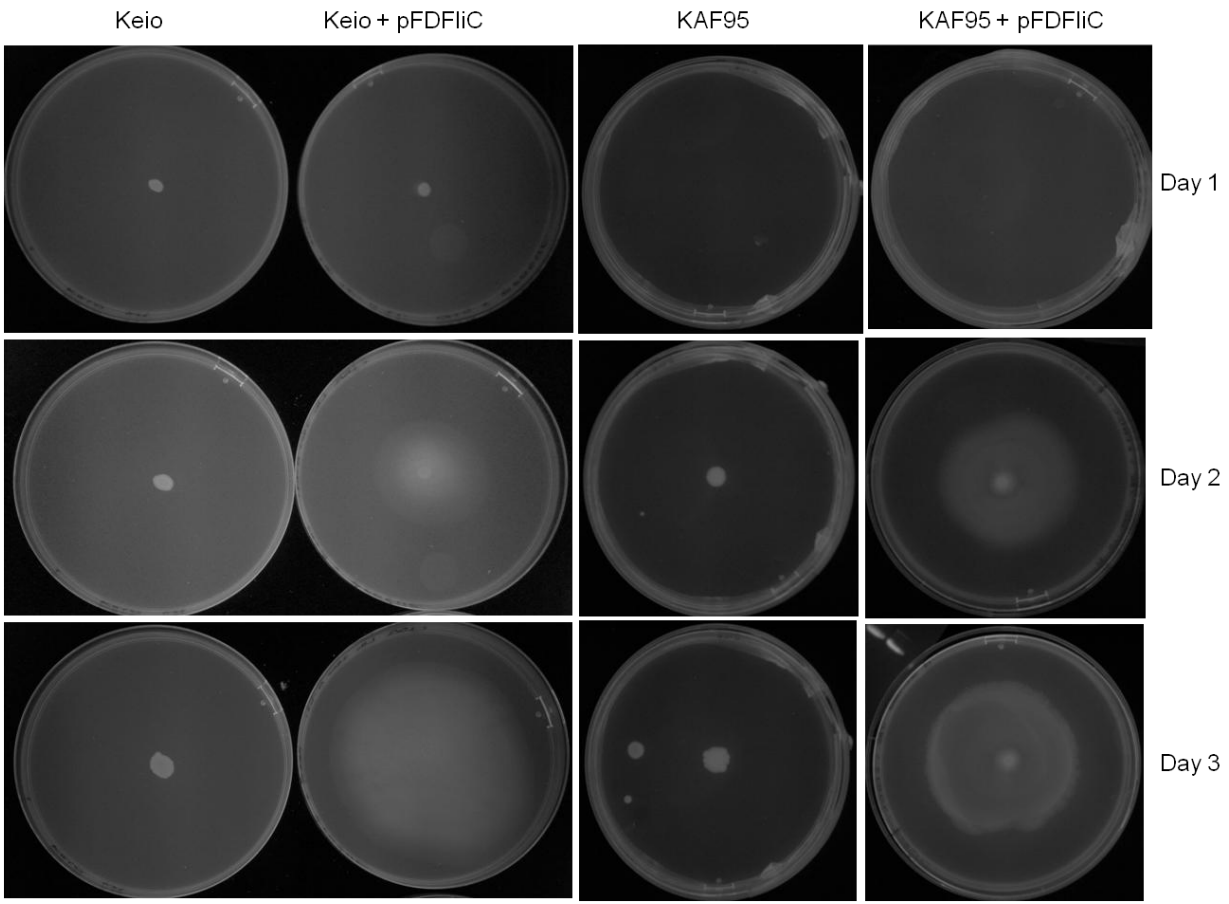


Figure 57. Comparison of the motility of Keio and KAF95 and Keio and KAF95 harbouring the plasmid pFDFliC. The two non-motile strains were transformed with pFDFliC, colonies were picked and used to inoculate LB cultures., a 1 µl drop of the saturated liquid cultures were used to inoculate 0.3 % motility assay plate (9 cm diameter). Photos were taken at specified timepoints.

4.2.4 Implementation of amber suppression in FliC

4.2.4.1 Introduction of TAG mutation into FliC

For the implementation of the amber suppression technology in the *E*. *coli* cells, a TAG mutation was required in the FliC gene. As addressed in Chapter 1, the flagellin protein is composed of domains, each with a specific function, therefore the position of the NAA substitution within the flagellin protein was considered.

4.2.4.1.1 Structural and positional considerations

The N- and C- terminal of the flagellin protein, domain 0 and 1, are required for formation of the functional flagellin tertiary structure (Figure 58). Although presence of a TAG mutation close to the N-terminal has benefits, namely, short truncated protein that should not confer function, mutation of codons within domain 0 and 1 was not immediately considered as perturbing the structure here could prevent formation of functional flagella. The variable region, domain 3, of the flagellin protein provides an ideal position for insertion of a NAA. As the name suggests, this region has been shown to possess a number of variations between species, which account for different serotypes (Schoenhals and Whitfield 1993). Also, numerous mutagenic studies have investigated the function of this variable region. In fact it has been shown that this region can be entirely deleted and the flagellin retains function (Kuwajima 1988a). Another point to note is that Lu et al (1995) have demonstrated that an entire protein can be inserted into the flagellin protein at position 239.

Therefore a site within the variable region of flagellin site was an ideal position for mutagenesis as there should be limited perturbation of the protein's ability to fold correctly. Specifically, position 239 was chosen as an ideal target. The threonine at codon 239 is within a loop and points into the solvent. This was an ideal position because, as the flagellin is created the NAA would be placed in a solvent exposed position along the entirety of the flagella.

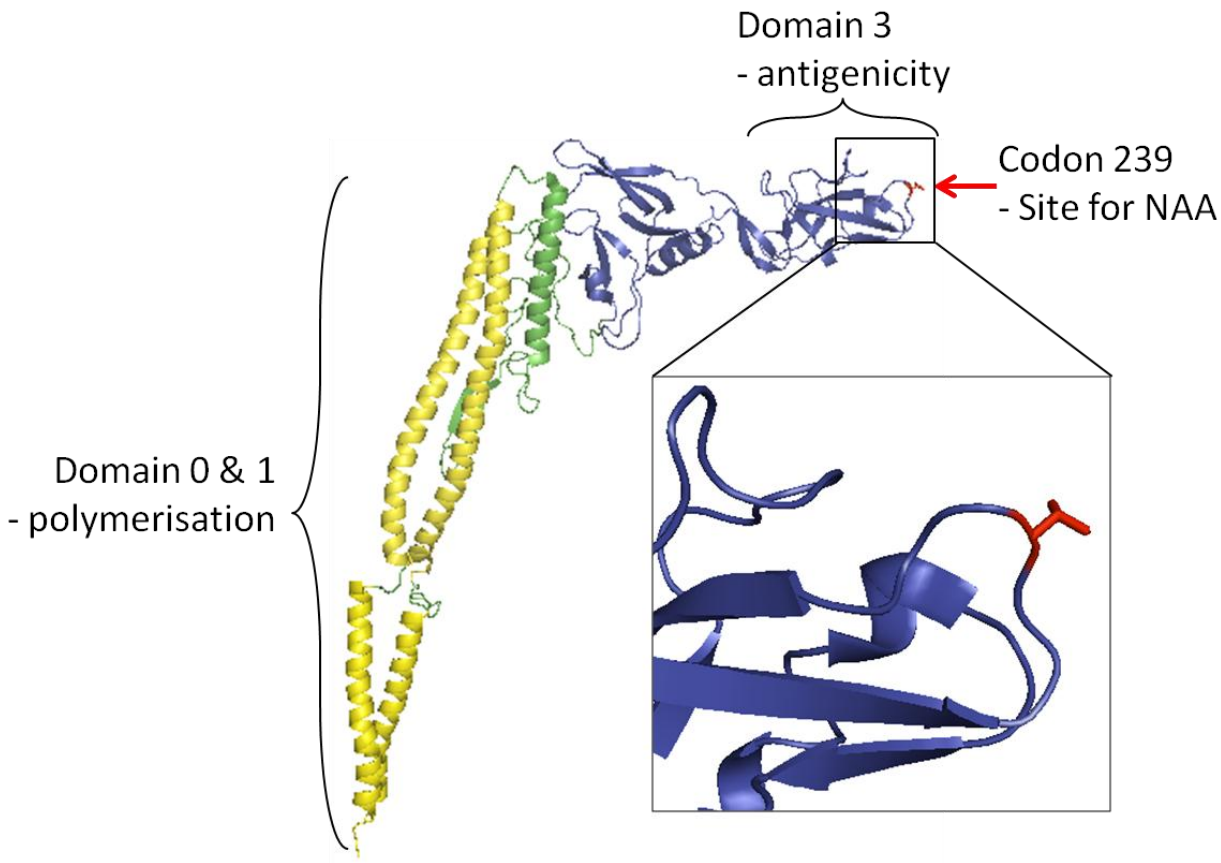


Figure 58. Structure of FliC, domains and putative site for NAA incorporation. PDB structure 1UCU was used (Yonekura et al. 2003), N- and C- terminal coils of domain 0 and 1 are coloured yellow, antigenic domain 3 is coloured blue and residue 239 is coloured red. Structure was produced using Pymol (DeLano and Lam 2006).

4.2.4.1.2 Creation of plasmid pFDFliC239TAG

The TAG mutagenesis was performed using site-directed mutagenic PCR, using primers pFD313frag16 TAG FP/RP that introduced a TAG codon at position 239 of FliC within plasmid pFDFliC (Appendix A2). Sequencing of the plasmid DNA confirmed that the Thr239→TAG mutagenesis had been successful. This created plasmid pFDFliC239TAG.

4.2.4.2 Introduction of pFDFliC239TAG in to KAF95 strain

The FliC gene contained an in-frame TAG codon within its sequence, therefore, full–length protein should not be produced, subsequently there should not be restoration of motility when the plasmid pFDFliC239TAG was introduced into the non-motile strains. When the pFDFliC239TAG plasmid was introduced into the KAF95 strain, motility was restored (Figure 59). This was an unexpected result. To eliminate the possibility that contamination of the culture with a cell strain displaying WT motility, the experiment was repeated several times, however, in each case the same result occurred.

To investigate the possibility that the incorporation of natural amino acid into position 239 was occurring and this resulted in full-length protein and functional flagella, a similar experiment was performed using minimal media for inoculation culture and motility assay plates. It was thought that limiting the concentration of nutrients and most importantly, amino acids, the cells would have limited natural amino acid to incorporate into the flagellin protein in response to the TAG codon. However, the cells still showed restoration of motility (Figure 59).

4.2.4.3 Introduction of pFDFliC239TAG in to Keio strain

When the pFDFliC239TAG plasmid was introduced into the Keio strain, the cells remained non-motile (Figure 60). The Keio strain was therefore taken forward for further investigations in to whether the amber suppression technology could be used to modulate the phenotype of a bacterial cell.

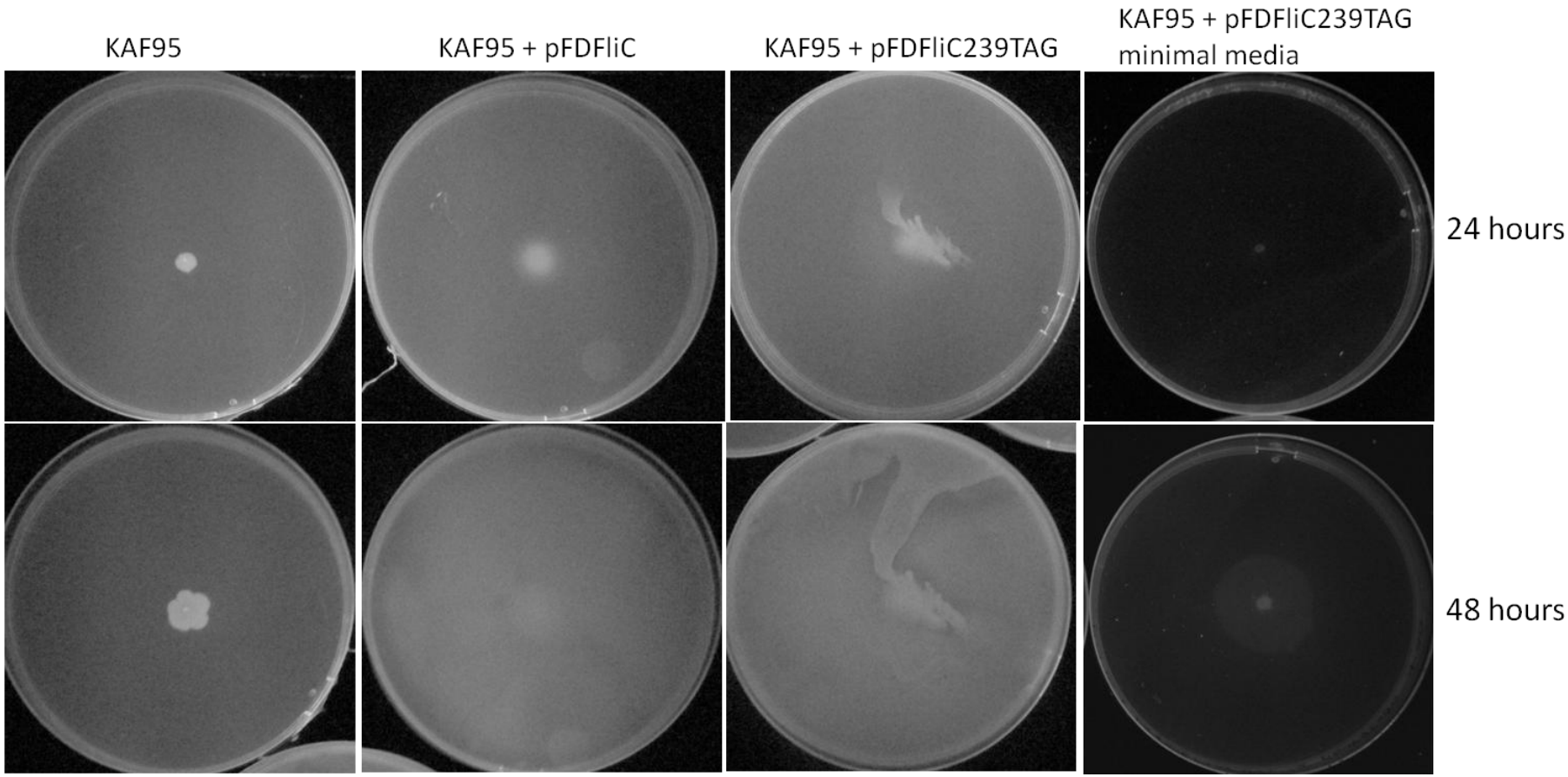


Figure 59. Motility assays investigating the ability of the plasmids pFDFliC and pFDFLIC239TAG to restore motility to KAF95. The non-motile strain KAF95 was transformed with either pFDFliC or pFDFliC239TAG. Colonies were picked and used to inoculate LB cultures., a 1 µl drop of the saturated liquid cultures were used to inoculate 0.3 % motility assay plate (9 cm diameter). A motility assay inoculated with untransformed KAF95 was also performed. Photos were taken at specified timepoints.

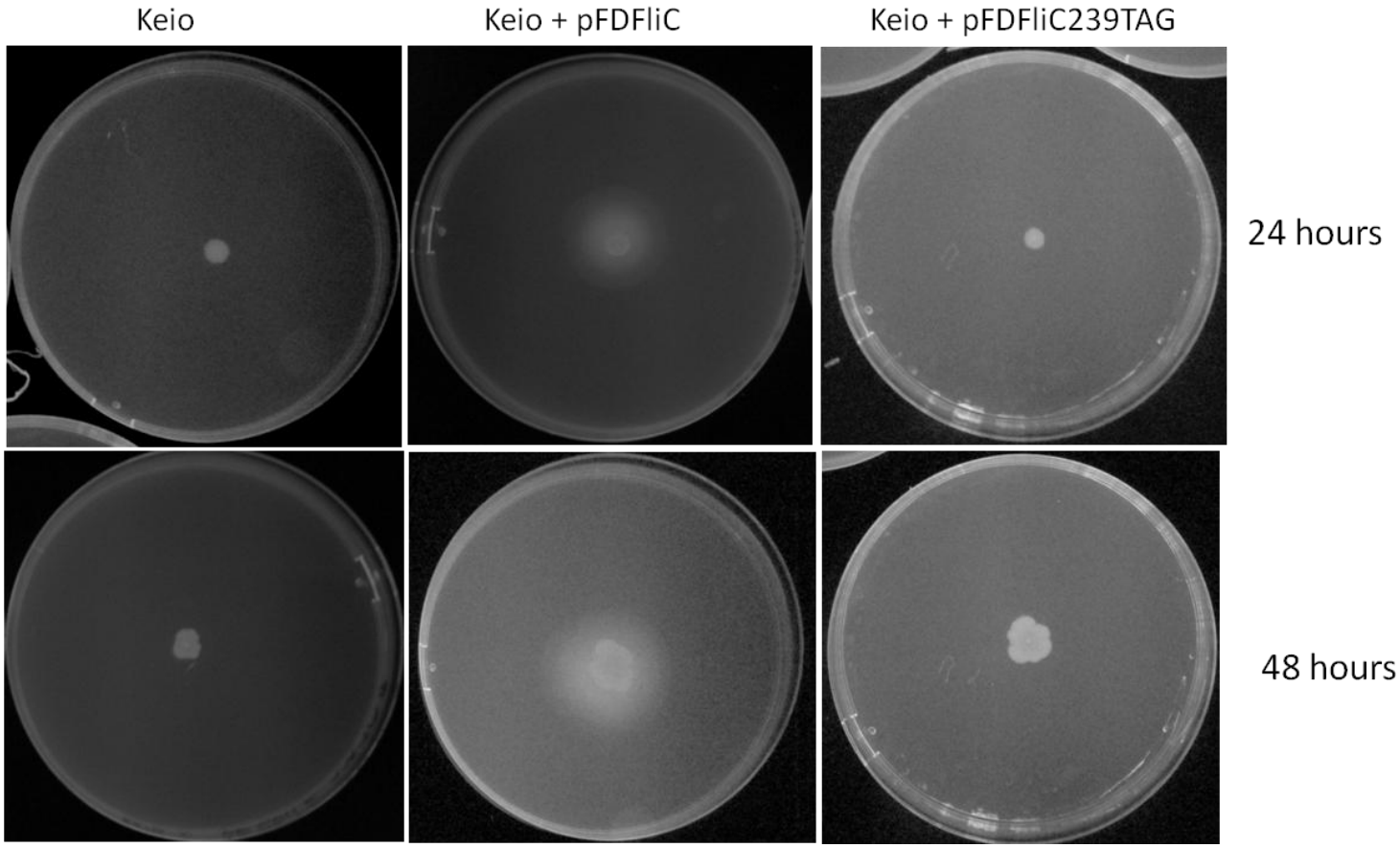


Figure 60. Motility assays investigating the ability of the plasmids pFDFliC and pFDFLIC239TAG to restore motility to Keio.. The non-motile strain Keio was transformed with either pFDFliC or pFDFliC239TAG. Colonies were picked and used to inoculate LB cultures., a 1 µl drop of the saturated liquid cultures were used to inoculate 0.3 % motility assay plate (9 cm diameter). A motility assay inoculated with untransformed Keio was also performed. Photos were taken at specified timepoints.

4.2.4.4 Investigations into the control of motility via amber suppression technology

To investigate whether the amber suppression technology could be used to control the motility phenotype of the *E*. *coli* cultures, the pFDFliC239TAG construct was co-transformed into the non-motile strains with one of the NAA-RS-encoding plasmids. The pAAazRS, pSUPiodoRS and pDULEazRS plasmids, all encoded a NAA-specific aaRS (iodoF or azF) and $tRNA_{CUA}^{Tyr}$, which provided the machinery for amber suppression and incorporation of NAA into protein.

Despite the utilisation of different aaRS/tRNA constructs, NAA used and position of TAG mutation within FliC, in each case the same result was obtained; KAF95 remained motile despite the absence of NAA (Figure 61) and motility could not be restored to the Keio strain in the presence of NAA (Figure 62). It was thought that possibly the truncated protein produced by the low efficiency of NAA incorporation was blocking the formation of functional NAA-containing full-length flagella. Therefore a construct harbouring a TAG mutation at Ile4TAG was created, plasmid pFDFliC4TAG. It was shown that this construct was unable to restore motility to KAF95 or Keio strains when introduced with a NAA-RS-containing plasmid.

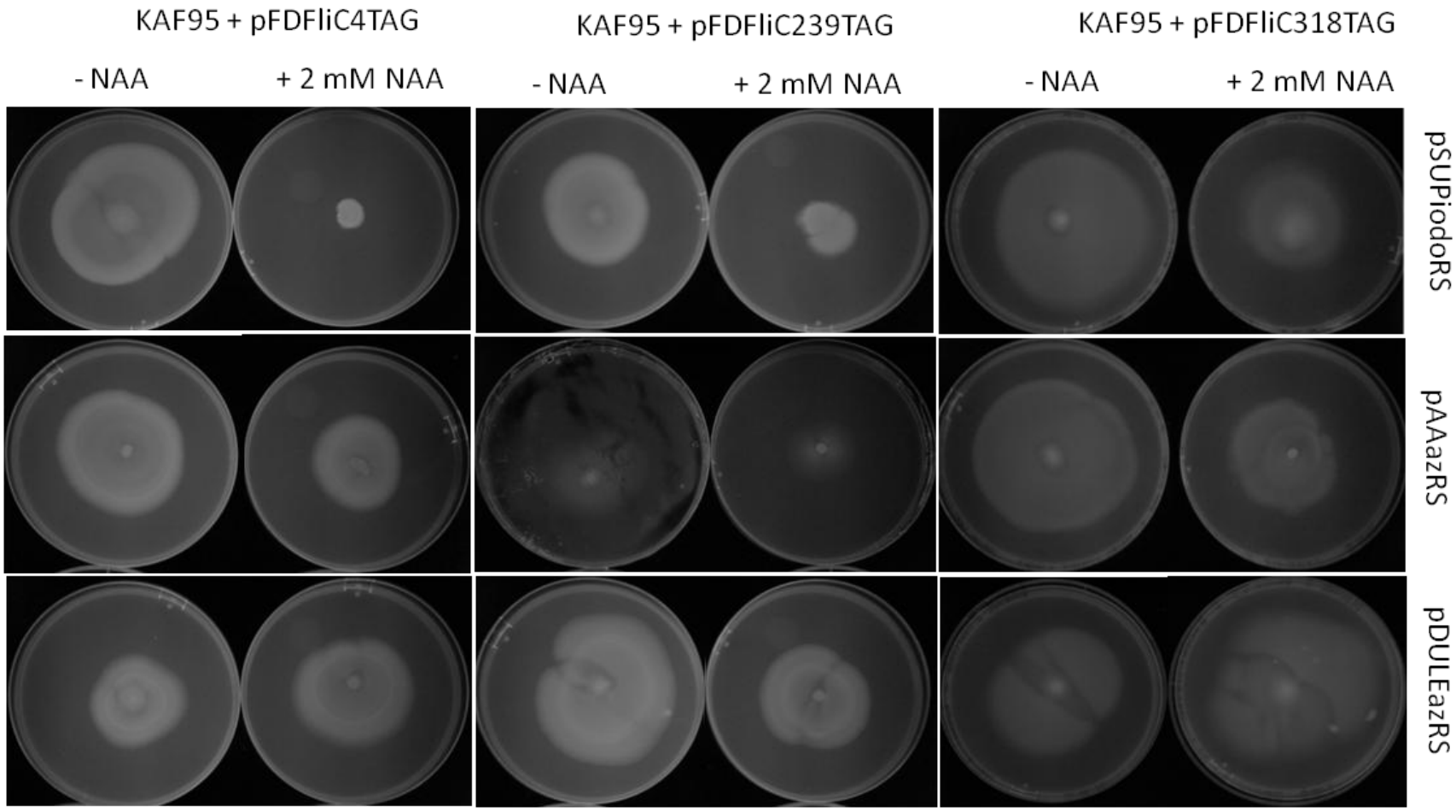


Figure 61. Motility assay of KAF95 co-transformed with either pFDFliC4TAG, pFDFliC239TAG or pFDFliC318TAG and either pSUPiodoRS, pAAazRS or pDULEazRS. Colonies were picked, grown to saturation in media either with or without NAA, and used to inoculate a 0.3 % agar motility assay plate (9 cm diameter) either with or without NAA.

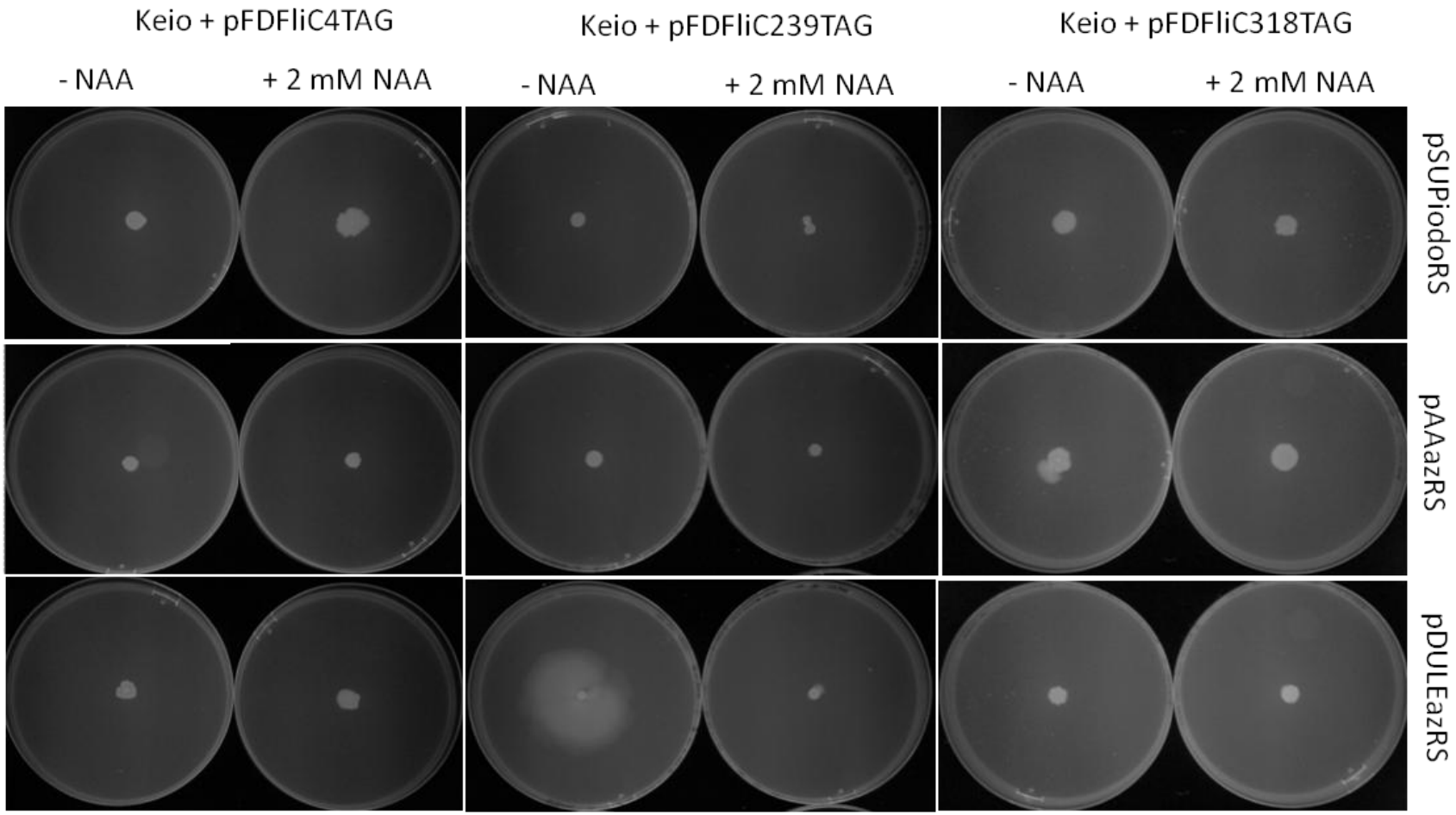


Figure 62. Motility assay of KAF95 co-transformed with either pFDFliC4TAG, pFDFliC239TAG or pFDFliC318TAG and either pSUPiodoRS, pAAazRS or pDULEazRS. Colonies were picked, grown to saturation in media either with or without NAA, and used to inoculate a 0.3 % agar motility assay plate (9 cm diameter) either with or without NAA.

4.3 Discussion

4.3.1 FliC gene is required for motility of *E. coli*

The motility assay, utilising 0.3 % agar plates (Kondoh and Ozeki 1976) was shown to be useful in the elucidation of the motility of *E*. *coli*. Motile cells, namely *E*. *coli* strain BL21(DE3), were shown to be motile due to the presence of a halo of swimming cells radiating out from an inoculation point on a 0.3 % agar motility assay plates. This halo was not present in the *E*. *coli* strains containing no FliC gene, Keio, and mutant FliC gene, KAF95. The Keio and KAF95 strains were therefore used in experiments to restore motility via plasmid encoded FliC.

4.3.2 Plasmids encoding FliC

A variety of plasmid constructs were created that encoded FliC in attempt to restore motility to the non-motile strains. The failure of the plasmids pAAFliC and pQE60FliC to restore motility to non-motile strains indicated that plasmid construction was important for expression of functional flagellin. Although the use of a T5 promoter should have allowed expression of protein within any *E*. *coli* strain via use of the endogenous RNA polymerase, it was possible that in this instance its use to control expression of flagellin was not ideal. The pAA plasmid has previously been shown to express protein (pAAazRS; within Chapter 3) and pQE60 was commercially available, therefore a more specific issue was considered. Possible explanations include that protein was not expressed from these plasmids; it was possible that the plasmids could overexpress protein, but cells could not export or form functional flagella using this expressed protein. This could be due to overexpression of flagellin, that blocks the export mechanism, or that the cell does not recognise the protein as WT flagellin and therefore was unable to export the protein or form functional flagella. It is possible that flagellin was expressed, flagella were produced, but these flagella were non-functional or too few formed for the cells to swim. Isolating overexpressed protein using these constructs could determine whether protein was, in fact, being expressed.

However, the use of exogenous promoters to express flagellin protein within *E. coli* has been shown to be possible. Lu et al (1995) produced flagellin monomers from a bacteriophage λ *pL* promoter that were exported to the cell surface and shown to form partially functional flagella (not WT motility due to insertion of thioredoxin) as determined by observing rotation of tethered cell to a microscope slide. Goudarzi et al (2009) have also expressed flagellin using an exogenous promoter on a pET plasmid (T7 polymerase). However, the flagellin was overexpressed and purified and therefore, whether functional flagella were formed was not ascertained.

pFDFliC, with the natural FliC promoter and terminator, was able to restore motility to both non-motile strains. This result was as anticipated and reinforced the idea that the exogenous promoters within pAAFliC and pQE60FliC were responsible for the failure of previous attempts to restore motility to non-motile strains via plasmid-encoded FliC. The use of the natural promoter allowed the formation of flagellin and production of flagella in response to cognate signalling. This may have explained why this construct produced functional flagella whereas the other constructs had not.

It has been introduced in Chapter 1 that flagellin monomers are exported from the cytoplasm to the polymerisation point on the flagella via a flagellum-specific pathway (Type III) with the export signal being present within the N-terminal portion of the flagellin protein itself. Therefore, it can be supposed that the export signal for flagellin would have been present regardless of plasmid construction. However, if not intact within the plasmids harbouring exogenous promoters this could explain why these plasmids were unable to restore motility to the non-motile strains. For the pFD plasmids, utilising the natural FliC 5' and 3' DNA sequence of the gene, should have provided any signal sequence if not present within the expressed protein. This construct was shown to produce functional WT flagella and this provides evidence that the export machinery within the cells was functional. This suggests that the issue with restoring motility to non-motile *E. coli* under the control of NAA technology was the non incorporation of NAA itself.

4.3.3 KAF95 could not be used to modulate motility via amber suppression

The production of the non-flagellate KAF strain was due to ethyl methanesulfonate mutagenesis of the FliC gene (Berg and Turner 1993). This is a position-non-specific chemical mutagenesis and therefore produces non-defined mutations within the genome (for review see Sega 1984). The KAF95 strain was therefore not ideal for genetic manipulation using plasmid technologies. This inactivation, and not deletion of FliC, could explain some of the unexpected data produced using plasmid-encoded FliC.

The use of KAF95 was initially promising as introduction of FliC encoded on the pFDFliC plasmid restored motility to the strain. However, when pFDFliC239TAG was introduced into the KAF95 cells, motility was still restored. This was not as anticipated because full-length flagellin protein and therefore functional flagella should not have been expressed. A likely explanation for this is that incorporation of natural amino acid at position 239 had occurred, typically Tyr, which allowed read-through of the entire FliC gene and production of functional protein (Nehring et al. 2012). This may be possible without the presence of NAA-RS or $tRNA_{CUA}^{Tyr}$-containing plasmid utilising the endogenous system entirely. Previous work within Chapter 3 of this Thesis has shown that incorporation of natural amino acid in response to the TAG codon is low due to the orthogonal nature of the *Mj* NAA-machinery; however, if any incorporation of natural amino acid has occurred, functional full-length flagellin would be produced. It could be argued that a sufficient amount of flagellin was produced and that functional flagella were formed. However, because the pFDFliC239TAG construct did not confer motility to the Keio cells other reasons are considered.

It was possible that full-length flagellin was produced as combination of the two aberrant FliC genes; some portion from the plasmid-encoded flagellin and the other from the inactivated endogenous FliC. It is also possible that during the transformation or motility assay process a motile strain had been selected from a mixed population or that spontaneous reversion of the mutation had occurred

rendering the cell motile once again (Armstrong et al. 1967; Lu et al. 1995). The presence of unequal motility rings on the motility assay plates inoculated with KAF95 (Figure 61) suggested that the population of cells was heterogeneous (i.e. cells had different propensity to swim). Also to note was that the KAF strain has no selection marker and therefore any contamination of the growth media with bacteria displaying WT-like motility would skew the results. The fact that when the pFDFliC239TAG plasmid was introduced into the Keio strain, the cells remained non-motile reinforced the supposition that it was in fact the FliC inactivation within the KAF strain which had somehow caused the aberrant results when introducing the FliC-TAG encoding plasmid.

When amber suppression technology was attempted in the KAF95 strain motility was observed when NAA was absent or present in the media. Interestingly, the addition of NAA appeared to reduce the motility observed when compared to when NAA was absent. This could be due to toxicity of the NAA which retarded culture growth either in liquid culture or during swim assay. Possibly, the incorporation of NAA at UAGs found at the termination of endogenous genes had reduced cell growth.

Due to these considerations, the results using the KAF95 strain cannot be considered with any certainty.

4.3.4 Keio had the potential to be used to modulate motility via amber suppression

The Keio strain had an entire FliC deletion and insertion of kanamycin resistance gene in place, which had two benefits. Firstly, the entire deletion of FliC provided a FliC deficient background in which results were more easily interpreted. More specifically, it could be ascertained more easily whether the plasmid encoded FliC was responsible for restoration of motility. Secondly, the resistance to kanamycin supplied within the growth media and swim plates helped to reduce the possibility that contamination of the culture was responsible for the motility observed.

The ability to insert plasmid-encoded FliC, via transformation, into this FliC knockout background meant that the FliC gene could easily be mutated prior to transformation. Also important was that the flagella would be entirely created from mutant flagellin derived from FliC. This is important in showing the motility phenotype of the cells is due to the plasmid-encoded FliC.

Motility assays demonstrated that introduction of WT FliC, via plasmid pFDFliC, restored motility to Keio. Therefore Keio was selected for further analysis using amber suppression technology.

4.3.5 Implementation of amber suppression technology

When the TAG-containing FliC gene construct (pFDFliC239TAG) was introduced into the Keio strain, the outcome was as anticipated. Namely, the cessation of restoration of motility to the cells. Therefore, the Keio strain, in combination with plasmid pFDFliC239TAG was selected to implement amber suppression technology in order to try to restore motility to the non-motile strain. However, when the orthogonal $tRNA_{CUA}^{Tyr}$ and NAA-RS-containing plasmids, along with NAA, were introduced into the Keio cells, motility could not be restored. This result suggested that the technology to incorporate NAA was non-functioning in the Keio cells. There are numerous suggestions as to why motility was not restored and these will be considered.

Three different plasmids were used in combination with pFDFliCTAG plasmids, namely pAAazRS, pSUPiodoRS and pDULEazRS. The use of different RSs developed to incorporate different NAAs allowed use of both iodoF and azF NAAs. This allowed investigations into whether the efficiency of uptake or incorporation of the NAA would affect the ability to produce function protein and therefore phenotype. However, neither RSs, tRNAs or NAA increased the ability to restore motility to Keio. Although these plasmids and NAAs have been used for amber suppression in *E*. *coli*, the use of specialist cell strains has not been investigated. Although differences in amber suppression efficiency have been noted between certain *E*. *coli* strains (O'Donoghue et al. 2012).

The incorporation of NAA could have perturbed the structure of flagellin so no functional flagella could be produced, however the position of incorporation of NAA was considered. The position initially chosen (codon 239) was within the variable region of flagellin, where variations are tolerated, and, in fact, entire protein insertions have been successfully performed and functional flagella produced (Lu et al. 1995). Also taken into account was that the chosen site was exposed to the solvent (Figure 58) and not crucial in structural integrity of the flagellin protein. To investigate whether the position of the incorporation of the NAA affected the folding or the ability of the protein to form functional flagella, three positions within the FliC gene were mutated to TAG. Positions Ile4TAG, Gly239TAG and 318TAG were used. It appeared that position of NAA incorporation had no effect on the application.

The efficiency of incorporation of NAA could have been so low that an excess of flagellin truncates were produced. If truncated protein were to accumulate, this could block the export machinery and hinder formation of functional flagella. However, the use of the pFDFliC4TAG construct provided evidence that truncated protein was not responsible because the three amino acid long truncate produced with this system should not have been long enough to block any export machinery.

Due to the reduced efficiency of protein expression during amber suppression, it is possible that this limited the production of functional flagellin and therefore not enough flagella were produced to restore motility to the non-motile cells. Higher concentrations of NAA could be utilised to investigate whether the presence of NAA was limiting protein production. However, previous work has shown that 2 mM NAA (as used in experiments) does not limit protein expression (Antonczak 2012 Thesis).

Salmonella typhimurium cells present one flagellum on their cell surface and have previously been used for bacterial motility analysis (Aizawa *et al.* 1990; Homma et al. 1987a; Homma et al. 1987b; Jones et al. 1989; Muramoto *et al.* 1999; Rosqvist *et al.* ; Yoshioka *et al.* 1995). If production of protein had limited the use of this application, using *Salmonella* would be beneficial because less NAA-containing flagellin would be required to restore motility.

The use of motility assay plate to determine restoration of motility requires that protein expression was high enough to produce functional flagella, and enough cells within the population were capable of creating a large enough response to move away from the inoculation point of the agar plate. Although motility plates were chosen as a simple assay for motility, the use of high magnification or electron microscopy could be used to visualise cells to ascertain ability to swim or the presence of flagella respectively. The use of microscopy to monitor motility of single cells would allow observation of slight movements if flagella production was low or motility limited.

Implementation of amber suppression to incorporate NAA into FliC required the presence of two plasmids within the cell. Both are required for the successful implementation of amber suppression. These were selected for using antibiotics. At each step of the motility assay, fresh antibiotics were used. However, after a certain time on the motility assay plate the cells could have lost either of the plasmids during replication and created satellite colonies, which would be unable to produce full-length flagellin (Sambrook *et al.* 1989). The presence of the pFDFliC239TAG plasmid was selected for using carbenicillin and although this antibiotic is more stable and less quickly broken down by β-lactamase enzyme (produced by the *E. coli* harbouring plasmid) than ampicillin, it is possible that secretion of β-lactamase into the surrounding agar could provide a habitat where *E. coli* not containing pFDFliC239TAG could grow. To circumvent this problem, a one plasmid system harbouring the target gene, the synthetase and tRNA would be beneficial. Although the formation of satellite colonies and loss of the plasmid could still occur, using a one plasmid system, it may be reduced. It also may be possible to reduce its occurrence by choosing a bactericidal antibiotic as a selection agent.

Production of a FliC-GFP fusion construct could have provided a mechanism to try to visualise the production of full-length protein. Although the presence of the GFP would probably perturb formation of functional flagella, it would allow simple visualisation of whether full-length NAA-containing protein was being expressed.

4.3.5.1 Improving amber suppression

The function of orthogonal aaRSs are important in achieving fidelity of amino acid incorporation, however these are just one of several components that contribute to the overall efficiency. In order to improve the efficiency of amber suppression, the effect certain factors involved in translation have on the efficiency of NAA incorporation have been considered. As mentioned in Chapter 1, release factor 1 (RF-1) is the only molecule that ultimately terminates protein synthesis due to recognition of the UAG codon, therefore this posed a possible target. Short 3rd et al (1999) examined the effect of RF-1 during incorporation of NAAs. The group demonstrated that the incorporation of NAA in response to UAG was increased when *in vitro* translation using *E. coli* extracts containing partially deactivated RF-1 was used. This effect was especially noted for NAA exhibiting low incorporation efficiency. Sakamoto and co-workers furthered this work by creating a conditional knockout of RF-1 in *E. coli* (Mukai *et al.* 2010a; Ohtake *et al.* 2012). Using this RF-1 deficient strain the group successfully reassigned the UAG stop codon to a universal sense codon. A similar approach by Johnson et al (2011) demonstrated that knockout of RF-1 allowed incorporation of NAA at multiple sites within a gene.

The work performed by Rackham and Chin (2005) and Wang et al (2007a) sought to develop an orthogonal ribosome that functions to incorporate NAA in response to UAG codons via orthogonal aaRS/tRNA systems which functions alongside the cognate ribosome/translation. Wang et al's orthogonal ribosome, ribo-X, was shown to increase the efficiency of site-specific NAA incorporation from ~20 % to >60 % using a single amber codon and from <1 % to >20 % for two amber codons within a gene. It was hypothesised that the decreased recognition of RF-1 by ribo-X accounted for the increase in NAA incorporation efficiency.

The elongation factor Tu (EF-Tu) exhibits reduced binding efficiency for tRNAs aminoacylated with large aromatic amino acids (Hohsaka *et al.* 1999; Nakata *et al.* 2006). This reduced binding efficiency suppresses the delivery of NAA-tRNA in to the ribosome and also shortens its lifetime due to spontaneous ester hydrolysis and this leads to inefficient incorporation into peptide. To address this inefficiency, Ohtsuki, Sisido and coworkers have created EF-Tu mutants that drastically

improved the efficiency of incorporation of large aromatic NAA for example, 1-pyrenylalanine and 2-anthraquinonylalanine (Doi *et al.* 2007; Ohtsuki *et al.* 2010). The E215A and D216 mutations appeared to enlarge the aa-tRNA binding pocket and remove unwanted contacts between EF-Tu and the large-NAA-tRNAs.

Using the improved systems outlined above, or investigating possible new targets within the translation machinery, the efficiency of NAA incorporation could be improved and may prove beneficial in the application discussed within this chapter.

4.4 Conclusion

The basis for creating a read-out system for controlling the motility of an *E. coli* cell has been created. Restoration of motility via plasmid-encoded FliC has been shown to be possible using the non-motile strain, Keio. However complexities regarding the implementation of amber suppression in order to control this motility via the presence/absence of NAA limited the functionality of this readout system. More detailed investigations into whether full-length NAA-containing protein was being expressed or functional flagella produced could aid in uncovering the specific issues with this application.

The use of another readout system, one that does not involve such complex protein export, quaternary structure or large amounts of NAA-containing protein could still be used to control cellular phenotype via presence of NAA using amber suppression.

5 CHAPTER 5. DEVELOPMENT OF A FUNCTIONAL MAMMALIAN CELL READ OUT SYSTEM

5.1 Introduction

As described in Chapter 4 using a prokaryotic system, the second novel application of amber suppression explored within this Thesis is the creation of a biological switch controllable by the presence of NAA. This, and the following chapter seek to develop this application using a eukaryotic system. The creation of a eukaryotic cellular readout system is described within this Chapter and the application of amber suppression technology utilising this readout system is described within Chapter 6.

The creation of a novel way to control eukaryotic phenotype has interesting applications. In theory this technology could be used to control the production of any protein and therefore provides control over display of the phenotype conferred by the target protein. This could include the production of a therapeutic protein, the levels of which could be tuned to provide optimal concentrations by varying NAA concentration or efficiency of aaRS charging or NAA incorporation. The ultimate goal would be to produce a self regulating mechanism where a negative feedback system can limit the over expression of target protein. This could be achieved by NAA-directed proteolysis which cleaves NAA-protein at high concentrations, releasing NAA for re-incorporation when protein levels are low. However, for a proof-of-principle application within this thesis, the modulation of eukaryotic motility was chosen.

For this novel application of the amber suppression technology, a mammalian cellular readout was required that could be easily assayed. Cellular motility was chosen as the specific phenotype to modulate as this can be easily and reproducibly monitored and assessed utilising live cell imaging. As part of the amber suppression control pathway Rac1 was chosen as the target gene as this gene product is critical in cellular signalling, culminating in production of lamellipodia and subsequent motility of the cells (Nobes and Hall 1995). As evidence of this, it has been reported that knockout of Rac1 results in reduced ability to produce membrane ruffles and lamellipodia and aberrant cell migration which is clearly discernible form the Rac1-wild type control (Guo et al. 2006; Monypenny et al. 2009; Vidali et al. 2006).

Therefore, the aim of this chapter was to develop this cellular readout system in controlling the expression of RAC1 and in turn control the phenotype of a mammalian cell.

Specifically, the investigations presented within this chapter sought to:

- Establish that there were differences in motility between the Rac1 KO cell line and Rac1 control fibroblasts.
- Investigate whether plasmid-encoded Rac1 could restore motility to Rac1 KO cells, via transient and stable transfection.

Firstly, the morphological characteristics and growth of the previously created Rac1 control and KO Rac1 fibroblast cell lines were evaluated. Utilising these cell lines, it was investigated whether scratch wound assays in conjunction with timelapse microscopy could be utilised to monitor the differences in migratory abilities of the two cell lines. The requirements of Rac1 for cellular migration was then investigated utilising these methods. Secondly, the re-introduction of Rac1 via plasmid was investigated. These investigations included optimisation of the transient transfection process utilising two GFP tagged Rac1 constructs, in conjunction with fluorescent microscopy. Stably expressing cell lines were also created utilising a third plasmid construct, harbouring a selection marker. The evaluation of motility restoration was investigated by quantitative analysis of scratch wound closure and individual cellular movement utilising fluorescent timelapse microscopy.

5.2 Materials and methods

5.2.1 Tissue culture plastic

All tissue culture plastic was obtained from Sarstedt. 24 well plates used for timelapse microscopy were obtained from BD Biosciences, Oxford.

5.2.2 Instruments

A Leica AF 6000 live cell imaging confocal laser scanning microscope was utilised for timelapse and fluorescent microscopy. The microscope was equipped with a BOX and CUBE2 temperature control system, a BRICK gas mixer (Life Imagine Services), a Leica CTR6500 electronics box (Leica) and an ebq100 powered UV box.

A BioDoc-It® Imaging System (UVP) was used to capture agarose gel images.

A Panasonic DMC-G1 camera was used to capture brightfield images.

5.2.3 Software

The Leica Application suite was used to capture images using the Leica AF 6000 live cell imaging confocal laser scanning microscope.

ImageJ was used to create and analyse images (Java 1.6.0_20, 64-bit, Wayne Rasband, National Institutes of Health, USA). Analysis included using the tools and plugins as indicated in the text.

Data collection and analysis was performed using Microsoft Excel.

5.2.4 Molecular biology

Molecular biology was performed as defined in Chapter 2. Details of plasmids can be found in Appendix A1. Details of Primers can be found in Appendix A2.

5.2.5 Mammalian Cell Culture

5.2.5.1 Cell lines

Rac1 knockout (Rac1 KO) and Rac1 control fibroblasts were a kind gift from Dr Klemens Rottner (Institute of Genetics, University of Bonn). Human Embryonic Kidney (HEK) 293 cells were a kind gift from Lea Bauer (Dr Vera Knauper's group, School of Dentistry, Cardiff University).

5.2.5.2 Media

All tissue culture media was obtained from Invitrogen. Various media were used as specified in Table 13. Media was prepared as needed, stored at 4 °C and pre-warmed using a water bath at 37 °C prior to use.

Table 13. Mammalian tissue culture media

Media name	Base	Additives	Use
Growth media	DMEM 41965 high glucose L-glutamine (580 mg / L) Phenol red	10 % (v/v) heat inactivated foetal bovine serum (FBS) 1 % (v/v) nonessential amino acids 1 mM sodium pyruvate	Routine culture of Rac1 KO, control fibroblasts and HEK293
Transfection media	DMEM 41965 high glucose L-glutamine Phenol red	10 % FBS 1 mM sodium pyruvate	At transfection
Selection media	DMEM 41965 high glucose L-glutamine Phenol red	10 % FBS 1 % nonessential amino acids 1 mM sodium pyruvate 100 µg/ml Zeocin	For selection of Rac1 KO stable colonies
Maintenance media for stably transfected cells	DMEM 41965 high glucose L-glutamine Phenol red	10 % FBS 1 % nonessential amino acids 1 mM sodium pyruvate 50 µg/ml Zeocin™	Maintenance media for stably transfected cells
Phenol red free	DMEM 21063 high glucose L-glutamine	10 % FBS 1 % nonessential amino acids 1 mM sodium pyruvate	Viewing GFP positive cells
Opti-MEM™	MEM 11058		Transfection complex formation
Freezing media	HI-FBS	10 % DMSO	Resuspension of cells for cryostorage

5.2.5.3 Recovery from cryostorage

The cells were recovered from cryostorage by thawing rapidly at 37 °C. The cells were added to approximately 10 ml of pre-warmed (37 °C) growth media and pelleted by centrifugation at 600 x *g* for 5 minutes (room temperature). The supernatant was discarded and the cells resuspended in approximately 10 ml growth media. A cell count and viability assay by trypan blue exclusion was performed in conjunction with a Neubauer counting chamber and 1 x 10^6 cells were added to a T75 flask. Cultures were incubated at 37 °C in a humidified 5 % CO_2/95 % air atmosphere.

5.2.5.4 Cell sub-culture

Every other day or at approximately 90% confluence, the cells were sub-cultured by decanting the media, washing in 10 ml Phosphate buffered saline (PBS) (which was subsequently removed) and then incubating the cells with 1 ml of Trypsin solution (0.05 % (w/v) trypsin, 0.53 mM EDTA) at 37 °C for 5 minutes. The cells were dislodged by gentle tapping and then suspended in 10 ml growth media. Cells were pelleted by centrifugation at 600 x *g* for 5 minutes (room temperature). A viable cell count was performed using trypan blue exclusion in conjunction with a Neubauer counting chamber. 1 x 10^6 cells were used to inoculate a sterile T75 flask containing approximately 10 ml growth media. Cultures were regularly tested for mycoplasma infection using the in-house service (School of Dentistry, Cardiff University). Only mycoplasma negative cells were utilised in experiments.

5.2.5.5 Preparation of cells for cryostorage

At intervals, aliquots of pelleted cells were resuspended in freezing media (Table 13) transferred into labelled cryovials and frozen by placing them in a Mr Frosty (Nalgene; containing isopropyl alcohol), which was then placed at -80 °C. After approximately 24 hours, the cryovials were moved to liquid nitrogen for long-term storage.

5.2.5.6 Cellular Transfection

Tissue culture multi-well plates (BD) were seeded with a specified cell number per well. During a typical transfection, a 24 well plate was used and seeded with 4×10^4 cells /well. The multi-well plates were then incubated overnight at 37 °C with 5 % CO_2 /95% air (humidified). The following day the media was removed, the wells washed with PBS and transfection media (Table 13) added.

Lipofectamine 2000 (Invitrogen) or FugeneHD (Roche, Promega) were utilised for transfection of cells. The reagents were stored at 4 °C and used following the manufacturers' protocols. Typically, Lipofectamine 2000 (2 : 1 (Lipofectamine : DNA ratio)) and DNA (1 µg / well) were independently diluted in OptiMEM media. After 5 minutes, the diluted Lipofectamine and DNA were mixed and incubated for 20 minutes to allow for the formation of complexes. The solution containing complexes was mixed by pipette and aliquots added to wells containing cells accordingly.

For a typical transfection using FugeneHD, DNA (1 µg / well) was added to Optimem, followed by Fugene reagent (typically 3 : 1 or 3.5 : 1 ratio Fugene : DNA). The solution was incubated for 5, 10 or 15 minutes and then this solution containing complexes was mixed by pipette and aliquots added to wells containing cells accordingly.

5.2.5.7 Creation of monolayer scratch wound

The *in vitro* scratch wound model assays were performed as described previously (Stephens *et al.* 2004). Twenty-four well plates were seeded with cells, as described above (4×10^4 / well), and allowed to reach 95 % confluency over a period of 24 hours. The monolayers were scratch wounded by scoring through the centre of the confluent monolayer with a 200 µl pipette tip either (a) immediately before transfection or (b) the day following transfection. Following scratch wounding the wells were washed twice with PBS and 1 ml growth media or 500 µl transfection media was added.

5.2.5.8 Stable selection

Zeocin™ antibiotic, a member of the glycopeptide bleomycin antibiotic family, was utilised for cell selection. It was defrosted on ice, diluted in growth media to the required concentration and added to the cells.

5.2.5.9 Creation of stable cell lines

To create stably transfected cell lines, 1 μg DNA per 2 cm^2 tissue culture plastic was used to transfect cells plated the previous day at an initial cell seeding of 4 x 10^4 per well of a 24 well plate. A DNA to Fugene ratio of 3.5 : 1 and a complex formation time of 5 minutes (room temperature) was utilised (as determined using results obtained during the chapter). Complexes were added to the well and were incubated overnight. The following day, the cells were removed from the plate by trypsinisation, diluted in growth medium and re-plated into, typically, 2 x 6 well plates. This was to obtain clones derived from one cell and allow easy picking of these clones. The following day, the growth medium was removed, the wells washed with PBS and selection media added (Growth media + 100 μg / ml Zeocin; Table 13). The selection media was removed as required, typically every other day, the wells washed with PBS twice and fresh selection media added. This was repeated until identifiable individual colonies were growing, this was typically approximately 1 - 2 weeks. The plates were then viewed using fluorescent microscopy and the GFP-positive colonies marked.

5.2.5.9.1 Picking GFP-positive stably transfected colonies

After the GFP-positive colonies were marked, the media was removed from the well, followed by PBS wash. Cloning rings were created using cut and autoclaved 200 μL tips. These cloning rings were placed over the marked colonies using high vacuum grease (Dow Corning) to create a seal. Trypsin (0.05 %; 50 – 100 μl) was pipetted into the cloning ring and left on the cells until the cells were rounded and starting to come off the well plate (approximately 5 minutes). The cells were resuspended in the trypsin, then pipetted into clean wells of a 24 well plate containing fresh stable colony maintenance medium (Table 13). The cloning ring was washed with medium to collect any remaining cells. The maintenance medium

was replaced as necessary and the colonies were expanded. Aliquots were prepared for cryostorage as described above and cells were used in subsequent experiments. Cells were monitored for GFP-positivity throughout growth, using fluorescence microscopy.

5.2.5.10 Live Cell Imaging

A Leica AF6000 microscope was used to monitor the scratch wounds in real time. The inverted microscope was held within a chamber that allowed conditions to be controlled for cell growth. The chamber was pre-warmed to, and maintained throughout experiments at, 37 °C, with 5 % CO_2 /95 % air (humidified). A 10 x magnification was used throughout timelapse.

Typically, positions in each well were chosen, and images taken every 15 minutes for 48 hours. For transfected cells, both brightfield and GFP fluorescence was monitored.

5.2.5.11 RNA extraction

All steps using RNA were performed at a designated "RNA only" lab bench, using "RNA only" reagents and plasticware. In between each step, the RNA-containing solutions were kept on ice. Centrifugation steps were performed at 4 °C.

Cells for RNA extraction were disrupted using TRIzol® Reagent (Invitrogen) typically with 1 ml TRIzol per approximately 1 x 10^6 cells. After incubation (approximately 5 minutes), the cells were scraped from the tissue culture plastic and the suspension collected into an eppendorf and frozen at -80 °C overnight or until required. The samples were defrosted and phase separation was performed by adding 200 µl chloroform per 1 ml TRIzol®, the solutions were mixed by inverting the tube several times and incubated for 3 minutes (RT). A phase lock gel tube (PLG; Eppendorf) was pelleted by microcentrifugation at 12,000 x *g* for 30 seconds. The TRIzol®/chloroform mixture was added to the PLG, mixed by inverting and then subjected to centrifugation at 12,000 x *g* for 5 minutes. The RNA-containing aqueous upper phase was removed by pipette into a fresh tube. 500 µl isopropanol was added to the RNA-containing aqueous phase, mixed by inverting and left at -20 °C overnight for the RNA to precipitate. The RNA was pelleted by centrifugation

at 12,000 × *g* for 10 minutes, the supernatant was decanted and the pellet washed using ice-cold 75 % ethanol, flicking the tube to remove pelleted RNA from the tube wall. RNA was re-pelleted by centrifugation at 7500 × *g* for 5 minutes. The ethanol was removed and the pellet left to air dry for < 1 hour. The RNA was resuspended in 50 µl Diethylpyrocarbonate (DEPC) treated H_2O (ddH_2O plus 0.1 % DEPC left overnight and autoclaved).

5.2.5.11.1 DNase treatment

Extracted RNA was subjected to DNase treatment (Promega). A typical 50 µl reaction contained 1 x buffer, 40 U RNAsin inhibitor (Promega) and 1 U DNase. The reactions were incubated at 37 °C for 15 minutes. The RNA was then isolated from the reaction using an RNeasy Minelute cleanup kit (Qiagen), utilising the proprietary buffers provided. Briefly, the samples were made up to 100 µl with DEPC water, 350 µl Buffer RLT (containing guanidine isothiocyanate) supplemented with 10 µl β-mercaptoethanol per 1 ml Buffer RLT was added to the sample; this promoted the selective binding of RNA to a silica-membrane RNeasy MinElute membrane. 250 µl ethanol was added to the diluted RNA and mixed. RNA-mixture was then applied to a RNeasy MinElute membrane by centrifugation at ≥8000 x *g* for 30 seconds. Contaminants were removed by washing the column twice with 500 µl Buffer RPE and RNA was eluted in 20 µl DEPC water.

5.2.5.11.2 Quantification of RNA

RNA concentration was measured using a NanoVue (GE Healthcare). 260/280 ratios were monitored for purity. RNA with values above 2 was used.

5.2.5.12 Reverse transcription polymerase chain reaction (RT-PCR)

5.2.5.12.1 cDNA synthesis

First-strand cDNA synthesis was performed using recombinant Moloney murine leukaemia virus reverse transcriptase (MMLV-RT; Promega), random hexamer primers (Promega) and total RNA prepared from cells (as described above). A typical reaction contained 0.5 µg RNA, 0.5 µl random hexamer primers (0.5 µg/µl) and DEPC water up to a volume of 6.5 µl. Reactions were incubated at 70 °C for

5 minutes, then ice for 5 minutes. 20 U RNAsin, 1 x MMLV buffer (50 mM Tris-HCl (pH 8.3 @ 25 °C), 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT), 2 mM dNTPs and 100 U MMLV RT were then added and a thermocycler used to cycle the reactions; 25 °C for 10 minutes, 42 °C for 60 minutes and 95 °C for 5 minutes.

5.2.5.12.2 PCR amplification

PCR reactions were performed using Platinum® Blue PCR Supermix (Invitrogen; 22 U/mL complexed recombinant *Taq* DNA polymerase with Platinum® *Taq* Antibody, 22 mM Tris-HCl (pH 8.4), 55 mM KCl, 1.65 mM $MgCl_2$, 220 µM dGTP, 220 µM dATP, 220 µM dTTP, 220 µM dCTP, and stabilisers). Amplification of β-actin, using primers β-actin forward and reverse, was used as a housekeeping gene control, to check to integrity of the cDNA. A typical reaction contained 20 µl Supermix, 200 nM forward primer, 200 nM reverse primer and 1 µl cDNA per reaction. All PCR reactions were set up on ice. To screen for Rac1 gene, primers racFP and racRP3 were used; to screen for azRS, primers EcTyrRS FP and EcTyrRS RP2 were used; to screen for Zeocin resistance gene primers zeoFP and zeoRP were used (see Appendix A2). A thermocycler (UnoCycler, VWR) was used to cycle between temperatures as indicated in (Table 14). Temperatures used for annealing were optimised for each primer combination.

Agarose gel electrophoresis, as described in Chapter 2, was used to visualise PCR products. Molecular weight was approximated by comparison to 100 bp ladder (Promega).

Table 14. PCR parameters for Platinum PCR supermix

Cycling step	Temperature	Time	# of cycles
Initial denaturation	94 °C	3 min	1
Denaturation	94 °C	20 s	30-40
Annealing	β-actin 60 °C Rac 1 60 °C azRS 56 °C zeo 55 °C	20 s	
Extension	72 °C	30 – 60 s	
Final extension	72 °C	5 - 10 min	1
Hold	4 °C	Indefinitely	1

5.2.6 Image Analysis

5.2.6.1 Creation of merged image

The brightfield channel and GFP fluorescence channel images of static and timelapse images were imported into ImageJ (Rasband 1997). The background of the GFP fluorescence channel was reduced using the *Process…Reduce Background* tool of ImageJ and then overlaid with the brightfield channel using the *Image…Colour…Merge Channels* tool.

5.2.6.2 Calculation of transfection efficiency

Approximately 24 hours post-transfection, a bright field and UV-fluorescence image was taken of the cells. Transfection efficiency was calculated by counting the number of total number of GFP positive cells within the frame of the image and the percentage transfection efficiency calculated by dividing this GFP positive cell count by the total number of cells per frame. Cell counting was performed using the *analyze…cell counter* tool in ImageJ. Where the number of cells exceeded a number that could be accurately counted, transfection efficiency was reported as number of GFP positive cells per frame.

5.2.6.3 Quantification of wound closure

Quantification of wound closure was obtained by measuring the remaining wound area over the time course of the wound experiment by drawing a line following the wound edge and using the *Measure* tool to calculate the area (ImageJ). The wound area was measured at specified timepoints throughout the 48 hour period and plotted. The regression line was used to calculate rate of wound closure. For those curves that followed a sigmoidal curve, the rate was obtained using those points that followed a straight line, the fastest rate for an experiment was always taken. This typically was the middle section of the graph. The R^2 value was used to establish whether the portions of the graph used were accurately fitted. An R^2 value above 0.9 was accepted.

5.2.6.4 Quantification of the total distance moved and directionality of cells

Quantification of the total distance, speed and direction moved by individual GFP positive cells was performed using the *Manual Tracking* plug-in of ImageJ. The position of individual cells was monitored throughout the course of the timelapse experiment, plotting the position of the cell at every frame (i.e. every 15 or 20 minutes). The *Measure* tool was then used to measure the total distance for the cell.

The directionality of the cell movement was visualised by plotting the cells' trajectories, as measured using the *Chemotaxis and Migration* plug-in of ImageJ, on a circular graph, where the initial position of the cell is taken as position 0 , 0. Using these data, the *Chemotaxis and Migration* plug-in was used to calculate average directionality of cells, as defined in Figure 63.

5.2.7 Statistical analysis

Statistical analyses were undertaken using Microsoft Excel. Data were compared using a Student's t-test, and a one-way ANOVA used for group analysis. Results were expressed as a mean average and standard deviation (S.D.). All p values were two-tailed. Statistical significance was considered at a probability of $p \leq 0.05$. To test that data was normally distributed and therefore could be subjected to these statistical tests, histograms were plotted using Excel.

$$Directionality = \frac{euclidean\ distance}{accumulated\ distance}$$

$$Directionality \rightarrow 1 : straight\ motion$$

$$Directionality \rightarrow 0 : non - straight\ motion$$

$$average\ Directionality = \frac{1}{n}\sum_{i=1}^{n} Directionality_i$$

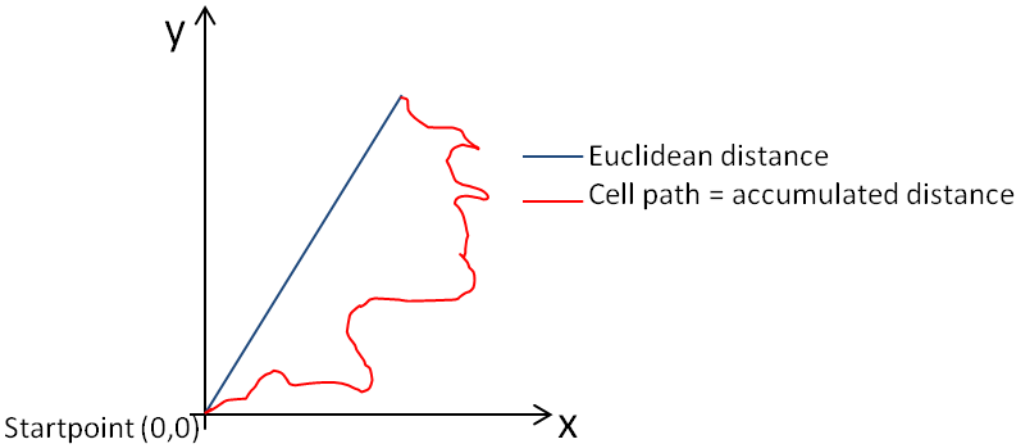


Figure 63. Calculation of directionality

5.3 Results

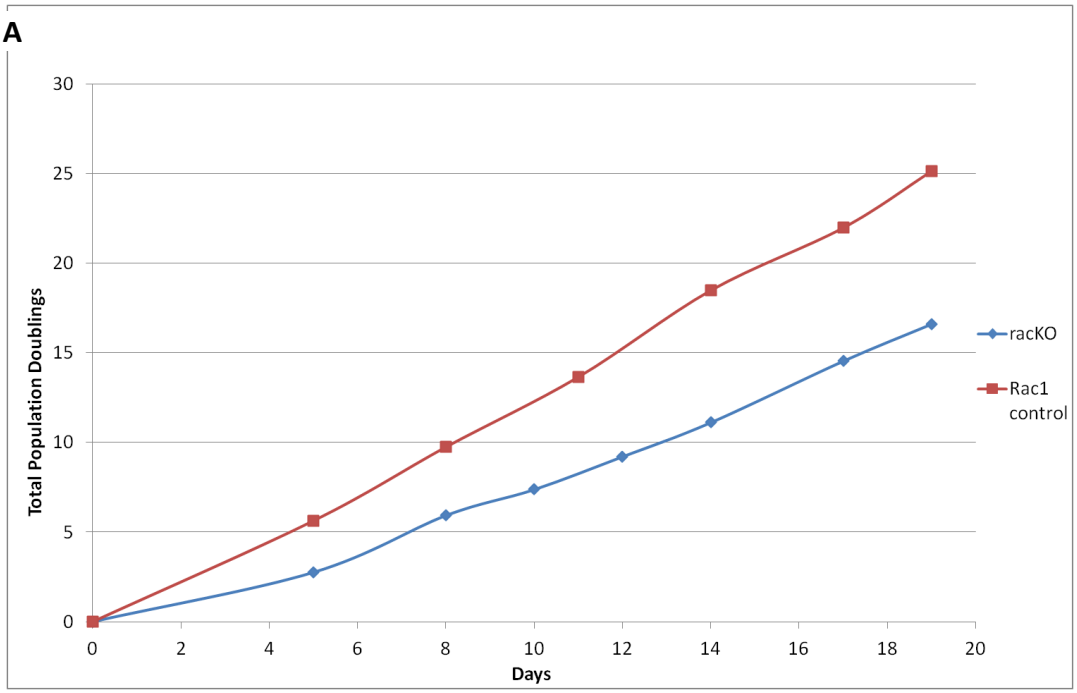
5.3.1 Analysis of control and Rac1 KO cell lines

5.3.1.1 Loss of Rac1 alters cell morphology and proliferation

For the development and testing of the non natural amino acid technological application, a cell-based instant readout system was required; namely the ability to restore cell migration. Specifically, a cell line was selected which had a gene knocked-out that was crucial for its motility which could then be re-introduced, encoded on a plasmid, to restore wild-type-like motility. The mouse fibroblast cell line, Rac 1A pEGFP-N3_puro was utilised as a control (referred to as Rac1 control cells in subsequent text) and a Rac1 knockout (Rac1 KO) as the test cell population. Both cell lines were a kind gift from Klemens Rottner, University of Bonn.

The two cell lines were regularly subcultured and population doublings determined in order to monitor the growth of these cells. Whilst cell proliferation was observed for both cell lines, it was clear that the Rac1 control cells grew faster than the Rac1 KO cell line (Figure 64A)

Typically, the control cells were an evenly distributed size and shape, whereas the Rac1 KO cells were more varied (Figure 64B). The control cell line was easily subcultured, and could be split to a low confluency. The cells were able to become an evenly spread confluent monolayer of cells. Unlike the control cells, the Rac1 KO cells tended to be more extended and did not form an evenly spread confluent monolayer of cells if the culture was initially seeded at low confluency. Once divided, the daughter cells tended to remain attached to the parental cell. This squashing could have caused some cell abnormalities or cell stress, therefore the Rac1 KO cell line was passaged every other day, maintaining a high seeding density, to reduce this effect.



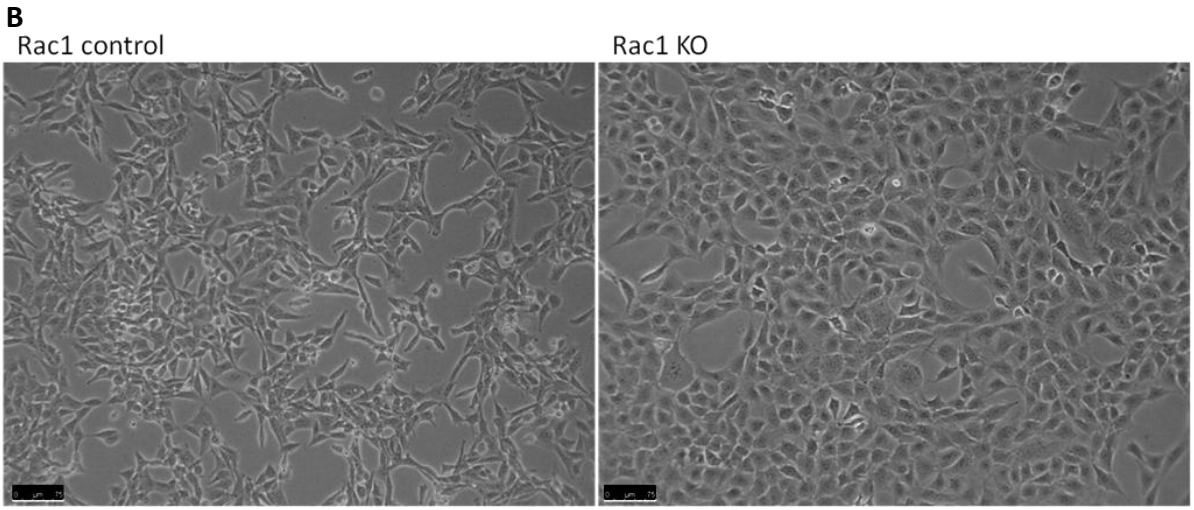


Figure 64. **(A) Population doubling levels of control and Rac1 KO fibroblasts**. Cultures were routinely passaged and a cell count performed. Population doubling was calculated and this was plotted against time.

(B) Brightfield images of Rac1 control cells and Rac1 KO cells. Scale bar represents 75 µm.

5.3.1.2 Optimisation of scratch wound assay

Wells of a 24-well plate were inoculated with differing numbers of Rac1 KO cells using serial dilution ($8x10^4$, $4x10^4$, $2x10^4$, $1x10^4$). One day post-inoculation the plates were scratched and imaged to ascertain the confluency and therefore the suitability for scratch assay (Table 15, Figure 65). For a successful scratch assay, it was shown that inoculating a well at $4x10^4$ gave a cell monolayer at 80 % confluency one-day post-inoculation. Scratching at this density provided optimal cell density to create a straight and complete scratch wound edge suitable for monitoring cellular motility via the rate of scratch wound repopulation.

5.3.1.3 Normal cell migration is dependent on the expression of Rac1

As the cell-based readout system required for future studies was centred on modulation of cell migration, firstly it was important to confirm that the Rac1 KO cells were in fact deficient in their ability to migrate.

Rac1 control mouse fibroblasts were grown to approximately 80 % confluency, a scratch wound created and monitored by time-lapse microscopy (Figure 66, Movie 1). 12 hours post wounding there was clear evidence of cell migration into the wound space. Over the following 24 hours the remaining wound space was progressively diminished due to further cell migration and proliferation such that it was closed by 36 hours. In contrast, for wounds made within the monolayer of Rac1 KO cells there was no evidence of independent cell migration (i.e. cells breaking free of the bulk cell population) over the entire time course of the experiment (Figure 66, Movie 2). Furthermore, at all time points analysed, it was clear that the Rac1 KO wounds were much larger than the control wounds such that by 48 hours the Rac1 KO wounds had still not closed. Some diminution of wound size was evident within Rac1 KO wounds presumably due to the contributions of cell proliferation at the edges of the wound.

Table 15. Optimisation of scratch wound assay.

Cell number at inoculation	Confluency one-day post-inoculation	Suitability for scratch assay	Scratch assay outcome
$8x10^4$	100 %	Cells already overgrown	Cells behind leading edge get very overconfluent
$4x10^4$	80 %	Good confluency	Leading edge of wound is straight and complete and cell migration represents that of wound repopulation
$2x10^4$	50 %	Sub-confluent	Scratch edge is not quite complete, resulting in uneven edge of wound
$1x10^4$	20 %	Cells too sparse to form even scratch	Scratch edge is not complete, cells grow as colonies instead of migrating into wound area

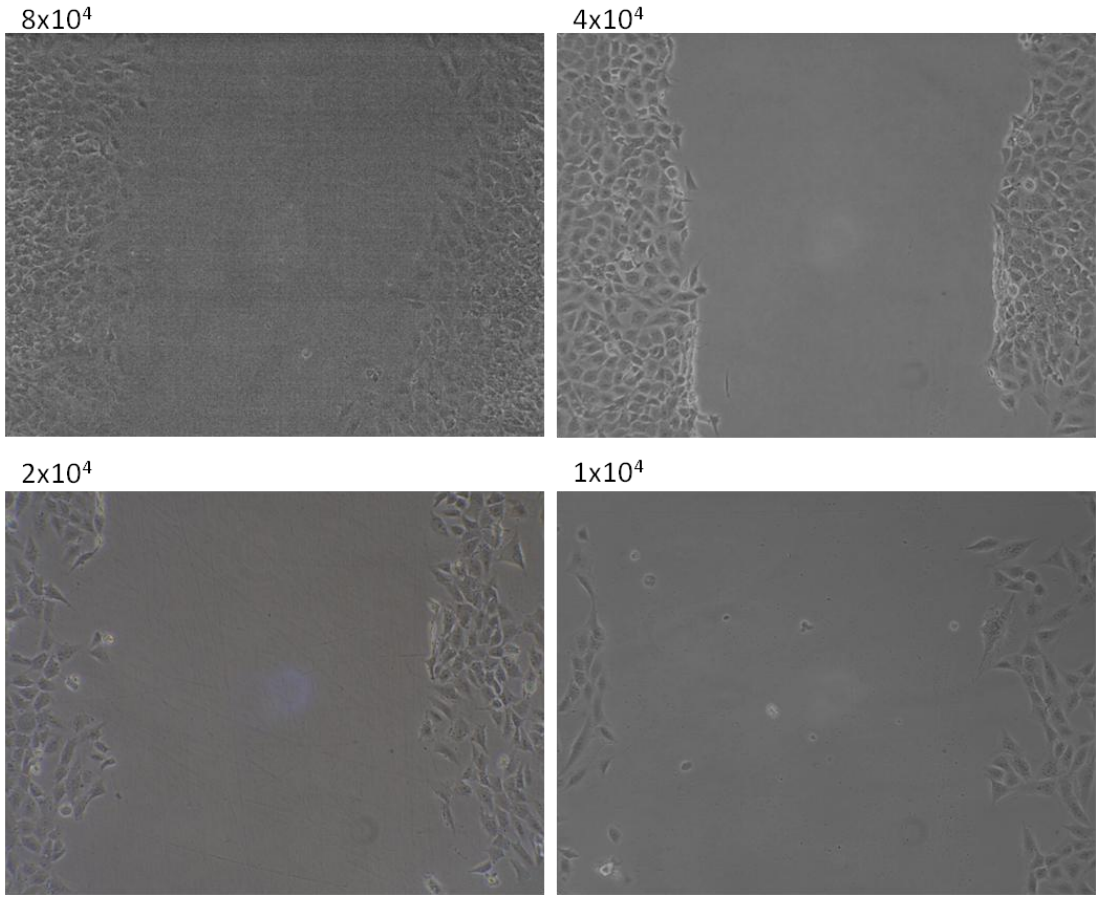


Figure 65 and Table 15. Optimisation of scratch wound assay. The number of cells used to inoculate wells of a 24-well plate was varied as indicated. One day post-inoculation the plates were scratched and imaged to ascertain the confluency and therefore the suitability for scratch assay.

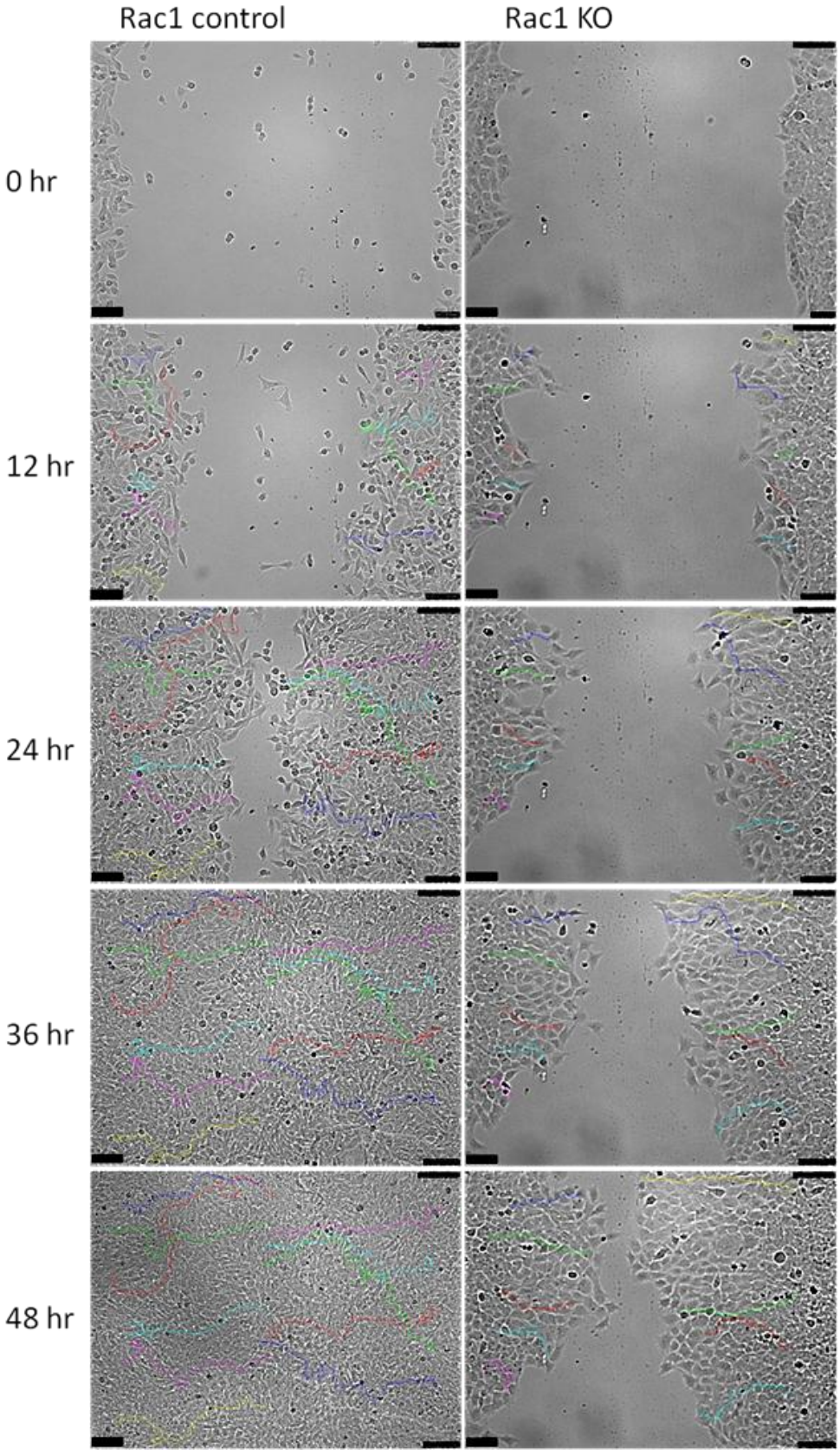


Figure 66. Brightfield photos of scratch wounds created in Rac1 control cells and Rac1 KO cells, at indicated time points. Coloured lines represent tracks moved by individual cells. Scale bar (lower left corner) represents 75 µm.

5.3.1.4 Quantitative analysis of wound closure

Quantitative analysis of the differences in wound closure was undertaken from time-related wound area measurements. The data demonstrated that the rate of wound closure was significantly faster for scratch wounds created in Rac1 control cells than in Rac1 KO cells ($p < 0.0001$; Figure 67A).

5.3.1.5 Quantitative analysis of motility of individual cells

Furthermore, to determine deficiencies in cell migration at the cellular level, the *Manual Tracking* tool along with the *Chemotaxis and Migration* tool (ImageJ) was used to analyse the migration of individual cells throughout the 48 hour timelapse period. The analysis demonstrated that the control cells migrated further than the Rac1 KO cells ($p<0.0001$; Figure 67B).

The trajectories of a number of selected cells throughout the 48 hour timelapse microscopy period were plotted (Figure 68). The Rac1 control fibroblasts clearly demonstrated tracks typical of directional movement (away from an overlaid common starting point) into the wound area post scratch. Conversely, the Rac1 KO cell line demonstrated significantly reduced movement as evidenced by their limited track lengths. However, the directionality of the movement shown by the control cells and Rac1 KO was similar (Figure 67C).

These results clearly demonstrate that knockout of the Rac1 gene has a detrimental effect on cellular migration and thus such a cell line is a suitable system for the further investigations within the Thesis.

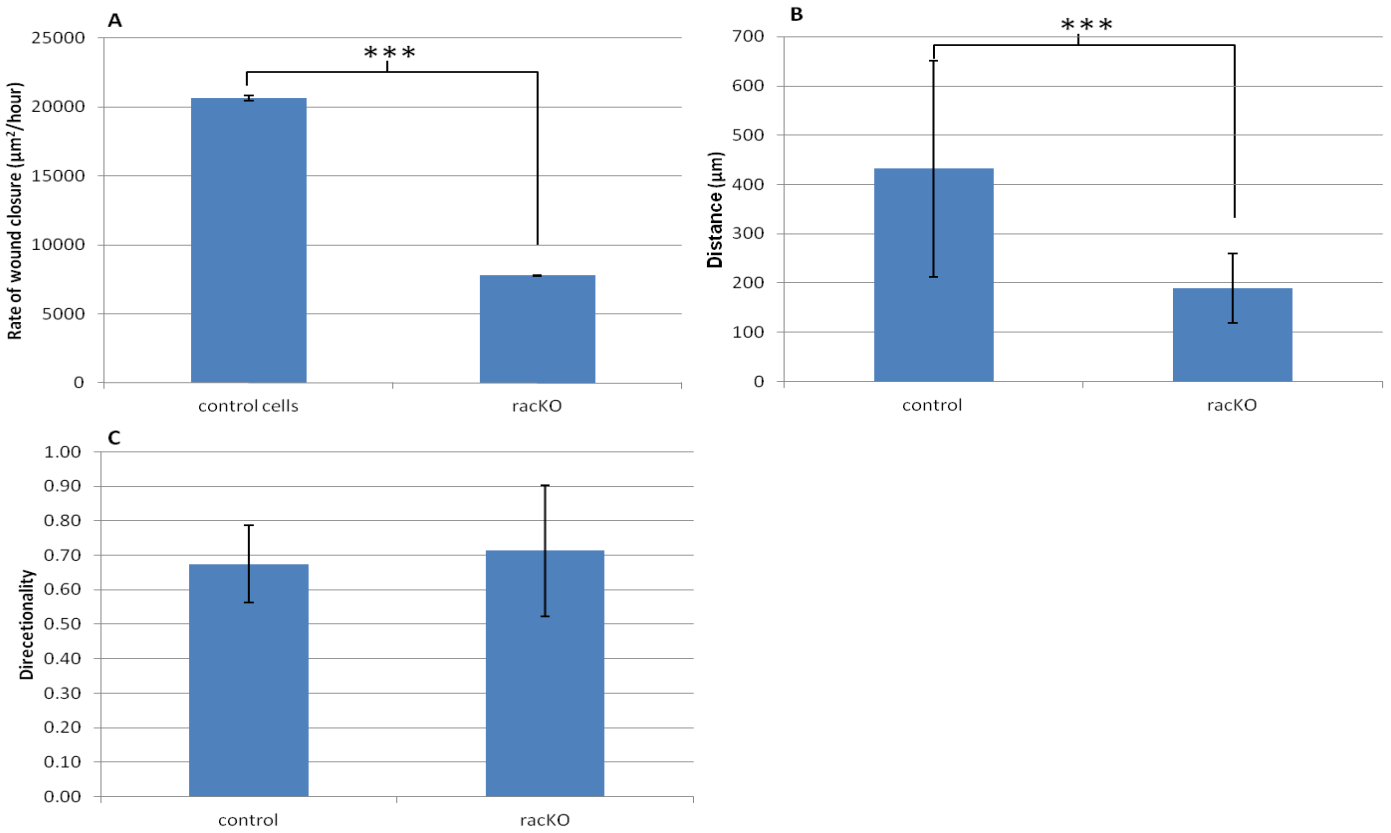


Figure 67. Graphs showing (A) averages of rate of wound closure (*n*=3), (B) distance travelled by individual cells (*n*=35), (C) directionality of individual cells (*n*=35). Scratch wounds were created in monolayers of either Rac1 control or KO cells. Timelapse microscopy was used to monitor the cells over a 48 hour period. Analysis of motility of individual cells was calculated using the *Manual Tracking* and *Chemotaxis and Migration* tool (ImageJ). Mean average values for each condition was calculated and plotted. Errors bars represent standard deviation.

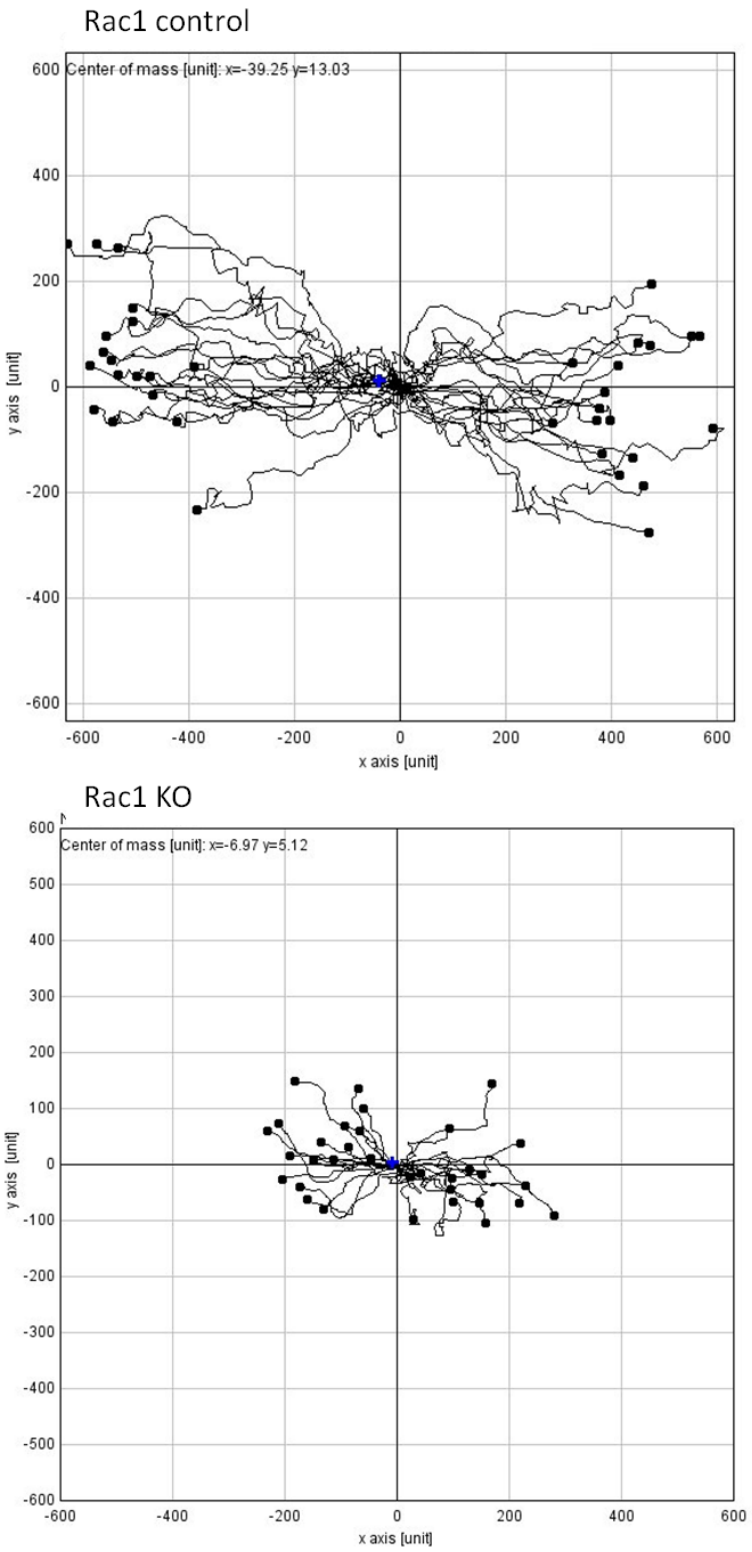


Figure 68. Trajectories of Rac1 Control and Rac1 KO mouse fibroblasts during scratch wound healing. Scratch wound assays were performed in monolayers of Rac1 control and Rac1 KO cells and monitored by timelapse microscopy. The trajectories of individual cells at the wound edge were followed using the *manual tracking* plugin and plotted using the *Chemotaxis and Migration* plugin of ImageJ. All cell tracks were overlaid onto a common starting point for clarity.

5.3.2 Restoration of motility to Rac1 KO cells

5.3.2.1 Plasmid constructs for transfection in to Rac1 KO cells

The next aim of the investigation was to restore motility to Rac1 KO cells. To do this plasmid constructs encoding Rac1 were required. The previously created peGFP-Rac1-C2 plasmid (based on the commercially available plasmid peGFP-C2 (Clontech); referred to as peGFPRac1 herein) was kindly donated by Prof. Klemens Rottner, University of Bonn. The plasmid encoded a GFP-Rac1 fusion protein under the control of a strong constitutive CMV promoter, and terminating with a SV40 polyA (Figure 70). The GFP tag was used to visualise the expression of the target protein.

5.3.2.2 Creation of peGFP plasmid for transfection into Rac1 KO cells

The plasmid peGFP was required as a control to complement experiments using peGFPRac1. Transfecting the cells with a plasmid not containing Rac1 (i.e. the peGFP) was an important control to (a) demonstrate transfection had successfully occurred and (b) that the GFP itself was not responsible for any alteration in cell motility.

Plasmid peGFP was created by digesting peGFPRac1 with the restriction enzyme *Eco*RI-HF, sites of which flanked the inserted Rac1 gene (Figure 70). The band corresponding to peGFP (4.7 kb) was extracted from a preparative agarose gel, the DNA purified and re-ligated. Comparison of prepared DNA on an analytical agarose gel suggested that peGFP had been isolated (Figure 69). The putative peGFP plasmid was smaller (4.7 Kb) than peGFPRac1 (5.3 Kb). Sequencing confirmed the isolation of peGFP plasmid.

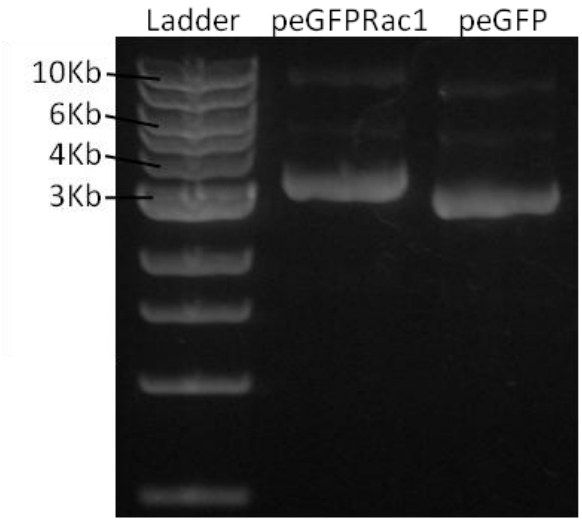


Figure 69. peGFPRac1 (5.3 Kb) and peGFP (4.7 Kb) plasmid DNA was isolated and analysed by agarose gel electrophoresis. The three bands reflect the three forms of DNA, supercoiled, nicked, relaxed (bottom to top). MW ladder represents 1 Kb ladder (NEB).

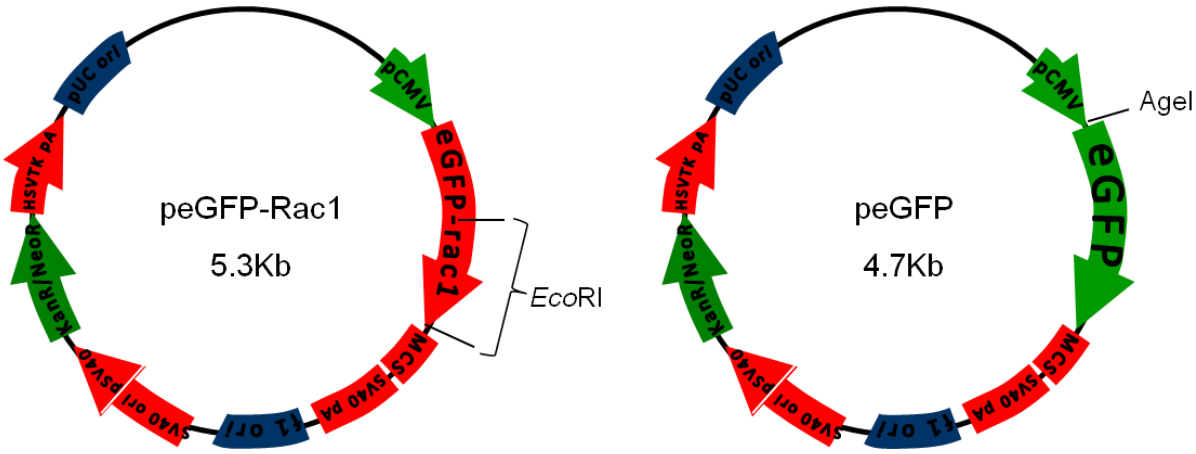


Figure 70. Plasmid maps of peGFPRac1 and peGFP.

5.3.2.3 Introduction of plasmid-encoded Rac1 to Rac1 KO cells

Experiments were performed to transiently transfect the Rac1 KO cell line with the plasmids peGFP and peGFPRac1. In order to maximise any potential restoration of migration it was decided that a high transfection efficiency would be needed. Therefore, two commercially available transfection reagents (Lipofectamine2000 [Invitrogen] and Fugene HD [Roche]) were utilised for optimisation of the transfection procedure.

5.3.2.3.1 Optimisation of transfection conditions – Lipofectamine 2000

Variables investigated included the amount of DNA added per well, the ratio of DNA : reagent and the cell confluency at transfection (Dalby *et al.* 2004). The data demonstrated that using a DNA : Lipofectamine ratio 1 : 2, adding 1 µg DNA per well containing cells seeded at 6×10^4 / well of 24 well plate, gave the highest number of GFP positive cells (Figure 71). Seeding the cells at 6×10^4 gave, not only the highest number of green cells, but also gave a useful cell confluency to perform the scratch assays approximately 24 hours post transfection (Figure 72).

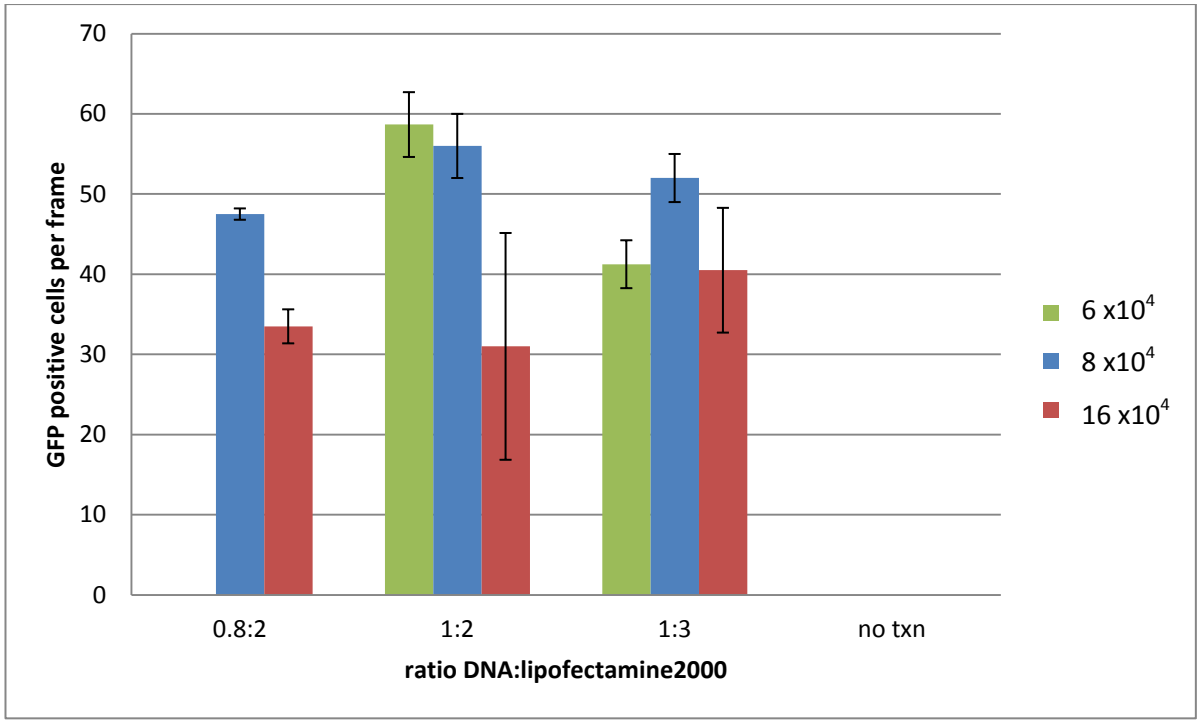


Figure 71. Graph comparing the transfection efficiency of transfecting Rac1 KO cells with the plasmid peGFPRac1 using Lipofectamine 2000 at varying ratios, 2 different DNA amounts and cell number per well. Rac1 KO cells were seeded in a 24 well plate at either 6, 8 or 16 x 10^4/well. A day later, the cells were transiently transfected with the plasmid peGFPRac1 using Lipofectamine. A day post transfection fluorescence microscopy was used to obtain images of the plates. The number of GFP positive cells per frame was counted and plotted. Error bars represent standard deviation of the mean of *n*= 3 repeats.

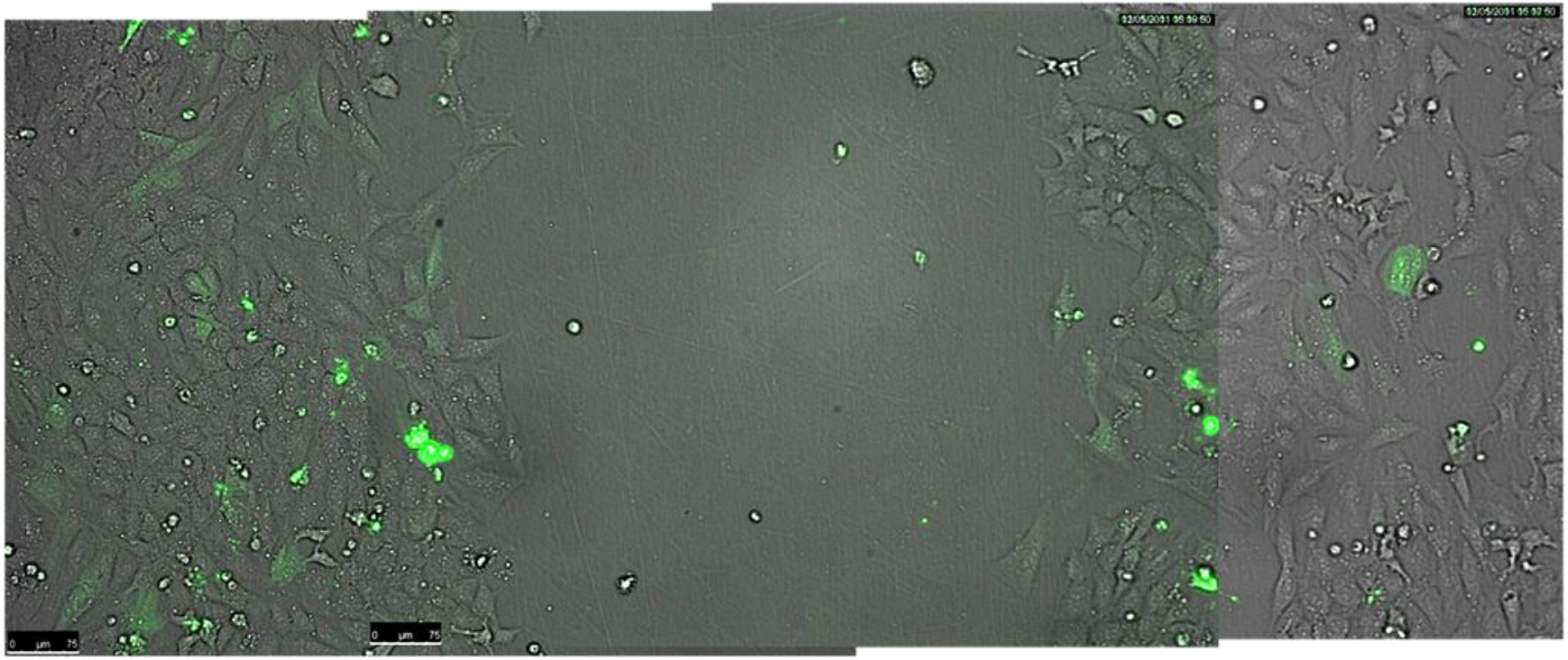


Figure 72. Composite brightfield/fluorescent image showing representative GFP fluorescence at scratch wound area. Rac1 KO cells were seeded at 6 x 10^4 / well 1 day before transfection, cells were transfected with 1 µg peGFPRac1 plasmid using 2 µl Lipofectamine. 3 frames / well were taken, using brightfield and fluoroscent microscopy, and overlaid. Scale bar represents 75 µm.

5.3.2.3.2 Optimisation of transfection conditions – Fugene

The formation of Fugene / DNA complexes was investigated. The size of complex formed, and therefore the ability to penetrate the cell and effectiveness of DNA release into the cell may depend on the ratio and the time the complex was allowed to form. Therefore, these variables were investigated. It was shown that transfection efficiency increased as FugeneHD : DNA ratio increased up to a maximum transfection efficiency seen with a 3.5 : 1 Fugene : DNA ratio. At a ratio of 4 : 1, the transfection efficiency was seen to decrease. Ratios lower than 3 : 1 appear ineffective at transfecting the fibroblast cells. It was shown that the optimal time of complex formation using a ratio 3.5 : 1 was 5 minutes.

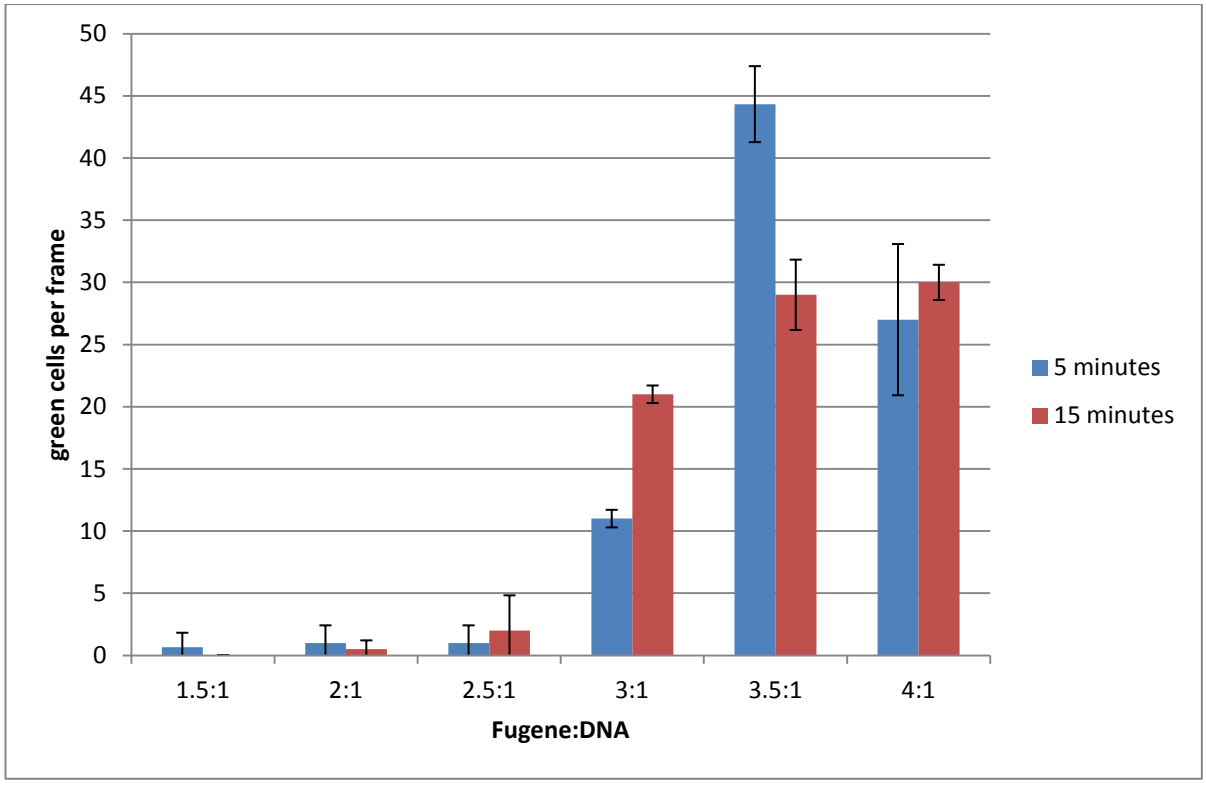


Figure 73. Graph comparing transfection efficiency of complexes formed using different Fugene:DNA ratios and complex formation times. Rac1 KO cells were seeded in a 24 well plate at either 4 x 10^4/well. A day later, the cells were transiently transfected with the plasmid peGFPRac1 using FugeneHD. A day post transfection fluorescence microscopy was used to obtain images of the plates. The number of GFP positive cells per frame was counted and plotted. Error bars represent standard deviation of the mean of *n*=3 repeats.

5.3.2.3.3 Comparison of Lipofectamine and Fugene transfection

Transfections performed using FugeneHD were compared to those using Lipofectamine. The previously optimised conditions for Lipofectamine 2000, namely 1 µg DNA / well, ratio 1 : 2 (DNA : Lipofectamine 2000), were used. Fugene was used at ratio 3.5 : 1, with 5 minute complex formation time. Brightfield and fluorescent microscopy was utilised to analyse the number of GFP positive cells (Figure 74). These data showed that although the Lipofectamine transfection gave higher number of GFP positive cells initially ($p < 0.01$), the number had decreased dramatically two days post transfection. Transfections performed using Fugene gave lower initial transfection efficiencies, and although the number of GFP positive cells decreased after two days post transfection, this reduction was smaller compared to the reduction seen using Lipofectamine 2000. After three days post-transfection, the number of GFP-positive cells for transfection using Lipofectamine 2000 had again reduced drastically. However, the number of GFP positive cells was maintained for transfections using FugeneHD. Another issue regarding comparison of transfections using Lipofectamine 2000 versus Fugene was that visual analysis of the photos indicated that Lipofectamine conferred toxicity. Therefore, FugeneHD was selected for further investigations to improve transfection efficiencies.

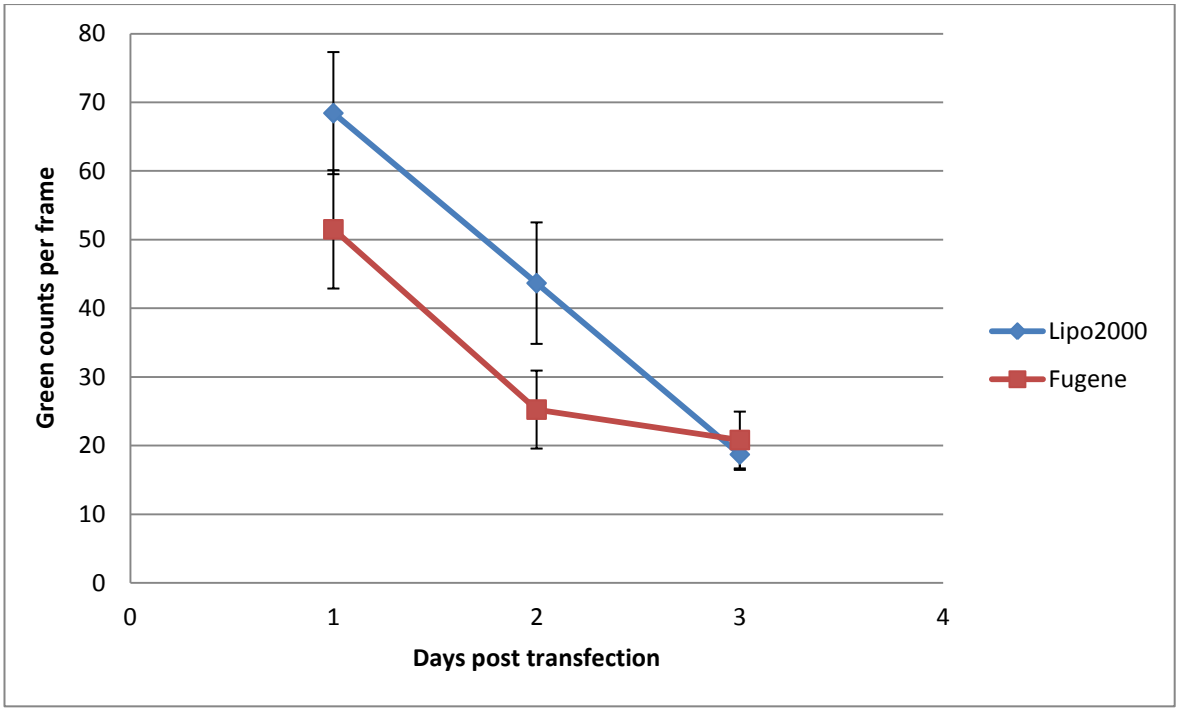


Figure 74. **Graph comparing transfection efficiency using either Lipofectamine or Fugene.** Cells were transfected with plasmid 1 µg peGFPRac using either Lipofectamine 2000 (Lipo2000) or Fugene HD (Fugene) using optimised protocols, as indicated in materials and methods. Brightfield and fluorescent microscopy was used to capture images of the cells at the indicated timepoints, the number of green cells per frame was counted and plotted. Error bars represent standard deviation of the mean of *n*=6 measurements

5.3.2.4 Optimisation of scratch wound assay

It was hypothesised that transfection efficiency could be increased by scratching the cell monolayer prior to transfection. It was thought that the scratch wounding process would increase the permeability of the cells, particularly those at the scratch edge (Abbaci *et al.* 2008; Geddes *et al.* 2003). It was demonstrated that scratch wounding the cell monolayer before transfection did increase the transfection efficiency (Figure 75, Figure 76). This effect was seen for both transfections with Lipofectamine 2000 and FugeneHD. There was significantly more GFP-positive cells in Rac1 KO cells scratched before transfection with peGFPRac1 ($p < 0.005$). In addition, there was significantly more GFP positive cells when transfected with peGFPRac using Fugene compared to Lipofectamine 2000 ($p < 0.01$).

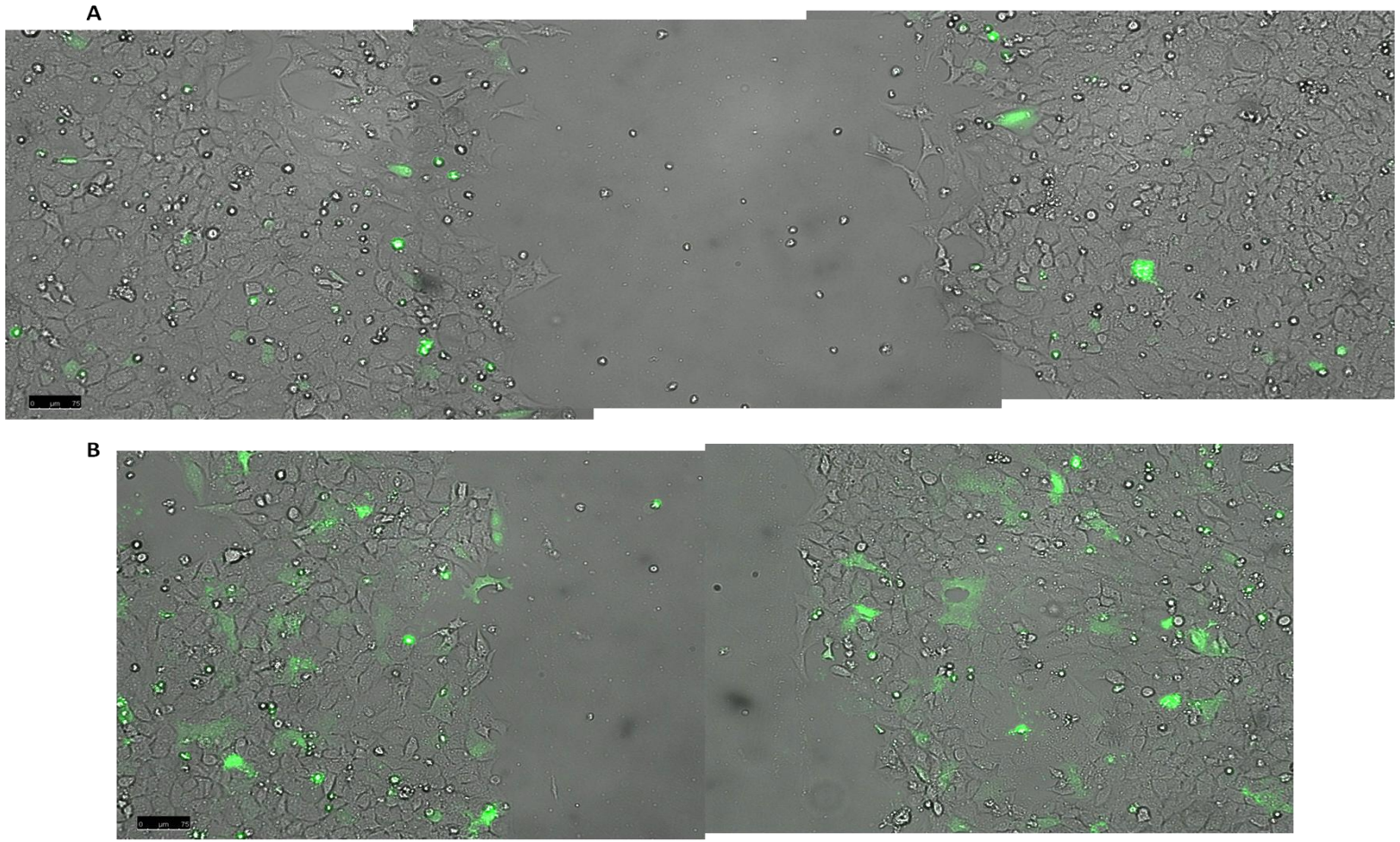


Figure 75. Composite images comparing transfection efficiency of transfections using FugeneHD, when the cells were transfected then scratched (A) or scratched then transfected (B). Scale bar represents 75 µm.

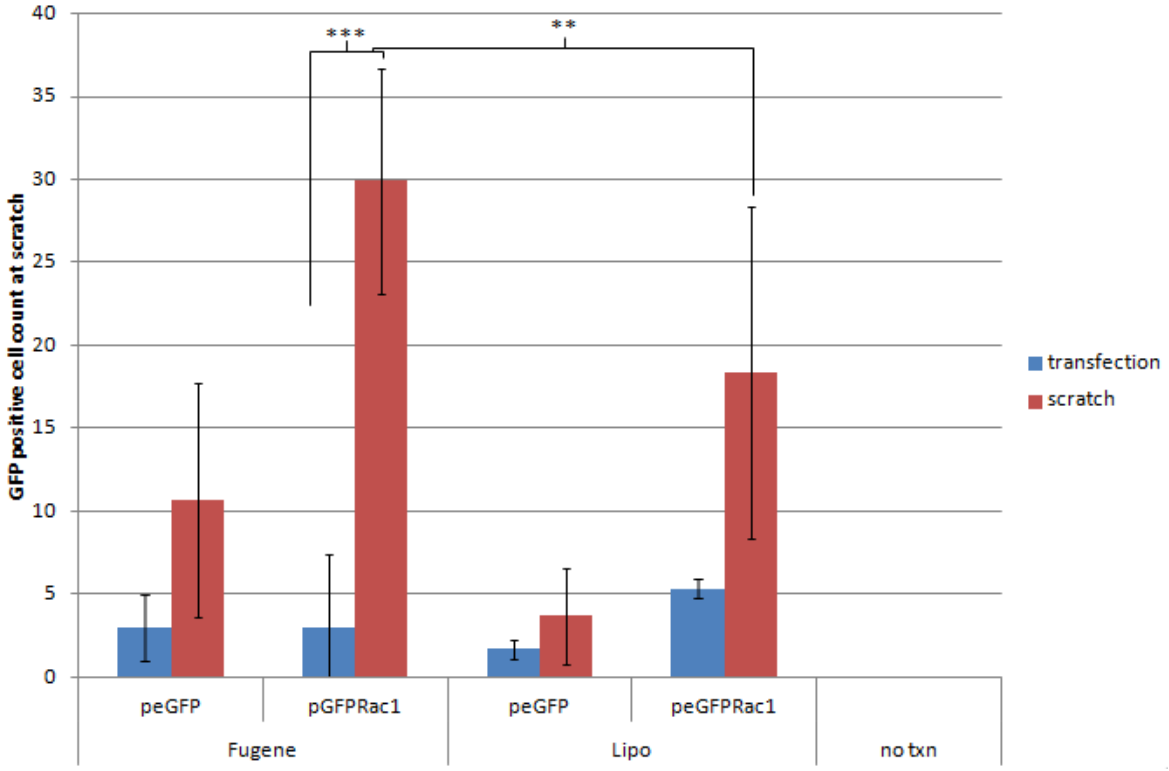


Figure 76. Graph comparing transfection efficiency of transfections using Lipofectamine2000 (Lipo) or FugeneHD, when the cells were transfected then scratched (transfection), or scratched then transfected (scratch). Error bars represent standard deviation of the mean of *n*=4 scratch wounds.

5.3.2.5 Restoration of motility via plasmid encoded Rac1

Using the optimised conditions, as described above, scratch wounds were made in monolayers of Rac1 KO fibroblasts which were then transfected with either peGFP or peGFPRac1. Despite thorough optimisation of the transfection procedure, not all cells were GFP-positive. Repopulation of the scratch wounds was then monitored by timelapse microscopy. Compared to the Rac1 KO-peGFP (control) cells, the Rac1 KO-peGFPRac1 cells demonstrated clear evidence of migration into the wound space (Figure 77; Movie 3) with repopulation and complete wound closure by 36 hours (similar to the Rac1 control fibroblast [Figure 66 above]). For cells transfected with GFP alone, wounds were still not fully closed at 48 hours (Figure 77; Movie 4); similar to untransfected Rac1 KO cells [Figure 66 above]).

It was noted that Rac1 KO cells transfected with the control plasmid, peGFP, migrated much as for untransfected Rac1 KO cells, namely the closure of the wound due to the force developed during cell division.

It was notable (particularly in Movie 3) that cells expressing GFPRac1 (green cells) were able to break away from the confluent cell mass at the wound edge and clearly migrate into the wound space. This movement provides good evidence that the introduction of plasmid encoded Rac1 can restore motility to Rac KO fibroblasts.

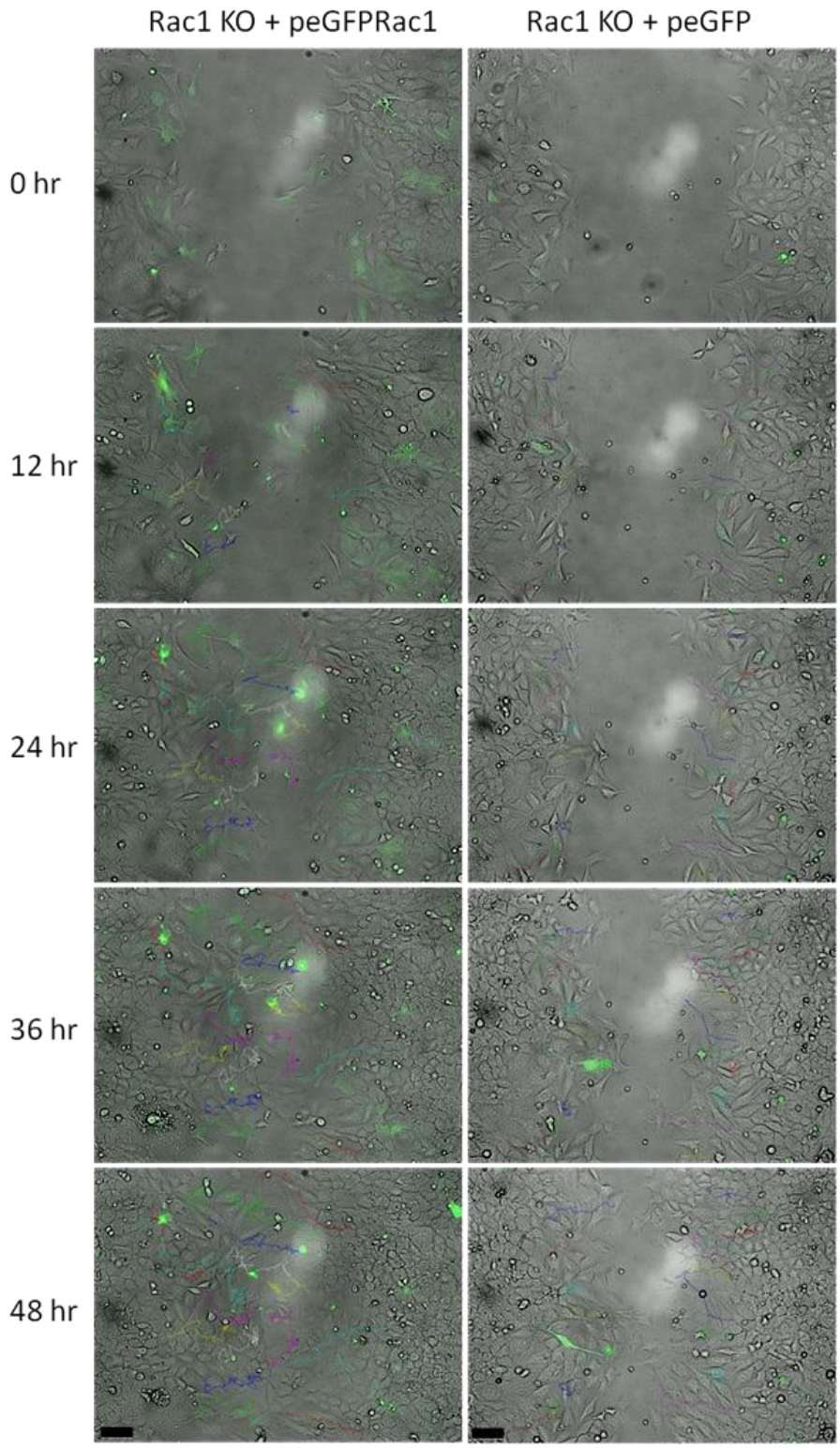


Figure 77. Scratch wound closure of Rac1 KO cells transiently transfected with either peGFPRac1 or peGFP. A monolayer of Rac1 KO cells were grown in 24 well plates, scratch wounded and then transfected with 1 µg peGFP or pGFPRac1 plasmid DNA using Fugene. The closure of the wound was monitored by fluorescent timelapse microscopy over a 48 hour period. Images at 20 minute intervals were captured; the brightfield and GFP fluorescent channels were overlaid using ImageJ. Coloured lines represent tracks moved by individual cells. Scale bar = 75 µm.

5.3.2.6 Quantitative analysis of wound closure

To quantify the rate of closure of the wound, the timelapse images were analysed by measuring the wound area at timepoints throughout the 48 hour image collection. The mean average rates of closure were calculated for the wells transfected with either plasmid peGFP or peGFPRac1 (Figure 78A). The rate of wound closure demonstrated that the Rac1 KO cells transfected with peGFPRac1 closed more efficiently than Rac1 KO cells transfected with peGFP ($p < 0.01$).

5.3.2.7 Quantitative analysis of motility of individual cells

The movement of the wound edge did not fully represent the restoration of motility observed, namely there was an increase in motility observed for single cells compared to the slow gradual movement of the wound edge. The analysis of individual cell movement demonstrated that Rac1 KO cells transiently transfected with peGFPRac1 migrated further than the Rac1 KO cells transfected with peGFP ($p < 0.0001$; Figure 78B).

The trajectories of a number of selected cells throughout the 48 hour timelapse microscopy period were plotted (Figure 79). The Rac1 KO cells transfected with the peGFP plasmid (encoding eGFP only), showed limited movement, similar to the Rac1 KO cells. There appeared to be similar directionality, in the cells moving into the wound area. Rac1 KO cells transfected with peGFPRac1 demonstrated increased motility and there was a general directionality, with the cells moving into the wound area. Although the motility was increased, the directionality of some of the cells was slightly erratic. However, this difference was not statistically significant (Figure 78C).

These results demonstrate that the reduced motility of the Rac1 KO cells can be partially restored via transient transfection of a plasmid encoding Rac1, using the plasmid peGFPRac1.

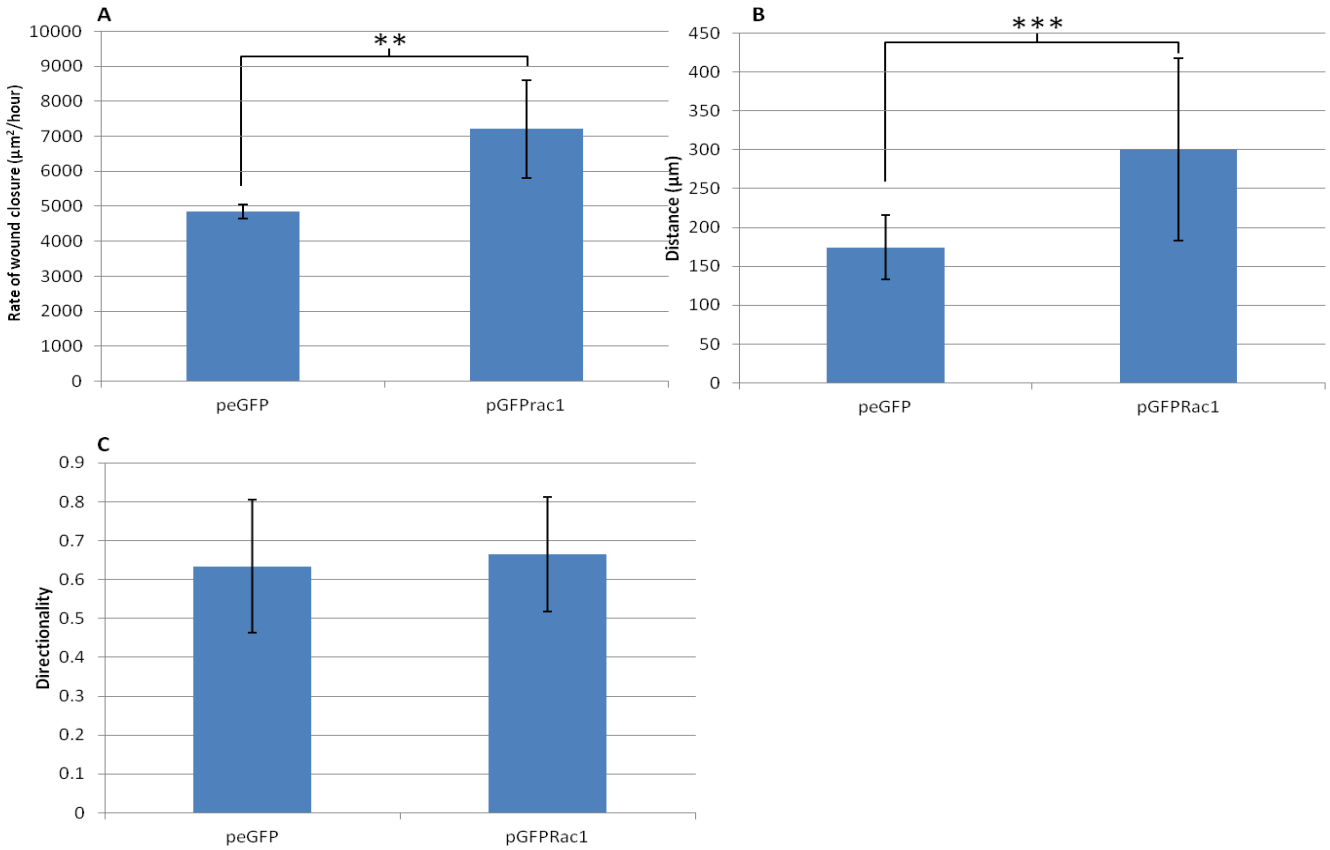


Figure 78. Graphs showing (A) averages of rate of wound closure (*n*=12), (B) distance travelled by individual cells (*n*=121), (C) directionality of individual cells (*n*=121). Scratch wounds were created in Rac1 KO cells transfected with either peGFP or peGFRac1, using Fugene. Fluorescent timelapse microscopy was used to monitor the cells over a 48 hour period. Analysis of motility of individual green cells was calculated using the *Manual Tracking* and *Chemotaxis and Migration* tool (ImageJ). Mean average values for each condition was calculated and plotted. Errors bars represent standard deviation.

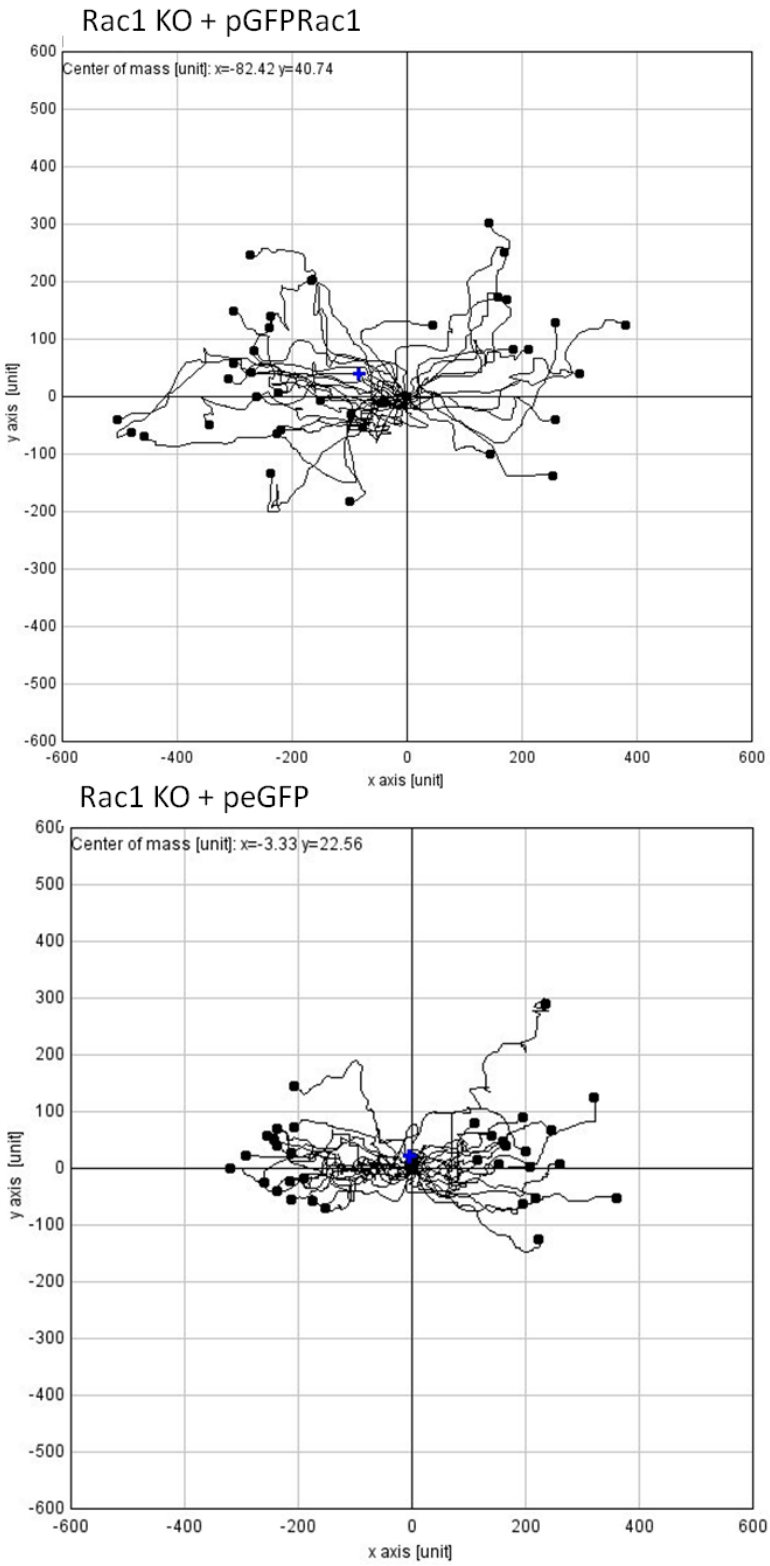


Figure 79. Trajectories of Control and Rac1 KO mouse fibroblasts during scratch wound assay. Scratch wound assays were performed in monolayers of control and Rac1 KO cells. The scratches were monitored by timelapse microscopy. And the trajectories of individual cells at the wound edge were followed using the *manual tracking* plugin and plotted using the *Chemotaxis and migration* plugin of ImageJ.

5.3.2.8 Creation of a Rac1-GFP fusion

The plasmid peGFPRac1 encodes a GFP-Rac1 fusion protein with the GFP N-terminal to the Rac1. To use GFP as a visualisation tool to monitor the expression of Rac1 under the control of the amber suppression technology it was required that a Rac1-GFP fusion was created (Figure 80). This would indicate whether a cell was expressing full-length (NAA containing) Rac1 as GFP would only also be expressed if read-through of the TAG occurred and therefore incorporation of the NAA. Protein products truncated at the TAG within the Rac1 gene would theoretically not produce GFP.

Furthermore, Rac1 contains a sequence (CLLL) that signals the prenylation and cleavage and methylation of the cysteine to allow plasma membrane association of Rac1 (see section 1.13.2.6.3). In addition, Rac1 contains a polybasic region just upstream of the CLLL signal. Due to the post-translational modification of Rac1 protein, i.e. the isoprenylation, proteolysis and methylation of the C-terminal CAAX motif (CLLL in Rac1), the position of this signal sequence during the design of the Rac1-GFP construct, was considered. It was thought that, to retain the ability of Rac1 to associate with the plasma membrane, the CLLL was required to be at the C-terminal of the fusion protein.

It was decided that two constructs would be made, one, that encoded Rac1-GFP-CLLL (prenylation, proteolysis and methylation of C (of CLLL) at C-terminal of GFP) and two, that encoded Rac1-GFP-polybasic (RK)-CLLL (polybasic and CLLL C-terminal to GFP).

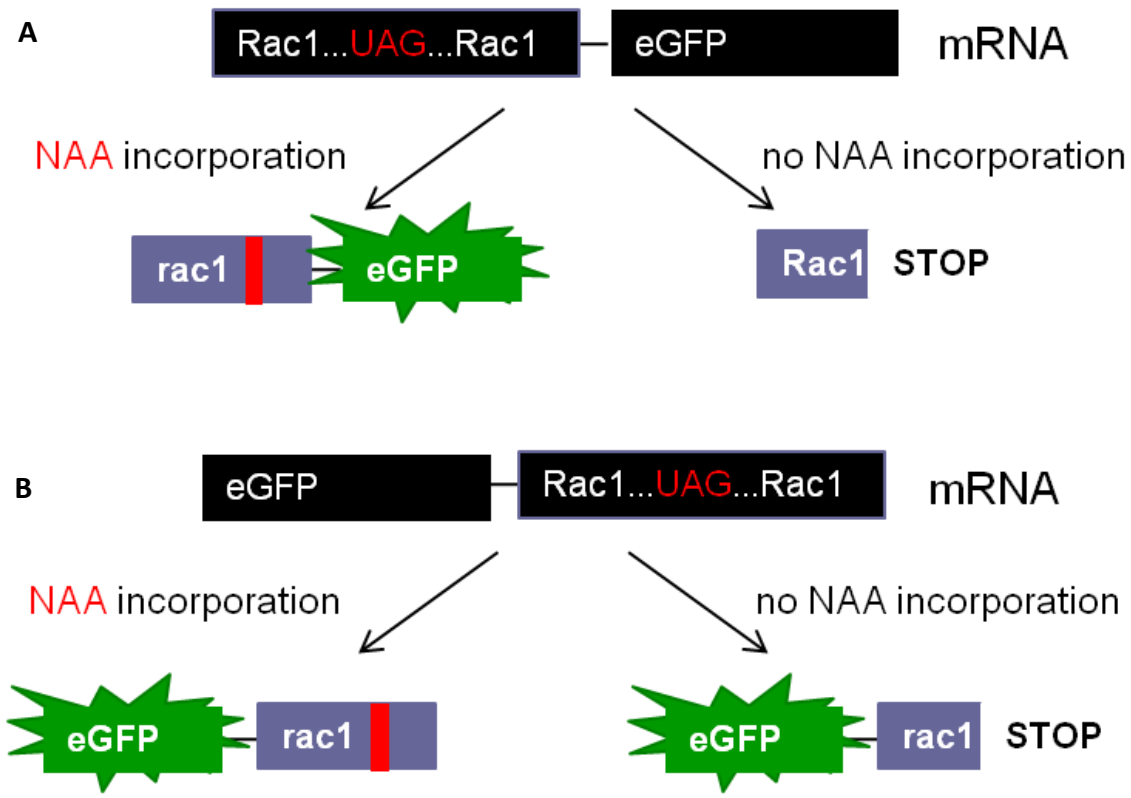


Figure 80. The outcomes when a Rac1-GFP fusion (A) or a GFP-Rac1 fusion (B) construct is used during amber suppression.

5.3.2.9 Creation of pRacGFPCL plasmid for transfection into Rac1 KO cells

To produce a Rac-GFP fusion, Rac1 was cloned into the *Age*I site upstream of eGFP in the plasmid peGFP (Figure 82). PCR was used to amplify the Rac1 gene from pCMVTnTRac (details found in Appendix A9) using Phusion DNA polymerase, using primers racGFP005 forward and either racGFP004, racGFP006 or racGFP007 reverse (Appendix A2). The reverse primers used in the PCR to isolate the Rac1 gene were designed to remove either the stop sequence to create a Rac-GFP fusion protein; the prenylation and stop sequence; or the polybasic region, prenylation and the stop sequence. The purified PCR products and peGFP plasmid were subjected to *Age*I-HF restriction digest and the vector was dephosphorylated. The purified products were ligated together, transformed into Top10 *E. coli* and plated onto agar containing kanamycin. Colonies harbouring the plasmid encoding RacGFP were identified using colony PCR (Figure 81). Positive colonies were determined by the presence of a band corresponding to 1.2 Kb. Rac1 insertion was confirmed sequencing (Figure 83).

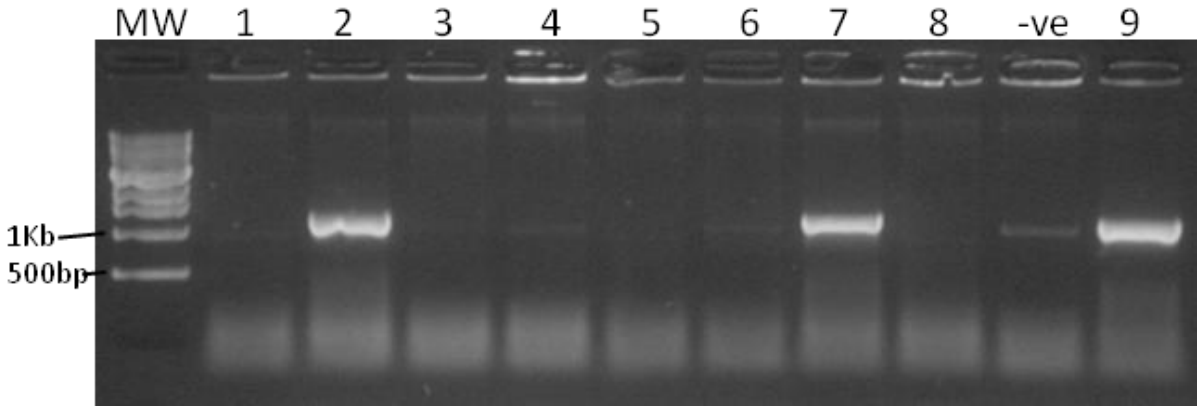


Figure 81. Agarose gel of PCR products from colony screen. PCR was performed using colonies (1 – 9) from plate containing *E. coli* transformed with ligation mixture of *AgeI* digested peGFP and Rac1 PCR product. –ve represents colony from a plate containing *E. coli* transformed with un-ligated mixture of digested peGFP and Rac1 PCR product.

After cloning of the RacGFP construct, the linker between Rac1 and GFP was modified to GGSGGS by site-directed mutagenesis (primers racGFP012 and either racGFP013 (for template minus polybasic and prenylation region) or racGFP014 (for template minus prenylation region)), as it was thought that this would enable some degree of flexibility to aid in correct protein folding and production of functional proteins and also, to ensure that the two genes were in-frame. This mutagenesis was confirmed by DNA sequencing (Figure 83).

In consideration of the signalling sequences located at the C-terminal of the Rac1 protein, two constructs were made, one that would encode a protein with the prenylation (CLLL) sequenced moved to the C-terminal end of the Rac1-eGFP fusion (RacGFPCL) and the other encoding a protein with the polybasic region and the prenylation sequence moved to the C-terminal end of the Rac1-eGFP fusion (RacGFPRKCL). These constructs were created by performing site-directed mutagenesis using primers that annealed to the 3' end of eGFP containing the relevant signalling sequences present at the 3' end of Rac1 (primers racGFP016 and either racGFP015 (addition of prenylation sequence) or racGFP017 (addition of polybasic and prenylation sequence)). The addition of these sequences was confirmed by DNA sequencing. This gave plasmid pRac1-eGFP (pRac1-eGFPCL (prenylation signal 3' of GFP), and pRac1-eGFPRK (polybasic region and prenylation signal 3' of GFP) (Figure 82). Figure 83 shows the protein sequence as determined using the protein translation tool (ExPASy) using the DNA sequencing, the difference in sequence between RacGFPCL and RacGFPRK are indicated.

It was then tested whether these constructs could be transfected and functional protein produced in the Rac1 KO cell line.

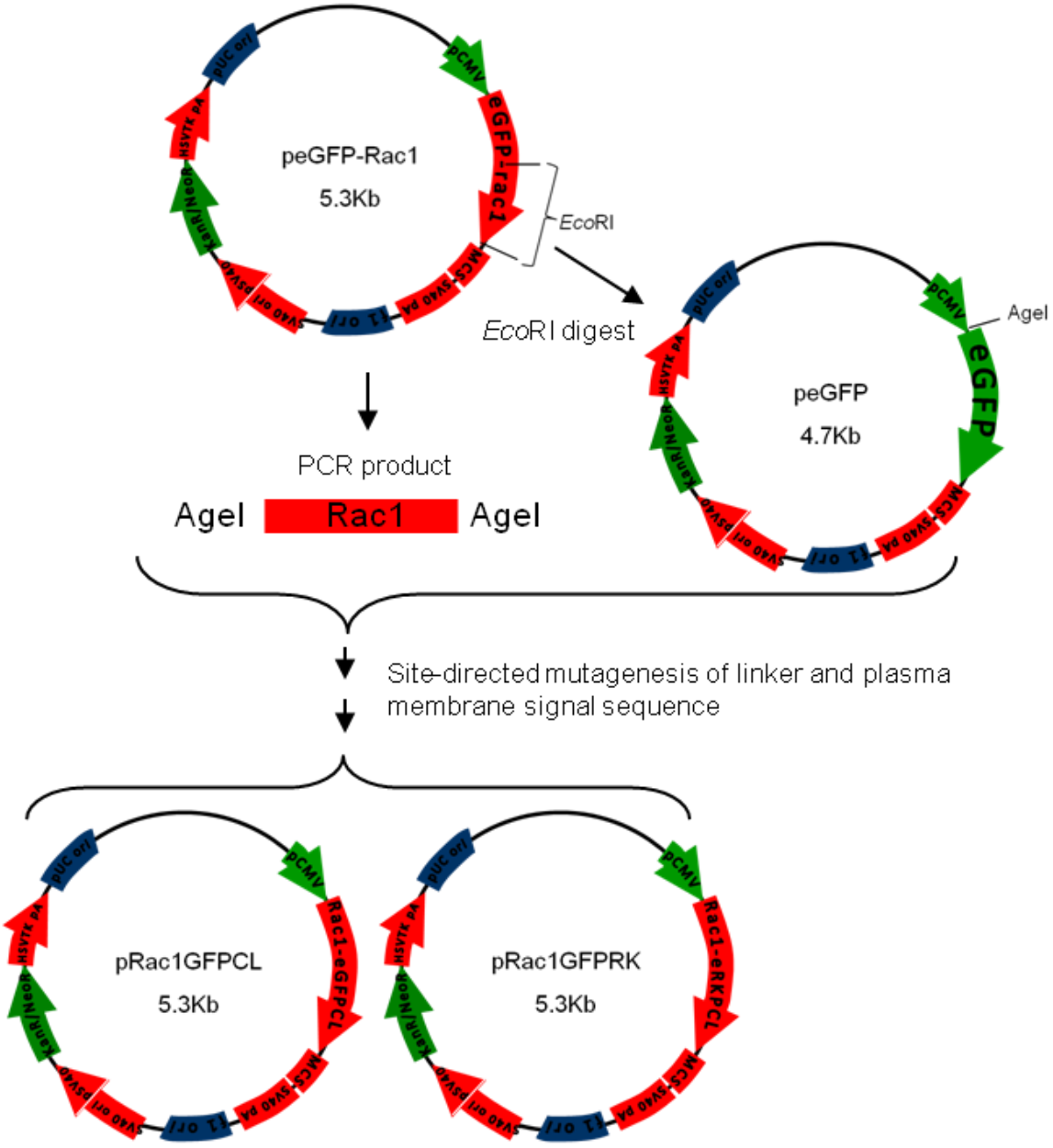


Figure 82. Cloning scheme to create plasmids peGFP and pRac1-eGFP (either pRac1GFPCL (CLLL sequence added to C-terminal of GFP) or pRac1GFPRK (polybasic region and CLLL signal sequence added to C-terminal of GFP).

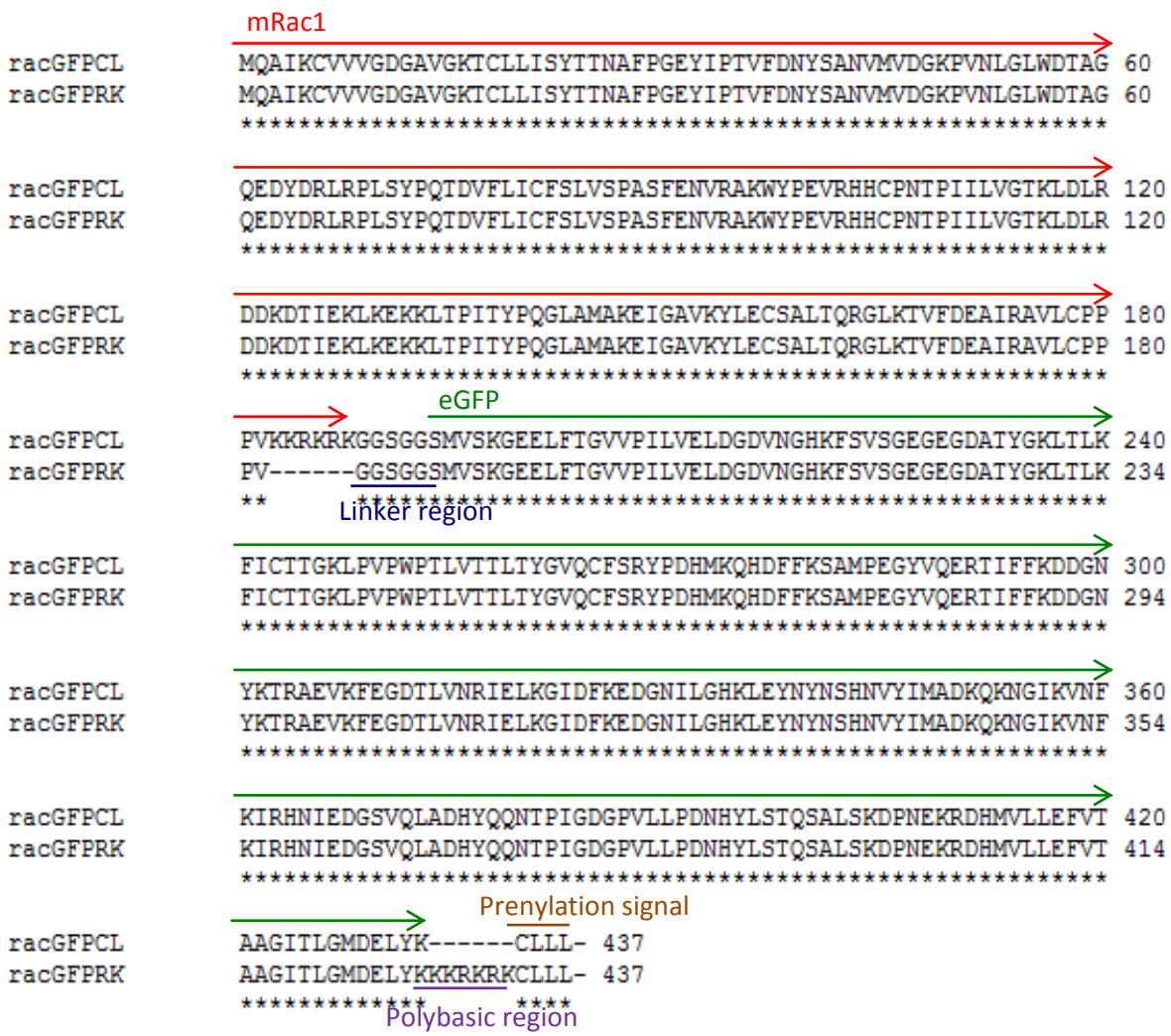


Figure 83. Protein sequence of RacGFP constructs. DNA sequence from RacGFPCL and RacGFPRK construct was translated to protein sequence (ExPASy, translate tool) and compared (ClustalW2).

5.3.2.10 Transient transfection of plasmids pRacGFPCL and pRacGFPRK

The construction of the Rac-GFP fusion gene was complex, namely, in the construction of the C-terminal plasma membrane signalling sequences. Also important was whether the GGSGGS linker region introduced as a linker region between the Rac1 and eGFP would provide the flexibility for the correct folding and function of the two proteins. Therefore, it was investigated whether these constructs would indeed be expressed and a functional product produced when transfected into the Rac1 KO cells. Fluorescent microscopy demonstrated that GFP-positive cells were obtained after transient transfection (optimised Fugene protocol) of the plasmid pRacGFPCL and pRacGFPRK into Rac1 KO cells (Figure 84). It appeared that the pRacGFPRK construct was not well tolerated, as over time the green cells appeared to undergo cell death. The pRacGFPCL construct was chosen as the preferred construct for motility assays.

5.3.2.11 Restoration of motility – transient expression of RacGFPCL

To ascertain whether the RacGFPCL construct could restore motility to the Rac1 KO fibroblasts, scratch wounds in conjunction with timelapse fluorescent microscopy was performed. The microscopy revealed that the RacGFPCL construct was capable of increasing motility when compared to cells expressing peGFP (Figure 85). The scratch wounds in the wells transfected with pRacGFPCL were closed by 37 hours (Movie 5); whereas the scratch wounds in the wells transfected with pGFP were still not closed by 48 hours (Movie 6). There was a visually apparent increased movement of GFP positive cells into the scratch wound area in those cells transfected with pRacGFPCL.

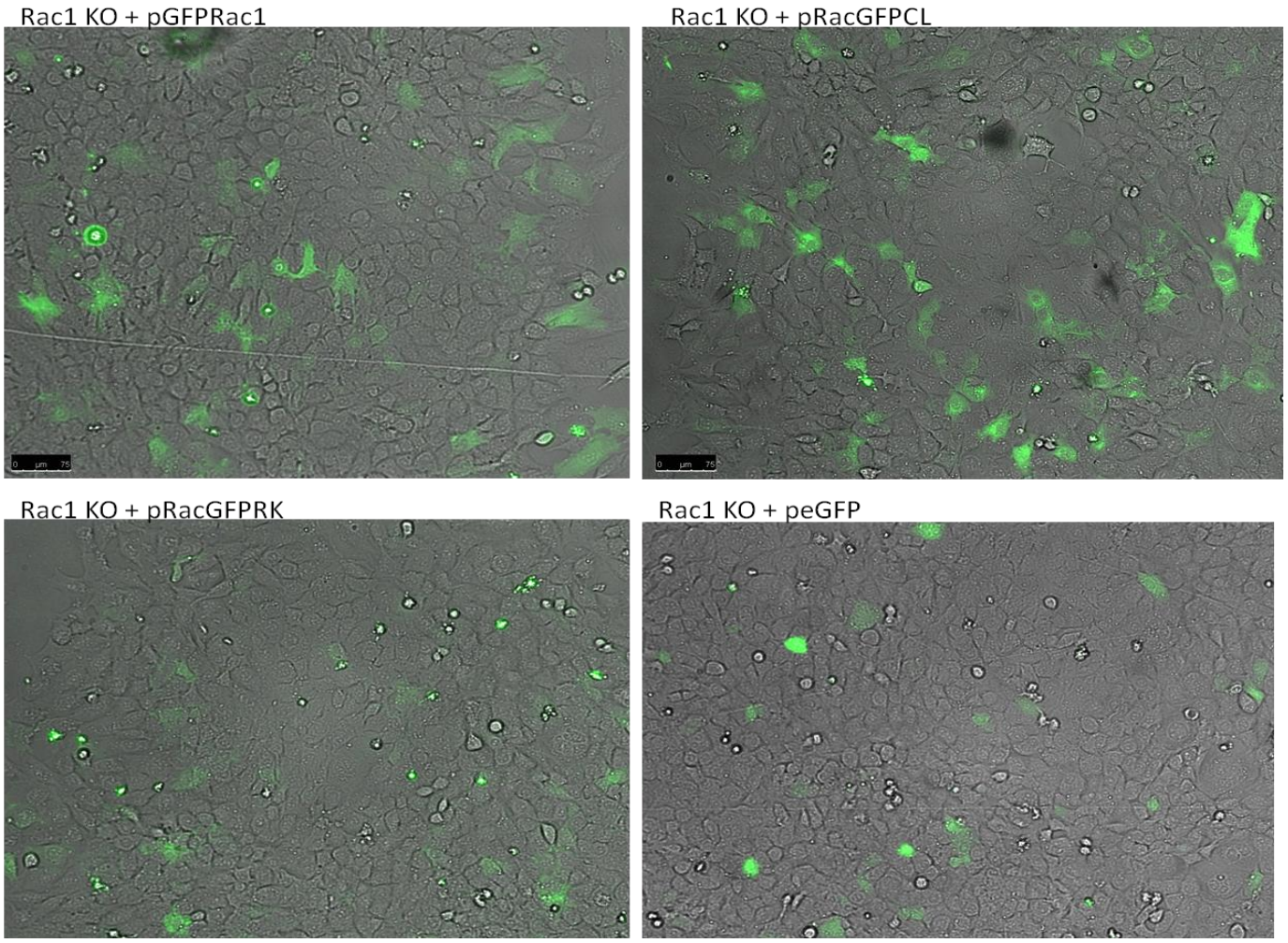


Figure 84. Transient transfection of Rac1 KO cells with pGFPRac, pRacGFPCL, pRacGFPRK and peGFP. Scale bar represent 75 µm.

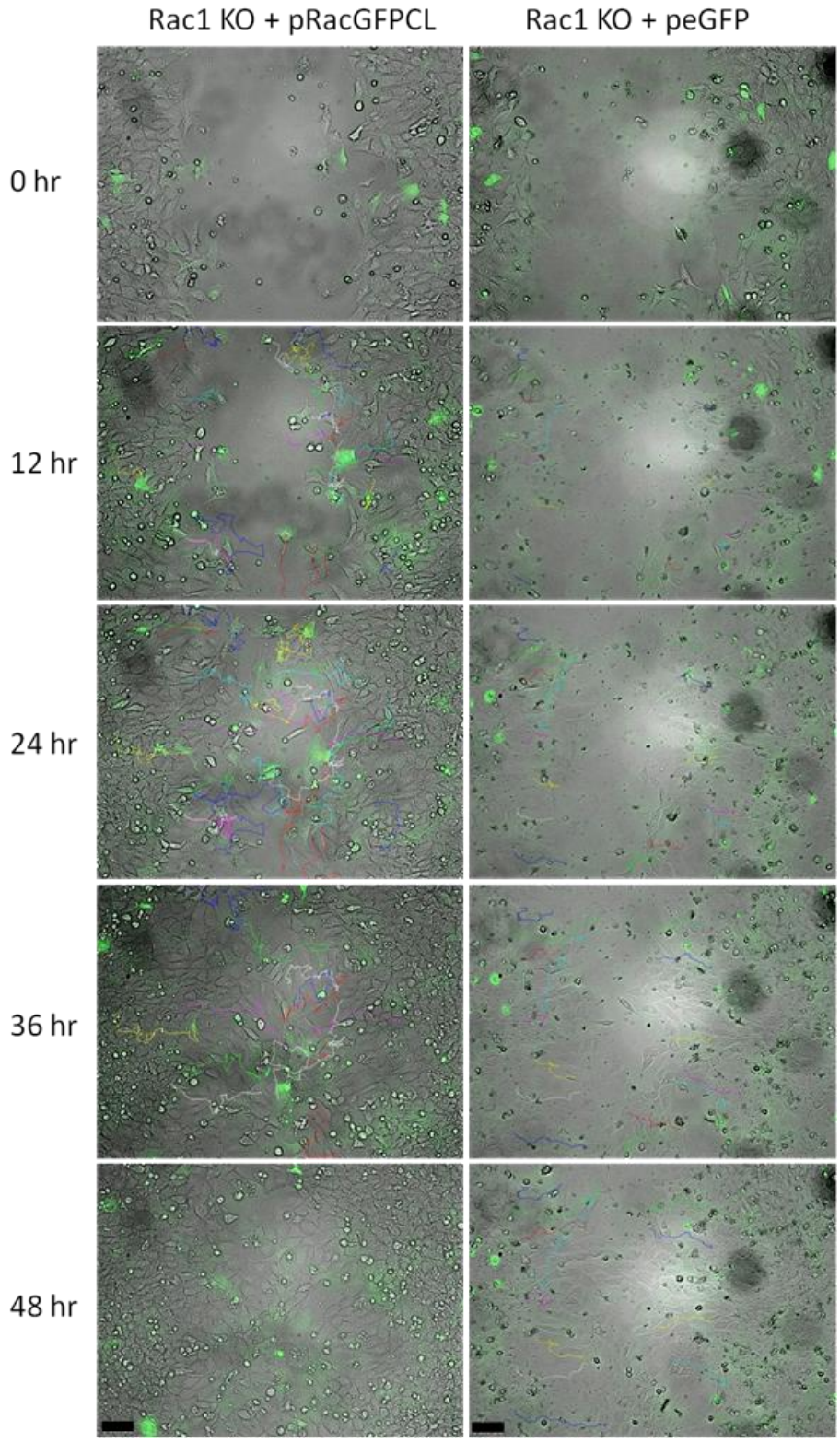


Figure 85. Scratch wound closure of Rac1 KO cells transiently transfected with either pRacGFPCL or pGFP. A monolayer of Rac1 KO cells were grown in 24 well plates, scratch wounded and then transfected with 1 µg peGFP or pGFPRac1 plasmid DNA using Fugene. The closure of the wound was monitored by fluorescent timelapse microscopy over a 48 hour period. Images at 15 minute intervals were captured; the brightfield and GFP fluorescent channels were overlaid using ImageJ. Coloured lines represent tracks moved by individual cells. Scale bar = 75 µm.

5.3.2.12 Quantitative analysis of wound closure

The mean average rates of closure were calculated for wells transfected with either plasmid peGFP or pRacGFPCL (Figure 86A). The rate of wound closure by Rac1 KO cells transiently transfected with pRacGFPCL was significantly faster than that of the Rac1 KO cells transfected with peGFP ($p < 0.001$).

5.3.2.13 Quantitative analysis of motility of individual cells

It was immediately apparent that the green cells at the wound edge had increased in motility, could break-away from the cell mass and move into the wound space freely. To try to capture the increase in motility of individual cells, the total distance moved by green cells in the 48 hour period were measured (Figure 86B). The analysis demonstrated that Rac1 KO cells transiently transfected with pRacGFPCL migrated further than the Rac1 KO cells transfected with peGFP ($p < 0.001$).

To further analyse the motility of the cell populations, the directionality of the cells' movements was analysed using data from the manual cell tracking. Plotting the trajectories demonstrated that the Rac1 KO-pRacGFPCL cells showed greatly increased motility, however the movement of the cells appeared to be predominantly erratic (Figure 87). Analysis of cell directionality demonstrated that the movement of the Rac1 KO-pRacGFPCL cells was more erratic than that displayed by Rac1 KO-peGFP cells ($p < 0.0001$; Figure 86C).

These results demonstrated that transient transfection with the Rac-GFPCL construct (prenylation sequence at the C-terminal of the GFP) could also restore motility to the Rac1 KO cell line. Therefore, the RacGFPCL construct could be used in further investigations in this study.

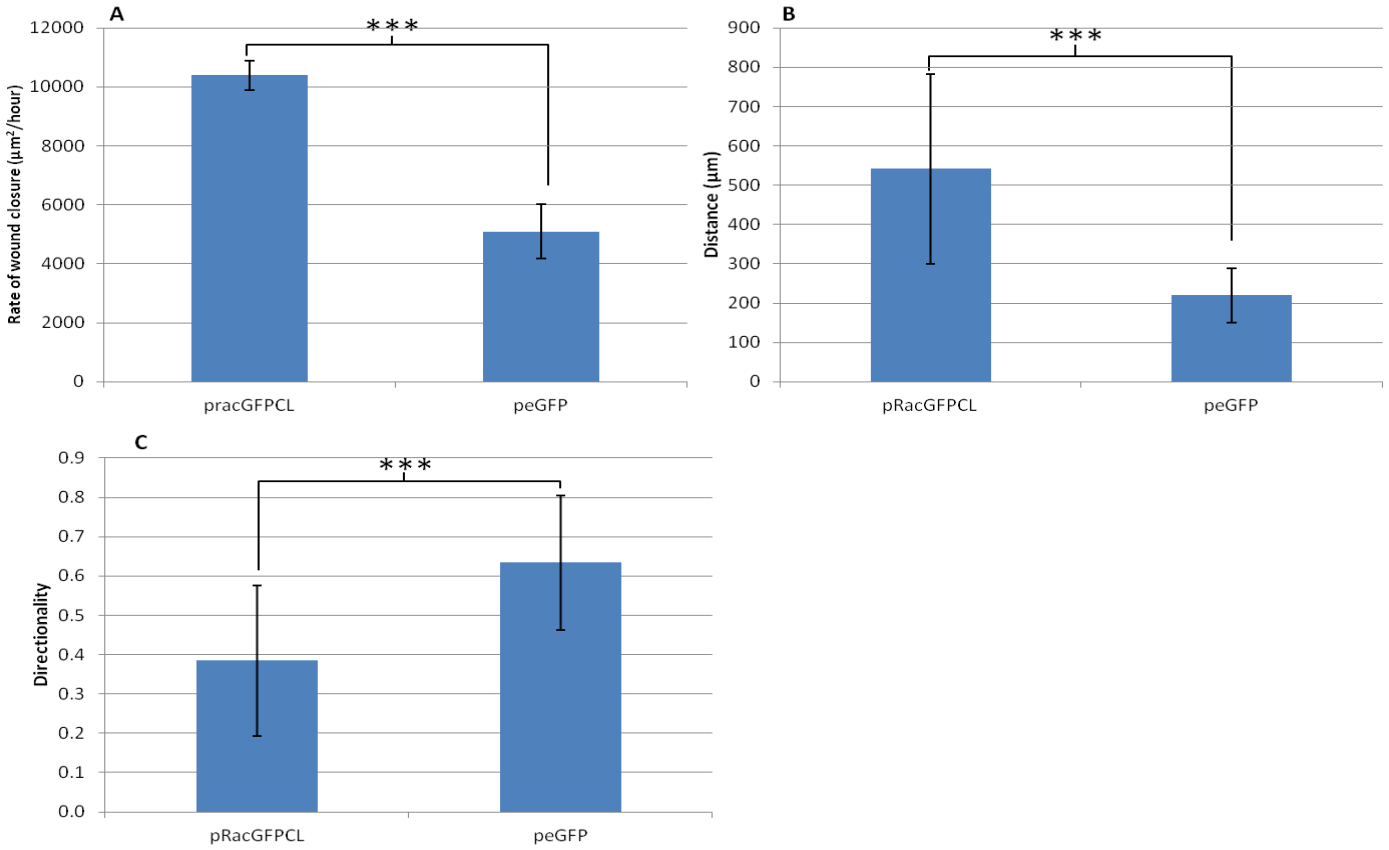


Figure 86. Graphs showing (A) averages of rate of wound closure (*n*=4), (B) distance travelled by individual cells (*n*=75), (C) directionality of individual cells (*n*=75). Scratch wounds were created in Rac1 KO cells transfected with either peGFP or pRacGFPCL, using Fugene. Fluorescent timelapse microscopy was used to monitor the cells over a 48 hour period. Analysis of motility of individual green cells was calculated using the *Manual Tracking* and *Chemotaxis and Migration* tool (ImageJ). Mean average values for each condition was calculated and plotted. Errors bars represent standard deviation.

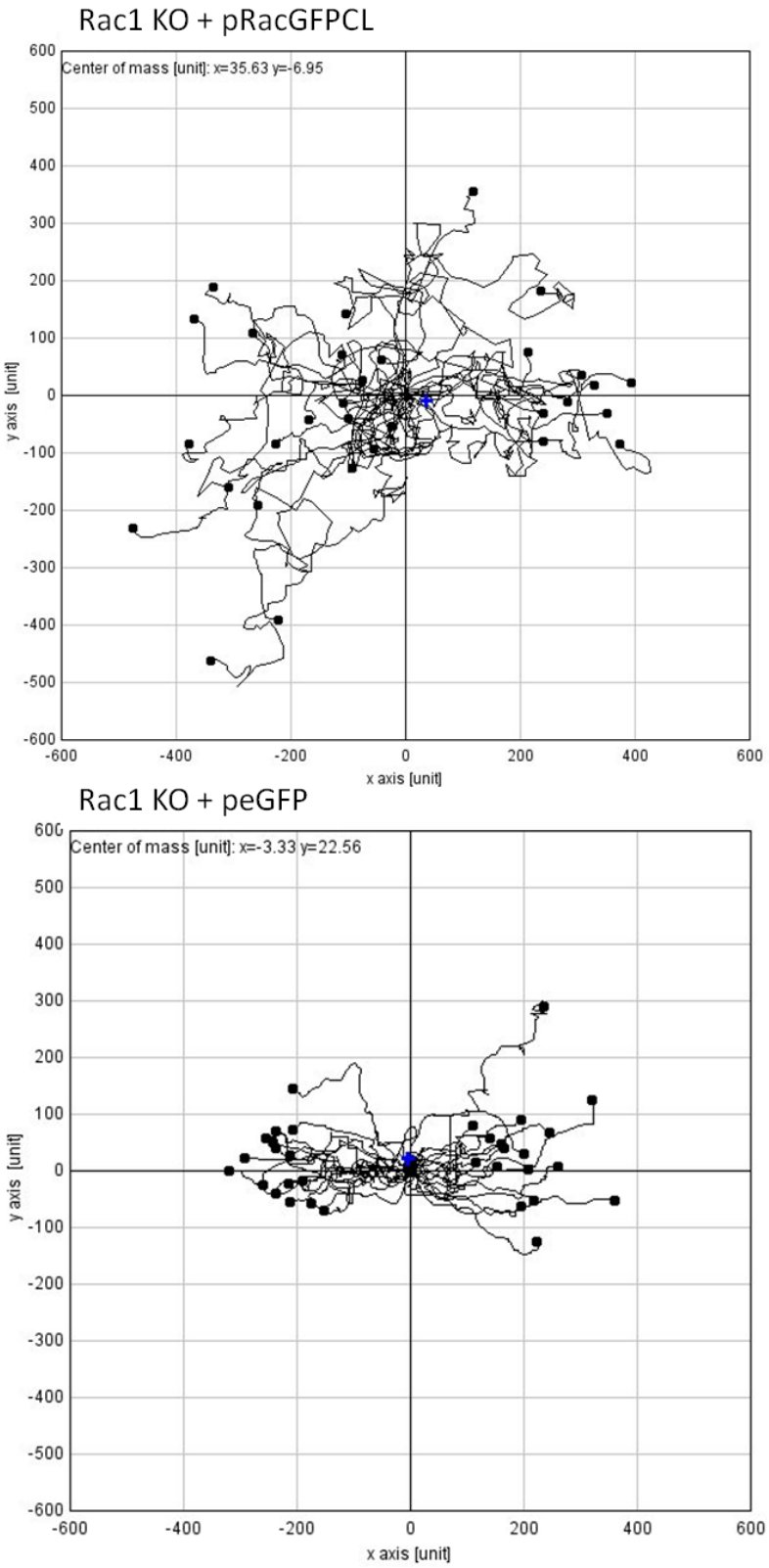


Figure 87. Trajectories of Rac1 KO mouse fibroblasts transiently transfected with pRacGFPCL. Scratch wound assays were performed in monolayers of Rac1 KO cells transiently transfected with pRacGFPCL. The scratches were monitored by timelapse microscopy and the trajectories of individual cells at the wound edge were followed using the *manual tracking* plugin and plotted using the *Chemotaxis and migration* plugin of ImageJ.

5.3.2.14 Evidence for expression of Rac1 gene

Although the visualisation of GFP during the transfection experiments provided some evidence that the Rac1 protein was expressed (as the expression was a GFP-fusion protein), further evidence for expression of Rac1 was sought. To do this, RNA was extracted from control cells (for positive control), Rac1 KO cells and Rac1 KO cells transfected with either peGFP, peGFPRac, pRacGFPCL or pRacGFPRK. Using reverse-transcription PCR (RT-PCR), cDNA was synthesised utilising the RNA extraction. PCR was performed using primers that annealed to the ubiquitous gene β-actin were used to ensure cDNA had been successfully isolated (Appendix A8).

Primers, racFP and racRP3 annealed to nucleotides 1 - 17 and 531 - 546 of the Rac1 gene respectively and were used to amplify Rac1 gene from the cDNA using PCR. An agarose gel of the products was analysed for correct sized products. The PCR products were purified and DNA sequencing was used to confirm amplification of the Rac1 gene. As expected the control cells and Rac1 KO cells transfected with plasmid encoding Rac1 gave a product of 546 bp. cDNA synthesised using RNA extracted from Rac1 KO cells and Rac1 KO cells transiently transfected with peGFP did not give product corresponding to Rac1 (Figure 88).

5.3.3 Stable transfection of RacGFP constructs

To further investigate the constructs' efficiency to restore Rac1 function and motility to the Rac1 KO cell line; it was thought that a plasmid could be used to create stable transfectants. This would provide a situation where scratch wounds would show maximum increase in motility due to the entire population expressing Rac1. Therefore, Rac1-eGFP was cloned into a Zeocin resistant plasmid, namely the pcDNA3.1/zeo plasmid (Invitrogen). This would provide a situation where successfully transfected cells could be selected, grown and used solely in scratch wound assays.

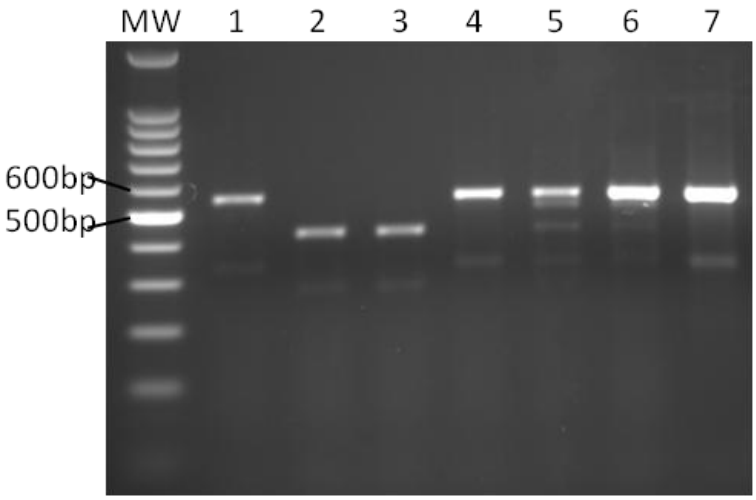


Figure 88. RT-PCR analysis of Rac1. Agarose gel (2%) of PCR to amplify Rac1 gene from cDNA synthesised from RNA extracted from control cells (lane 1), Rac1 KO cells (lane 2), and Rac1 KO cells transfected with peGFP (lane 3), peGFPRac1 (lane 4), pRacGFPCL (lane 5), pRacGFPRK (lane 6) and plasmid pRacGFPCL (lane 7).

Zeocin was chosen for the selection marker for a number of reasons. Firstly, the Rac1 KO cell line was made geneticin (G418) resistant during the production of the cell line (confirmed by growth in < 500 mg / ml G418), making this selection unsuitable. Secondly, puromycin was used transiently during the production of the cell line, thereby making this antibiotic unfavourable. Thirdly, Zeocin uses a different mode of action to G418 and puromycin, namely the inhibition of protein synthesis by disruption of translocation and promotion of mistranslation at the 80S ribosome. Fourthly, the previously created pSWAN plasmids (Liu et al. 2007) used in application of the multi-plasmid system for incorporation of NAAs in mammalian cells are also based on the pcDNA3.1/zeo plasmid. Therefore, it was also possible to create a selectable plasmid containing all the components needed for the incorporation of NAA (*Bs*tRNA, aaRS, and RacGFPCL). In consideration of these factors, the pcDNA3.1/zeo plasmid was used as a basis for the cloning of a Zeocin selectable, Rac1-expressing construct for use in this study.

5.3.3.1 Creation of a selectable plasmid expressing Rac1-GFP

The RacGFPCL cassette was cloned into the previously created plasmid, pcDNA3.1/zeo-RS, which is based on the commercially available plasmid, pcDNA3.1/zeo (Invitrogen). PCR was used to amplify the RacGFPCL construct, using primers that placed a *Not*I (racGFP021) and *Age*I site (racGFP-IRES007) at 5' and 3' terminal, respectively, of RacGFPCL (Appendix A2). Digested PCR product was ligated into complementarily digested, dephosphorylated and gel extracted plasmid pcDNA3.1/zeo. A colony PCR screen, using primers racGFP021 and BGHpA RP identified colonies that harboured the RacGFPCL containing pcDNA3.1/zeo plasmid (Figure 89). A PCR product of 1.4 Kb corresponded to insertion of RacGFPCL in to pcDNA3.1/zeo and DNA sequencing confirmed the insertion. This produced a Zeocin selectable plasmid expressing RacGFPCL, plasmid pzeoCL (Figure 90).

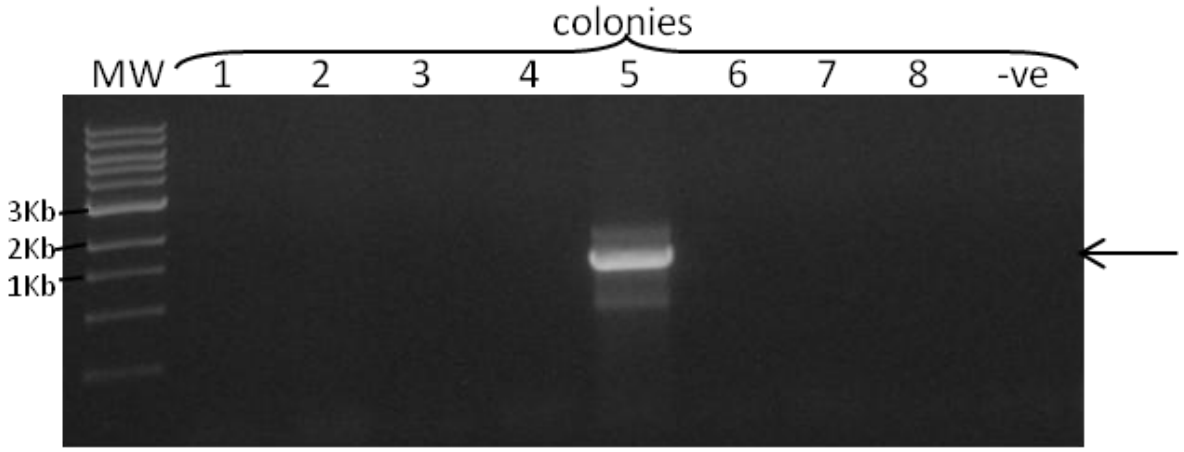


Figure 89. Agarose gel of colony PCR, screening colonies for insertion of RacGFPCL into plasmid pcDNA3.1/zeo. PCR was performed using primers racGFP021 forward and BGHpA reverse. A PCR product of 1.4 Kb corresponded to insertion of RacGFPCL in to pcDNA3.1/zeo.

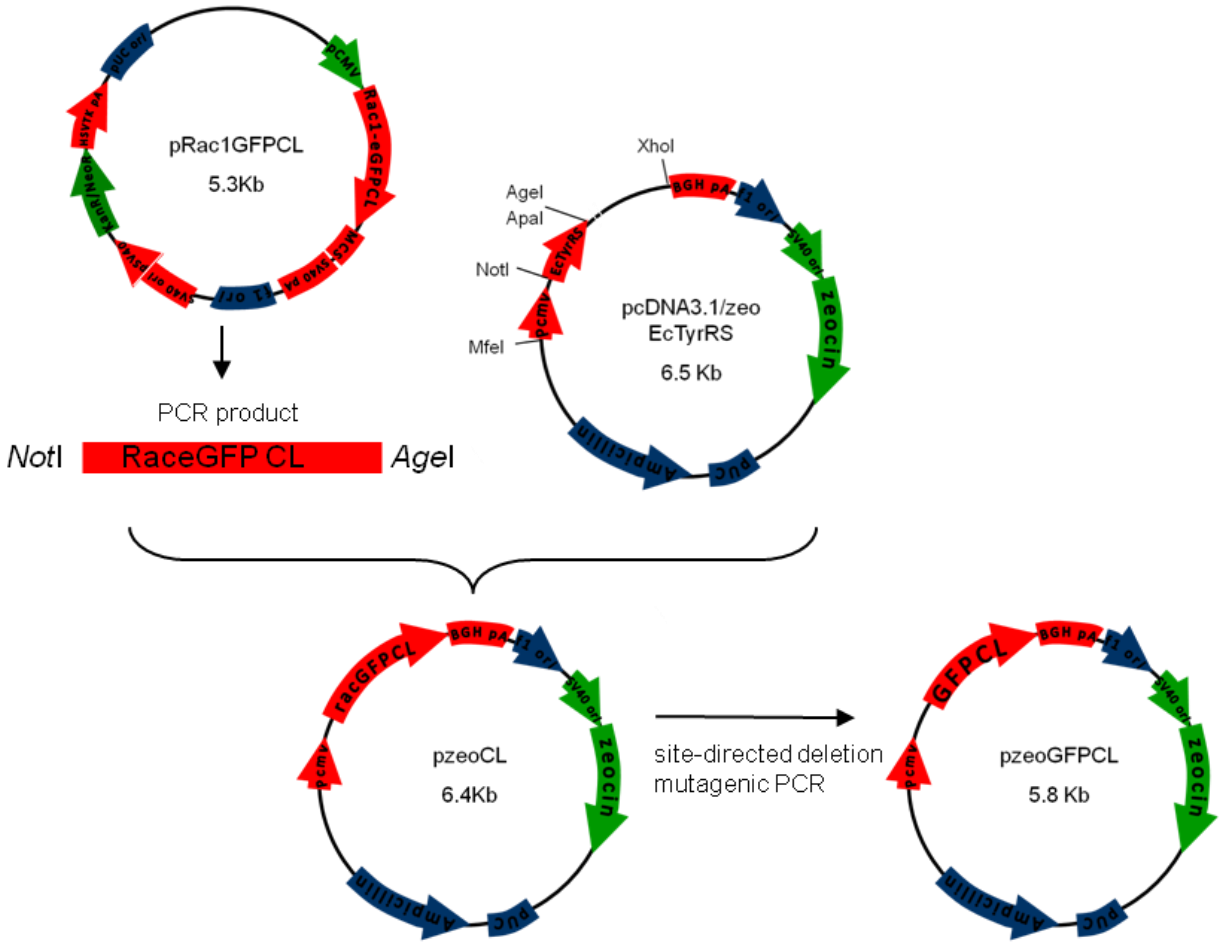


Figure 90. Plasmid cloning scheme to create pzeoCL and pzeoGFPCL.

5.3.3.2 Creation of a selectable plasmid expressing eGFP

To create a population of stably transfected cells that were not expressing Rac1 that could act as a control, a plasmid that had the same backbone as pzeoCL, but encoded GFPCL instead of RacGFPCL was created. This was important to ensure the re-introduction of Rac1 was responsible for the restoration of motility and not the process of producing a stably transfected cell line.

The Rac1 gene was removed from the pzeoCL using site-directed deletion mutagenic PCR with primers pUAA001-2 and pzeo002, which annealed to the beginning of the GFP gene and plasmid sequence upstream of the Rac1 gene respectively (Appendix A2). The kozak sequence (ACCATG) was preserved at the GFP start site and the prenylation sequence was conserved at the 3' terminal to ensure localisation of expressed products was similar to RacGFP construct. DNA sequencing confirmed the deletion of Rac1. This created a Zeocin selectable plasmid encoding GFPCL, pzeoGFPCL (Figure 90).

5.3.3.3 Zeocin kill curve

Because intrinsic Zeocin sensitivity varies among cell lines, the concentration for selection in the Rac1 KO cell line was optimised. Concentrations ranging from 1000 – 25 µg / ml in DMEM growth media were used to create a kill curve (Figure 91). A concentration of 100 µg/ml was chosen for selection of stably transfected Rac1 KO cells because little growth of wild type Rac1 KO cells occurred after addition and all were killed after 7 days incubation.

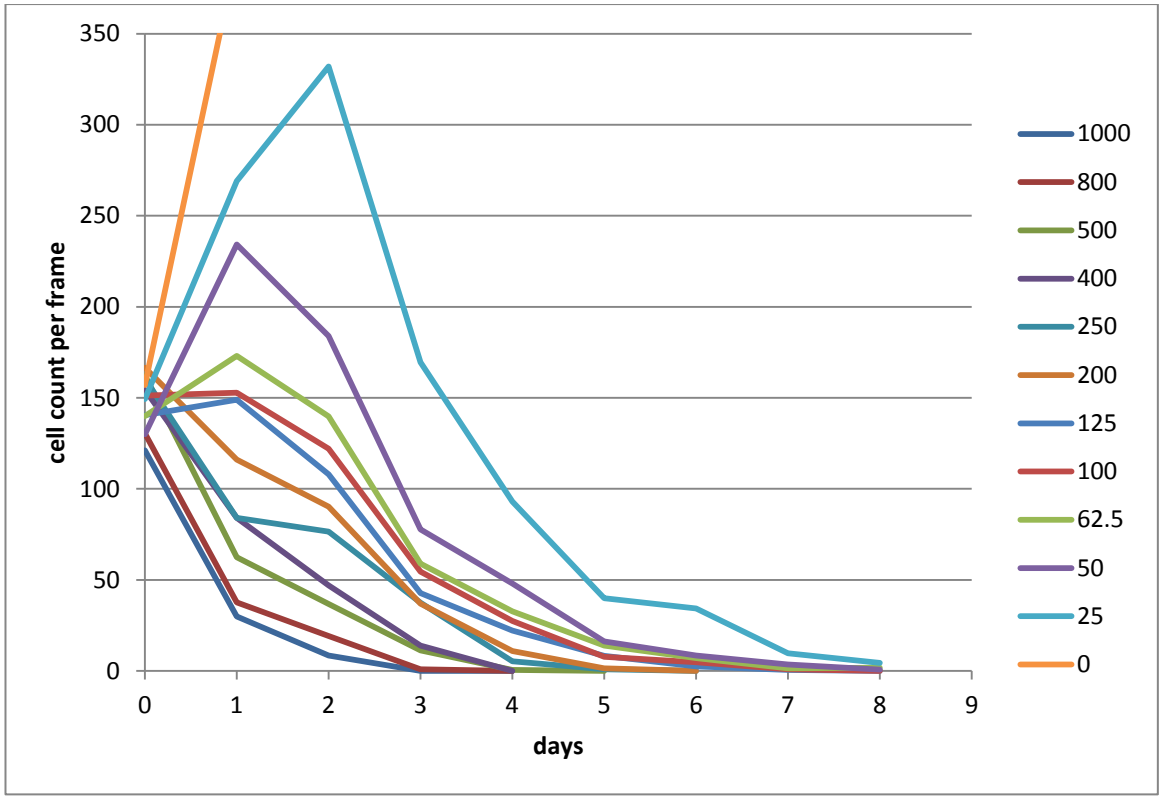


Figure 91. Zeocin selection kill curve for Rac1 KO. 24 well plates were inoculated with Rac1 KO cells at 2 x 10^4 / well. One day post inoculation, media was replaced with media containing Zeocin at the specified concentration. Every two days the selection media was replaced. Images were captured every day and the number of cells per frame was counted and plotted.

5.3.3.4 Rac1 KO cells stable expression

5.3.3.4.1 Expression of target genes from Zeocin-resistant plasmids

To investigate whether the plasmids pzeoCL and pzeoGFPCL would express in Rac1 KO, the plasmids were transiently transfected into the cells. It was shown that the plasmids pzeoCL and pzeoGFPCL could express GFP after transient transfection in Rac1 KO cells. Therefore, transfections were performed to create stably transfected cells, as described in materials and methods.

It was demonstrated that not all the stable clones were positive for GFP expression; therefore the colonies were screened using fluorescence microscopy. A number of GFP-positive clones were picked and grown. Each clone was assessed for their rate of growth, cell health as determined visually, and checked for GFP expression (via fluorescence microscopy) periodically throughout cellular amplification. This was undertaken to ensure that the expression of target genes was retained throughout cell amplification; that silencing had not occurred; and that the population was not contaminated with non-GFP positive cells that may out-compete the GFP-positive cells. Two clones expressing RacGFPCL, pzeoCL1.3 and pzeoCL10, and two clones expressing GFP, pzeoGFPCL1 and pzeoGFPCL3, were chosen and used in subsequent experiments. After analysis of wound closure, the clones pzeoCL1,3 and pzeoGFPCL3 were used in further analysis and will subsequently be referred to as Rac1 KO + pzeoCL and Rac1 KO + pzeoGFPCL.

5.3.3.4.2 Restoration of motility – stable expression of RacGFPCL

The resistant clones, harbouring either pzeoCL (Rac) or pzeoGFPCL (GFP) were grown, aliquots were cryopreserved and were used in scratch wound assays. It was apparent from microscopy that the Rac-expressing stable cell line had altered visual appearance and increased motility (Figure 92; Movie 7). The cells were more able to form projections, break away from the cell mass and could be seeded at lower densities than the Rac1 KO cell line. Timelapse microscopy of the scratch wounds showed that the Rac1 KO cells stably expressing Rac1 were more motile, readily moved away from the cell mass and into the wound space and therefore closed the

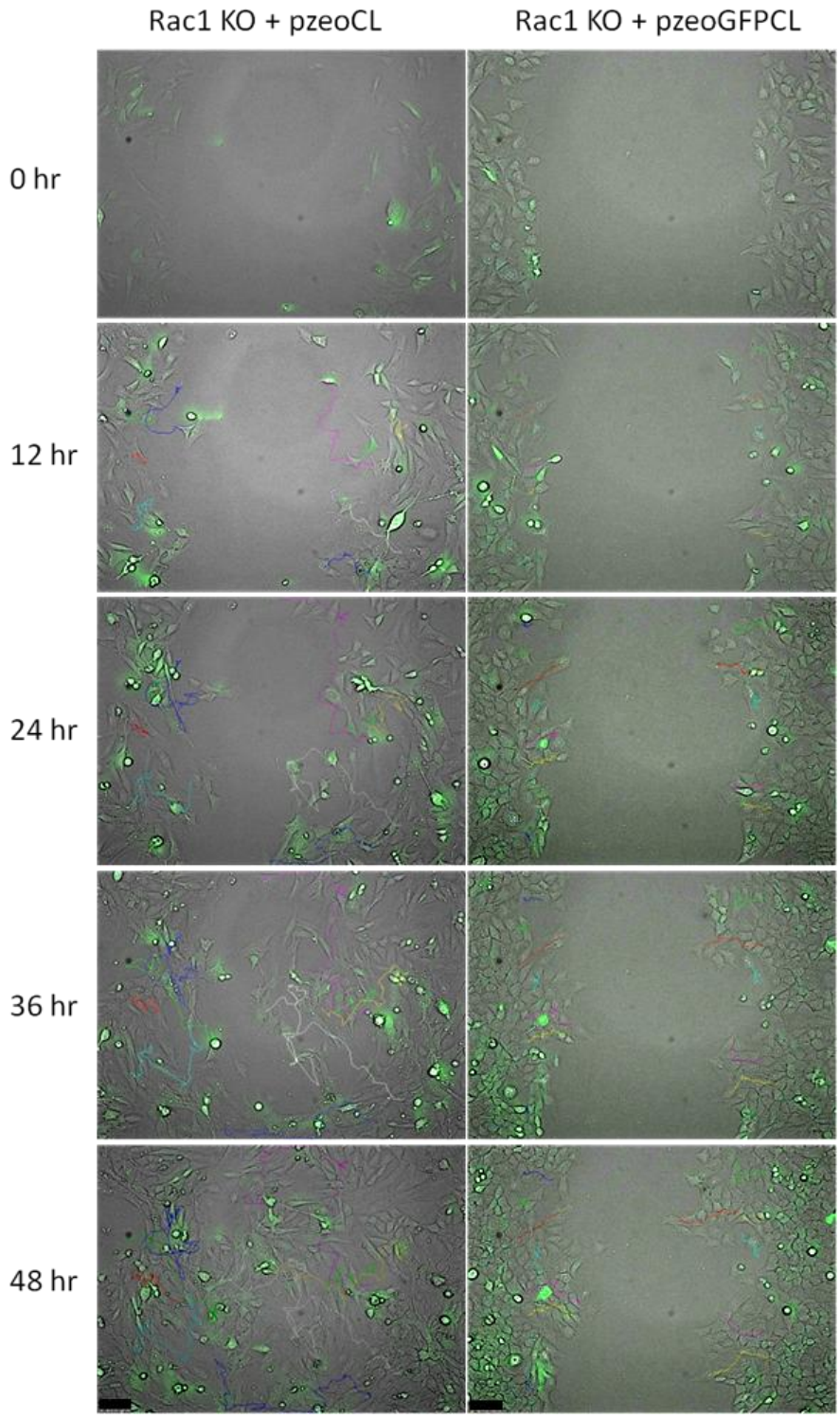


Figure 92. Scratch wound closure of stable cell lines expressing RacGFPCL (Rac1 KO + pzeoCL) and GFPCL (Rac1 KO + pzeoGFPCL). Rac1 KO cells were transfected with plasmids pzeoCL or pzeoGFPCL; resistant and GFP-positive colonies were selected using 100 µg/ml Zeocin antibiotic and fluorescent microscopy respectively. Stable cells were used to inoculate 24 well plates, scratched and viewed using timelapse microscopy over a 48 hour period. Images at 15 minute intervals were captured; the brightfield and GFP fluorescent channels were overlaid using ImageJ. Coloured lines represent tracks moved by individual cells. Scale bar = 75 µm.

scratch wound after approximately 45 hours (Figure 92; Movie 7). Whereas the GFP transfected cells still had a sizeable scratch wound after 45 hours (Movie 8). The cells stably expressing GFP had a visual appearance and motility much like that of the parent Rac1 KO cell line.

5.3.3.4.3 Quantitative analysis of wound closure

The rate of wound closure demonstrated that cells stably expressing Rac1 closed more efficiently than Rac1 KO cells stably expressing eGFP (Figure 93A; $p < 0.0001$).

5.3.3.4.4 Quantitative analysis of motility of individual cells

To try to capture the increase in motility of individual cells, total distance and directionality moved by individual cells in the 48 hour period were analysed. The analysis demonstrated that Rac1 KO cells stably expressing Rac1 migrated further than the Rac1 KO stably expressing eGFP ($p < 0.0001$; Figure 93B).

Plotting the trajectories demonstrated that the Rac1 KO stably expressing Rac1 showed greatly increased motility however, the movement of the cells appeared to be predominantly erratic (Figure 94). Analysis of cell directionality demonstrated that the movement of the Rac1 KO stably expressing Rac1GFPCL was more erratic than that displayed by Rac1 KO-peGFP cells ($p < 0.01$; Figure 93C).

These results demonstrated that the Rac1 KO cells could be used to create stably transfected cell lines using the pcDNA3.1/Zeo plasmid (denoted pzeo-based plasmid). Using stable expression of the Rac1-GFPCL construct, it was demonstrated to restore motility to the Rac1 KO cell line.

5.3.3.4.5 Expression of target genes

To verify that components of the plasmid were indeed being transcribed, RT-PCR was performed. RNA was extracted from the stable cell lines and PCR was performed to check for transcription of β-actin, Rac and Zeocin resistance genes (zeo^r). Agarose analysis of PCR products showed that β-actin, Rac and zeo^r were transcribed (Figure 95).

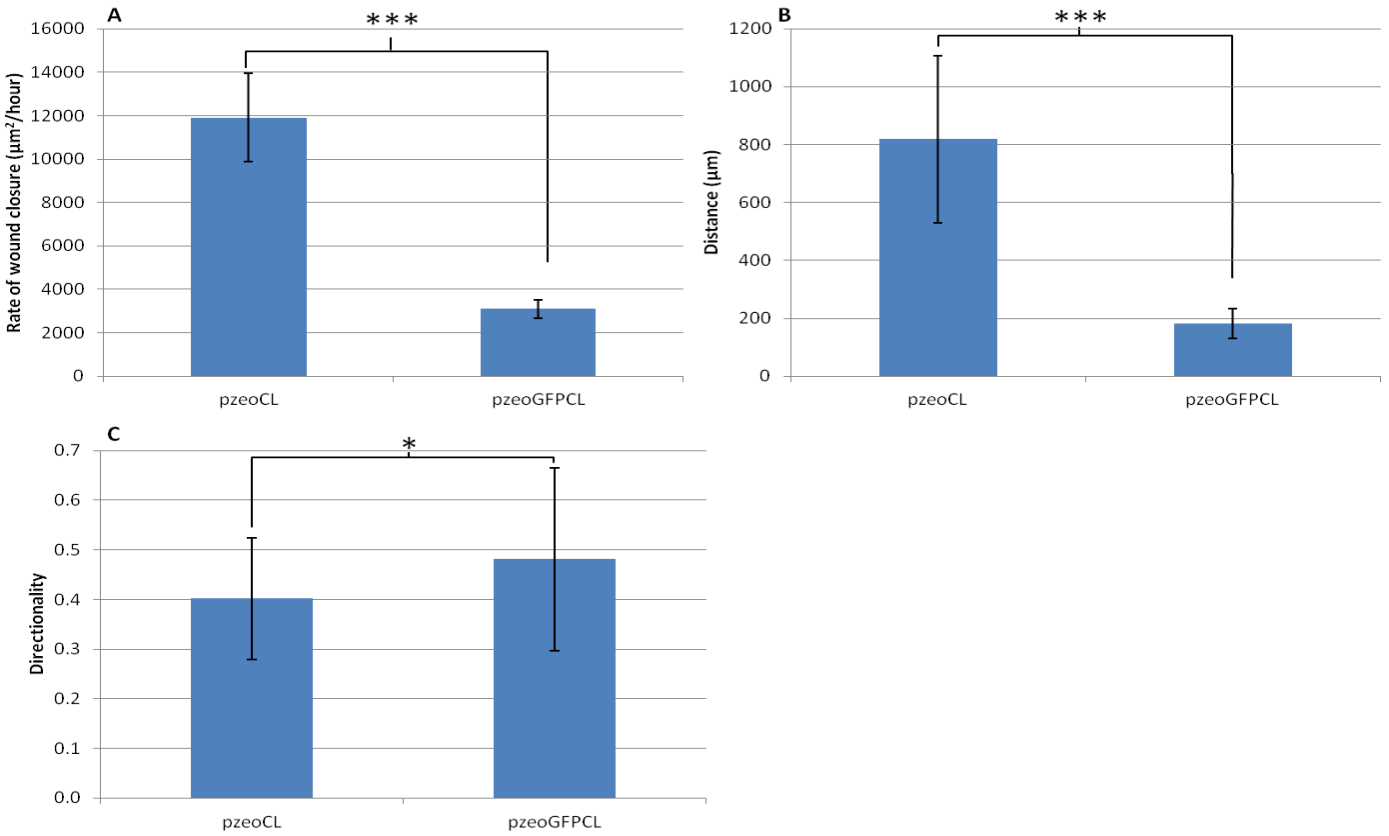


Figure 93. Graphs showing (A) averages of rate of wound closure (*n*=6), (B) distance travelled by individual cells (*n*=141), (C) directionality of individual cells (*n*=141). Scratch wounds were created in Rac1 KO cells stably transfected with either pzeoCL (Rac1) or pzeoGFPCL (GFP). Fluorescent timelapse microscopy was used to monitor the cells over a 48 hour period. Analysis of motility of individual cells was calculated using the *Manual Tracking* and *Chemotaxis and Migration* tool (ImageJ). Mean average values for each condition was calculated and plotted. Errors bars represent standard deviation.

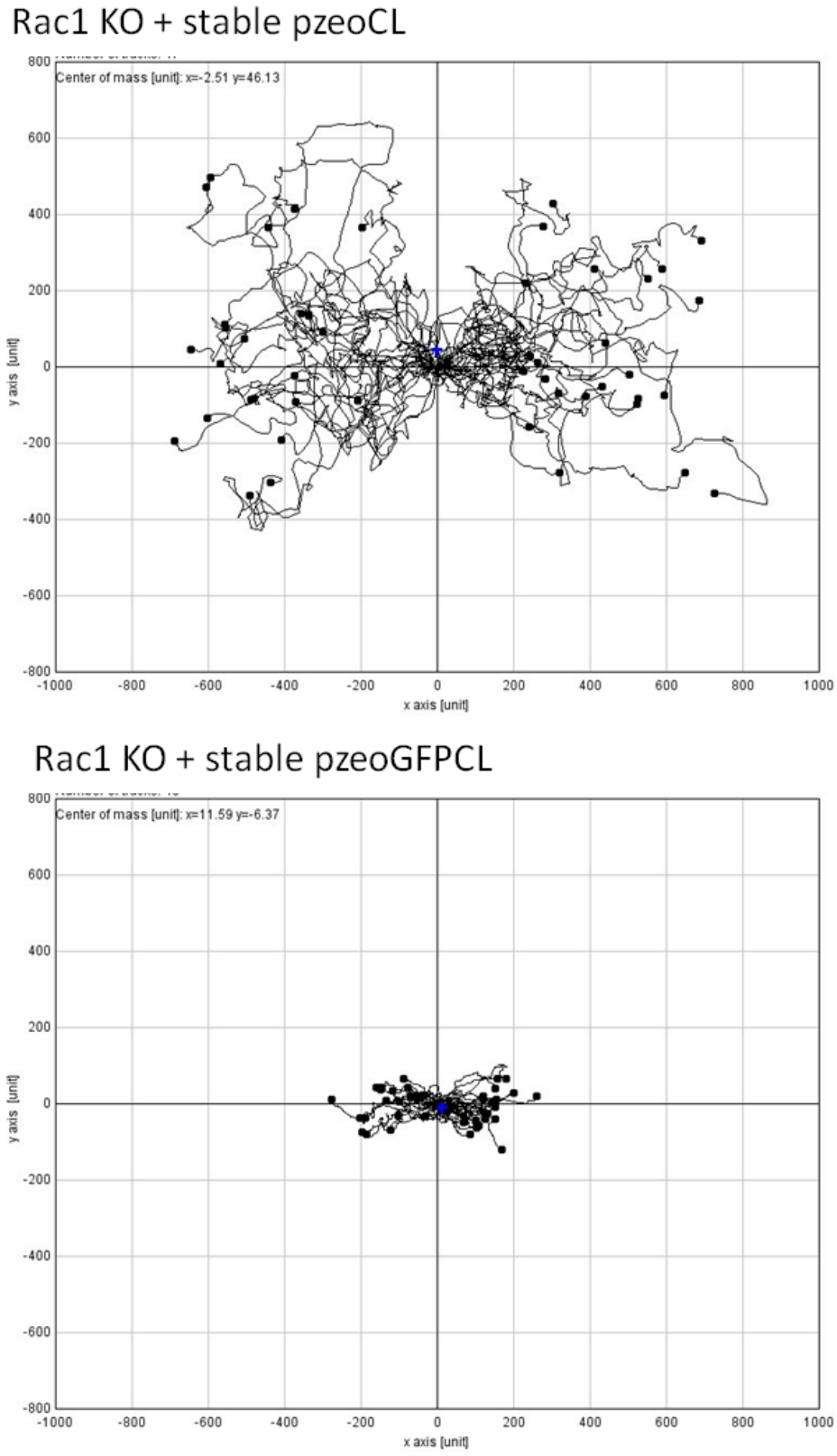


Figure 94. Trajectories of Rac1 KO mouse fibroblasts stably transfected with pzeoCL or pzeoGFPCL, during scratch wound assay. Scratch wound assays were performed in monolayers of Rac1 KO cells stably transfected with either pzeoCL or pzeoGFPCL. The scratches were monitored by timelapse microscopy. The trajectories of individual cells at the wound edge were followed using the *manual tracking* plugin and plotted using the *Chemotaxis and Migration* plugin of ImageJ.

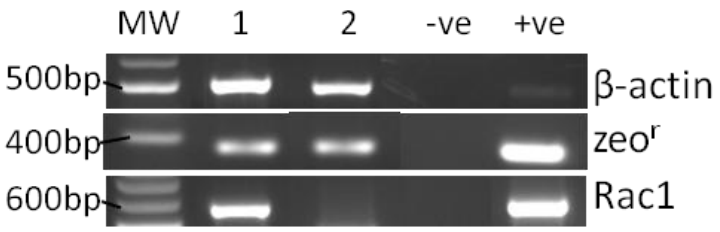


Figure 95. Agarose gel (2%) of PCR to amplify β-actin, Zeocin resistance gene (zeo^r) and Rac1 genes from cDNA synthesised from RNA extracted from Rac1 KO cells stably transfected with pzeoCL (lane 1), Rac1 KO cells stably transfected with pzeoGFPCL (lane 2), negative water control (lane 3), positive control (lane 4).

5.3.4 Comparison of motility between WT Rac1 expression and plasmid encoded Rac1

Plasmid encoded Rac1 was not able to restore WT-like wound closure such as observed in Rac1 control cells ($p < 0.0001$; Figure 96A). The Rac1 KO cells transiently (pRacGFPCL) and stably-expressing Rac1-GFPCL (pzeoCL) gave the highest rate of wound closure, however this rate was only approximately half of the rate seen for the Rac1 control cells. The Rac1 KO transiently expressing GFP-Rac1 (peGFPRac1) demonstrated a slower rate of wound closure than that demonstrated by the cells expressing Rac1-GFPCL ($p > 0.0001$).

While none of the plasmid constructs restored WT motility (comparable to that seen by Rac1 control cells) to the Rac1 KO cells with regards to rate of wound closure, the Rac1 KO cells transfected transiently and stably with plasmids expressing the construct Rac1-GFPCL showed increased motility at the individual cell level compared to Rac1 control cells (Figure 96B). Rac1 KO cells transiently expressing Rac1-GFPCL showed a moderate increase in individual cell motility compared to Rac1 control cells ($p < 0.05$), but when this construct was stably expressed there was an approximate two fold increase in distance travelled by individual cells ($p < 0.0001$). Interestingly, the distance travelled by Rac1 KO cells transiently expressing GFP-Rac1 was lower than that observed for Rac1 control cells ($p < 0.005$); however, the average distance was greater than that observed for the Rac1 KO cells (Figure 78).

The Rac1 control cells, Rac1 KO cells and the Rac1 KO cells transiently transfected with plasmids peGFPRac1 and peGFP all showed similar directionality (Figure 97). However, the Rac1 KO cells transiently and stably transfected with plasmids expressing Rac1-GFPCL showed decreased directionality ($p < 0.0001$).

Rac1 KO cells, when expressing plasmid-encoded Rac1 demonstrated the presence of lamellipodia in the direction in which they appeared to be moving, which is another good indication that Rac1 has restored motility to the fibroblasts via the canonical lamellipodia pathway and not some other signalling pathway (Figure 98).

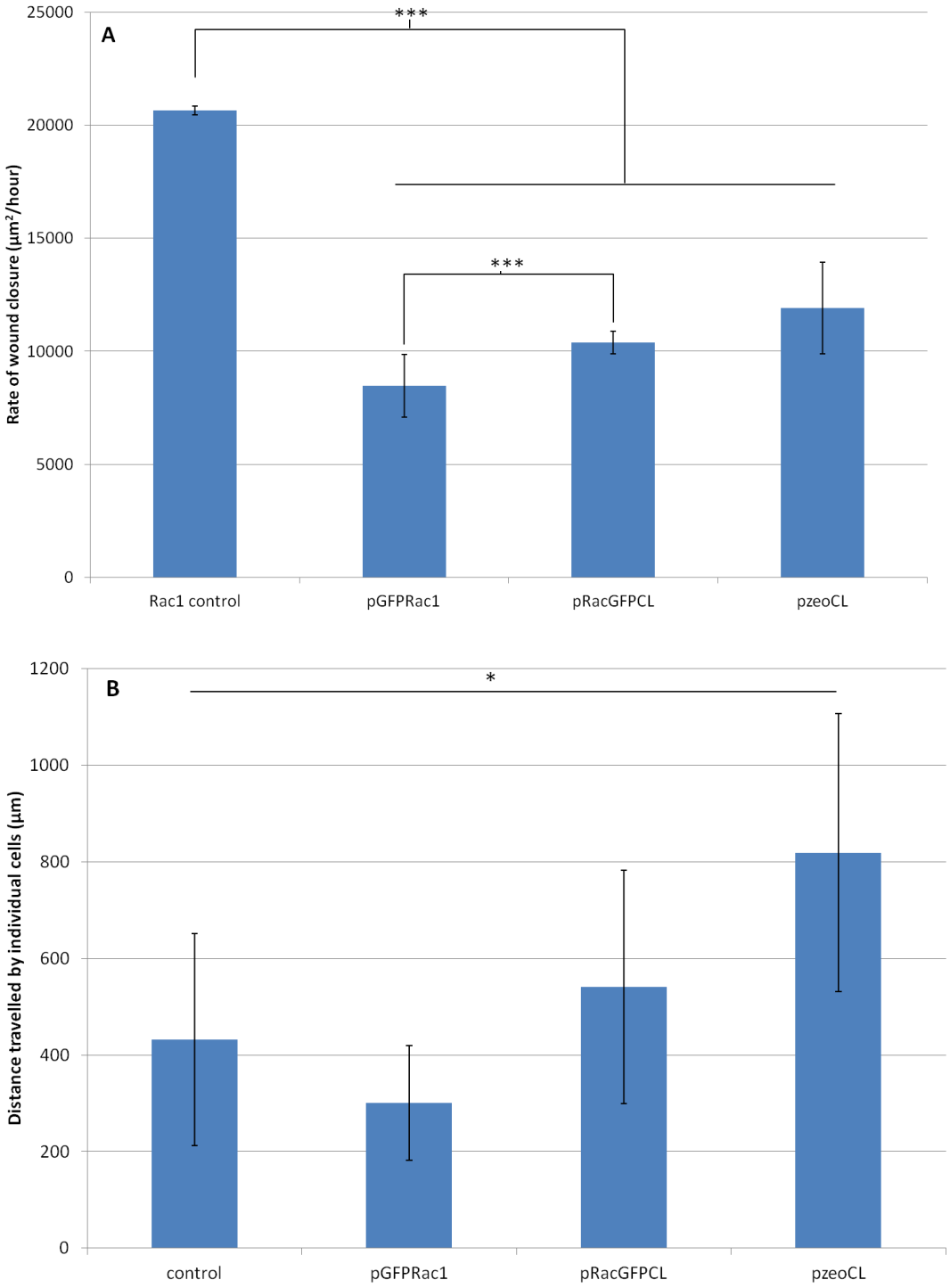


Figure 96. A) Graph comparing of rate of wound closure between Rac1 control cells and Rac1 KO cells transfected with Rac1-containing plasmids.

B) Graph comparing distance travelled by individual cells between Rac1 control cells and Rac1 KO cells transfected with Rac1-containing plasmids.

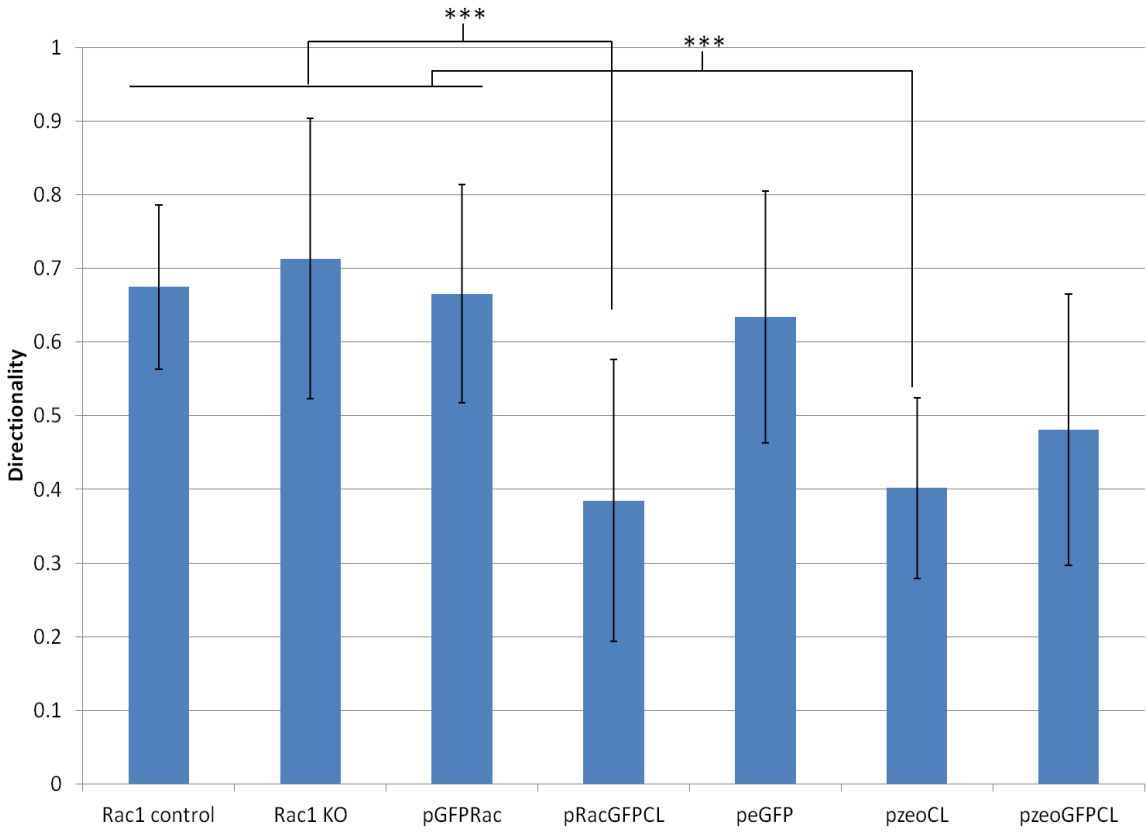


Figure 97. Graph comparing directionality of Rac1 control cells and Rac1 KO cells transfected with Rac1-containing plasmids.

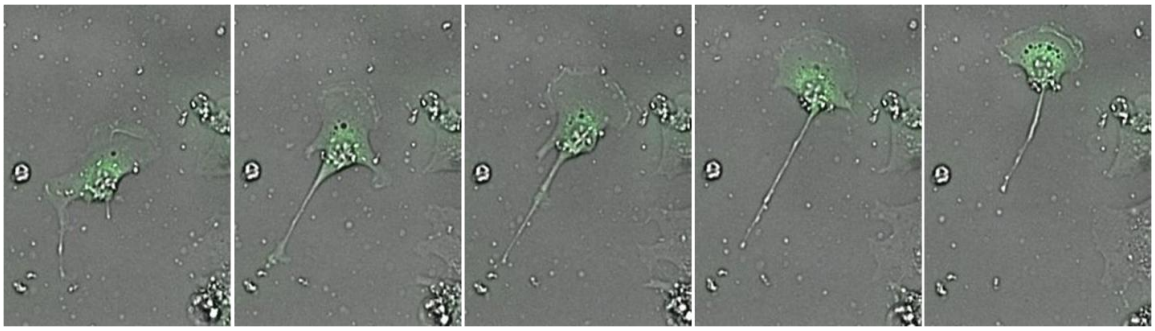


Figure 98. Lamellipodia formation. Rac1 KO cells were transiently transfected with pRacGFPCL and monitored using timelapse fluorescent microscopy.

5.4 Discussion

5.4.1 Affect of Rac1 knockout

5.4.1.1 Growth rate

The Rac1 KO fibroblast cell line used within this Chapter demonstrated reduced growth rate when compared to the Rac1 control cell line. Utilising DN mutants, Rac1 has previously been implicated in the regulation of cell cycle progression (Joyce *et al.* 1999; Olson *et al.* 1995; Westwick *et al.* 1997) and the reduction in proliferation in Rac1-null fibroblasts as seen within this work compared to previously obtained results (Guo et al. 2006; Vidali et al. 2006).

Vidali et al (2006) suggested that the reduced growth observed for both primary and immortalised Rac1 null Mouse embryonic fibroblasts (MEF) cells was due to reduced cell proliferation and not increased apoptosis rates (as determined using TUNEL and annexin V staining). Utilising Bromodeoxyuridine (BrdU) labelling, the group determined that this was due to reduction in DNA synthesis and that cells were shifted to the G1 phase with a reduction of cells in S phase, when compared to Rac1-WT control cells. The group determined that this observation was due to the depletion of Rac1 and not due to the treatments used within the study.

Guo et al (2006), also demonstrated that Rac1 deletion in primary MEFs resulted in slowed cell growth. However, the group demonstrated that the reduction in growth rate was related to increased apoptosis rates, as determined using annexin V staining in conjunction with FACS analysis. Further analysis utilising immunofluorescence staining of active caspase 3 confirmed this phenotype and also, re-introduction of Rac1 reconstituted normal growth rates. Guo et al suggested that the increase in apoptosis might be attributed to both anoikis-dependent (because Rac1 deletion reduced cell adhesion to fibronectin and the focal adhesion plaque formation) and anoikis independent pathways involved in cell survival.

5.4.1.2 Morphology

The Rac1 KO fibroblast cells used within this Chapter exhibited altered morphology when compared to the Rac1 control fibroblasts. Namely, the Rac1 KO cells were larger, and were more rounded in appearance. This was in contrast to that of other Rac1 null fibroblasts used within other studies (Guo et al. 2006; Monypenny et al. 2009; Vidali et al. 2006). Vidali et al's (2006) Rac1 null MEFs were shown to have different morphology when compared to the typical well-spread polygonal shape of Rac1-WT cells. The Rac1 null fibroblasts were generally smaller, most probably due to the lack of lamellipodia, and displayed predominantly elongated morphology, but triangular and small cell morphology was also observed. These different morphologies could represent the mixed population present as a result of the preparation of the conditioned Rac1 knockout. The group reported a 90 % depletion of Rac1 during the Cre-treatment, and therefore 10 % could display WT morphology and could account for the mixed morphological population. The Rac1 null MEFs studied by Guo et al (2006) and Rac1 knocked-down MEFs studied by Monypenny et al (2009) also displayed a contracted and elongated morphology, consistent with that reported by Vidali et al (2006). However, this morphology was in contrast to that observed within the work presented within this Chapter. It is possible that during the process of immortalisation of the Rac1 KO cell line utilised within this study, a specific clone with altered morphology had been selected and this differed to the MEFs used within these previous studies. In addition, genomic changes that occur throughout repetitive cell passage could have resulted in a Rac1 null cell line with altered morphology.

5.4.1.3 Motility

Despite the difference in morphology shown in Rac1-null MEFs used in previous studies and Rac1 KO cells used within this work, the reduction in Rac1 had similar effects on motility. Specifically, a significant difference in cell migration rates and rate of wound closure between Rac1-deficient and control fibroblasts (Monypenny et al. 2009; Vidali et al. 2006).

Vidali et al (2006), utilised scratch wound assays to establish the migratory ability of their Rac1 null MEFs. Although, the method used to measure rate of wound closure

differed between work presented within this chapter and that of Vidali et al's, in both cases the rate of wound closure for Rac1-deficient fibroblasts was approximately half of the rate demonstrated by Rac1-WT cells. As demonstrated within this study and that of Vidali et al, the tracking of individual cells also demonstrated that Rac1 null fibroblasts displayed reduced motility. When the velocity of individual cell migration was compared, the similarities were notable (Table 16). Similarly, it was also noted that tethering of cells to one another retarded migration.

Table 16. Velocity of individual cell movement, comparing Rac1 WT and Rac1 KO cells

	Individual cell velocity (μm/min)	
	Rac1 WT	Rac1 KO
This study	0.15 ± 0.076	0.07 ± 0.025
Vidali et al	0.16 ± 0.022	0.07 ± 0.004

Monypenny et al (2009), utilised short interfering RNAs (siRNA) to knockdown Rac1 in MEFs. The affect on the individual cell motility was assessed using chemotactic responses to a platelet derived growth factor (PDGF) gradient within a Dunn direct viewing chamber by timelapse microscopy. The group demonstrated that knockdown of Rac1 reduced the migratory capabilities of the cells to approximately 30 % of Rac1-WT cells. This reduction is in concurrence with the finding within this study, however, the reduction observed within this study was over 50 %. The reduction by 30 % could be attributed to that fact that some Rac1 might still be present in the cells where Rac1 had been knocked down using siRNA. Also to note is that the response to PDGF induced chemotaxis may be different to that of scratch wound induced migration.

During the scratch wound assay, the Rac1 KO cells demonstrated some migration into the wound space. This was in contrast to the function of Rac1 as determined by Nobes and Hall (1999) utilising microinjection of either DN Rac1 mutant protein

or expression vector encoding DN mutant. The group showed that cell movement of fibroblasts (rat EF) was inhibited by 98 % using DN mutants and concluded that Rac activity was essential for motility. Although the limited migration demonstrated within the work presented within this Chapter appeared to be due to cell proliferation at the wound edges, the cells may have displayed limited motility due to non-lamellipodial motility. Limited migration and chemotaxis was also observed in the more recent studies utilising Rac1-null MEFs (Monypenny et al. 2009; Vidali et al. 2006). It has been suggested that when Rac1 is depleted, the cells were still able to migrate via compensatory mechanisms, namely by the extension of filopodia and other thicker membrane extensions mediated by other GTPases, such as Cdc42 (Allen *et al.* 1998; Chen *et al.* 2000; Pankov *et al.* 2005) or RhoG. These observations were similar to that seen in dermal fibroblasts from gelsolin (downstream effector of Rac1) null mice (Azuma *et al.* 1998).

The results using Rac1 knockout should be more indicative of the function of Rac1 compared to studies using DN T17N mutants, which has reduced affinity for GTP and sequesters GEFs (Feig 1999). This is because these DN proteins may bind to multiple GEFs in a promiscuous manner, which may result in impaired general GTPase function, for example, other Rac isoforms, Cdc42 and other rho family members (Guo *et al.* 2003; Schmidt and Hall 2002).

The effect of Rac1 deficiency on other cell types has also been important in elucidating its function. Although Rac1-null macrophages are unable to produce membrane ruffles, they were able to migrate probably due to the presence of Rac2 which may contribute to motility mediation (Wells *et al.* 2004; Wheeler *et al.* 2006). On the other hand, it has been shown that neutrophils exhibit much reduced motility, reduced chemotaxis and no lamellipodia formation when both Rac1 and Rac2 were knocked-out (Sun *et al.* 2004).

5.4.1.4 Directionality

The work presented within this Chapter demonstrated no difference in the directionality of cell motility into the wound space displayed by the Rac1 KO compared to the Rac1 control fibroblast cells. No difference in directionality was

demonstrated by Monypenny et al's (2009) Rac1 null fibroblasts in response to PDGF chemotaxis. In contrast, Vidali's results indicated that Rac1-null MEFs demonstrated reduced directionality into a scratch wound compared to cells WT for Rac1. The work by Pankov et al (2005) sought to explore the differences in directional migration, where the cell rapidly moves between points and random migration, where the cells are allowed to explore their local environment. The group suggested a Rac1-based mechanism distinct from the mechanism controlling cellular chemotaxis that allows the cells to switch between the two modes of migration. The work used RNA interference to modulate the levels of Rac1 activity in fibroblasts (50 – 80 %). It was shown that lowering Rac1 activity switched the migration from random to directionally persistent in a process that did not require phosphatidylinositol 3 kinase activity. It was suggested that by decreasing Rac1 activity the production of peripheral lamellae was suppressed therefore causing the leading lamellae to become more pronounced and the subsequent mediation of directional migration. To explain the decrease in directional migration, as demonstrated by Vidali's Rac1 null cells, it was suggested that the complete loss of Rac1 had abolished the presence of a leading lamellae therefore motility preceded via an alternative mechanism. In the case of the work presented in this Chapter, neither an increase or decrease in directionality was observed between Rac1 KO and control cells therefore a combination of the two or a third hypothesis could be suggested. The use of a chemotactic gradient within the motility assay would affect the cellular responses involved in motility when compared to scratch wound induced motility, which could alter the need for specific directional movement. Differences in migration could be explained by differences in local chemotactic gradients that may have been present during the scratch wound assays, the difference in scratch wound assay themselves and differences in media or the cells used. The directionality is considered further below in respect to the introduction of plasmid-encoded Rac1.

5.4.1.5 Biochemical analysis

Previous work looked deeper into the cellular and biochemical differences in the Rac1 null cells and Rac1-WT fibroblast cells, namely, in the visualisation of F-actin,

spreading assays, and response to chemotactical gradients (Guo et al. 2006; Monypenny et al. 2009; Vidali et al. 2006). Work performed using Rac1 null fibroblasts (Guo et al. 2006; Monypenny et al. 2009; Vidali et al. 2006) and studies utilising DN Rac1 protein (Nobes and Hall 1999; Ridley et al. 1992) demonstrated that Rac1 is an absolute requirement for the formation of lamellipodia and dorsal and lateral ruffles. However, in contrast to work performed using DN mutants (Anand-Apte *et al.* 1997; Berrier *et al.* 2000; Clark *et al.* 1998; Fenteany et al. 2000; Nobes and Hall 1999; Price *et al.* 1998), the ability of Rac-null cells to perform chemotaxis, spread on fibronectin and ability to close wounds was only reduced (Guo et al. 2006; Vidali et al. 2006). It would be interesting to further investigate these points using the Rac1 KO cells utilised within this study and therefore compare these findings to previous work, however, for the application required within this Thesis this in-depth characterisation was considered surplus.

5.4.2 Restoration of motility via plasmid encoded Rac1

5.4.2.1 Transient transfection of Rac1 plasmids

Optimisation of transient transfection was performed to try to obtain high transfection efficiency. A high transfection efficiency was required to obtain a high number of Rac1-expressing cells and therefore enable visualisation of a large difference in rate of scratch wound closure. Optimisation was performed using two commercially available transfection reagents, Lipofectamine 2000 and FugeneHD. Although Lipofectamine 2000 typically gave higher transfection efficiencies, FugeneHD was utilised for transfections as the toxicity conferred was lower. Survival of the cells and sustained expression of Rac1 was vital during the timecourse of the timelapse experiment to maintain the phenotype conferred by the transfected gene. After optimisation, a sufficiently high transfection efficiency (approximately 40 %) was achieved. This transfection efficiency was high enough to visualise difference in cell motility for experiments using plasmids harbouring Rac1.

Interestingly, the transfection efficiency was higher in monolayers which had recently been scratched when compared to transfection of a typical monolayer. The scratching process may have caused mechanical deformation of the cell

membranes increasing their permeability to exogenous particles, namely the transfection complexes (Abbaci et al. 2008; Geddes et al. 2003). Another possibility was that the uncovering of the negatively charged tissue culture treated plastic, due to the scratching, attracted the plasmid complexes, which are surrounded by cationically charged transfection reagent. The attraction of plasmid complexes into the scratch wound area would have allowed the cells adjacent to the scratch to uptake the plasmid more easily.

5.4.2.2 Stable transfection of Rac1 plasmids

The use of stable cell lines sought to remove the heterogeneity observed in cell monolayers transiently transfected. By selecting cells which have integrated the plasmid DNA into their genome, a population of cells expressing target gene (Rac1 or eGFP) were obtained. As a consequence of using Rac1 stably transfected cell lines an increased rate of migration and wound closure was observed.

However, during the selection of the stably expressing cell lines, not all of the resistant colonies appeared to express GFP. The linearisation of the plasmid for incorporation into the host genome occurs randomly; therefore, in those stable cells not expressing GFP, the GFP could have been transected, thus allowing incorporation of full-length Zeocin resistance gene, but not the target gene. Linearisation of the plasmids by restriction digest could have decreased transection of target gene and increased efficiency of stable selection. Also to consider is that incorporation of the plasmids into the host genome occurs randomly and therefore other genes could have been disrupted during the incorporation process, this could have affected growth and ability to survive the selection procedure. Also to note is that during the incorporation into the genome parts of the plasmid can be replicated, cleaved and this may have affected the efficiency of GFP-expression.

5.4.2.3 Comparison of GFP-Rac1 and Rac1-GFP constructs

Two Rac1-containing plasmid constructs were tested for their ability to restore motility to Rac1 KO fibroblasts, namely pGFPRac1 and pRacGFPCL. pGFPRac1 was initially used as this construct had previously been used in transient transfection of Rac1 KO cells (Personal communication, Klemens Rottner). The Rac1-GFP construct

was created in anticipation of implementation of the amber suppression technology. The Rac1-GFP fusion would be a readout system to visualise whether read-through of the TAG codon had occurred. However, the function of this construct required testing prior to use with TAG mutagenesis. Two constructs were made, one where the polybasic and prenylation sequence were moved to the C-terminal of the Rac-GFP fusion (RKCL) and another where only the prenylation sequenced were moved (CL). It was hoped that moving the prenylation sequence to C-terminal of the Rac1-GFP fusion protein, would allow the Rac1-GFP fusion protein to be correctly tagged for association with the plasma membrane where the Rac1 activity is required for the actin-based formation of lamellipodia and consequent cell motility. If the CLLL signal was maintained at the C-terminal of Rac1 within the Rac1-GFP fusion, cleavage of GFP would occur and therefore GFP presence would not be representative of Rac1 levels. In addition, the C-terminal extension (i.e. GFP) might inhibit processing of the CLLL sequence. It was shown that the presence of the polybasic region and prenylation sequence at the C-terminal was not optimal. However, it was shown that the Rac1-GFPCL was functional and able to restore motility to the Rac1 KO fibroblasts.

The technique of targeting proteins to plasma membrane using prenylation sequences has been used previously (Aronheim *et al.* 1994; Hancock *et al.* 1990). In addition, a commercially available plasmid peGFP-F has been constructed with the 20 amino acid farnesylation signal from c-Ha-Ras fused to the C-terminus of eGFP (Clontech). This plasmid produces a C-terminal GFP tagged protein of interest which is directed to the plasmid membrane.

5.4.2.4 Rate of wound closure

The rate of wound closure using plasmid-encoded Rac1 (for both GFPRac1 and RacGFPCL constructs) was increased when compared to cells expressing eGFP. However, using plasmid-encoding Rac1 did not fully restore WT-like motility to the Rac1 KO cells concerning rate of wound closure. The rates obtained were 41 %, 50 % and 58 % of that observed for Rac1 control cells, for pGFPRac1, pRacGFPCL and stable pzeoCL respectively. The difference could be attributed to the different transfection efficiencies of the two transiently transfected plasmids, levels of

expressed protein or effectiveness of the fusion protein produced. It is interesting that the C-terminal GFP fusion protein was most successful because the construction of this fusion was most complex. It was encouraging that the stably expressing cell lines gave the greatest rate of wound closure, most probably because the entire cell population were contributing to the closure of the wound.

This difference in observed motility could be explained by the use of the noncognate promoter driving the expression of the Rac1 construct. The expression of Rac1 was constitutive and was therefore not subject the cognate transcriptional control seen for the Rac1 gene within the Rac1 control cells. The Rac1, both endogenously and exogenously produced, should have been subject to protein-based regulation, however, the presence of the GFP tag could have perturbed the recognition of Rac1 by GEFs and other interacting proteins; altered the GTPase's ability to bind or hydrolyse GTP; perturbed PM localisation and thereby altered the action and subsequent signalling related to Rac1 protein.

Guo et al (2006) have demonstrated that reconstitution of Rac1 into a Rac1 null fibroblast via a retroviral construct was possible. Although not a direct fusion protein, the expression of Rac1 was linked to expression of eGFP for visualisation via a bicistronic approach. The group showed that the loss of lamellipodia, reduction in cell spreading and increase in spontaneous apoptosis due to Rac1 knockout could be reconstituted via exogenous Rac1 expression. However, the group did not perform any motility assays to ascertain whether Rac1 deficient or Rac1 reconstituted cells were capable of migration.

5.4.2.5 Individual cell motility

While none of the plasmid constructs restored WT motility (comparable to that seen by Rac1 control cells) to the Rac1 KO cells with regards to rate of wound closure, the Rac1 KO cells transfected with plasmids expressing the construct Rac1-GFPCL demonstrated increased motility at the individual cell level. More specifically, 125 % and 190 % of that of Rac1 control cells, for transient and stable expression respectively. Rac1 KO cells transiently transfected with pGFPRac1

demonstrated individual cell motility approximately 70 % of that of Rac1 control fibroblasts.

The measurement of rate of wound closure and individual cell motility captured different modes of motility. The results suggest that the motility observed in individual cells was not efficiently translated into the mass movement of the monolayer to close the scratch wound. The movement of individual cells away from wound edge was not captured when measuring rate of wound closure, this was because measurement was taken using the wound front that was created by the mass of cells of the monolayer. This could explain the difference in rate of wound closure and the motility observed at the individual cell level. As mentioned previously, the use of a noncognate promoter to control expression of Rac1 would have perturbed the regulation of Rac1. This could have upset the ability of the cells to orchestrate the directed cell movement to close the wound; this loss of directionality is discussed further below in section 5.4.2.6.

Also to consider is that during the experiments using transient expression of Rac1-containing constructs, not all the cells were GFP-positive therefore the population of cells was heterogeneous. There were GFP-negative cells, which would probably display migration similar to Rac1 KO cells, cells expressing non-detectable levels of GFP and those that were clearly GFP-positive, each displaying various abilities to migrate. In addition to the limited wound closure ability of the GFP-negative cells, the cells could have interrupted the migration of those expressing Rac1.

5.4.2.6 Directionality

The directionality of migration for Rac1 KO, Rac1 control cells and cells transiently expressing GFP-Rac1 or eGFP were similar, more specifically ~0.7. This indicated that the movement of the cells was directed into the wound space. However, the increase in individual cell motility when compared to rate of wound closure indicated the movement of the cells expressing Rac1-GFPCL was not directed into the wound space. This was also suggested by the reduced directionality shown by the Rac1 KO cells transiently and stably expressing the Rac1-GFPCL construct (~0.4; Figure 97). As mentioned above, the constitutive promotion of Rac1 expression

could result in very high expression levels of Rac1 compared to WT expression. In addition, the expression was not under control of the cognate transcriptional regulation. If the random migration seen was due to overexpression of Rac1, this may support Pankov's hypothesis that decreased Rac1 levels induces directional migration in fibroblasts (Pankov et al. 2005). Using Pankov's reasoning, it can be suggested that cells containing high levels of Rac1 are not responsive to polarisation of the cell, lots of alternative lamellae initiation sites are created, therefore movement occurs in varying directions depending on which lamellae is predominant. Although measurements of levels of Rac1 expression would need to be ascertained to determine whether this hypothesis can be extrapolated.

It has been suggested that during chemotaxis to PDGF stimulation an initial upregulation of Rac1 activity occurs at PDGF stimulation, causing mobilisation of the actin polymerisation machinery which allows the cells to make the transition from quiescent state to a motile state (Monypenny et al. 2009). After this initial burst of Rac1 activity, the levels are downregulated to prevent excessive activation, and subsequent local alterations in Rac1 activity maintain the steady treadmilling of actin cytoskeleton during migration. Although the creation of a wound does not initiate the chemotactic response, it is possible that a similar Rac1 regulation occurs in response to the wound which enables the cells to migrate into the wound space. Therefore, the initial and continued expression of Rac1 from the plasmid constructs could cause excessive Rac1 signalling. Thus, causing non-WT migration activity into the wound and this could explain the increase in individual cell motility compared to the rate of wound closure.

5.5 Conclusion

The Rac1 KO fibroblast cell line was shown to have decreased motility when compared to Rac1 control fibroblasts. This was evident by the decreased rate of wound closure and decreased motility of individual cells, as determined by scratch wound assays. These data addressed the first aim of this work. To address the second aim, a selection of plasmid constructs were designed, created and used to transfect the Rac1 KO cells. These constructs encoded either a GFP-Rac1 or a Rac1-GFP fusion protein. Both via transient and stable transfection, it was demonstrated that plasmid-encoded Rac1 could restore motility to the Rac1 KO cells. The restoration of motility was evident by an increased rate of wound closure and motility of individual cells. When the Rac1-GFP construct (RacGFPCL) was introduced stably into Rac1 KO cells, the cells displayed the largest restoration of cell motility. However, the rate of wound closure was only 58 % of that demonstrated by endogenous Rac1 expression, as seen with the Rac1 control cells. Interestingly, the motility of individual Rac1 KO cells stably expressing Rac1 was shown to be 190 % of that displayed by Rac1 control cells and this was as a result of a loss of directionality (0.4 compared to 0.7 displayed by Rac1 control cells).

The work presented within this chapter has demonstrated the development of a functional mammalian cell readout system. This provided the basis for further investigation into whether this readout system could be used in the application of amber suppression to control the motile phenotype of a mammalian cell using the supplementation of NAA. The development of this technique is the subject of Chapter 6.

6 CHAPTER 6: UTILISATION OF THE AMBER SUPPRESSION TECHNOLOGY TO CONTROL A FUNCTIONAL, MAMMALIAN CELL READ OUT SYSTEM

6.1 Introduction

Chapter 5 demonstrated the effective utilisation of scratch wound assays in conjunction with fluorescent timelapse microscopy to monitor cellular motility. It was also demonstrated within Chapter 5 that Rac1 KO fibroblasts display reduced motility (compared to Rac1 control cells) and that plasmid-encoded Rac1 could restore this motility, both transiently and using stably transfected cell lines. These results formed a cellular readout system, namely, the modulation of cellular motility controlled by the absence (Rac1 KO cells, limited motility) and presence of Rac1 (encoded via plasmid, increased motility). Chapter 6 aims to utilise the cellular readout system established within Chapter 5 and apply this method in an attempt to modulate cellular motility via implementation of amber suppression technology, as described in Chapter 4.

Therefore, the work presented within this Chapter sought to:

- Produce a selectable single plasmid that encoded the orthogonal machinery (*Bs*tRNA$^{Tyr}_{CUA}$, azido specific aminoacyl tRNA synthetase (azRS)) and target gene (RacGFPCL construct containing a TAG mutation) required in the implementation of amber suppression technology within mammalian cells
- Investigate whether it was possible to control a mammalian cellular phenotype (i.e. migration) as a function of the presence/absence of non-natural amino acid utilising the amber suppression technology

Firstly, a complex multi-step cloning strategy was designed and implemented in order to create a one-plasmid system required for the application of amber suppression within mammalian cell lines. More specifically, a plasmid harbouring a bicistronic cassette encoding the two genes (azRS and racGFP) separated by an Internal Ribosome Entry Site (IRES) within the multi-cloning site (MCS) of a Zeocin-selectable plasmid, and the insertion of the *Bs*tRNA$^{Tyr}_{CUA}$ (along with internal promoter) into a site upstream of the MCS. Secondly, this plasmid was tested for function within the Rac1 KO cell lines (along with Rac1 control cells and HEK293 cells), utilising GFP tagged protein and fluorescent microscopy to ascertain function (along with RT-PCR). A TAG mutation was introduced in to a chosen position within the Rac1 gene and the ability of amber suppression to control cellular motility was ascertained.

6.2 Materials and methods

6.2.1 Molecular biology

Molecular biology was performed as defined in Chapter 2

6.2.2 Mammalian cell culture

Mammalian cell culture was performed as defined in Chapter 5.

6.2.3 Medium supplementation with non-natural amino acids

Iodophenylalanine (iodoF; BACHEM) or azidophenylalanine (azF; BACHEM) were prepared either by addition to growth media using sonication to dissolve or pre-dissolved in NaOH as set out below.

NAA were dissolved (to varying final concentrations) in growth medium prior to addition to the wells transfected with TAG-containing plasmids. Sonication was performed in order to dissolve the NAA and was thought to sterilise the solution. This was typically 3 x 5 minute sonications, with mixing in between sonication steps. Sonication was performed at 37 °C. Dissolved NAA was used immediately and made fresh for each experiment.

Stock solutions of 100 mM iodoF or azF were created by dissolving in 200 mM NaOH. Solutions were sterile filtered (0.2 µm syringe sterile filters) and stored at 4 °C wrapped in aluminium foil to reduce exposure to light and subsequent photoreactions.

Addition of NAA during experiments occurred 2 – 24 hours after transfection. The volume of media in the 24 wells was increased to 1 ml and the specified concentration of NAA added. Concentrations used in experiments ranged from 0.5 – 7.5 mM. For experiments minus NAA, the same volume of 200 mM NaOH was added as a control.

6.2.4 DNA Electroporation

For electroporation of DNA into mammalian cells an ECM 630 Exponential Decay Wave Electroporation System (BTX) was used. Cells to be electroporated were removed from culture flask, pelleted and counted as described in Chapter 5. The cells were resuspended at 4 x10^6 / ml and 30 µg DNA was added to the cells. 800 µl of cells/DNA was added to a 4 mm cuvette (BTX) on ice. Cells were electroporated using 250 – 350 V, for approximately 30 ms using 75 Ω resistance/2000 µF capacitance. Cells were then recovered from the cuvette and added directly to well plates. The cuvettes were washed out with medium to ensure all cells were collected.

6.2.5 Liposomal transfection

Liposomal transfections were performed using Lipofectin®. Lipofectin® reagent is a 1:1 (w/w) liposome formulation of the cationic lipid N-[1-(2,3-dioleyloxy)propyl]-n,n,n-trimethylammonium chloride (DOTMA) and dioleoyl phophotidylethanolamine (DOPE) and was purchased from Invitrogen. The DNA/liposomal mixture was prepared as suggested by Manufacturers. Briefly, DNA (1 µg /well) and Lipofectin (2 -20 µl) were diluted in 100 µl OptiMEM medium separately. The diluted DNA and Lipofectin were then combined at a ratio 1 : 2 to 1 : 20 µg DNA to µl Lipofectin and incubated for 15 – 45 minutes (RT). The cells to be transfected were washed with 2 x 2 ml PBS and then 0.8 ml growth medium minus FBS was added. Complexes (200 µl) were then added to cells. Cells plus Lipofectin/DNA were then incubated at 37°C. Approximately 5 hours post transfection an equal volume of growth medium containing 20 % FBS was added.

6.3 Results

6.3.1 Creation of a single plasmid system for use in amber suppression in mammalian cells

To optimise transfection and ultimately the success of amber suppression in Rac1 KO cells a single plasmid system harbouring all genes required was designed (Figure 99). After consideration it was decided that the azF-specific synthetase (azRS) and Rac1-GFP construct would be expressed using the same promoter (constitutive CMV promoter). The two genes were separated by an internal ribosome entry site (IRES) which created a ribosome binding site within the mRNA transcript, thereby allowing expression of the two genes under the same promoter (Houdebine and Attal 1999). The Zeocin resistance gene would be driven by a separate promoter and used to select stably transfected cell lines. The *Bs*tRNA$^{Tyr}_{CUA}$ contained an internal promoter in its sequence and therefore could be placed within any position within the plasmid sequence. The *Bs*tRNA$^{Tyr}_{CUA}$ sequence was preceded by the 5'addition of the human tRNA^{Tyr} leader sequence which was included to improve its expression (Ye et al. 2009; Ye et al. 2008).

The pBluescript-IRES plasmid (a gift from Dr Xiao-Qing Wei, School of Dentistry, Cardiff University), contained the IRES sequence inserted into the middle of MCS between the *Bam*HI and *EcoR*I sites (Figure 100). To create the azRS-IRES-RacGFP construct, the RacGFP construct was inserted into the *EcoR*I and *Xho*I sites of pBluescript-IRES, followed by the addition of azRS into the *Xba*I and *Not*I sites (see section 6.3.1.1, 6.3.1.2 and Figure 101).

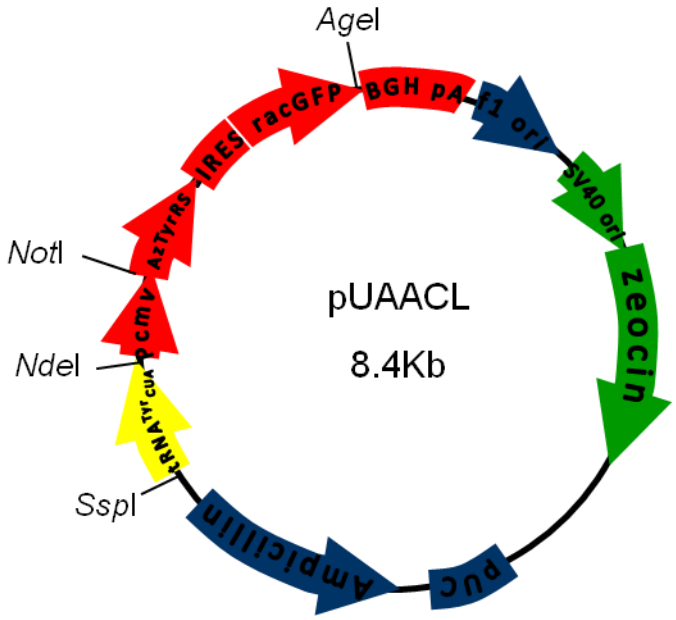


Figure 99. Plasmid map of a single plasmid system harbouring all genes required for amber suppression, pUAACL. The CMV promoter (pCMV, red) drives the expression of the azFs-specific aminoacyl tRNA synthetase (azTyrRS, red) and the target Rac-GFP construct (red). The orthogonal *Bs*tRNA (*Bs*tRNA$^{\boldsymbol{Tyr}}_{\boldsymbol{CUA'}}$, yellow) contains an internal promoter sequence and was preceeded by the human tRNA$^{\boldsymbol{Tyr}}$ leader sequence.

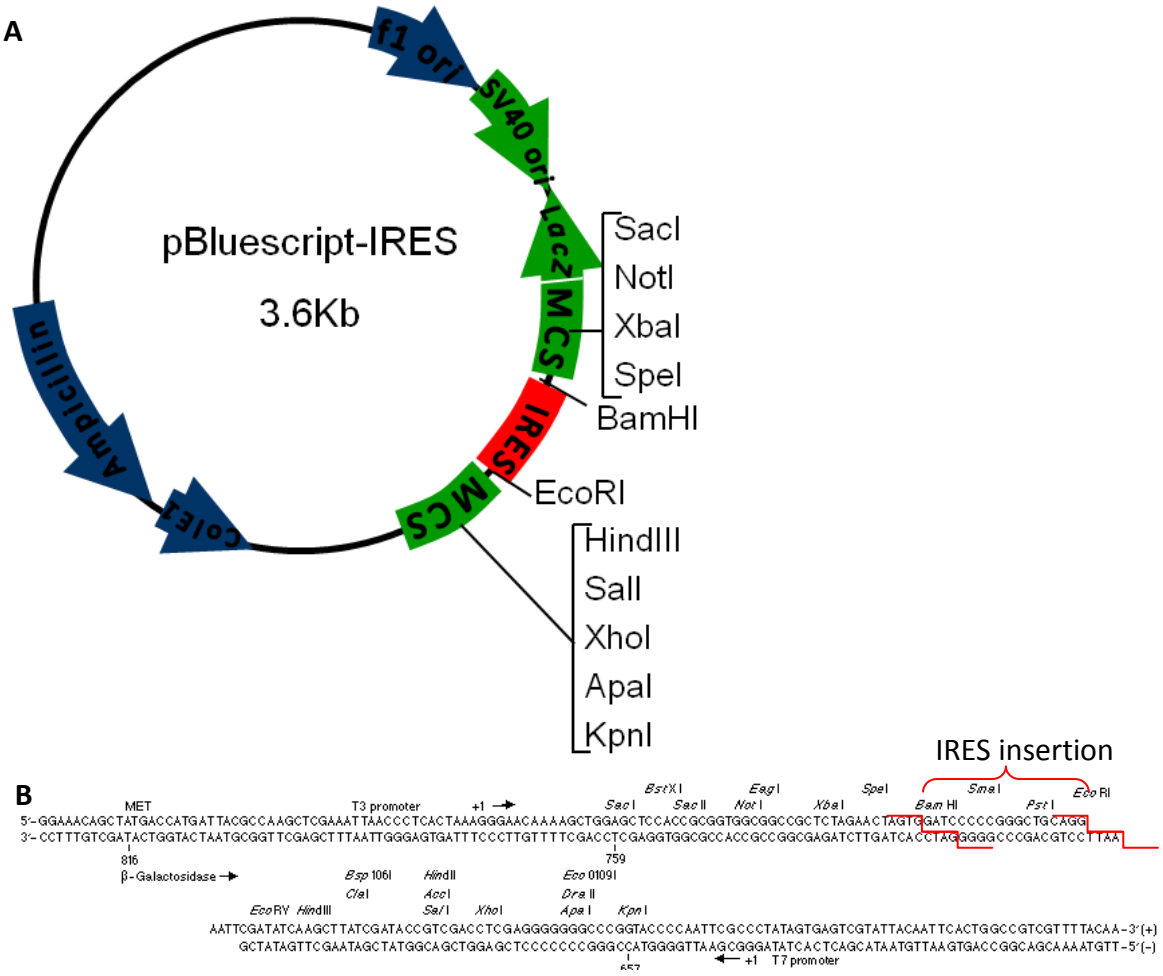


Figure 100. A) Plasmid map of pBluescript-IRES

B) MCS showing insertion of IRES sequence.

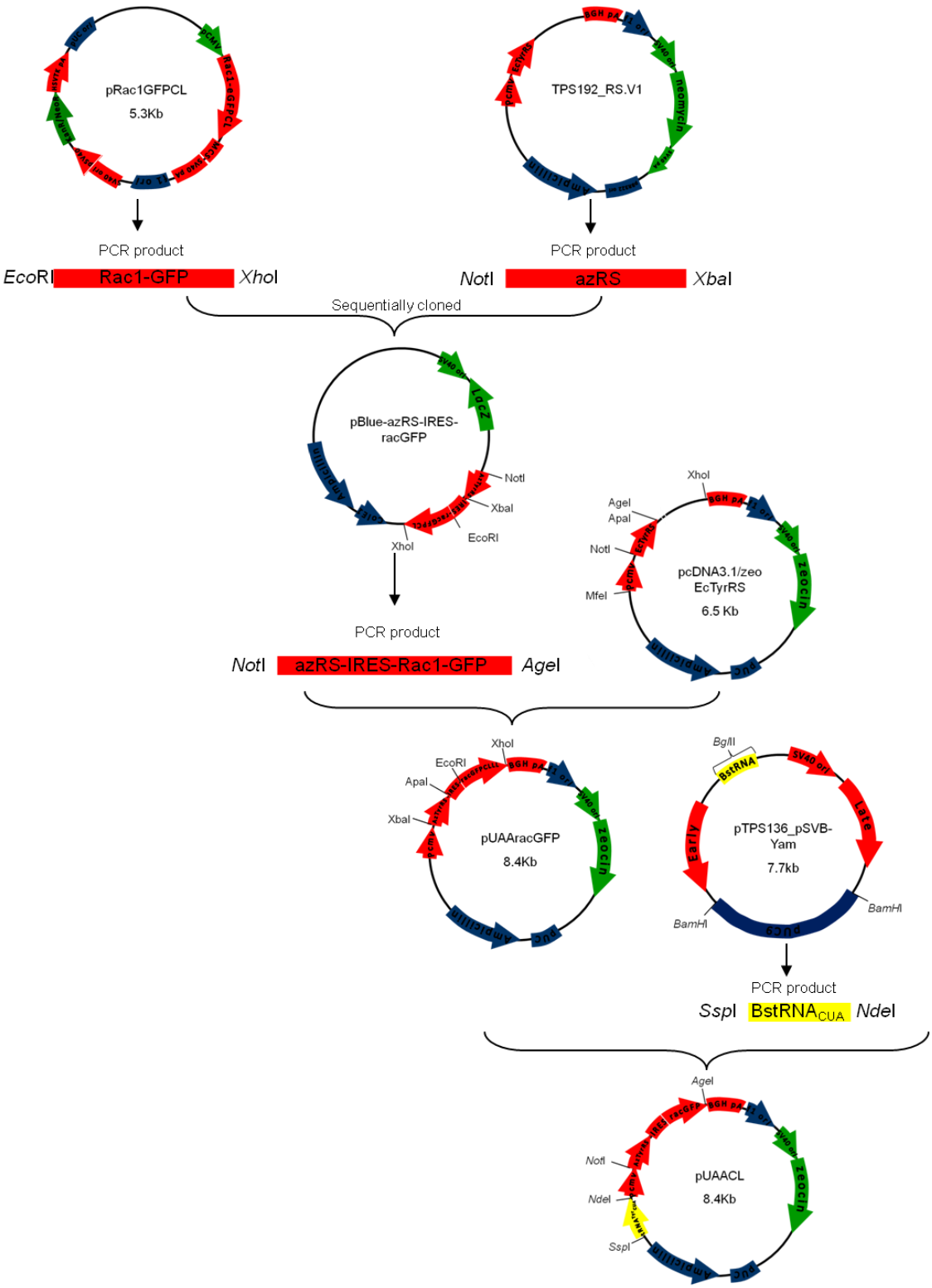


Figure 101. Plasmid cloning scheme for creation of a single plasmid system harbouring all genes required for amber suppression, pUAACL. To create the azRS-IRES-RacGFPCL construct, the RacGFPCL construct (from pRac1GFPCL) was inserted into the *EcoR*I and *Xho*I sites of pBluescript-IRES, followed by the addition of azRS into the *Xba*I and *Not*I sites. This azRS-IRES-RacGFPCL cassette was then inserted into the *Not*I and *Age*I sites of pcDNA3.1/zeoRS, followed by insertion of the *Bs*tRNA$^{\boldsymbol{Tyr}}_{\boldsymbol{CUA}}$ between the *Ssp*I and *Nde*I sites.

6.3.1.1 Cloning RacGFPCL into pBluescript-IRES

The RacGFPCL construct was amplified from pRacGFPCL using primers that added a 5' *Eco*RI site (racGFPIRES004) and a 3' *Xho*I (racGFPIRES005) to the PCR product (Appendix A2). The digested PCR product was ligated into complementarily digested and dephosphorylated pBluescript-IRES. Colonies were screened for insert using colony PCR with the primers rac FP and eGFP RP. Products of size 1.4 Kb corresponded to successful incorporation of RacGFPCL into pBluescript-IRES (Figure 102). Colonies giving no PCR product did not harbour the RacGFPCL insert and therefore were not used in further analysis. The DNA from four positive clones was isolated and the insertion position was sequenced to confirm successful insertion of the RacGFPCL construct. One plasmid, harbouring the correct sequence was used in subsequent cloning. This created plasmid pBluescript-IRES-RacGFPCL (Figure 101).

6.3.1.2 Cloning azRS into pBluescript-IRES-RacGFP

The synthetase gene specific for azF (azRS) was amplified from the plasmid TPS192_RS-V1 (a kind gift from Dr Thomas Sakmar, Rockefeller University) using primers that added a *Not*I (azRSIRES001) and *Xba*I (azRSIRES002) site to the 5' and 3' ends respectively (Appendix A2). The digested purified PCR product and complementarily digested, dephosphorylated and gel extracted pBluescript-IRES-RacGFPCL plasmid were ligated together. Colonies were screened for azRS insert using T7 and T3 primers, sites of which flank the MCS in pBluescript-IRES. A PCR product of 2.7 Kb indicated the insertion of azRS into pBluescript-IRES-RacGFPCL plasmids (Figure 103). A product smaller than the expected product of 2.7 Kb (2 Kb) indicated that these colonies harboured plasmid without insertion of RacGFPCL (most probably amplification of IRES sequence present in original pBluescript-IRES plasmid). The DNA from the four clones was isolated and the insertion position was sequenced to confirm successful cloning. One plasmid, harbouring the correct sequence was used in subsequent cloning. This created plasmid pBluescript-azRS-IRES-RacGFPCL (Figure 101).

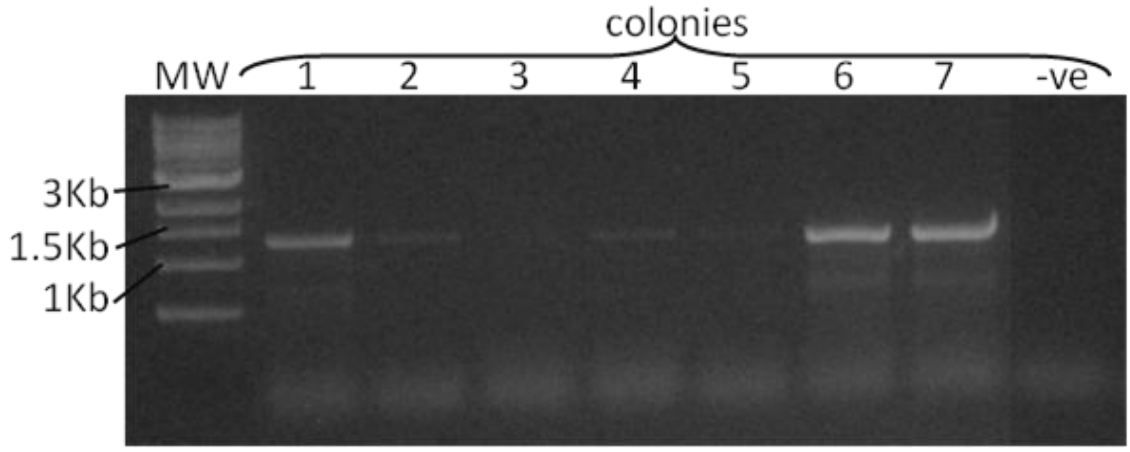


Figure 102. Agarose gel of colony PCR to screen for insertion of RacGFPCL into pBluescript-IRES. 7 randomly picked colonies (lanes 1-7) were screened for insert using primers rac FP and eGFP RP. Product of size 1.4 Kb corresponded to successful incorporation of RacGFP into pBluescript-IRES. MW was 1 Kb ladder. –ve control = colony from plate transformed with un-ligated plasmid + DNA insert.

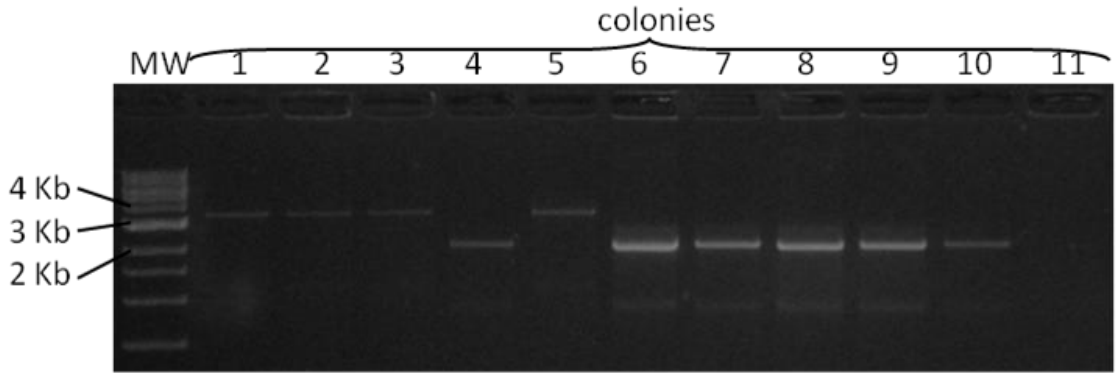


Figure 103. Agarose gel of colony PCR to screen for insertion of azRS into pBluescript-IRES-RacGFPCL. 10 randomly picked colonies (lanes 1-10) and negative control colony (lane 11; colony from plate transformed with un-ligated plasmid + DNA insert) were screened for azRS insert using T7 and T3 primers, sites of which flank the MCS in pBluescript-IRES. A PCR of product of 2.7 Kb indicated the insertion of azRS into pBluescript-IRES-RacGFPCL. MW was 1 Kb ladder.

6.3.1.3 Cloning azRS-IRES-RacGFPCL into pcDNA3.1/zeo-RS

This entire azRS-IRES-RacGFPCL construct was then cloned into the *Not*I and *Age*I site of pcDNA3.1/zeo-RS. PCR was used to amplify the azRS-IRES-RacGFPCL construct using primers (AzRSIRES001 and racGFPIRES007) that added a *Not*I and *Age*I to the 5' and 3' ends of the PCR product respectively. Complementarily digested and prepared PCR products and pcDNA3.1/zeo-RS were ligated. Colonies were screened for the azRS-IRES-RacGFPCL insert using primers CMV for (plasmid specific) and eGFP RP (insert specific). A PCR of product of approximately 3.28 Kb indicated the insertion of azRS-IRES-RacGFPCL into pcDNA3.1/zeo-RS plasmid (Figure 104). Colonies that gave PCR products of lower molecular weight and those that gave no product were not taken forward for further analysis. DNA from four colonies determined to harbour the azRS-IRES-RacGFPCL construct (by colony PCR) was isolated and subjected to DNA sequencing. Sequencing confirmed the insertion of azRS-IRES-RacGFPCL into pcDNA3.1/zeo-RS.

The sequencing for each plasmid demonstrated the presence of a few mutations in the DNA sequences. For the plasmid harbouring the least mutations, analysis of the translated protein sequence (ExPASy Translate tool) demonstrated that the mutations were a silent glycine (GGT→GGC) mutation at position 191 and a silent leucine (TTA→CTA) mutation at position 410 within azRS and Ser3→Gly (AGC→GGC) mutation in GFP (Figure 105). The silent mutations were not detrimental to protein sequence and the Ser3Gly mutation within GFP was tolerated therefore, this plasmid DNA was utilised for subsequent cloning.

This gave a plasmid conferring Zeocin resistance, the synthetase specific for azF and the RacGFPCL construct, pUAAracGFP (Figure 101).

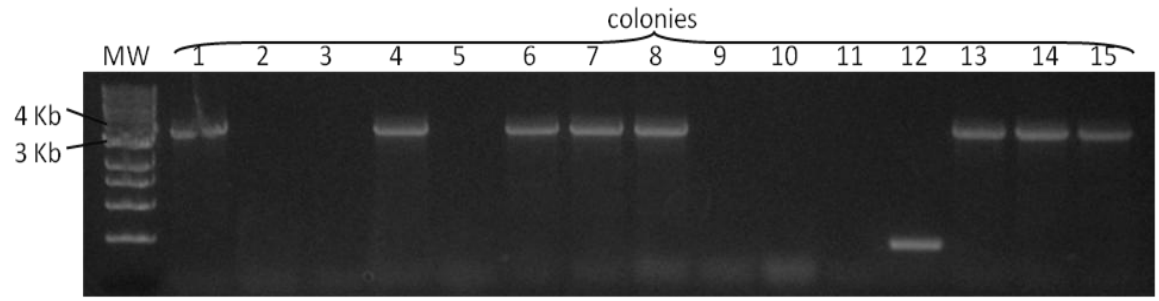


Figure 104. Colony PCR of azRS-IRES-RacGFPCL insertion into pcDNA3.1/zeo. 15 randomly picked colonies (lanes 1-15) were screened for azRS-IRES-RacGFPCL insert using primers CMV for (plasmid specific) and eGFP RP (insert specific). A PCR product of approximately 3.28 Kb indicated the insertion of azRS-IRES-RacGFPCL into pcDNA3.1/zeo-RS plasmid. No colonies were present on the agar plate containing *E. coli* cells transformed with un-ligated digested pcDNA3.1/zeo + azRS-IRES-RacGFPCL insert, therefore there was no –ve control colony. MW was 1 Kb ladder (NEB).

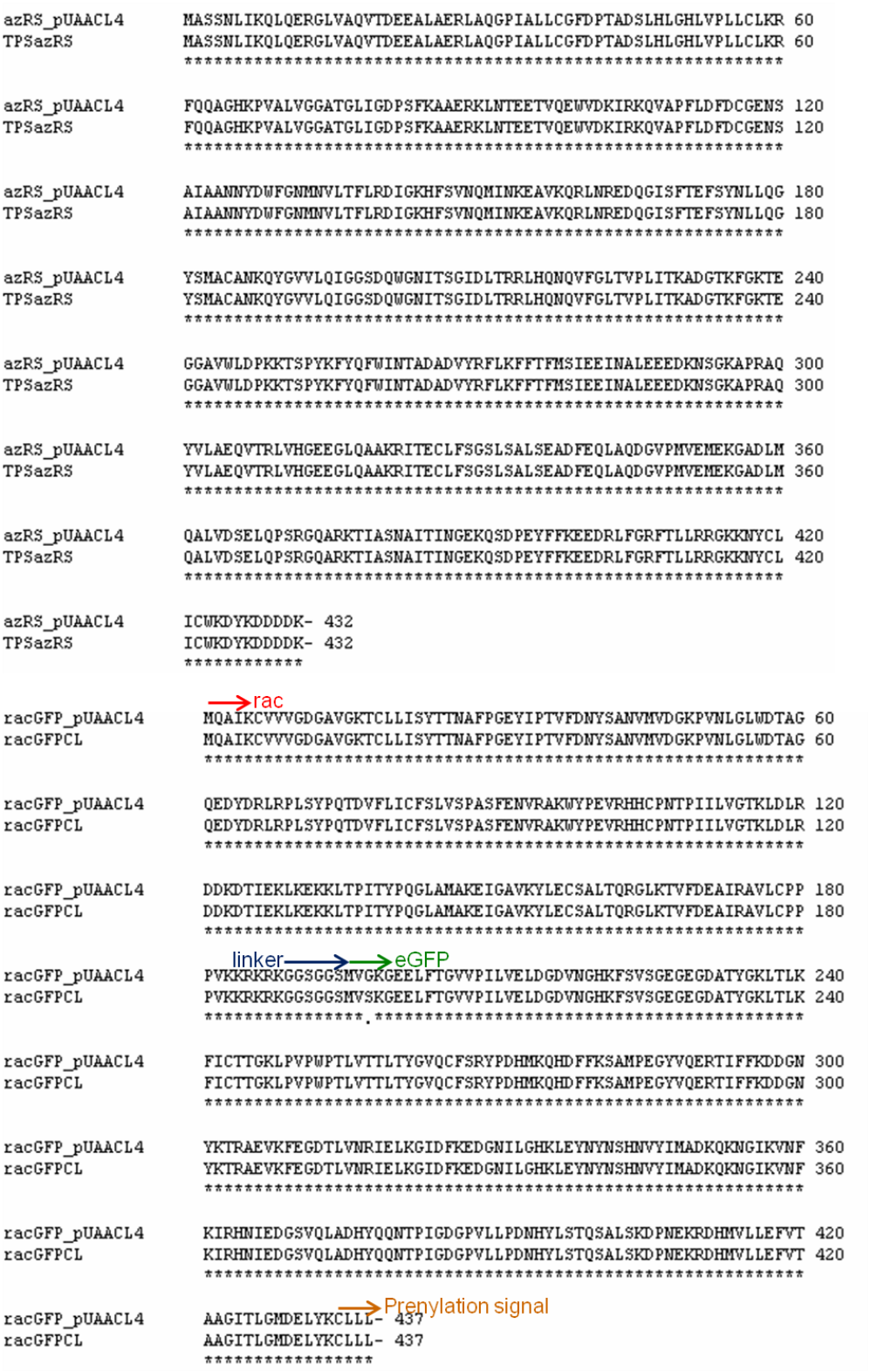


Figure 105. Analysis and comparison of azRS-IRES-RacGFP protein sequence determined from DNA sequence in cloned construct pUAAracGFP and wild type protein sequences (Clustalw2, ExPASy)

6.3.1.4 Cloning *Bst*RNA$^{Tyr}_{CUA}$ into pUAAracGFP

The *Bacillus stearothermophilus* tRNA$^{Tyr}_{CUA}$ with the human tRNA^{Tyr} leader sequence was amplified by PCR from the plasmid TPS136_pSVB-Yam (a kind gift from Dr Thomas Sakmar, Rockefeller University) using primers *Bst*RNA FP and *Bst*RNA RP that added an *Ssp*I site and *Nde*I site to the 5' and 3' ends of the PCR product respectively (Appendix A2). The PCR product was ligated into complementarily digested, dephosphorylated and gel extracted plasmid pUAAracGFP. Insertion of *Bst*RNA$^{Tyr}_{CUA}$ into pUAAracGFP was determined using colony PCR, using primers CMV RP and BstRNA FP (Figure 106). A product of 500 bp indicated the insertion of *Bst*RNA$^{Tyr}_{CUA}$ into pUAAracGFP. DNA from the one colony determined to harbour the *Bst*RNA insertion (by colony PCR) was isolated and subjected to DNA sequencing. Colonies that gave PCR products of higher molecular weight and those that gave no product were not taken forward for further analysis.

Sequencing confirmed the insertion of *Bst*RNA$^{Tyr}_{CUA}$ into pUAARacGFP. This produced a plasmid conferring Zeocin resistance, the synthetase specific for azF, the RacGFPCL construct and the *Bst*RNA $^{Tyr}_{CUA}$, plasmid pUAACL (Figure 101).

6.3.1.5 Creation of pUAAGFP

To create a complementary plasmid encoding only eGFP, the Rac1 gene was removed from the pUAAracGFP plasmid using Phusion site-directed deletion mutagenic PCR. The PCR was performed using primers, pUAA001-2 and pUAA002-3, which annealed to the beginning of the GFP gene and 3' end of the IRES sequence respectively. PCR product of the expected size (7.9 Kb) was obtained, as determined by agarose gel electrophoresis (Figure 107), purified, ligated and used to transform Top10 *E. coli* cells. Four colonies were picked, grown and the DNA prepared and sequenced. One plasmid encoding the azRS-IRES-GFPCL construct was selected and utilised in further experiments. This created a Zeocin selectable plasmid encoding azRS-IRES-GFPCL and *Bst*RNA$^{Tyr}_{CUA}$, pUAAGFPCL.

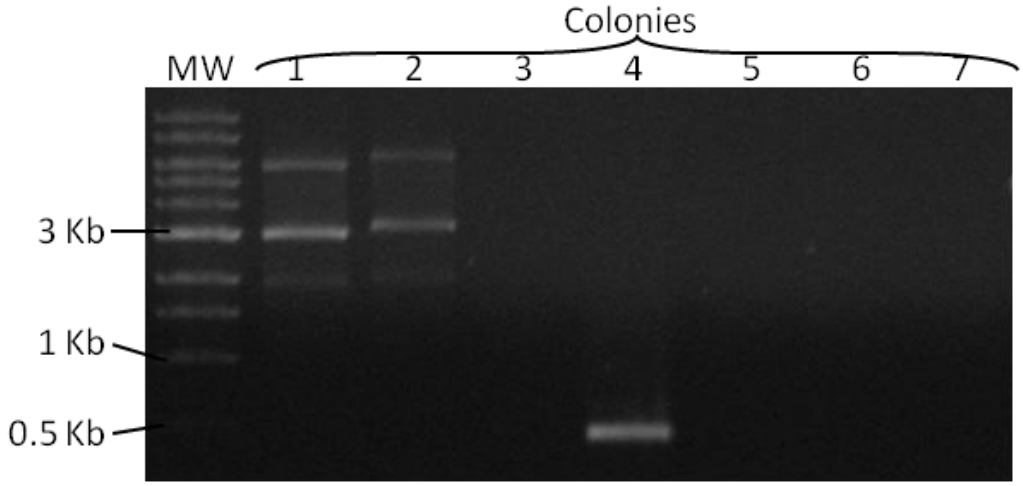


Figure 106. Agarose gel of colony PCR, screening for insertion of *Bst*RNA$^{Tyr}_{CUA}$ into pUAAracGFP. Insertion of *Bst*RNA into pUAAracGFP (product ~ 500 bp) was determined using colony PCR (7 randomly selected colonies), using primers CMV RP and BstRNA FP. Colony 7, a colony from a transformation of un-ligated digested pUAARacGFP + insert, represents –ve control. MW was 1 Kb ladder (NEB).

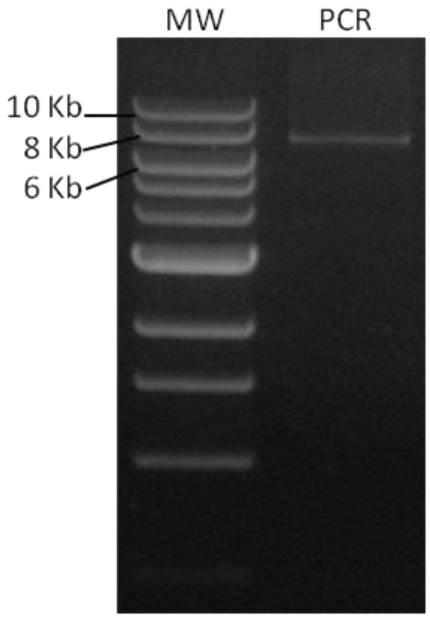


Figure 107. Agarose gel of purified PCR product from PCR to remove Rac1 from plasmid pUAACL, to give plasmid pUAAGFPCL. Phusion site-directed deletion mutagenic PCR was performed using primers pUAA001-2 and pUAA002-3, which annealed to the beginning of the GFP gene and 3' end of the IRES sequence respectively. The PCR product of 7.9 Kb indicated the amplification of pUAACL minus the Rac1 gene. MW was 1 Kb ladder (NEB).

6.3.2 Mutagenesis of Rac1

6.3.2.1 Analysis of Rac1 protein - placement of non-natural amino acid

TAG mutations were required in the Rac gene in order to implement the incorporation of non-natural amino acids using the amber suppression technology. Because Rac1 is involved in a variety of signalling pathways, it has to make numerous interactions with a wide variety of proteins. It was important that incorporation of a NAA did not significantly interfere with any of these interactions. A literature search was used to ascertain some of the key signalling interactions that Rac1 makes (outlined in Chapter 1) (for review see Bishop and Hall 2000).

The PDB structure, 2RMK (Modha et al. 2008), was analysed for suitable speculative positions for the incorporation of the non-natural amino acid. It is optimal to position the TAG codon, and therefore the NAA, near the 5' end of the gene. This means that there are limited possibilities of the truncated protein remaining active or indeed inhibitory, due to its short length, by a competitive manner or other, to its cognate signalling, or acting in another manner. Due to the similarities between the NAAs, namely, azF, and tyrosine, it was thought that a Tyr would be a good substitution position for the NAA. Phenylalanines were also considered as substitution positions, for the same reason. The Phe and Tyr residues were highlighted in the PDB structure (Figure 108). Those N-terminal (residues 1 - 70), in loops, not present within important regions (as outlined in Chapter 1) and pointing away from hydrophobic core were most highly considered. It was demonstrated that tyrosine 23 fulfilled these criteria and was not present within the regions important for GTPase function and was therefore selected for mutagenesis. However, because Tyr residues are often phosphorylated and the presence of an azido group would absolutely perturb this modification, other bulky, hydrophobic amino acid substitutions were considered.

Figure 108. Ribbon structure of Rac1 highlighting Phe and Tyr residues within N-terminal 70 residues, PDB structure 2RMK (deposited by Modha et al. 2008). The switch I, switch II and important GTP-binding residues are indicated in blue colours. Phe and Tyr residues are shown in colour indicated by area. Tyr, Y23 is indicated in red. Polybasic region – pink, prenylation sequence (CLLL) – light pink. Figure was produced using Pymol (DeLano and Lam 2006).

6.3.2.2 Creation of plasmids with TAG mutation

A TAG mutation was made at position Y23 within Rac1 utilising plasmid pUAACL. Phusion polymerase mutagenic PCR was used to mutate Y23 in Rac1 to TAG codon using primers mRac1Y23TAG FP and mRac1Y23TAG RP (Figure 109). After mutagenesis, the DNA from 3 colonies was purified and sequenced. Analysis of the DNA sequences demonstrated that all 3 clones harboured the required TAG mutation. One clone harbouring the TAG mutation along with no other mutations was used in further experiments. This produced plasmid pY23TAG.

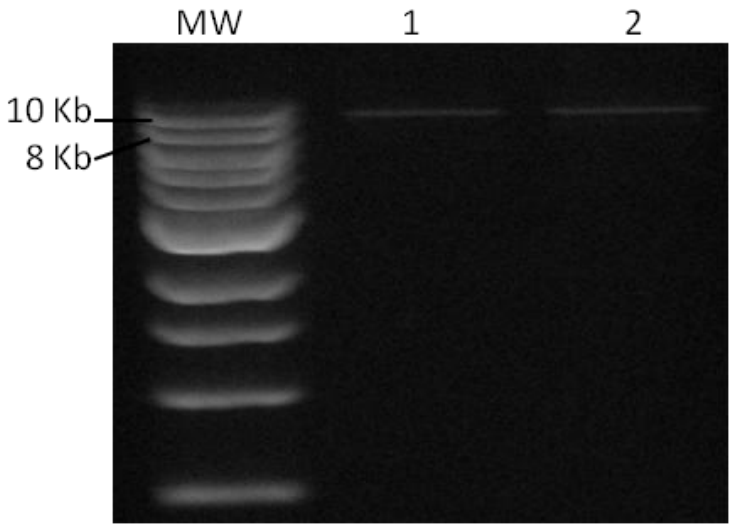


Figure 109. Agarose gel of purified PCR product from PCR to introduce a TAG mutation at Y23 within Rac1 gene, to give plasmid pY23TAG. Phusion site-directed deletion mutagenic PCR was performed using primers mRac1Y23TAG FP and mRac1Y23TAG RP, which annealed to the sequence flanking codon Y23. The PCR product of approximately 8.4 Kb indicated the amplification of the entire plasmid introducing the Y23TAG mutation Lane 1 and 2 represent two PCR reactions. MW was 1 Kb ladder (NEB).

6.3.3 Introduction of plasmids into Rac1 KO cells

6.3.3.1 Transient transfection of pUAACL, pUAAGFPCL and pY23TAG into Rac1 KO

The plasmids, pUAACL, pUAAGFPCL and pY23TAG were used to transiently transfect Rac1 KO cells to ensure that the plasmids were capable of expressing target genes. Fluorescent microscopy showed that cells transiently transfected with pUAACL and pUAAGFPCL showed GFP fluorescence (Figure 110), whereas Rac1 KO cells transiently transfected with pY23TAG did not. This result demonstrated that the constructed plasmids, pUAACL and pUAAGFPCL, were capable of expressing GFP, indicating the use of the IRES sequence allowed expression of two genes from the same promoter. Importantly, when a TAG mutation was introduced into the racGFPCL construct, GFP fluorescence was no longer seen (Figure 110).

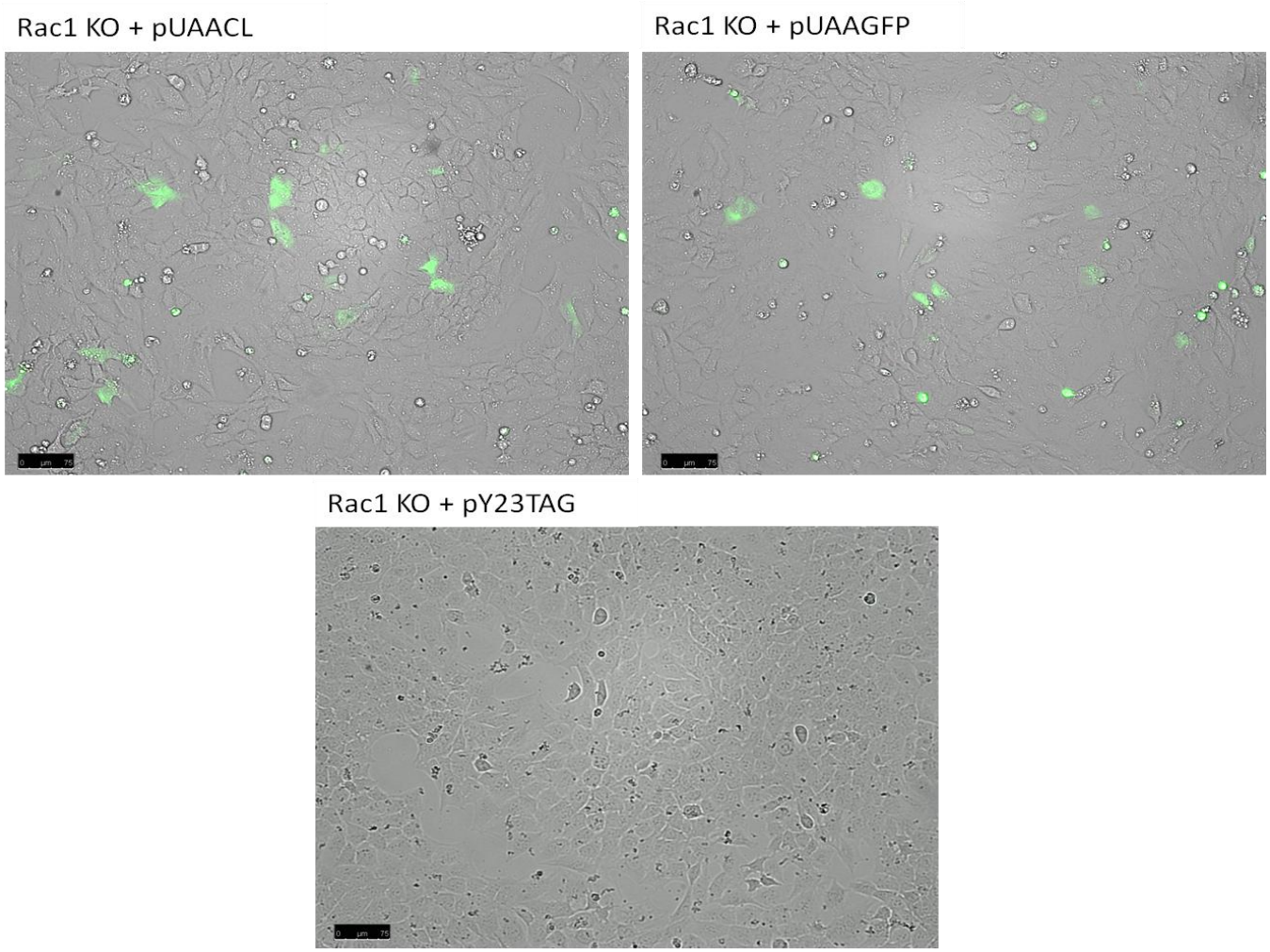


Figure 110. Fluorescent microscopy images of Rac1 KO cells transiently transfected with plasmids pUAACL4, pUAAGFP or pY23TAG. Scale bar = 75 µm.

6.3.3.2 Transcription of genes Rac1, zeo^r and azRS

To ensure that RacGFP, azRS and the Zeocin resistance gene (zeo^r) were successfully transcribed from the created plasmid constructs, RNA extraction, cDNA synthesis, PCR for these genes and agarose gel electrophoresis was performed using Rac1 KO cells that had been transiently transfected with either pUAACL, pUAAGFPCL or pY23TAG (Figure 111).

PCR products corresponding to approximately 550 bp corresponded to Rac1 and were obtained from cellular extracts transfected with plasmids pUAACL and pY23TAG. Cells transfected with pUAAGFP showed no production of Rac1 transcript (Figure 111).

The cells transfected with plasmids pUAACL, pUAAGFPCL and pY23TAG showed production of zeo^r transcript (PCR products approximately 375 bp) and azRS transcript (PCR products approximately 1.3 Kb; Figure 111).

6.3.3.3 Stable transfection of pUAACL

The production of a stably expressing cell line was undertaken in an attempt to achieve optimal numbers of cells expressing NAA-containing RAC1 and therefore give a significant response during wound healing. As demonstrated in Chapter 5, the production of stable cells lines using Zeocin selection was shown to be possible by the creation of Rac1 KO cell lines stably expressing Rac1 and eGFP from plasmids using the commercially available plasmid, pcDNA3.1/zeo. The plasmids pUAACL and pUAAGFPCL (also based on the commercially available plasmid, pcDNA3.1/zeo) were transfected into Rac1 KO cells and selection pressure (Zeocin) applied. One resistant, GFP-positive colony harbouring the plasmid pUAACL was isolated, however, the growth of the colony was slow and the GFP fluorescence was lost over time. For the other plasmid constructs (pUAAGFPCL, pY23TAG), there were many resistant colonies, but none of the colonies displayed GFP fluorescence, as determined by fluorescent microscopy.

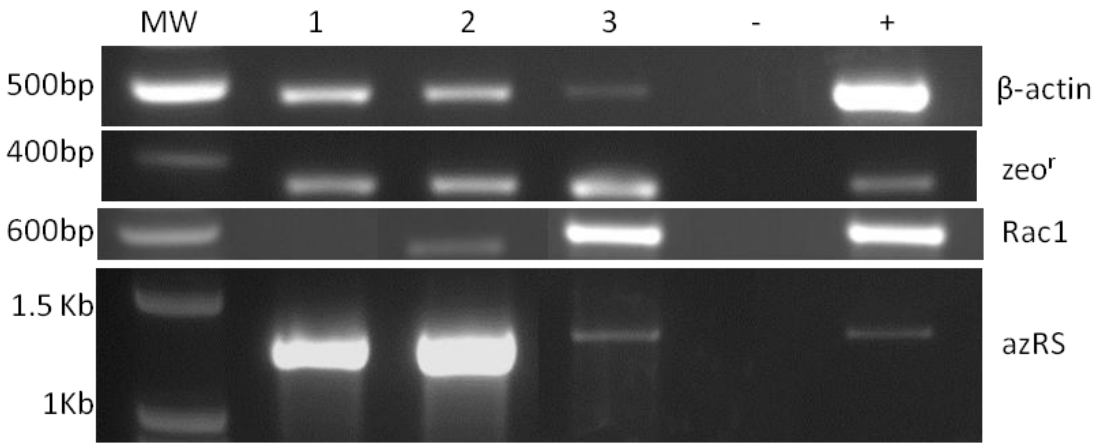


Figure 111. Agarose gel of RT-PCR screening for β-actin Rac1, zeo^r, azRS. cDNA was synthesised from RNA extracted from Rac1 KO cells transiently transfected with pUAAGFPCL (lane 1), pY23TAG (lane 2), pUAACL (lane 3); -ve water control (lane 4), plasmid (pUAACL) positive control (lane 5). MW is 100 bp ladder (Promega).

6.3.3.4 Amber suppression in Rac1 KO cells

The plasmid pY23TAG was used to transfect Rac1 KO cells utilising the optimised Fugene transfection method as determined in Chapter 5, and the medium was supplemented with NAA. Whilst there was visualisation of some GFP-positive cells (Figure 112; indicating that amber suppression had occurred) the efficiency was too low to produce enough cells to perform a scratch wound assay (the functional cellular output assay).

Electroporation was therefore, utilised in an attempt to improve the efficiency of NAA incorporation (it was postulated that electroporation would allow increased uptake of NAA) and thereby increase production of NAA-containing RAC1. Electroporation was performed either in the presence of NAA and pY23TAG DNA or in the presence of NAA followed by subsequent transfection of DNA by Fugene. In each case no improvement in the number of GFP positive cells was observed. Following a similar hypothesis, liposomal transfection using Lipofectin was utilised. However, still no improvement in number of GFP positive cells was observed.

In another attempt to improve the uptake of NAA, the Rac1 KO cells were cultured in medium without the supplementation of non-essential amino acids. It was hypothesised that when NAA was supplemented into the media the 'starved' cells would increase the uptake of amino acids and thereby increase uptake of NAA. The time at which NAA was added to the cells was varied (3 hours before transfection, at transfection or 3 hours post-transfection). However, no improvement in number of GFP positive cells was observed after starving the cells of non-essential amino acids.

These results demonstrated that the plasmid constructs without TAG mutation (pUAACL, pUAAGFPCL) produced GFP positive cells when introduced in to Rac1 KO cells. However, when amber suppression was attempted using the TAG containing plasmid, pY23TAG, the results indicated that the Rac1 KO fibroblasts were refractory to amber suppression. Also demonstrated was that NAAs, iodoF and AzF, appeared to have limited toxicity to the cells. There was no immediate extensive

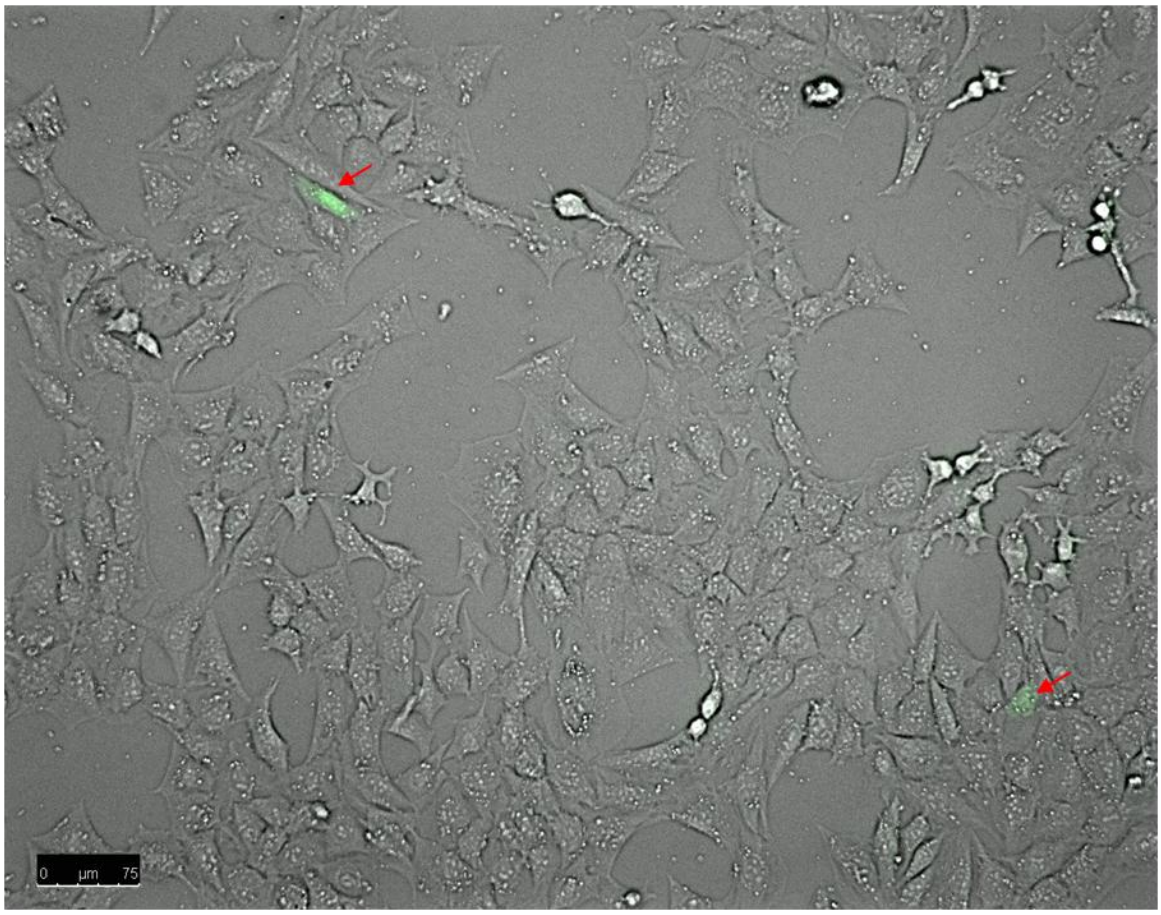


Figure 112. Brightfield/fluorescence composite image demonstrating amber suppression in Rac1 KO cells. Rac1 KO cells were transiently transfected with pY23TAG and the growth media was supplemented with 1 mM IodoF dissolved in NaOH. Images were capture two days post transfection. Red arrows indicate GFP-positive cells, that are expressing NAA-containing Rac1-GFP. Scale bar = 75 µm.

death, in fact, very little cell death, and the cultures seemed to tolerate the NAA for the duration of the experiments.

6.3.4 Amber suppression in HEK293 cells

It had been demonstrated (see above) that the amber suppression had low efficiency in the Rac 1 KO cell line. Therefore, a number of variables were investigated to try to identify the issue and subsequently increase the incorporation of NAA in to RAC1. The HEK293 cell line was selected to investigate these considerations because these cells have previously been used successfully in the implementation of amber suppression. HEK293 cells were kindly donated by Lea Bauer of the Tissue Engineering and Reparative Dentistry group, School of Dentistry, Cardiff University.

6.3.4.1 Amber suppression using previously created plasmids pSWANGFP37TAG and TPS92azRS

Firstly, a GFP37TAG construct was used to investigate whether amber suppression could be implemented in the HEK293 cell line. The previously created pSWANGFP37TAG (Liu et al. 2007) plasmid encoded GFP with a TAG mutation at position 37, under the control of a CMV promoter and the orthogonal *Bs*tRNA$^{Tyr}_{CUA}$ (three copies; see Appendix A1). This was co-transfected with another previously created plasmid that encoded the orthogonal azF-specific synthetase under a CMV promoter (TPS192azRS_V1) (Ye et al. 2009). Fluorescence microscopy was used to analyse expression of GFP. It was shown that there were green cells in wells transfected with the plasmids, plus azF; and no green cells in wells transfected, but not containing azF (Figure 113). This result indicates that cells in wells containing non-natural amino acid were expressing GFP37azF. It was shown that 0.5 mM azF was sufficient to allow expression of GFP37azF.

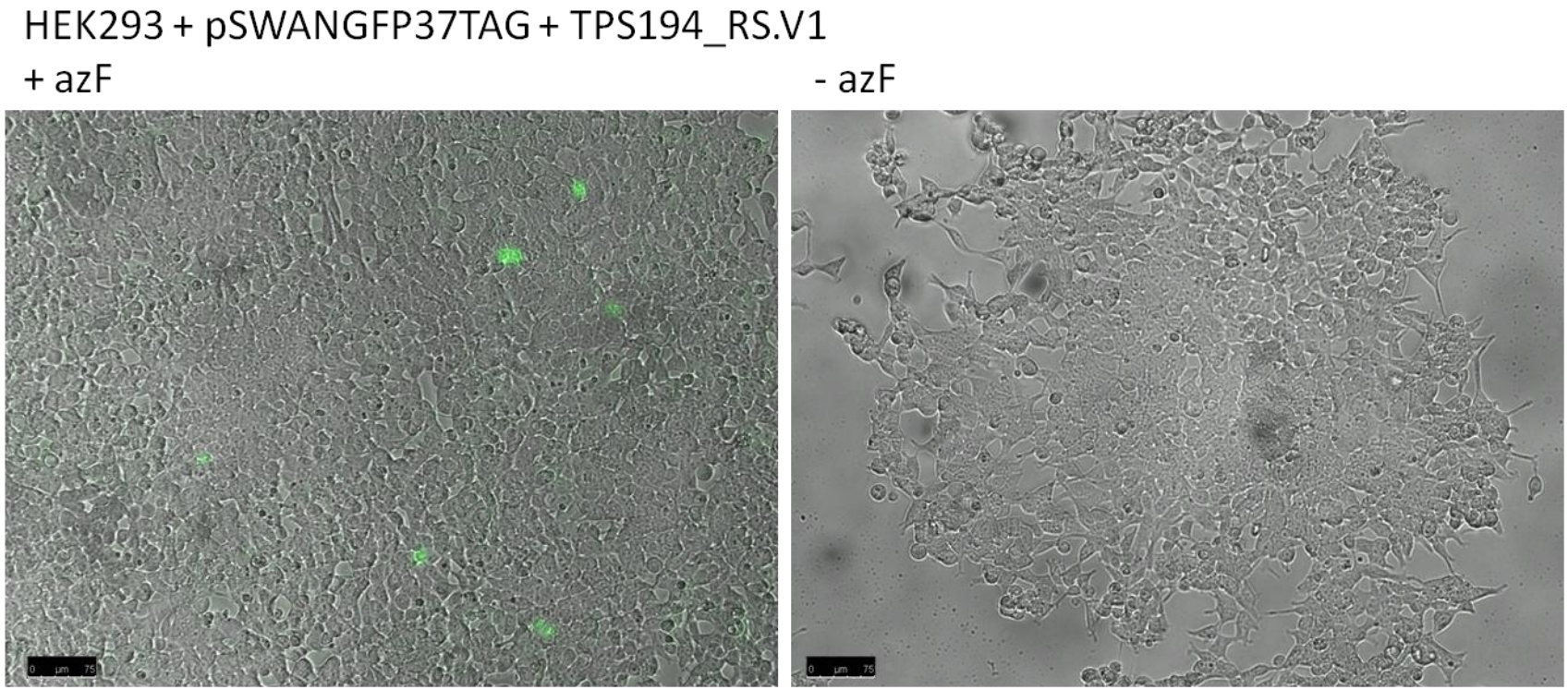


Figure 113. Composite brightfield/fluorescence microscopy images of HEK293 cells performing amber suppression. HEK23 cells were transiently transfected with previously created plasmids pSWANGFP37TAG and TPS194_RS.V1 using Fugene. 0.5 mM azF was added or withheld from the growth media. Scale bar = 75 µm**.**

The consideration that either azRS protein, *Bs*tRNA$^{Tyr}_{CUA}$, azF or GFP37azF protein might confer toxicity to the cells was addressed by carrying out timelapse microscopy on cells transfected with pSWANGFP37TAG and TPS192azRS_V1, with the addition of 0.5 mM azF. Although the transfection efficiency was low, the timelapse microscopy showed that GFP fluorescence was observed approximately 9 hours after addition of NAA with the GFP fluorescence intensifying up to approximately 24 hours (Figure 114, Movie 9). Throughout this time it appeared that the cells were healthy as they were not balled up or floating. The timelapse analysis also indicated that the GFP-positive cells were also able to divide. At 41 hours the addition of a second green cell in close proximity to the first, which appears to bleb from the original cell, indicates that the original GFP-positive cells had undergone division. After 48 hours, the cell still appeared healthy.

These experiments demonstrated that HEK293 cells were capable of incorporating NAA via the amber suppression method. In addition, azF and the expression of azRS and GFP37azF were well tolerated as demonstrated by the timelapse microscopy of GFP positive cells.

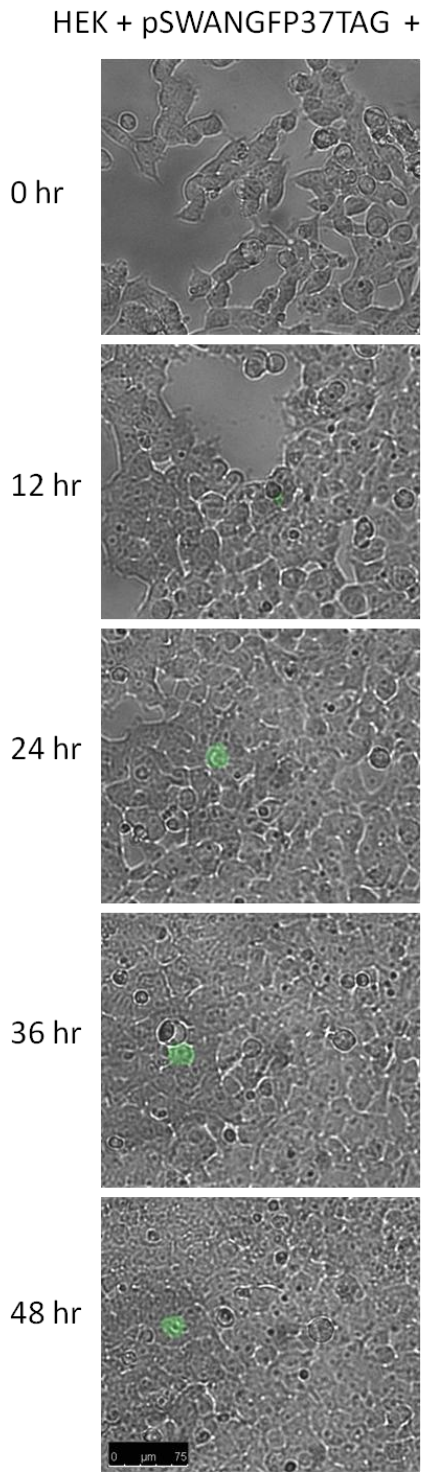


Figure 114. Composite brightfield/fluorescence timelapse images of HEK293 cells performing amber suppression to produce GFP37azF. HEK293 cells were transiently transfected with previously created plasmids pSWANGFP37TAG and TPS194_RS.V1, and 0.5 mM azF was supplemented in the media. Images were captured at specified timepoints. Scale bar = 75 µm.

6.3.4.2 Transient transfection of plasmids pUAACL and pUAAGFPCL into HEK293

The plasmids pzeoCL, pzeoGFPCL, pUAACL, pUAAGFPCL were used to transiently transfect HEK293 cells to ascertain whether these plasmids would express protein. It was demonstrated by fluorescent microscopy that all the plasmids were capable of expressing GFP (Figure 115).

The transfection efficiency was lower for the pUAA-based plasmids than the pzeo-based plasmids ($p < 0.01$; Figure 116). However, the transfection efficiency was higher than that obtained using the Rac1 KO cells (see above); therefore using HEK293 cells makes the investigations into the functionality of the plasmids more accessible.

These results demonstrate that HEK293 cells were able to express racGFP or GFP from the plasmids constructed as described within this Chapter.

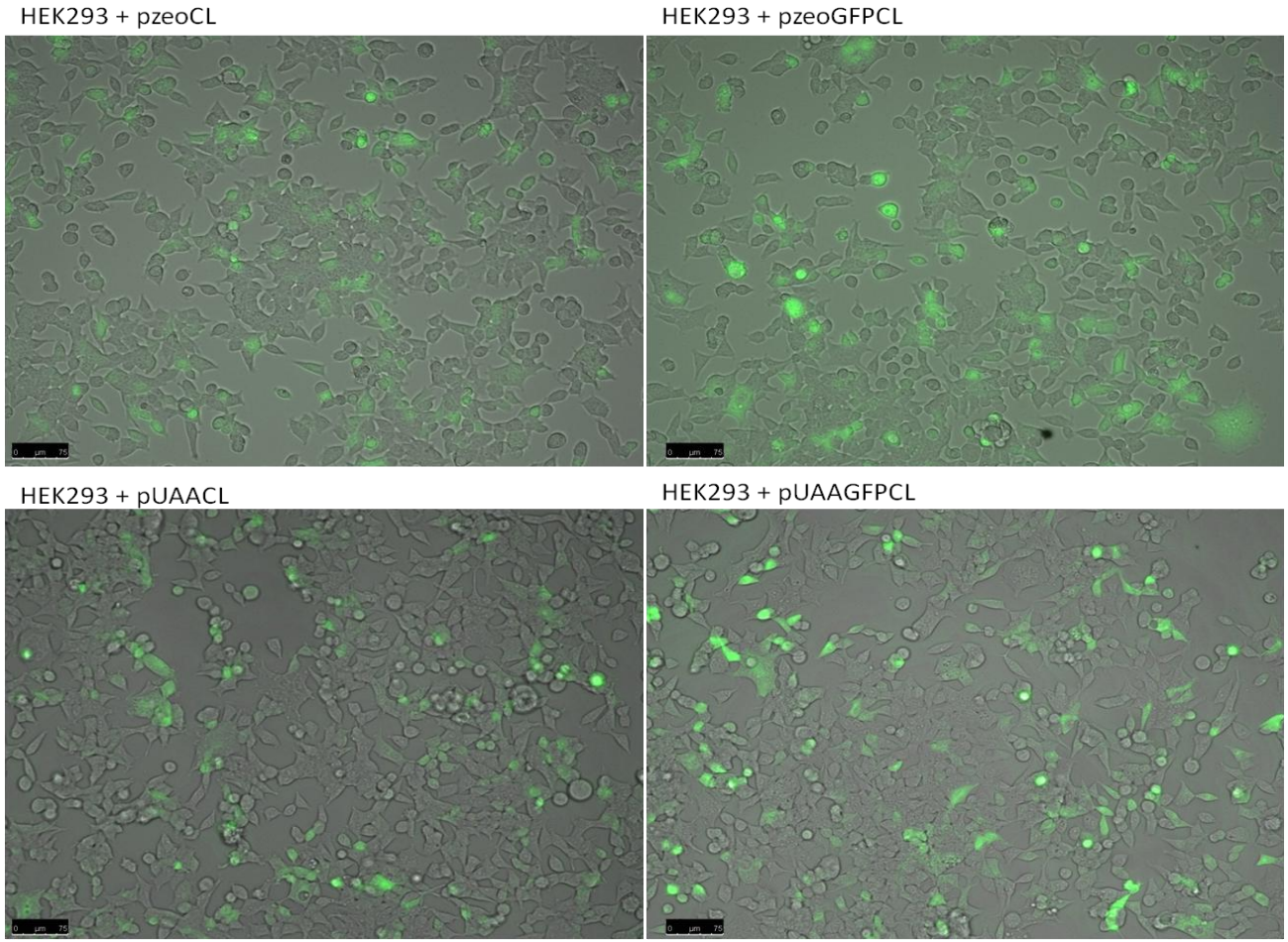


Figure 115. Composite brightfield/fluorescent microscopy images of HEK293 cells transient transfected with plasmids pzeoCL, pzeoGFPCL, pUAACL4 or pUAAGFP. Scale bar = 75 µm.

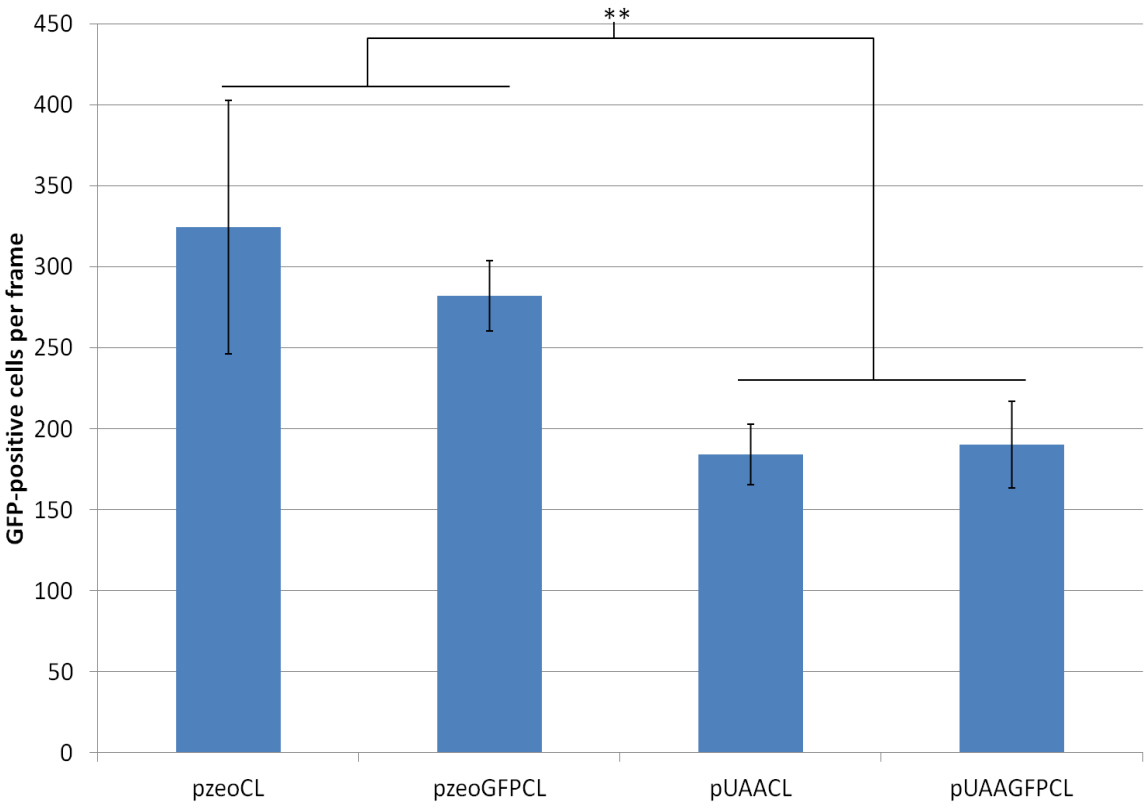


Figure 116. Graph showing transfection efficiency of HEK293 cells transfected with either pzeoCL, pzeoGFPCL, pUAACL or pUAAGFPCL, using Fugene. Images were captured using fluorescent microscopy and number of GFP-positive cells counted. Errors represent average of *n*=4 transfections.

6.3.4.2.1 Optimisation of transfection

To try to give the best possible chance of NAA incorporation, a variety of conditions were varied to try to optimise the transfection of HEK293 cells. Namely, the Fugene : DNA ratio and the time allowed for Fugene : DNA complexes to form. Plasmid pzeoGFPCL was used to optimise the transfection. It appeared that to obtain a significant number of GFP-positive cells a Fugene : DNA ratio of 3 : 1 was required (Figure 117). However, using a ratio of 4 : 1 did not increase the transfection efficiency further. A ratio of 3 : 1 was therefore used for subsequent transfections. The time for Fugene/DNA complex formation determines the size of the complexes and this can affect the uptake of the complexes into the cells. The optimal size of complex can vary between cell lines, therefore this was also investigated. It was demonstrated that using a Fugene : DNA ratio 3 : 1, a complex formation time of 5, 10 or 15 gave equivalent transfection efficiency (Figure 16).

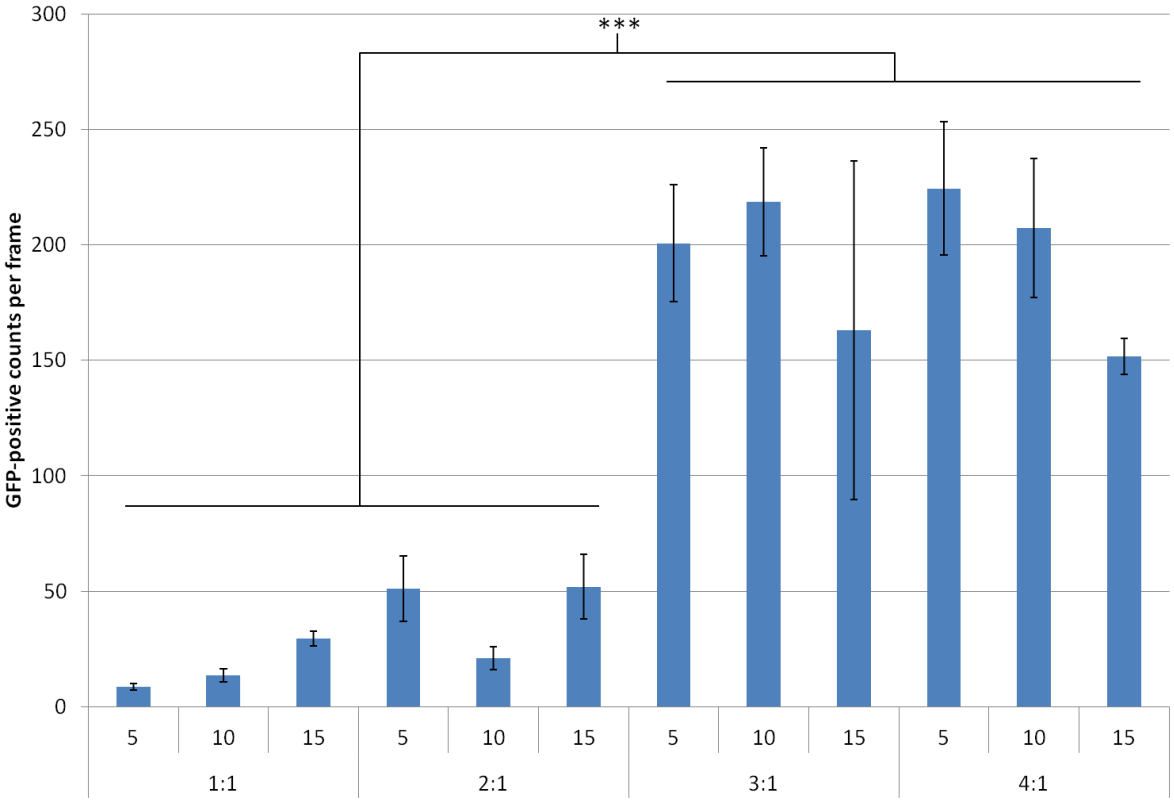


Figure 117. Graph comparing transfection efficiency of Fugene : DNA complexes formed using varying amount of DNA, ratio Fugene : DNA and complex formation time. 1 µg DNA was added per well. The errors bars represent the standard deviation of average (*n*=3 experiments).

6.3.4.3 Amber suppression in HEK293 cells using one plasmid system, pY23TAG

To investigate the functionality of the one plasmid system created as described within this Chapter (expressing azRS, *Bs*tRNA$^{Tyr}_{CUA}$ and target gene, RacGFPCL harbouring at TAG mutation at Y23), the HEK293 cells were transiently transfected with the pY23TAG plasmid using Fugene. The media was supplemented with 0, 0.5 or 1 mM azF and fluorescent microscopy was used to investigate whether NAA incorporation had occurred. GFP-positive cells were visualised, whereas there were no GFP-positive cells in the wells transfected with pY23TAG where no NAA had been added to the media (Figure 118). This indicated the high fidelity of the azRS to incorporate NAA. The TAG codon was located within the Rac1 gene, which was upstream of the GFP, therefore expression of full-length GFP provided some evidence that read-through of the TAG codon, and subsequent incorporation of NAA into protein, had occurred.

6.3.4.4 Promiscuity of NAA recognised by azRS

As mentioned previously (Chapter 3, 4), the *Mj* aaRS used in *E. coli* amber suppression are promiscuous towards the NAA that they charge. No previous work has been performed to investigate the promiscuity of the engineered amber suppressor *E. coli* synthetases used in the mammalian system however, it seems likely that the same promiscuity is evident within the evolved *E. coli* aaRSs. Therefore, it was investigated whether the *E. coli* azRS could incorporate iodoF during amber suppression. Also of note is that azF is notoriously both thermo- and photoreactive and therefore the use of iodoF in place of azF would dispel any issues regarding breakdown of the azF during preparation or storage prior to addition to the cells. Therefore, HEK293 cells were transiently transfected with the plasmid pY23TAG and the media was supplemented with iodoF. Fluorescent microscopy showed GFP-positive cells (Figure 118).

Therefore, this experiment demonstrated that the *E. coli* aaRS developed for incorporation of azF via amber suppression in mammalian cells also shows

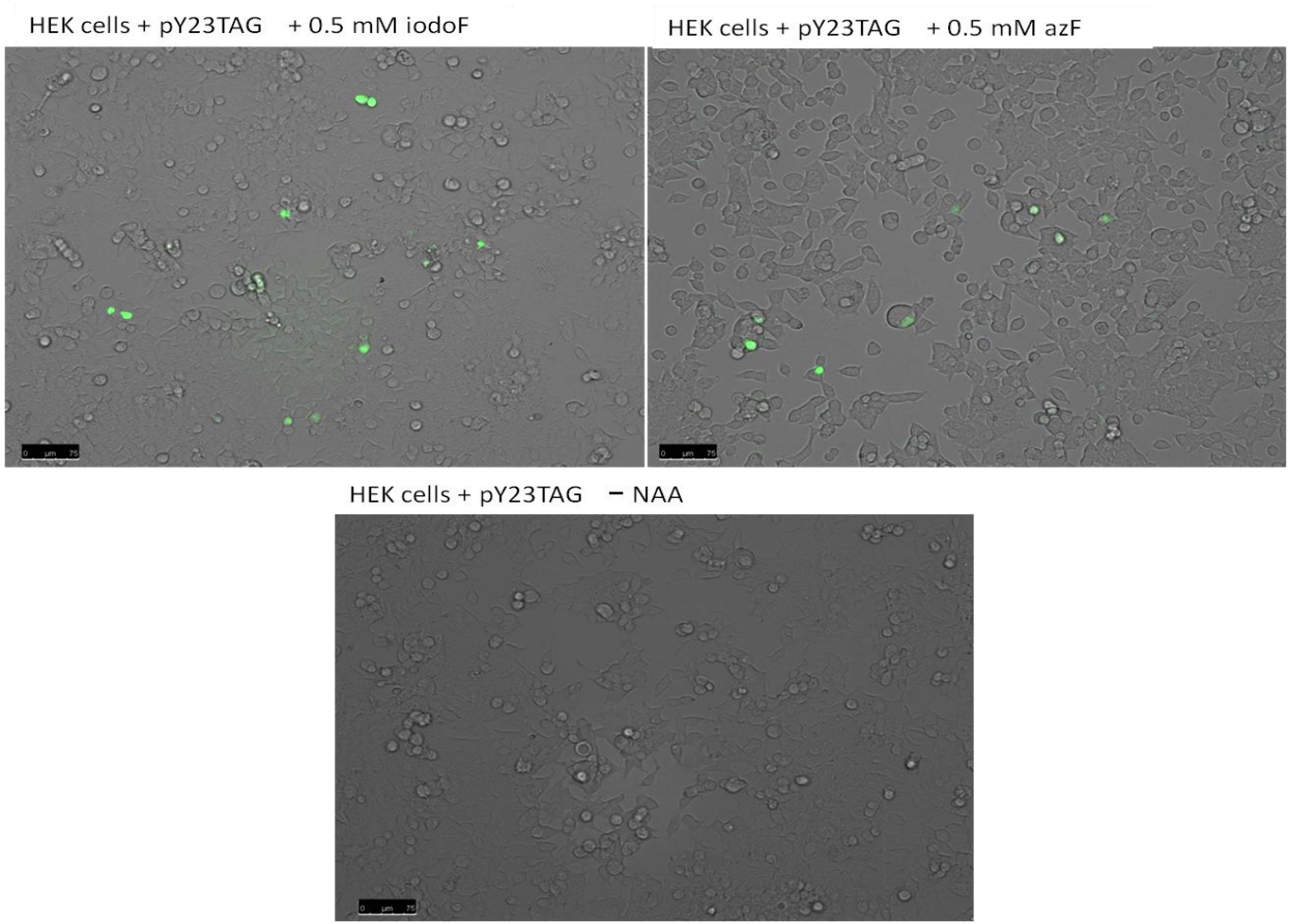


Figure 118. Composite brightfield/ fluorescent images of HEK293 cells transiently transfected with plasmid pY23TAG, plus/minus nonnatural amino acid. Scale bar = 75 µm.

promiscuity towards the NAA it recognises and successfully charges. Therefore, iodoF could be used in subsequent experiments.

6.3.4.5 pY23TAG expresses sufficient levels of azRS and *Bs*tRNA$^{Tyr}_{CUA}$ for amber suppression

To investigate whether efficiency of NAA incorporation using the plasmid pY23TAG was limited by the levels of azRS or *Bs*tRNA$^{Tyr}_{CUA}$ present in the cell, additional azRS and/or *Bs*tRNA$^{Tyr}_{CUA}$ was supplied by transfecting the HEK293 cells with plasmid pY23TAG plus plasmid TPS192azRS_V1 and/or TPS136_pSVB-Yam respectively. The data demonstrated that there was some increase in number of green cells when both azRS- and *Bs*tRNA$^{Tyr}_{CUA}$-containing plasmids were added to the transfection, however the difference was not significant (Figure 119; $p > 0.05$). No increase in GFP-positive cells was observed when either extra *Bs*tRNA$^{Tyr}_{CUA}$ or azRS–containing plasmid was added to the cells. There appeared no benefit in adding 1 mM NAA as opposed to 0.5 mM. Some GFP-positive cells were visualised in the wells containing no NAA however, the GFP expression was much lower than that seen in the well containing NAA ($p < 0.05$). No GFP-positive cells were observed in wells not transfected but supplemented with NAA.

These experiments demonstrated that the pY23TAG plasmid was sufficient to allow amber suppression in mammalian cells. Stable transfections could not be performed using this cell line as the genetic manipulation of this HEK293 cell line, rendered the cell line resistant to Zeocin antibiotics and the pY23TAG plasmid confers Zeocin resistance.

To further investigate expression of azRS, pY23TAG plasmid was transiently transfected into control cells that had a weak Kozak consensus at the azRS translational start site (Appendix A1). This was compared to transfections using the pY23TAG plasmid with the strong Kozak consensus sequence at azRS (Figure 120). There was no statistically significant difference between the number of GFP-positive cells obtained when the Kozak sequence at the azRS was altered ($p > 0.05$).

It was therefore demonstrated that the pY23TAG plasmids, along with the HEK293 were capable of expressing NAA-containing RAC1 protein.

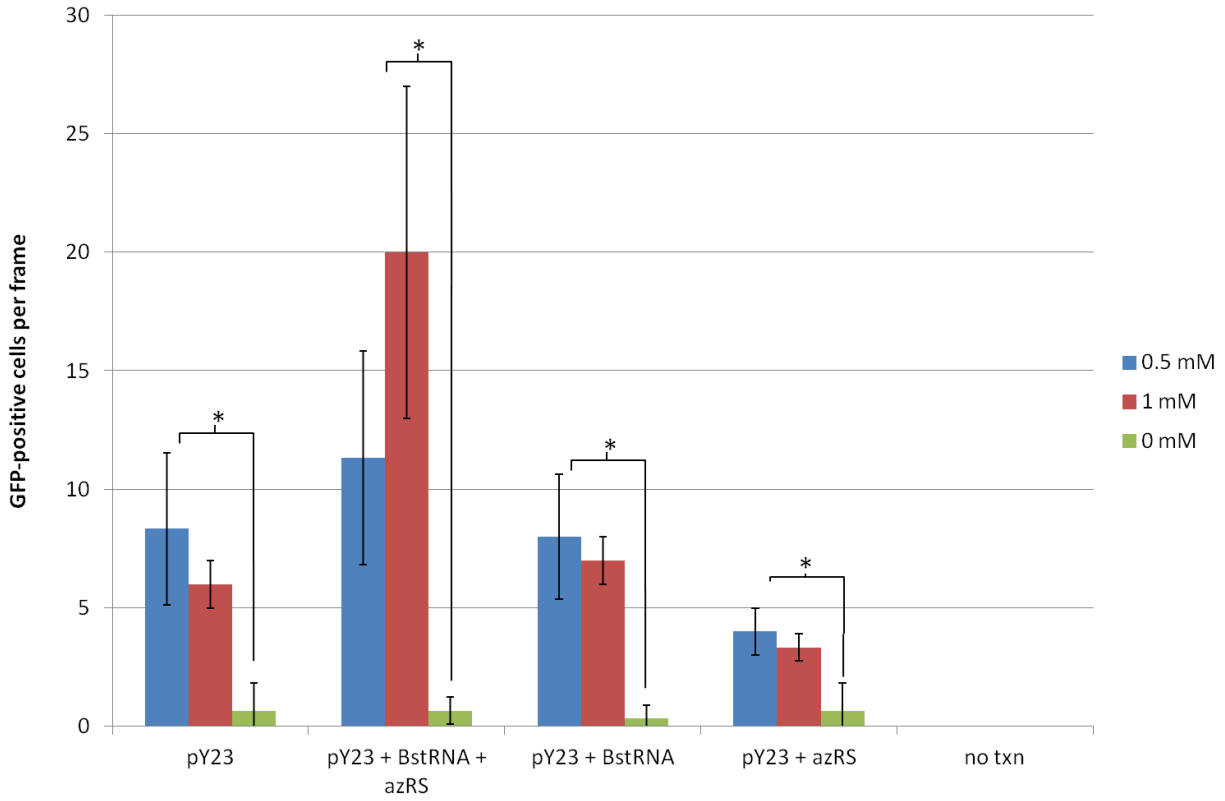


Figure 119. Graph comparing the number of GFP-positive cells per frame of HEK293 cells transiently transfected with pY23TAG, and supplemented with plasmids containing BstRNA (TPS136) or azRS (TPS192). HEK293 cells were transfected with either plasmid pY23TAG, pY23TAG + TPS136 + TPS192, pY23TAG + TPS136, or pY23TAG + TPS192. 3 hours post-transfection 0.5, 1 or 0 µM iodoF was supplied to the cells in the growth media. 2 days post transfection fluorescent microscopy was used to image the wells. Number of GFP-positive cells was calculated by counting the number of green cells per frame. The errors bars represent the standard deviation of average of *n*=3.

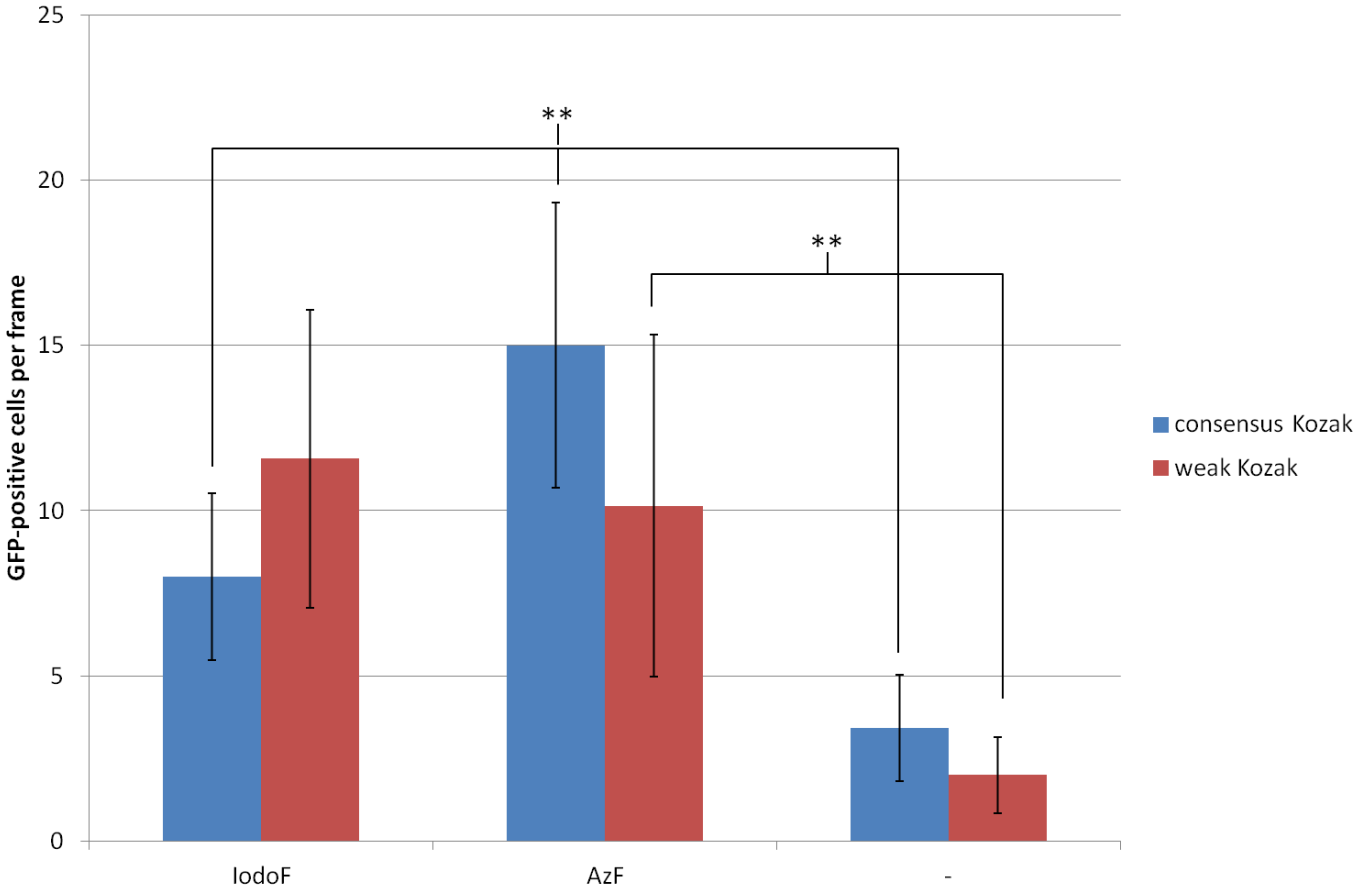


Figure 120. Graph comparing the number of GFP-positive cells obtained when plasmid harbouring strong or weak Kozak consensus sequence control translation of azRS. Error bars represent standard deviation of average of *n*= 6.

6.3.5 Amber suppression in Rac1 control cells

Using HEK293 cells, it was demonstrated that the pY23TAG plasmid was functional (see above). It was thought that the issues regarding the use of Rac1 KO cells might be because the experiments were conducted utilising fibroblasts. Therefore, it was investigated whether the pY23TAG plasmid was functional in the Rac1 control cells. The control cells were the cells from which the Rac1 KO cell line was initially derived (kind gift from Klemens Rottner, University of Bonn). Whether or not the pY23TAG plasmid was functional in this cell line would help to elucidate whether the issues regarding the implementation of the NAA technology in Rac1 KO cells was due to the refractory nature of the fibroblast cell line; or whether during the creation of these Rac1 KO fibroblasts something had occurred that had rendered the KO cell line specifically refractory to implementation of NAA technology.

6.3.5.1 Transient transfection of plasmids pUAACL and pUAAGFPCL into Rac1 control cells

Firstly, the plasmids pzeoCL, pzeoGFPCL, pUAACL, pUAAGFPCL were introduced transiently into Rac1 control cells to test the function of these plasmids. Fluorescent microscopy showed that all were capable of expressing GFP (Figure 121). Analysis of transfection efficiency demonstrated that the transfection efficiency for pzeo-based plasmids was higher than that for pUAA-based plasmids (Figure 122, $p < 0.001$).

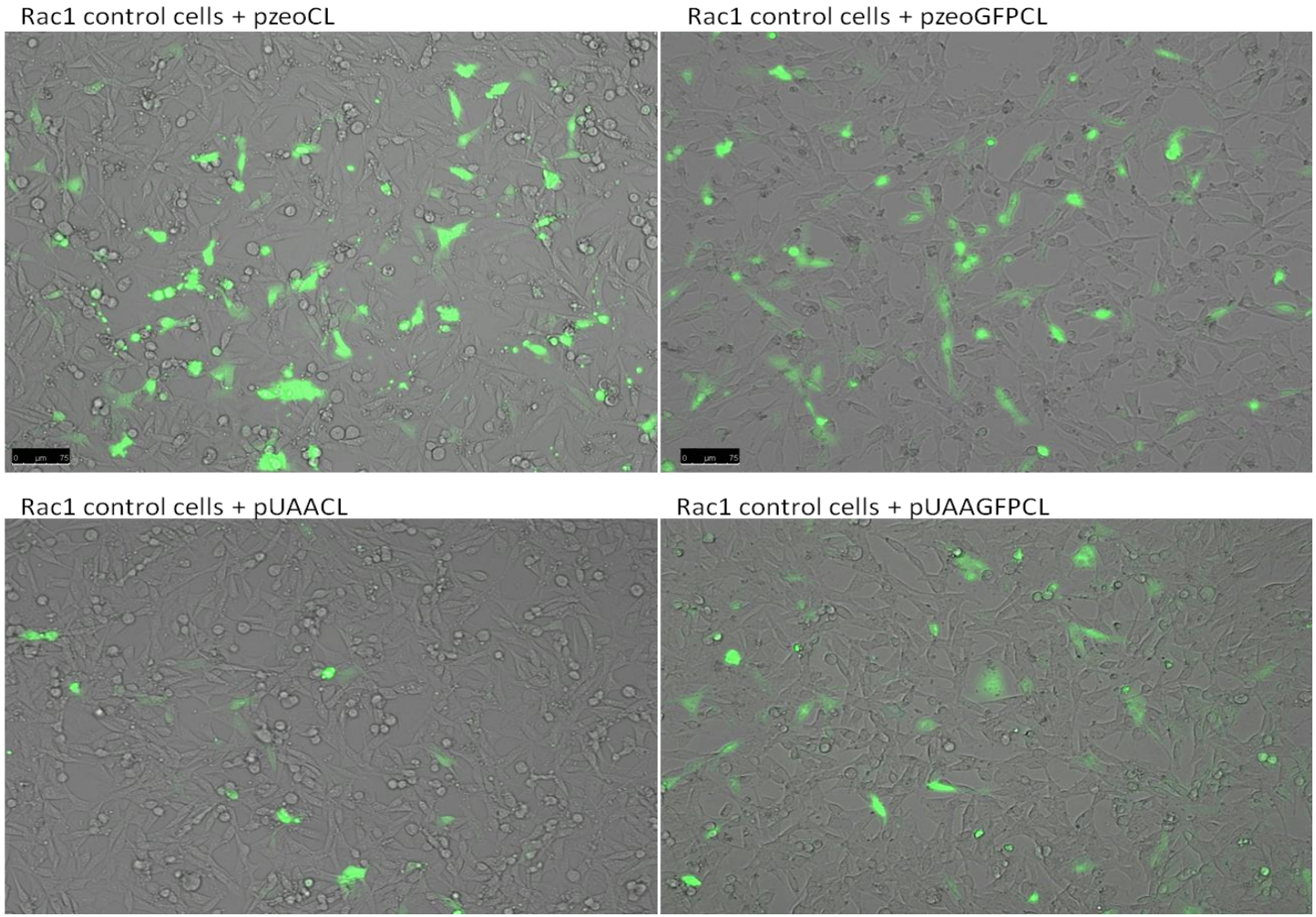


Figure 121. Fluorescent microscope images of control cells transiently transfected with pzeoCL, pzeoGFPCL, pUAACL, or pUAAGFPCL. Scale bar = 75 µm.

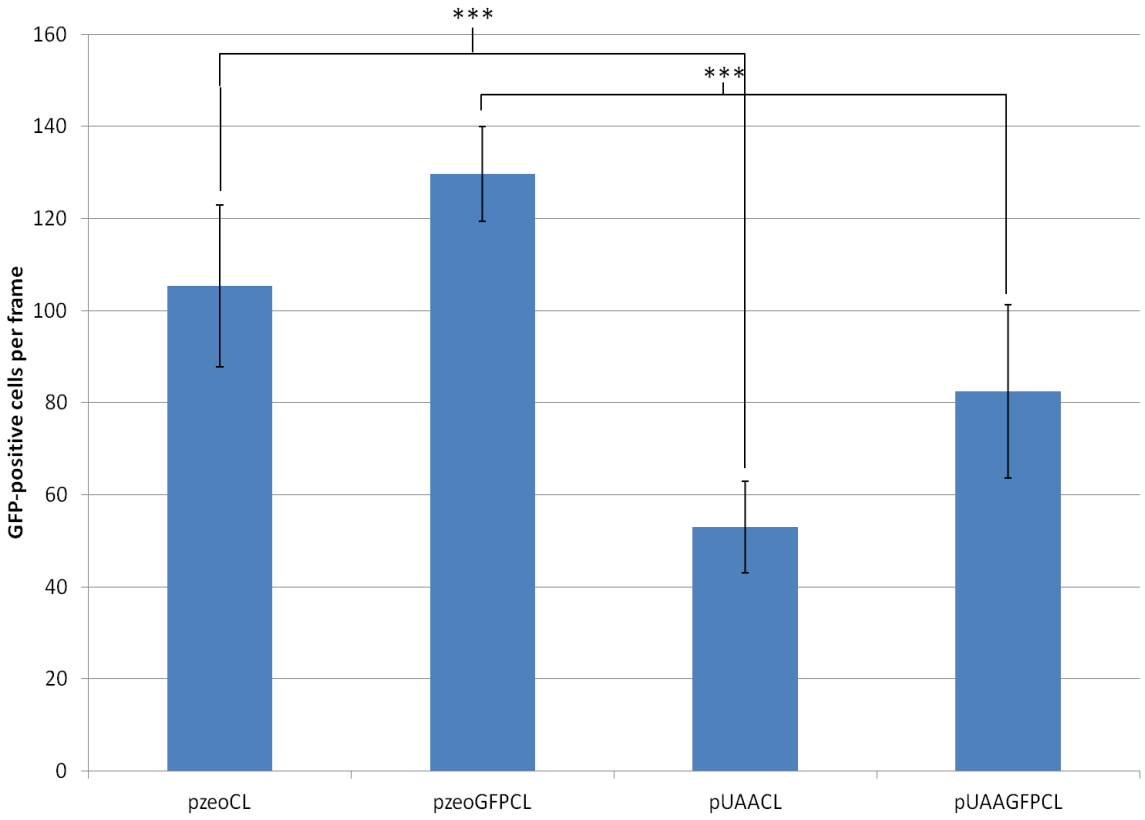


Figure 122. Graph showing transfection efficiency of Rac1 control cells transfected with either pzeoCL, pzeoGFPCL, pUAACL or pUAAGFPCL, using Fugene. Images were captured using fluorescent microscopy and number of GFP-positive cells counted. Errors represent average of *n*=6 transfections.

6.3.5.2 Rac1 control fibroblast cells are capable of amber suppression

The ability of control cells to produce NAA-containing gene using transient transfection of plasmid pY23TAG, was investigated. Quantitative analysis of fluorescent microscopy demonstrated that some Rac1 control cells transiently transfected with pY23TAG plasmid were GFP-positive when the medium was supplemented with either iodoF or azF (Figure 123). This experiment demonstrated that the fibroblasts were capable of taking up the NAA and performing amber suppression.

6.3.5.3 NAA can be dissolved by sonication or NaOH

For previous experiments, the NAA had been dissolved in growth medium and sonicated to ensure the NAA was fully dissolved. This was undertaken to limit the addition of substances to cell culture to NAA only. The NAA did not dissolve easily in growth medium and therefore this limited the concentrations of NAA that could be used. Previous work (Ye et al. 2008) had shown that NAA dissolved in NaOH could be used in amber suppression in mammalian culture. Therefore, iodoF dissolved in NaOH were added at varying concentrations to control cells transiently transfected with pY23TAG and compared to NAA prepared by sonication. It was demonstrated that using NaOH dissolved iodoF, a higher number of GFP-positive cells were obtained than for the same concentration using sonicated NAA (Figure 124; $p < 0.0001$). Adding 1 mM iodoF was optimal for both sonicated and NaOH dissolved iodoF ($p < 0.0001$). GFP-positive cells were also obtained when NaOH dissolved azF was used, although the efficiency was lower than that when using iodoF. This experiment demonstrated that NaOH dissolved NAA could be used in mammalian cell culture.

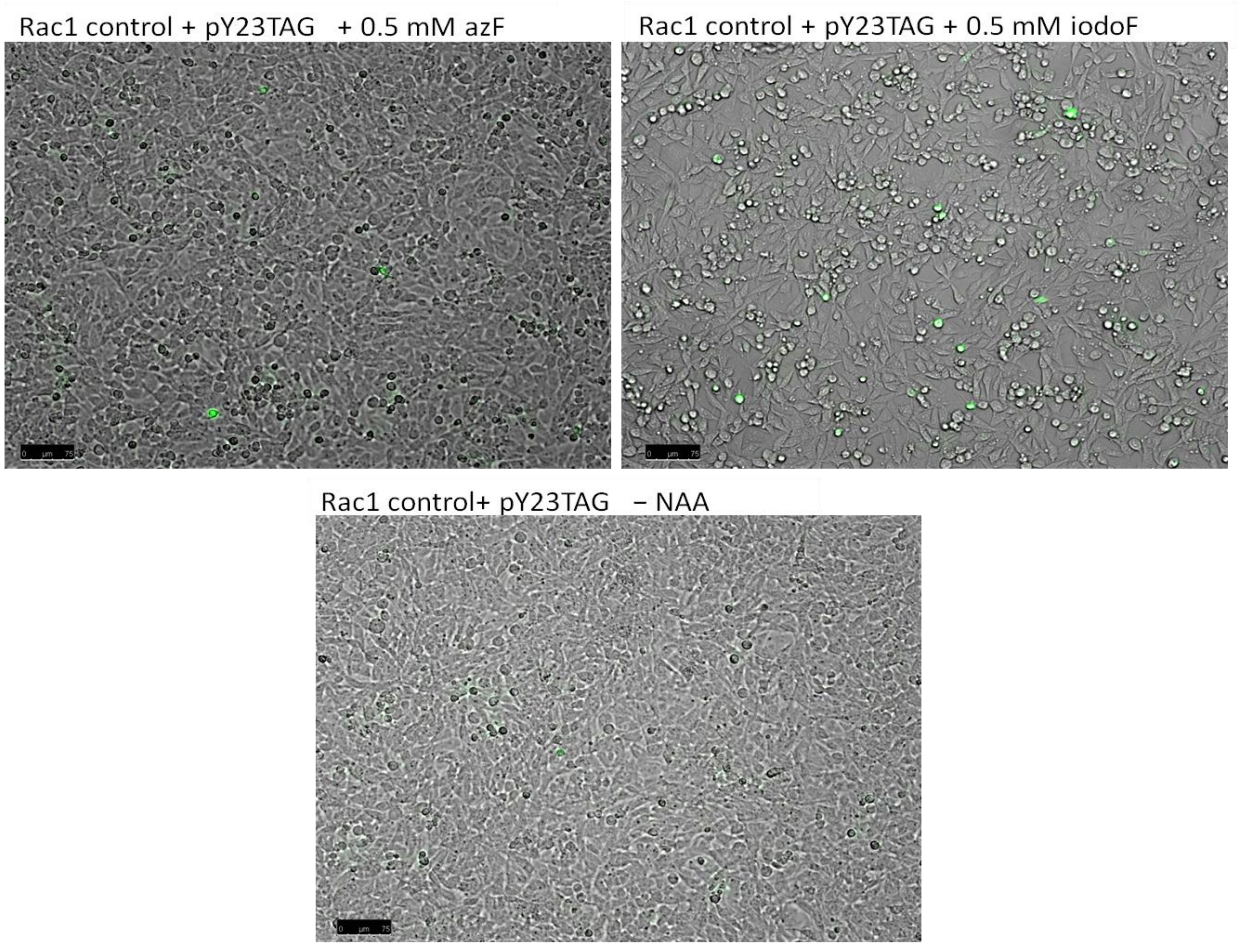


Figure 123. Composite brightfield/ fluorescent images of Rac1 control cells transiently transfected with plasmid pY23TAG, plus/minus nonnatural amino acid. Scale bar = 75 µm.

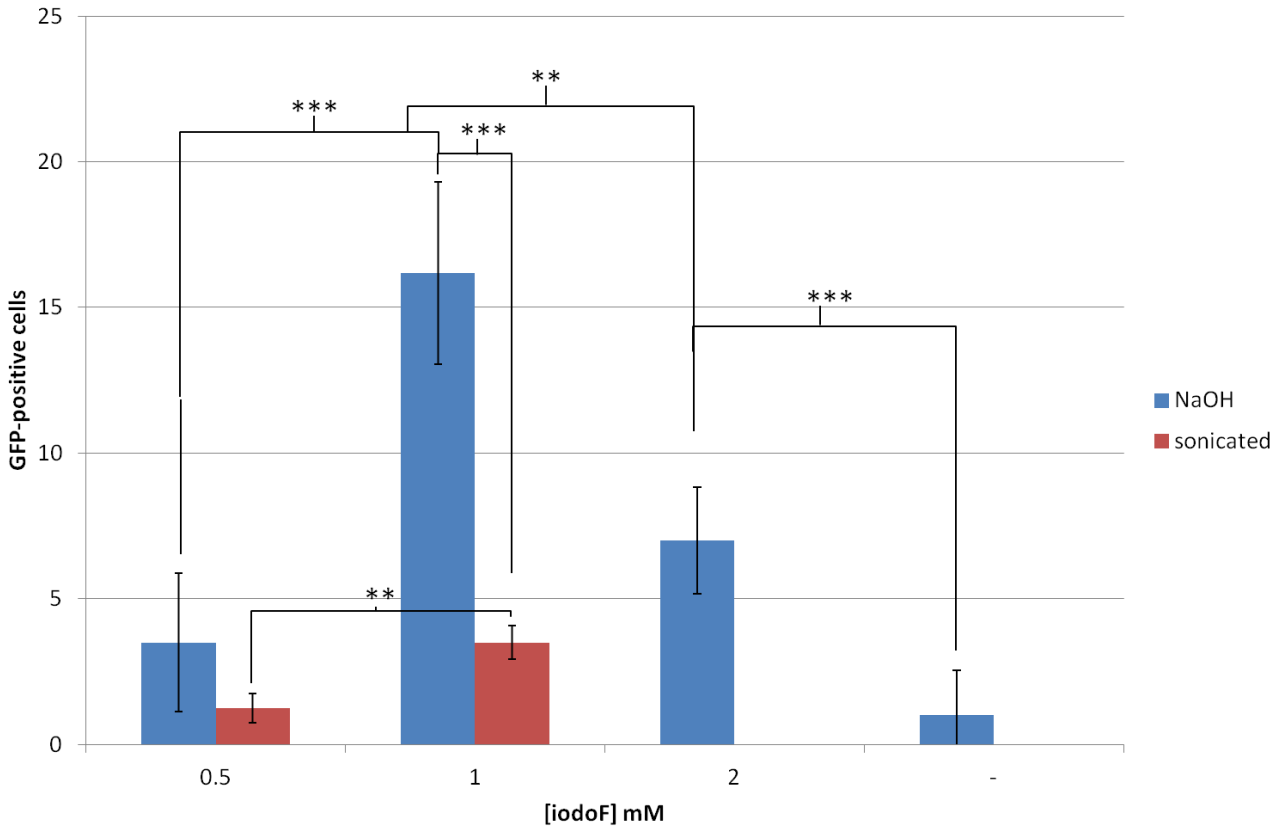


Figure 124. Graph comparing number of GFP-positive cells obtained when control cells were transiently transfected with plasmid pY23TAG and either sonicated or NaOH dissolved iodoF added to the media. Error bars represent standard deviation of average from *n*=4 experiments.

6.4 Discussion

6.4.1 Selectable one-plasmid for amber suppression

Amber suppression technology has notoriously low efficiency in mammalian cells (Hino et al. 2005; Hino et al. 2011; Liu et al. 2007; Sakamoto et al. 2002; Shen et al. 2011; Wang et al. 2007b; Ye et al. 2008). This coupled with the low transfection efficiency of fibroblast cells acts in combination to drastically reduce the efficiency of production of NAA-protein, and therefore any phenotype observed. To try to limit these effects, a selectable one plasmid construct was created, based on the pcDNA3.1/zeo plasmid. In theory only cells successfully transfected with the plasmid and have successfully integrated the DNA into their genome, survive growth in medium containing Zeocin antibiotic, therefore it is only the efficiency of amber suppression that limits the success of the technology.

6.4.1.1 Transient transfection

It was demonstrated that the one-plasmid amber suppression system (pUAA-based plasmids) could be used transiently to express non-TAG target gene, both by GFP fluorescence and RNA analysis. This demonstrated that one plasmid system, pUAA, was functional. The transient expression, however, was less efficient than that from the pRacGFPCL and pzeoCL plasmids. This was most probably due to the more complex nature of the plasmids, and the presence of *E*. *coli* genes, the expression of which are not well tolerated in mammalian culture. It was also demonstrated that transfection efficiency varied among the cell lines tested; efficiency was lowest for the Rac1 KO cells, followed by the Rac1 control fibroblasts and the HEK293 cells showed highest efficiency. The effect of transfection efficiency was important in the application of amber suppression technology and will be discussed in section 6.4.2.

6.4.1.2 Stable transfection

The experiments in Chapter 5 demonstrated that stably transfected Rac1 KO cells expressing Rac1-GFPCL or eGFP only could be produced by selecting those cells that

were Zeocin resistant (at 100 µg / ml) and were positive for GFP (as ascertained by fluorescent microscopy). This was encouraging and therefore provided the rational to attempt to create a stably transfected cell line expressing all the components required for amber suppression using the pUAA-based plasmids. It was important to note that, because only some of the pzeoCL and pzeoGFPCL Zeocin resistant colonies actually expressed GFP, it was expected that only a proportion of the pUAA stable cell lines would express GFP. However, none of the resistant colonies (using plasmids pUAACL, pUAAGFPCL) were shown to be GFP-positive, therefore were probably not expressing Rac1 and therefore were not used for further analysis.

This inability to create stable cell colonies could have been due to toxicity of the encoded genes. Although the cells tolerated short-term transient expression of the genes, long-term expression, as required during the production of stable cell lines, could have overloaded the cell with toxic protein and caused subsequent cell death or silencing of the genes. It was later discovered that the cell line expressed the large T antigen (Personal communication, Klemens Rottner) therefore making the production of stable cell lines using plasmids encoding a SV40 promoter difficult (Sambrook et al. 1989) Cells expressing the large T antigen are capable of replicating SV40-containing plasmids and after the insertion of the plasmid into the genome the replication of the plasmid still occurs (Sambrook et al. 1989). This attempted replication may duplicate and insert parts of DNA sequence into the surrounding DNA sequence and therefore hinder the expression of target protein. It is possible that because the pzeo-based plasmids as used in Chapter 5 expressed only one gene of interest, this replication may not have perturbed expression of the gene. However, the pUAA plasmids are large and complex and the attempted replication of this plasmid sequence may have had more of a consequence on the target gene expression or the genome integrity. Replacing the SV40 sequence in the pUAA plasmids, with, for example the Ef1a promoter, may make the stable transfection of these plasmids possible (Gopalkrishnan *et al.* 1999). It is also possible that some population of the Rac1 KO displayed some intrinsic Zeocin resistance and during selection this sub-population was selected and out-competed those cells that had stably integrated the target plasmid. However, this was

unlikely because there were no resistant colonies isolated during the production of the kill curve.

It is also possible that after integration of plasmid DNA into the host genome, the genes had been silenced by being incorporated into silenced parts of the genome or silenced post-integration due to the foreign nature of the target genes (Doerfler *et al.* 1997; Lichtenberg *et al.* 1988). The successful isolation of one GFP-positive, resistant colony that subsequently lost GFP-expression during growth suggests that post-integration events had affected the expression of protein from the plasmid integration site. The fact that not all resistant colonies expressed target gene would make things even more difficult for selection of the TAG-containing plasmids, as these cells would not be GFP-positive in the absence of NAA. Attempts were made to select resistant colonies in the presence of NAA however; no colonies were GFP-positive. The issues discussed above could be responsible for this finding.

6.4.2 Amber suppression

6.4.2.1 Cell lines successfully utilising amber suppression

Amber suppression technology using previously designed plasmids (Liu et al. 2007; Ye et al. 2009) and the one-plasmid system created within this study was successfully applied to the HEK293 cell line. Although the exact cell type of human embryonic kidney cell line is unknown (Graham *et al.* 1977), this HEK293 cell line has been widely used for a number of applications due to its robust nature, quick growth, easy maintenance, amenability to high transfection efficiency using a wide variety of methods; high protein production with faithful translation and processing, especially using constructs with CMV promoters (Thomas and Smart 2005). It is therefore an optimal cell type in which to perform amber suppression. This is evidenced by the fact that previous applications of amber suppression studies have utilised this cell line (Köhrer *et al.* 2001; Shen et al. 2011; Takimoto et al. 2009; Wang et al. 2007b; Ye et al. 2008; Ye et al. 2010).

Other mammalian cell lines have been used for amber suppression and these include CHO cells (Hino et al. 2005; Liu et al. 2007; Sakamoto et al. 2002; Taki et al.

2002) and HeLa cells (Takimoto et al. 2009); used, once again, for their high transfection efficiencies and high fidelity recombinant protein expression (Jayapal *et al.* 2007). It is important to demonstrate that different cell lines can be used for amber suppression, as the applications of this technology will inevitably expand and the use of specialised cell lines will become increasingly important. When this work was initially undertaken, the cell lines routinely successfully utilised for amber suppression were limited to those cell lines mentioned above and one instance successful in neurons (Wang et al. 2007b).

Wang et al (2007b), successfully performed amber suppression in neurons at an efficiency that could be detected by fluorescence microscopy. The group subsequently studied NAA-containing integral membrane protein Kv1.4 that was expressed using HEK293 cells. The group reported that amber suppression efficiencies in their study ranged from 13% to 41% relative to Tyr incorporation by the WT TyrRS. It is worth noting that the efficiency reported was "relative to Tyr incorporation by wild-type TyrRS" and therefore takes into account the efficiency of amber suppression via the orthogonal tRNA and aaRS, therefore, a lower empirical efficiency can be proposed. The efficiency of amber suppression in producing azF-luciferase was reported by Ye et al (2009) to be 13 %. During the research presented within this Thesis, efficiency of incorporation of Tyr was not ascertained and therefore efficiencies cannot be compared, however, this demonstrates that using specialised cell lines presents efficiency issues.

Subsequent to the initiation of this Thesis investigation, Shen et al (2011) successfully performed amber suppression in neurons differentiated from the neural stem cell line, HCN-A94. Amber suppression was utilised to incorporate a fluorescent NAA into the voltage-sensitive domain (VSD) of *Ciona intestinalis* voltage-sensitive phosphatase. The NAA was then used to report the conformational changes in the VSD in response to membrane depolarisation. The group stably transfected the cells with the machinery for amber suppression using lentiviral vectors. Shen et al's work demonstrated that amber suppression can be used in defined cell types, namely neuronal cells. The efficiency of NAA incorporation was not reported however, the efficiency must have been high

enough to produce enough for protein for subsequent studies. Therefore, the use of lentiviral vectors could be considered for application of amber suppression in the Rac1 KO fibroblasts because it may act to increase the ability to create stable cell lines and also improve amber suppression.

The application of amber suppression to a variety of cell lines is important in expanding the applications of this technology, for example in studying cell line specific events. The use of fibroblasts, as attempted within this Thesis, has been informative in developing this technology within more specialised cell lines.

6.4.2.2 Amber suppression in fibroblasts

Although it was demonstrated that amber suppression in Rac1 KO cells was successful, the efficiency was too low for further analysis. It is possible that NAA incorporation had occurred in more cells but if expression of protein was too low to be determined via GFP fluorescence, the efficiency would most probably not be high enough to visualise cellular motility via scratch wound assay, which would require higher levels of functional protein. However, during this study, amber suppression, using the pY23TAG plasmid, was demonstrated to be possible in HEK293 cells and Rac1 control fibroblasts. These experiments were crucial to show that amber suppression was successful in fibroblasts. Along with the observation that specialised cell lines, namely neurons, can be utilised for amber suppression (Shen et al. 2011), these considerations indicated that an issue specifically regarding the knock-out of Rac1 had perturbed the cells' ability to perform amber suppression and demonstrated the complexities that can arise in using specialised cell lines for amber suppression. Numerous possibilities for this will be considered below.

6.4.2.2.1 Transfection efficiency

It was demonstrated that the transfection efficiency was very low in Rac1 KO cells when compared to Rac1 control fibroblasts and HEK293 cells. In addition, the pUAA-based plasmids showed lower transfection efficiencies compared to pzeo-based or pRacGFPCL plasmids in all the cell lines transfected. These considerations suggest that transfection efficiency using pY23TAG (a pUAA-based plasmid) was

very low in the Rac1 KO cells. Therefore only a small population of the cells were in fact capable of amber suppression. This combined with the low efficiency of amber suppression probably accounted for the very low numbers of GFP-positive cells during the experiments using NAA. To address this issue, methods to improve transfection efficiency could be employed, for example, utilisation of other transfection methods (for example, other proprietary transfection reagents, $CaPO_4$, PEI, physical methods (For review see Luo and Saltzman 2000)); the use of stable cell lines (discussed above in section 6.4.1.2); and the use of a lentiviral-based system (as used by Shen et al (2011)). To ascertain transfection efficiency the inclusion of another plasmid encoding a fluorescent protein (for example mCherry) during transfection would allow visualisation of transfection efficiency separate from the GFP used in elucidating amber suppression efficiency.

6.4.2.2.2 Efficiency of amber suppression

Due to the way in which the Rac1-GFP construct was designed and created, GFP fluorescence should have been indicative of whether NAA incorporation had occurred. This is because read-through of the TAG should ideally only occur when NAA has been incorporated. However, there is always a chance that a natural amino acid, most likely Tyrosine, could be incorporated and in this case green cells would be seen (Nehring et al. 2012). GFP positive cells were visualised in the experiments when NAA was withheld from the medium, however, there were at statistically much lower levels when compared to when NAAs were included in the medium. To ensure that full-length Rac1 was being expressed, a Western blot could have been performed using an antibody specific to full-length Rac1. In addition, using an antibody for the N-terminal of Rac1, the ratio of full-length to truncated protein could be estimated. To ensure NAA incorporation was occurring, target protein could be expressed, purified and analysed by mass spectroscopy (Sakamoto et al. 2002).

In an attempt to improve formation of stable cell lines or expression of foreign DNA, sodium butyrate could have been utilised (Gorman *et al.* 1983; Kruh 1981). It has been suggested that sodium butyrate enhances protein expression by

hyperacetyltion of histones and inhibition of histone deacetylases (HDACs), and therefore acts to increase the availability of the DNA for transcription.

Specificity could be achieved by engineering a specific promoter that would drive the expression of mutant protein and/or orthogonal synthetase at the required time and in response to a specific signal. This would limit any toxicity due to aaRS, mutant or truncated target protein and could help improve efficiency of amber suppression. Control over expression of target genes has been utilised previously in amber suppression applications and is discussed further in section 6.4.2.2.6. It could be beneficial if the Rac1 gene was controlled by the Rac1 endogenous promoter (rather than via CMV in these investigations), as a result the signalling cascade that culminates in upregulation of Rac1 could act to limit over-expression of Rac1 limiting production of truncated protein before NAA was present. In addition, upregulation of Rac1 as a result of polarisation in response to scratch wounding and subsequent motility in to the wound could be brought about.

Expression of orthogonal amber suppression machinery in the mammalian cell lines in the presence of NAA could also result in NAA being incorporated at TAG codons used at the translational termination position of endogenous genes. Read-through of these TAGs could produce elongated proteins which could have altered function and signalling, which may affect the cell growth or viability (Köhrer et al. 2001; Sakamoto et al. 2002). Control over the expression of the orthogonal machinery would aid to limit this effect by limiting amber suppression to a specified time frame.

Similar investigations to those performed utilising the amber suppression *Mj* machinery (used in *E. coli* culture) as described in section 4.3.5.1, namely, the optimisation of amber suppression by creating orthogonal ribosomes or mutation of other factors involved in translation, could increase amber suppression efficiency in mammalian cells.

6.4.2.2.3 NAA uptake and transport

It is considered that NAAs are taken up by mammalian cells via the endogenous amino acid uptake machinery (Wang et al. 2007b). The system L amino acid

transporter which specifically facilitates uptake of aromatic amino acids for example, Tyr, Phe, is the most likely candidate due to structural similarities between NAAs and amino acids recognised by this transporter (Giese *et al.* 2008; Malandro and Kilberg 1996). However, uptake and metabolism of NAAs has not been rigorously investigated. Further investigations into uptake of NAA and improvement of this will be beneficial to amber suppression technology and its applications.

Giese et al (2008) demonstrated fluorinated tryptophan analogues are taken up by human breast cancer cells (MCF-7) through active transport using the system L amino acid transporter. It was also demonstrated that NAA accumulated at a 70-fold excess of intracellular over extracellular concentrations. Nehring et al (2012) suggested that it is this accumulation of NAA at high intracellular concentrations that provides the means for the measurable efficiency of amber suppression observed. Without this accumulation, the recognition of tyr, the original substrate for the orthogonal NAA-aaRSs, would occur more frequently (Nehring et al. 2012). However, the group also demonstrated that these NAAs effectively and irreversibly inhibited cell growth. The toxicity arising from the indiscriminate substitution of the amino acid analogues into proteins, which disrupts their function. Limited studies in to the uptake of p-benzoylphenylalanine (Bpa) by yeast have been performed. It was demonstrated that at concentrations of <2 mM a methionine–Bpa dipeptide was shown to increase the delivery of Bpa into yeast, while at 2 mM, both dipeptide and Bpa were equally effective (Huang *et al.* 2008).

It has been demonstrated that uptake of NAAs that vastly deviate structurally from canonical amino acids in side chain could be improved by esterification of the carboxyl group (For example, DanAla, as performed by Takimoto *et al.* 2010). Takimoto's argument was that the addition of an ester would increase the percentage of neutral moieties and increase lipophilicity and thereby increase translocation across the plasma membrane. Once inside the cell, intracellular esterases would cleave the ester and regenerate the original NAA for incorporation. The efficiency of uptake was quantified using incorporation of NAA versus esterified-NAA in GFP in conjunction with flow cytometry; and HPLC (Takimoto et

al. 2010). If uptake of NAA into Rac1 KO was limited due to the charged nature of the NAA, esterification or another modification may improve uptake and therefore efficiency of amber suppression.

The work by Huang et al (2008) and Takimoto (2010) hint that uptake of NAA could be improved by modifying the NAA or the formation of a dipeptide. Further investigations in to the mechanism of cellular uptake of NAA or augmentation of the NAA to improve uptake could improve amber suppression efficiency. Future work could include analysis of uptake of NAA using technologies as developed by Takimoto (2010) or using a Northern blot-based assay as utilised by Ye et al (2008). This method allows separation and visualisation of two species of tRNA; those unbound and those bound to NAA.

To overcome issues regarding NAA uptake via cognate amino acid transporters, a method using electroporation (Neumann *et al.* 1982) was attempted to try to increase uptake of NAA. Although, typically, it is reported that electroporation can be used to introduce nucleic acid molecules, peptides and dyes in to mammalian cells, it was thought that the pores created during electroporation could facilitate uptake of NAA. Electroporation had previously been used in NAA-incorporation technology to deliver aminoacylated-tRNAs into cells that were capable of expressing orthogonal aaRS and TAG-containing target genes (Monahan *et al.* 2003). Therefore, this provided the rational for attempting electroporation as a method to deliver NAA in to the cells. However, the data provided in this Chapter demonstrated no improvement in amber suppression efficiency; however, further optimisation could be useful in ascertaining whether electroporation could be used to deliver particularly complex NAAs that might not be recognised by the endogenous amino acid uptake machinery.

Liposomal transfection (Felgner *et al.* 1987) was also considered as a method to attempt to improve uptake of NAA. Although, once again, typically reported for use with nucleic acids, it was hypothesised that if the NAA was present during liposome formation it would also be encapsulated and delivered to the cell. The results indicated that liposomal transfection did not increase the production of NAA-containing Rac1. As further discussed below, Rac1 deficient cells may have limited

ability to traffic vesicles therefore it is possible that exogenous liposomes would be affected by this misregulation.

Starvation of Rac1 KO of non-essential amino acids normally supplemented into the growth media was another method to try to increase uptake of NAA. In this situation, the hypothesis was that at the point of re-supplementing the cells with natural amino acids or NAAs, uptake of amino acids would be upregulated because the cell required them for protein synthesis. The similarities between natural and NAA would mean that the cell would take up both in equal quantity and this would increase the concentration of NAA within the cell for amber suppression. However, this starvation did not appear to increase amber suppression.

As amber suppression had been shown to be possible using the Rac1 control fibroblasts, it was considered that the knockout of Rac1 had implications beyond the limited ability of the cells to migrate. The physical changes in the cells when Rac1 had been knocked out were evident by their different appearance and growth rate (as demonstrated in Chapter 5). It could have been this difference in morphology that hinted at limitations in the cells ability to either uptake exogenous factors such as plasmid DNA (as observed in reduced transfection efficiency) or NAA, or in their ability to perform amber suppression.

Rac1 has been demonstrated to be involved in trafficking to and from the plasma membrane (Benjamin *et al.* 2011; De Curtis 2001; Radhakrishna *et al.* 1999). Interruption of this trafficking by knockout of Rac1 could have the potential to affect the cells ability to take up plasmids and NAAs. Although typically associated with endosomal or vesicular trafficking, the irregular signalling regarding the actin cytoskeleton and associated cargo movement due to the Rac1 KO could also have accounted for why uptake of NAA could have been a problem in the application of amber suppression.

6.4.2.2.4 Gene context of the amber stop codon

It has been reported previously that the gene context of the amber codon can affect the efficiency of amber suppression and therefore levels of protein produced (Antonczak 2012 Thesis; Antonczak *et al.* 2011; Bossi 1983; Normanly *et al.* 1986).

It is possible that the position of NAA incorporation within Rac1 was not optimal. A directed evolution-type method to sample possible NAA incorporation sites could be used to optimise insertion position and protein expression efficiency (Daggett *et al.* 2009). In fact, the application of directed evolutionary approach to sampling TAG insertion position has been reported to be successful (Baldwin *et al.* 2009). This method would screen unlikely and un-anticipated successful positions for insertion of NAA and by testing function of Rac1 protein; an ideal position could therefore be identified.

In those cells where NAA incorporation was successful, it was possible that expressed protein would not be functional. The mutation of the residue could have had a deleterious effect on Rac1 folding or function; or a gain of toxic function. The removal of the Tyr could also have potential loss of function due to the abolition of the phosphorylation site. This consideration would not pose a significant problem in Rac1-WT cells (HEK293 and Rac1 control fibroblasts) because functional Rac1 would be present within the cells. However, within the Rac1 KO cells, the presence of non-functional Rac1 could interfere with a variety of signalling cascades and result in severe alteration of phenotype. This could have resulted in cell death before accumulation of NAA-containing RAC1 protein to visible and functional levels.

Also worthy of consideration is that the build up of truncated protein where NAA incorporation had failed had a deleterious effect on the health or viability of the cell. Therefore cells successfully transfected with the pY23TAG plasmid were less likely to survive and grow. The insertion of the TAG mutation close to the N-terminal of the protein was used in an attempt to limit the production of toxic truncated protein by producing very short truncates.

In mammalian cells, another factor limiting the efficiency of NAA incorporation is the stability of the UAG-containing mRNA. Nonsense mediated decay (NMD) orchestrates the degradation of mRNAs that contain premature stop codons and presents a significant problem when performing NAA incorporation via amber suppression. Wang and Wang (2008) have addressed this issue in yeast, by developing a yeast strain lacking an essential component of the NMD mechanism,

UPF1. Strains lacking UPF1 exhibited a two-fold increase in target gene expression when compared to wild type cells. The development of a similar system within higher eukaryotes could help to improve the efficiency of NAA incorporation using amber suppression by increasing the mRNA levels.

6.4.2.2.5 Expression of suppressor tRNA

The number of *Bs*tRNA$^{NAA}_{CUA}$ copies used in the application of amber suppression varies from one copy (as used within this study and Wang et al. 2007b), three (Liu et al. 2007), four (Shen et al. 2011), nine (Hino et al. 2011; Sakamoto et al. 2002) and to a ten-fold excess of tRNA$^{NAA}_{CUA}$ containing plasmid (Ye et al. 2009). It was considered that more than one copy of orthogonal tRNA was required for efficient amber suppression due to the efficiency of expression and function of a bacterial gene in mammalian cells. However, during this study, plasmid pY23TAG contained one copy of *Bs*tRNA$^{NAA}_{CUA}$ and this was demonstrated to be sufficient to allow amber suppression in HEK293 cells and Rac1 control fibroblasts. In fact the supplementation of extra plasmid containing *Bs*tRNA$^{NAA}_{CUA}$ in to the transfection did not appear to increase the efficiency of amber suppression. However, this could not be considered a direct comparison as the *Bs*tRNA$^{NAA}_{CUA}$ copies were not in tandem as for the other applications but supplemented on a separate plasmid. Unequal transfection efficiency of either of the plasmids or difference in levels of expression from different plasmid backbones could account for this difference. To check the expression of tRNA$^{NAA}_{CUA}$ and whether the tRNA$^{NAA}_{CUA}$ was efficiently loaded with NAA, a Northern blot assay could be utilised (Ye et al. 2008).

Mammalian cells stably expressing amber suppressing tRNA$^{NAA}_{CUA}$s have been created and shown not to perturb cell growth (Hudziak *et al.* 1982). On the other hand, it has been suggested that constitutive expression of suppressor tRNA is detrimental to mammalian cells and that regulation of expression of the tRNA$^{NAA}_{CUA}$ could aid in the efficiency of amber suppression in mammalian cells by limiting the toxicity conferred by tRNA$^{NAA}_{CUA}$ (Köhrer et al. 2001).

Both *E. coli* tRNA$^{NAA}_{CUA}$ (*Ec*tRNA$^{NAA}_{CUA}$) and *Bacillus stearothermophilus* tRNA$^{NAA}_{CUA}$ (*Bs*tRNA$^{NAA}_{CUA}$) have been utilised for amber suppression in mammalian cells.

Although *Ec*tRNA$^{NAA}_{CUA}$ would be complementary to the *Ec*TyrRS used within this study, the *Ec*tRNA$^{NAA}_{CUA}$ does not express well in mammalian cells due to the absence of an A-box which is required for efficient transcription by RNA polymerase III. Therefore the *Bs*tRNA$^{NAA}_{CUA}$ has previously (Kiga et al. 2002; Sakamoto et al. 2002), and within this study, been utilised in amber suppression in mammalian cells as it contains a natural A-box within its sequence (Bedouelle 1990; Sakamoto et al. 2002). However, various ways to increase the transcription of *Ec*tRNA$^{NAA}_{CUA}$ have been investigated. Wang et al (Wang et al. 2007b), have demonstrated that the use of the H1 promoter at the 5' end of the tRNA gene allows efficient expression of *Ec*tRNA$^{NAA}_{CUA}$ without the need of an intact A-box. Therefore, it could be beneficial to use the *Ec*tRNA$^{NAA}_{CUA}$ in conjunction with *Ec*TyrRS in this study, because the recognition of *Ec*tRNA$^{NAA}_{CUA}$ by *Ec*TyrRS is cognate and may improve efficiency of aminoacylation and therefore efficiency of amber suppression. By using either *Bs*tRNA$^{NAA}_{CUA}$ or *Ec*tRNA$^{NAA}_{CUA}$, efficiency of amber suppression could be improved by further engineering the tRNA$^{NAA}_{CUA}$ to improve expression, recognition by TyrRS or function during amber suppression.

6.4.2.2.6 Expression of aaRS

In order to ensure the expression of azRS from the pY23TAG plasmid was sufficient, supplementation of azRS was achieved by the transfection of extra azRS-containing plasmid (TPS192asRS_V1). It was demonstrated that this supplementation did not improve the amber suppression efficiency in HEK293 cells, indicating that expression of azRS from pY23TAG was occurring at sufficient levels. To perform a more detailed analysis of azRS expression, a Western blot could have been performed.

NAA-specific synthetases are produced from a library of synthetase genes harbouring mutations at five defined positions that have been rationally selected. These positions were chosen based on the crystal structure of the homologous *Bacillus stearothermophilus* TyrRS and are within 6.5 Å of the para position of the aryl ring of the bound tyrosine (Chin et al. 2003a). Takimoto et al (2009) recognised that the mutation of the anticodon of orthogonal tRNA$^{Tyr}_{GUA}$ to CUA changes the

affinity of the tRNA towards its synthetase. The group therefore addressed this issue by attempting to improve the recognition of the *Ec*TyrRS towards the $tRNA_{CUA}^{NAA}$ by engineering the anticodon-binding region of the aaRS. The work demonstrated that an Asp265Arg mutation in multiple NAA-specific aaRSs increased the activity toward $tRNA_{CUA}^{NAA}$, when compared to the original aaRS. However, it has been considered previously that a rational design approach may not be ideal when engineering enzymes for altered functions. Instead, the use of a directed evolutionary method may sample mutations not considered during rational design (Arnold and Volkov 1999; Dalby 2003). Therefore, the Asp265Arg mutation along with further engineering of the aaRSs used in amber suppression may improve efficiency and fidelity of NAA recognition.

Control over the expression of NAA-synthetases has been implemented using Tet*O* regulation in conjunction with the T-REx-CHO cells (Invitrogen), which constitutively express the tetracycline repressor (Hino et al. 2005; Hino et al. 2011; Liu et al. 2007; Sakamoto et al. 2002). The addition of tetracycline to the growth media induces expression of the NAA-aaRS, thereby allowing temporal control over the expression of the potentially toxic NAA-aaRS. However, this method of control is limited, at this time, to the use of the T-REx-CHO cell line which contains the tetracycline repressor gene. Introduction of this gene into other cell lines or expression from a plasmid could expand this application to more useful cell lines. Control over expression of orthogonal machinery would limit the incorporation of NAA at TAGs at termination points of endogenous genes, which may increase health and viability of the cells and also improve, and add more control over, amber suppression.

6.4.2.2.7 Promiscuity of azRS

Six synthetase variants have been evolved that recognise azF (Chin et al. 2003a) (Table 17). Of these six, Ye et al (2009), further analysed their ability to efficiently incorporate azF. The group demonstrated that azRS_V1 was most effective at amber suppression and this azRS was therefore utilised in the work within this Chapter. As shown in Chapter 3 and previous work (Antonczak 2012 Thesis), the *Mj* aaRSs evolved for certain NAAs for amber suppression in *E. coli* are promiscuous

towards the NAA that they can recognise and aminoacylate tRNA$_{CUA}^{NAA}$ with. During the work presented within this Chapter, it was ascertained that the *E. coli* aaRS developed for incorporation of azF via amber suppression in mammalian cells also showed promiscuity towards the NAA it recognised and successfully aminoacylated. Namely, in the incorporation of iodoF in to target protein. This finding was useful because iodoF is more stable compared to azF. azF has been demonstrated to be thermo- and photo-reactive (Chapter 3 and references within); therefore storage of the diluted amino acid was unwise. Using iodoF in place of azF has the benefits that stored iodoF and iodoF used throughout the culture period should remain chemically unchanged.

In consideration of this promiscuity, the residues within various mutant aaRSs were compared (Table 17) (Brick et al. 1989; Chin et al. 2003a) . As with mutant *Mj*aaRSs, the residue at 37 is crucial in distinguishing cognate Tyr binding from NAA binding. In each of the iodoF and azF mutants the hydrogen bonding performed by Tyr37 and Asp182 to cognate Tyr is abolished. Instead a hydrophobic residue resides in place of Tyr37 and Ser (or Asn) in place of Asp182 (Table 17). These residues therefore reduce recognition of Tyr and expand the binding cavity to allow accommodation of the bulkier side chains of the NAAs. The other mutated residues are similar between aaRSs evolved for iodoF and azF. Within the group of aaRSs for azF there is represented a residue that is also seen within the group of iodoF-specific aaRSs (Table 17). Therefore, these similarities may account for why iodoF is recognised by the azF-specific aaRS.

Table 17. Important residues in mutant aaRSs for amber suppression in mammalian cells (Chin et al. 2003a). azF 1 was used in this study.

	Residue				
NAA-specific aaRS	**37**	**126**	**182**	**183**	**186**
WT *Ec* TyrRS	Tyr	Asn	Asp	Phe	Leu
iodoRS 1	Val	Asn	Ser	Tyr	Leu
iodoRS 2	Iso	Asn	Ser	Met	Leu
iodoRS 3	Val	Asn	Ser	Met	Ala
azF 1	Leu	Asn	Ser	Met	Ala
azF 2	Val	Asn	Ser	Ala	Ala
azF 3	Leu	Asn	Ser	Ala	Ala
azF 4	Val	Asn	Ser	Ala	Val
azF 5	Ile	Asp	Asn	Phe	Val
azF 6	Thr	Asn	Ser	Ala	Leu

6.4.3 Controlling phenotype

The aim of the work presented within this Chapter and Chapter 4 was to control the motile phenotype of a mammalian and bacterial cell respectively. Although the application demonstrated limited success, the amber suppression technology has been explored previously to produce a cell readout system controllable by the presence of NAA. In addition other methods to control phenotype have been utilised, some of these applications that relate to amber suppression or Rac1 protein will be discussed below.

Liu et al (2011), noted that bacteria control the expression of some of their genes through the use of 5' transcriptional regulation units called *cis*-regulatory leader peptides. The translation of this leader peptide determines whether the downstream genes are transcribed efficiently. The group inserted a TAG mutation into the sequence of this leader peptide, so that when amber suppression machinery was introduced into the cell, it was capable of inserting NAA within the peptide. The group produced switches that when NAA was supplemented; the leader peptide was produced, and orchestrated the production (ON) or the hindrance (OFF) of the downstream gene, which, in this case was a test protein, GFP.

Huang et al (2008), have also investigated whether amber suppression can be utilised to control phenotype of a cell, in this case using *Saccharomyces cerevisiae*. The group investigated the production of NAA-containing (Bpa) Ste2p, a GPCR which arrests growth of cells in response to tridecapeptide pheromone α factor. The group explored whether NAA-containing Ste2p could bind ligand (α factor) and ultimately transduce the signal across the cell membrane, and in turn cause arrest of growth, as determined by formation of a clear halo (of no growth) around a disk impregnated with α factor when placed on top of an agar lawn of the yeast cells. The group demonstrated that in the presence of NAA, Ste2p was produced and the phenotype was observed (arrest of growth), whereas when NAA was absent the phenotype was reduced. The group noted the importance of TAG read-through, as some functional receptor was produced when NAA was absent from the culture

which resulted in the specific phenotype being observed. Also noted was the expression level and tolerance of NAA within insertion position affected whether functional receptors were produced and whether the phenotype was observed.

Another application to attempt to control eukaryotic phenotype via amber suppression was described by Edwards et al (2009). The group created a photocontrollable Cre-recombinase protein. Cre protein catalyses the DNA strand exchange at loxP sites introduced in to flanking regions of a target gene allowing excision. Cre mediates this process utilising nucleophilic tyrosine-324. The group replaced tyrosine-324 with a photocaged o-nitrobenzyl tyrosine (ONBY) using amber suppression in *E. coli* cells. Protein was introduced in to HEK293T cells and the photocaged group was unmasked by irradiation. This resulted in subsequent function of the Cre protein, removal of target gene and loss of phenotype.

A further report described the development of a genetically encoded photoactivatable derivative of Rac1 (Wang *et al.* 2010b; Wu *et al.* 2009). The group created a Rac1-LOV (photoreactive light oxygen voltage domain from phototropin) construct, in which Rac1 is inactive (in the dark) due to steric blocking by the LOV domain. Irradiation of the LOV domain induced a reversible conformational change which restored Rac1 function, resulting in Rac1-induced directed cell movement via lamellipodia formation. This work demonstrated that cell migration could be controlled spatially and temporally, via the intracellular regulation of enzymatic activity of Rac1 using a reversible light-induced reaction.

6.5 Conclusion

The aim of the work presented within this Chapter was to apply the amber suppression technology to control the mammalian cell readout system developed within Chapter 5. To develop this technology, a one-plasmid system harbouring the machinery for amber suppression was designed, created and demonstrated to be functional. However, amber suppression in the Rac1 KO cell line demonstrated low efficiency and despite optimisation of amber suppression using HEK293 cells and Rac1 control fibroblasts, a high enough efficiency to effect a change in phenotype could not be realised. The results suggested that the specific nature of the Rac1 KO had rendered the cells refractory to amber suppression; therefore further work to circumvent these problems could still allow amber suppression control of cell motility.

7 CHAPTER 7: GENERAL CONCLUSIONS

The site-specific incorporation of non-natural amino acids (NAAs) into proteins is a powerful tool used in modern chemical biology. There have been a variety of methods explored to incorporate novel reactivities into protein. The development of the amber suppression method provides a site-specific method that utilises the endogenous translation machinery along with exogenous orthogonal aminoacyl tRNA synthetase/tRNA pairs (aaRS/tRNA) to incorporate NAAs.

This work presented within this thesis has evaluated the amber suppression method and the various applications it has been used for. The use of orthogonal Tyr-derived machinery was the main focus as this machinery was used in the applications presented within this thesis. Namely, the *Methanococcus jannaschii* TyrRS/tRNA for application in *E. coli* cells; and *E. coli* TyrRS/*Bacillus stearothermophilus* tRNA for application in mammalian cells. Established applications of amber suppression include installation of post-translational modification mimics, modulation of protein function, protein structural determination, introduction of novel reactivity and probes. The aims within this Thesis were to expand upon these current applications and to apply amber suppression technology to two novel applications. Specifically, the study of a photoreactive moiety within a defined environment within a protein and the creation of a biological switch controlled by the presence or absence of NAA.

The first novel application of amber suppression, presented within Chapter 3, successfully demonstrated generation of reactive intermediates derived from NAAs within the hydrophobic pocket of a protein interior. Namely, the incorporation of the aryl azide-containing NAA, azF, into the hydrophobic pocket of T4 Lysozyme protein. Photoreaction of the azide group was studied by EPR, which detected the production of triplet phenyl nitrene. This work demonstrated the use of a novel, but also practical technique to generate and study highly reactive intermediates. The work presented within Chapter 3 has been accepted for publication by ACS Organic Letters.

The second novel application, presented within Chapter 4, 5 and 6, was to develop a cellular readout system that could ultimately be controlled by the presence/absence of NAA. Here, the amber suppression method itself was the

basis for the novel application rather than the chemical reactivity of the NAA. The hypothesis was to exploit the nature of the method that two scenarios occur depending on the presence/absence of the non-natural amino acid. The production of NAA-containing full-length protein would occur in the presence of NAA and would result in observation of phenotype, whereas when NAA was withheld, only truncated non-functional protein would be produced, and the phenotype would not be observed.

Chapter 4 explored this duality of amber suppression in an attempt to control the motility of an *E. coli* cell by the modulation of flagellin (and therefore flagella) production. It was hoped that in developing this readout system, a new method to screen NAA-specific aaRSs could be developed. The selection of efficient aaRSs could be achieved by picking *E. coli* cells that display efficient swimming ability after the supplementation of NAA, namely, those from the outside of the swimming halo on a motility assay plate. A cell readout system was successfully created using a FliC (flagellin gene) knockout strain and plasmid-encoded FliC (with or without TAG mutation). However, when the amber suppression machinery was introduced in to the cell along with NAA, motility could not be restored to the non-motile strain. The complex nature of the flagellin subunit export or formation into functional flagella could have accounted for this; or possibly that the efficiency of NAA incorporation was too low to produce enough flagellin monomers to form functional flagella.

Chapter 5 demonstrated the successful development of a cellular readout system for mammalian cells. Once again, motility of the cell was chosen as a visual phenotype and the target gene was Rac1 GTPase, which is vital for the production of lamellipodia in fibroblasts and without it limited motility is observed. In cells harbouring Rac1 knockout, this limited motility was observed and recovery of motility was observed after the re-introduction of Rac1 via plasmid.

Chapter 6 sought to develop the readout system developed in Chapter 5 and apply amber suppression to control the readout. A selectable one-plasmid system, harbouring the target gene and orthogonal machinery for amber suppression was designed and created. However, after introduction into the Rac1 KO fibroblasts,

the motility could not be restored due to the low efficiency of target protein expression, despite optimisations. Although, amber suppression was not successful in Rac1 KO fibroblasts, it was successful in Rac1 control fibroblasts and HEK293 cells. This demonstrated that the plasmid specially created for the purpose was functional. Therefore, it was surmised that the specific nature of the Rac1 KO had significantly reduced transfection efficiency or rendered the cells refractory to amber suppression. Further optimisation of amber suppression in *E*. *coli* and mammalian culture could improve the efficiency of expression of NAA-containing target gene and therefore improve the success of these applications. In addition, the selection of less complex read-out systems could also be considered.

In summary, the work presented within this Thesis has attempted to expand the applications of the amber suppression method and has therefore contributed to this expanding scientific field.

8 References

Abbaci, M., Barberi-Heyob, M., Blondel, W., Guillemin, F. and Didelon, J. (2008). Advantages and limitations of commonly used methods to assay the molecular permeability of gap junctional intercellular communication. *Biotechniques* **45:**33.

Abercrombie, M. (1980). The Croonian lecture, 1978: the crawling movement of metazoan cells. *Proceedings of the Royal Society of London. Series B, Biological Sciences* **207:**129-147.

Abercrombie, M., Heaysman, J. E. M. and Pegrum, S. M. (1970). The locomotion of fibroblasts in culture:: III. Movements of particles on the dorsal surface of the leading lamella. *Experimental cell research* **62:**389-398.

Abercrombie, M., Heaysman, J. E. M. and Pegrum, S. M. (1971). The locomotion of fibroblasts in culture:: IV. Electron microscopy of the leading lamella. *Experimental cell research* **67:**359-367.

Aizawa, S., Vonderviszt, F., Ishima, R. and Akasaka, K. (1990). Termini of Salmonella flagellin are disordered and become organized upon polymerization into flagellar filament* 1. *Journal of molecular biology* **211:**673-677.

Alben, J. O. and Fager, L. Y. (1972). Infrared studies of azide bound to myoglobin and hemoglobin temperature dependence of ionicity. *Biochemistry* **11:**842-847.

Alfonta, L., Zhang, Z., Uryu, S., Loo, J. A. and Schultz, P. G. (2003). Site-Specific Incorporation of a Redox-Active Amino Acid into Proteins. *Journal of the American Chemical Society* **125:**14662-14663.

Allen, W. E., Zicha, D., Ridley, A. J. and Jones, G. E. (1998). A role for Cdc42 in macrophage chemotaxis. *The Journal of cell biology* **141:**1147-1157.

Anand-Apte, B., Zetter, B. R., Viswanathan, A., Qiu, R. G., Chen, J., Ruggieri, R. and Symons, M. (1997). Platelet-derived growth factor and fibronectin-stimulated migration are differentially regulated by the Rac and extracellular signal-regulated kinase pathways. *Journal of Biological Chemistry* **272:**30688-30692.

Antonczak, A. (2012). Site-specific incorporation of non-standard amino acids : optimisation and applications. Thesis PhD, Cardiff University.

Antonczak, A., Simova, Z. and Tippmann, E. (2009). A critical examination of Escherichia coli esterase activity. *Journal of Biological Chemistry* **284:**28795.

Antonczak, A. K., Morris, J. and Tippmann, E. M. (2011). Advances in the mechanism and understanding of site-selective noncanonical amino acid incorporation. *Current opinion in structural biology* **21:**481-487.

Ardern, H., Sandilands, E., Machesky, L. M., Timpson, P., Frame, M. C. and Brunton, V. G. (2006). Src-dependent phosphorylation of Scar1 promotes its association with the Arp2/3 complex. *Cell Motility and the Cytoskeleton* **63:**6-13.

Armstrong, J. B., Adler, J. and Dahl, M. M. (1967). Nonchemotactic Mutants of Escherichia coli. *Journal of Bacteriology* **93:**390-398.

Arnold, F. H. and Volkov, A. A. (1999). Directed evolution of biocatalysts. *Current opinion in chemical biology* **3:**54-59.

Aronheim, A., Engelberg, D., Li, N., Al-Alawi, N., Schlessinger, J. and Karin, M. (1994). Membrane targeting of the nucleotide exchange factor Sos is sufficient for activating the Ras signaling pathway. *Cell* **78:**949-961.

Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, Duvaud S *et al.* (2012). ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res*: pp. W597 - W603.

Ayers, B., Blaschke, U. K., Camarero, J. A., Cotton, G. J., Holford, M. and Muir, T. W. (1999). Introduction of unnatural amino acids into proteins using expressed protein ligation. *Peptide Science* **51:**343-354.

Azuma, T., Witke, W., Stossel, T. P., Hartwig, J. H. and Kwiatkowski, D. J. (1998). Gelsolin is a downstream effector of rac for fibroblast motility. *The EMBO journal* **17:**1362-1370.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. *et al.* (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular systems biology* **2**.

Bain, J. D., Switzer, C., Chamberlin, A. R. and Benner, S. A. (1992). Ribosome-mediated incorporation of a nonstandard amino-acid into a peptide through expansion of the genetic code. *Nature* **356:**537-539.

Baldini, G., Martoglio, B., Schachenmann, A., Zugliani, C. and Brunner, J. (1988). Mischarging Escherichia coli tRNAPhe with L-4'-[3-(trifluoromethyl)-3H-diazirin-3-yl]phenylalanine, a photoactivatable analog of phenylalanine. *Biochemistry* **27:**7951-7959.

Baldwin, A. J., Arpino, J. A. J., Edwards, W. R., Tippmann, E. M. and Jones, D. D. (2009). Expanded chemical diversity sampling through whole protein evolution. *Mol. BioSyst.* **5:**764-766.

Beatty, K. E., Xie, F., Wang, Q. and Tirrell, D. A. (2005). Selective dye-labeling of newly synthesized proteins in bacterial cells. *Journal of the American Chemical Society* **127:**14150-14151.

Bedouelle, H. (1990). Recognition of tRNA(Tyr) by tyrosyl-tRNA synthetase. *Biochimie* **72:**589-598.

Bedouelle, H., Guez-Ivanier, V. and Nageotte, R. (1993). Discrimination between transfer-RNAs by tyrosyl-tRNA synthetase. *Biochimie* **75:**1099-1108.

Benjamin, S., Weidberg, H., Rapaport, D., Pekar, O., Nudelman, M., Segal, D., Hirschberg, K. *et al.* (2011). EHD2 mediates trafficking from the plasma membrane by modulating Rac1 activity. *Biochemical Journal* **439:**433-442.

Berg, H. and Turner, L. (1993). Torque generated by the flagellar motor of Escherichia coli. *Biophysical journal* **65:**2201-2216.

Berrier, A. L., Mastrangelo, A. M., Downward, J., Ginsberg, M. and LaFlamme, S. E. (2000). Activated R-ras, Rac1, PI 3-kinase and PKCε can each restore cell spreading inhibited by isolated integrin β1 cytoplasmic domains. *The Journal of cell biology* **151:**1549-1560.

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

Bianco, A., Townsley, F. M., Greiss, S., Lang, K. and Chin, J. W. (2012). Expanding the genetic code of Drosophila melanogaster. *Nat Chem Biol* **8:**748-750.

Billsten, P., Wahlgren, M., Arnebrant, T., McGuire, J. and Elwing, H. (1995). Structural changes of T4 lysozyme upon adsorption to silica nanoparticles measured by circular dichroism. *Journal of colloid and interface science* **175:**77-82.

Binamé, F., Pawlak, G., Roux, P. and Hibner, U. (2010). What makes cells move: requirements and obstacles for spontaneous cell motility. *Mol. BioSyst.* **6:**648-661.

Biolabs, N. E. (Date of access 2009 - 2013). Double digest finder. http://www.neb.com/nebecomm/doubledigestcalculator.asp#.UJzhlYdg-GO: New England Biolabs.

Bishop, A. L. and Hall, A. (2000). Rho GTPases and their effector proteins. *Biochemical Journal* **348:**241.

Boeck, A., Forchhammer, K., Heider, J., Leinfelder, W., Sawers, G., Veprek, B. and Zinoni, F. (1991). Selenocysteine The 21st amino acid. *Molecular Microbiology* **5:**515-520.

Bogumil, R., Hunter, C. L., Maurus, R., Tang, H.-L., Lee, H., Lloyd, E., Brayer, G. D. *et al.* (1994). FTIR Analysis of the Interaction of Azide with Horse Heart Myoglobin Variants. *Biochemistry* **33:**7600-7608.

Bokoch, G. M. (2003). Biology of the p21-activated kinases. *Annual review of biochemistry* **72:**743-781.

Borden, W. T., Gritsan, N. P., Hadad, C. M., Karney, W. L., Kemnitz, C. R. and Platz, M. S. (2000). The interplay of theory and experiment in the study of phenylnitrene. *Acc Chem Res* **33:**765-771.

Bose, M., Groff, D., Xie, J., Brustad, E. and Schultz, P. G. (2005). The Incorporation of a Photoisomerizable Amino Acid into Proteins in E. coli. *Journal of the American Chemical Society* **128:**388-389.

Bossi, L. (1983). Context effects: translation of UAG codon by suppressor tRNA is affected by the sequence following UAG in the message. *Journal of molecular biology* **164:**73-87.

Brawerman, G. and Yčas, M. (1957). Incorporation of the amino acid analog tryptazan into the protein of Escherichia coli. *Archives of Biochemistry and Biophysics* **68:**112-117.

Bray, D. and Thomas, C. (1975). The actin content of fibroblasts. *Biochemical Journal* **147:**221.

Brick, P., Bhat, T. N. and Blow, D. M. (1989). Structure of tyrosyl-tRNA synthetase refined at 2.3 Å resolution: Interaction of the enzyme with the tyrosyl adenylate intermediate. *Journal of Molecular Biology* **208:**83-98.

Brustad, E., Bushey, M. L., Lee, J. W., Groff, D., Liu, W. and Schultz, P. G. (2008a). A Genetically Encoded Boronate-Containing Amino Acid. *Angewandte Chemie International Edition* **47:**8220-8223.

Brustad, E. M., Lemke, E. A., Schultz, P. G. and Deniz, A. A. (2008b). A General and Efficient Method for the Site-Specific Dual-Labeling of Proteins for Single Molecule Fluorescence Resonance Energy Transfer. *Journal of the American Chemical Society* **130:**17664-17665.

Bucher, G., Tönshoff, C. and Nicolaides, A. (2005). Photochemistry of an azido-functionalized cryptand: Controlling the reactivity of an extremely long-lived singlet aryl nitrene by complexation to alkali cations. *Journal of the American Chemical Society* **127:**6883-6892.

Buchmueller, K., Hill, B., Platz, M. and Weeks, K. (2003). RNA-tethered phenyl azide photocrosslinking via a short-lived indiscriminant electrophile. *J. Am. Chem. Soc* **125:**10850-10861.

Budyka, M. F., Kantor, M. M. and Alfimov, M. V. (1992). Photolysis of phenyl azide. *Russian Chemical Bulletin* **41:**590-591.

Capecchi, M. R. and Gussin, G. N. (1965). Suppression in vitro: Identification of a Serine-sRNA as a "Nonsense" Suppressor. *Science* **149:**417-422.

Carpenter, C. L., Tolias, K. F., Van Vugt, A. and Hartwig, J. (1999). Lipid kinases are novel effectors of the GTPase Rac1. *Advances in enzyme regulation* **39:**299.

Cellitti, S. E., Jones, D. H., Lagpacan, L., Hao, X., Zhang, Q., Hu, H., Brittain, S. M. *et al.* (2008). In vivo incorporation of unnatural amino acids to probe structure, dynamics, and ligand binding in a large protein by nuclear magnetic resonance spectroscopy. *Journal of the American Chemical Society* **130:**9268-9281.

Chapman, O. and Le Roux, J. (1978). 1-Aza-1, 2, 4, 6-cycloheptatetraene. *Journal of the American Chemical Society* **100:**282-285.

Chen, F., Ma, L., Parrini, M. C., Mao, X., Lopez, M., Wu, C., Marks, P. W. *et al.* (2000). Cdc42 is required for PIP2-induced actin polymerization and early development but not for cell viability. *Current Biology* **10:**758-765.

Chen, Y. and Ebrigh, R. H. (1993). Phenyl-azide-mediated photocrosslinking analysis of Cro-DNA interaction. *Journal of molecular biology* **230:**453-460.

Chen, Y., Ebright, Y. W. and Ebright, R. H. (1994). Identification of the target of a transcription activator protein by protein-protein photocrosslinking. *Science (New York, N.Y.)* **265:**90-92.

Chew, T. L., Masaracchia, R. A., Goeckeler, Z. M. and Wysolmerski, R. B. (1998). Phosphorylation of non-muscle myosin II regulatory light chain by p21-activated kinase (γ-PAK). *Journal of muscle research and cell motility* **19:**839-854.

Chin, J., Santoro, S., Martin, A., King, D., Wang, L. and Schultz, P. (2002a). Addition of p-azido-L-phenylalanine to the genetic code of Escherichia coli. *J. Am. Chem. Soc* **124:**9026-9027.

Chin, J. W., Cropp, T. A., Anderson, J. C., Mukherji, M., Zhang, Z. and Schultz, P. G. (2003a). An Expanded Eukaryotic Genetic Code. *Science* **301:**964-967.

Chin, J. W., Cropp, T. A., Chu, S., Meggers, E. and Schultz, P. G. (2003b). Progress Toward an Expanded Eukaryotic Genetic Code. *Chemistry & Biology* **10:**511-519.

Chin, J. W., Martin, A. B., King, D. S., Wang, L. and Schultz, P. G. (2002b). Addition of a photocrosslinking amino acid to the genetic code of Escherichia coli. *Proceedings of the National Academy of Sciences* **99:**11020-11024.

Chin, J. W. and Schultz, P. G. (2002). In Vivo Photocrosslinking with Unnatural Amino Acid Mutagenesis. *ChemBioChem* **3:**1135-1137.

Clark, E. A., King, W. G., Brugge, J. S., Symons, M. and Hynes, R. O. (1998). Integrin-mediated signals regulated by members of the rho family of GTPases. *The Journal of cell biology* **142:**573-586.

Clarke, S. (1992). Protein isoprenylation and methylation at carboxyl-terminal cysteine residues. *Annual review of biochemistry* **61:**355-386.

Cohen, G. N. and Munier, R. (1956). Incorporation of structural analogues of amino acids in bacterial proteins. *Biochimica et biophysica acta* **21:**592-593.

Cory, G. O. C., Cramer, R., Blanchoin, L. and Ridley, A. J. (2003). Phosphorylation of the WASP-VCA domain increases its affinity for the Arp2/3 complex and enhances actin polymerization by WASP. *Molecular cell* **11:**1229-1239.

Cowie, D. B. and Cohen, G. N. (1957). Biosynthesis by Escherichia coli of active altered proteins containing selenium instead of sulfur. *Biochimica et Biophysica Acta* **26:**252-261.

Cowie, D. B., Cohen, G. N., Bolton, E. T. and De Robichon-Szulmajster, H. (1959). Amino acid analog incorporation into bacterial proteins. *Biochimica et Biophysica Acta* **34:**39-46.

Crick, F. (1955). On degenerate templates and adaptor hypothesis. *RNA Tie Club*.

Crick, F. (1958). On protein synthesis. *Symp. Soc. Exp. Biol.* **12:**138-163.

Daggett, K. A., Layer, M. and Cropp, T. A. (2009). A general method for scanning unnatural amino acid mutagenesis. *ACS chemical biology* **4:**109.

Dalby, B., Cates, S., Harris, A., Ohki, E. C., Tilkins, M. L., Price, P. J. and Ciccarone, V. C. (2004). Advanced transfection with Lipofectamine 2000 reagent: primary neurons, siRNA, and high-throughput applications. *Methods* **33:**95-103.

Dalby, P. A. (2003). Optimising enzyme function by directed evolution. *Current opinion in structural biology* **13:**500-505.

Datsenko, K. A. and Wanner, B. L. (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences* **97:**6640-6645.

Dawson, P. E., Muir, T. W., Clark-Lewis, I. and Kent, S. B. (1994). Synthesis of proteins by native chemical ligation. *Science (New York, N.Y.)* **266:**776-779.

Dayel, M. J. and Mullins, R. D. (2004). Activation of Arp2/3 Complex: Addition of the First Subunit of the New Filament by a WASP Protein Triggers Rapid ATP Hydrolysis on Arp2. *PLoS Biol* **2:**e91.

de Boer, P. A. J., Crossley, R. E. and Rothfield, L. I. (1989). A division inhibitor and a topological specificity factor coded for by the minicell locus determine proper placement of the division septum in E. coli. *Cell* **56:**641-649.

De Curtis, I. (2001). Cell migration: GAPs between membrane traffic and the cytoskeleton. *EMBO reports* **2:**277-281.

Deiters, A., Cropp, T. A., Mukherji, M., Chin, J. W., Anderson, J. C. and Schultz, P. G. (2003). Adding amino acids with novel reactivity to the genetic code of

Saccharomyces cerevisiae. *Journal of the American Chemical Society* **125:**11782-11783.

Deiters, A., Cropp, T. A., Summerer, D., Mukherji, M. and Schultz, P. G. (2004). Site-specific PEGylation of proteins containing unnatural amino acids. *Bioorganic & Medicinal Chemistry Letters* **14:**5743-5745.

Deiters, A., Geierstanger, B. H. and Schultz, P. G. (2005). Site-Specific in vivo Labeling of Proteins for NMR Studies. *ChemBioChem* **6:**55-58.

Deiters, A. and Schultz, P. G. (2005). In vivo incorporation of an alkyne into proteins in Escherichia coli. *Bioorganic & Medicinal Chemistry Letters* **15:**1521-1524.

DeLano, W. and Lam, J. (2006). Pymol. DeLano Scientific.

Dixon, M. M., Nicholson, H., Shewchuk, L., Baase, W. A. and Matthews, B. W. (1992). Structure of a hinge-bending bacteriophage T4 lysozyme mutant, Ile3→ Pro. *Journal of molecular biology* **227:**917-933.

Doerfler, W., Schubbert, R., Heller, H., Kämmer, C., Hilger-Eversheim, K., Knoblauch, M. and Remus, R. (1997). Integration of foreign DNA and its consequences in mammalian systems. *Trends in Biotechnology* **15:**297-301.

Doi, Y., Ohtsuki, T., Shimizu, Y., Ueda, T. and Sisido, M. (2007). Elongation factor Tu mutants expand amino acid tolerance of protein biosynthesis system. *Journal of the American Chemical Society* **129:**14458-14462.

Dougherty, D. (2000). Unnatural amino acids as probes of protein structure and function. *Current opinion in chemical biology* **4:**645-652.

Döring, V., Mootz, H. D., Nangle, L. A., Hendrickson, T. L., de Crécy-Lagard, V., Schimmel, P. and Marlière, P. (2001). Enlarging the Amino Acid Set of Escherichia coli by Infiltration of the Valine Coding Pathway. *Science* **292:**501-504.

Edmundson, D. M. (2010). *Plasmid pAAazRS had previously been shown to successfully incorporate azF into proteins using amber suppression* to: Morris, J.

Edwards, H. and Schimmel, P. (1990). A bacterial amber suppressor in Saccharomyces cerevisiae is selectively recognized by a bacterial aminoacyl-tRNA synthetase. *Molecular and cellular biology* **10:**1633-1641.

Edwards, W. F., Young, D. D. and Deiters, A. (2009). Light-activated Cre recombinase as a tool for the spatial and temporal control of gene function in mammalian cells. *ACS chemical biology* **4:**441-445.

Eftink, M. R. (1991). Fluorescence Techniques for Studying Protein Structure.*Methods of Biochemical Analysis*. John Wiley & Sons, Inc., pp. 127-205.

Eggertsson, G. and Soll, D. (1988). Transfer ribonucleic acid-mediated suppression of termination codons in Escherichia coli. *Microbiology and Molecular Biology Reviews* **52:**354.

Eriksson, A., Baase, W. and Matthews, B. (1993). Similar hydrophobic replacements of Leu99 and Phe153 within the core of T4 lysozyme have different structural and thermodynamic consequences. *Journal of molecular biology* **229:**747-769.

Eriksson, A. E., Baase, W. A., Wozniak, J. A. and Matthews, B. W. (1992). A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene.

Etienne-Manneville, S. and Hall, A. (2003). Cell polarity: Par6, aPKC and cytoskeletal crosstalk. *Current Opinion in Cell Biology* **15:**67-72.

Fairn, G. D., Hermansson, M., Somerharju, P. and Grinstein, S. (2011). Phosphatidylserine is polarized and required for proper Cdc42 localization and for development of cell polarity. *Nature cell biology*.

Farooqui, R. and Fenteany, G. (2005). Multiple rows of cells behind an epithelial wound edge extend cryptic lamellipodia to collectively drive cell-sheet movement. *Journal of cell science* **118:**51-63.

Farrell, I. S., Toroney, R., Hazen, J. L., Mehl, R. A. and Chin, J. W. (2005). Photo-cross-linking interacting proteins with a genetically encoded benzophenone. *Nature methods* **2:**377-384.

Fechter, P., Rudinger-Thirion, J., Théobald-Dietrich, A. and Giegé, R. (2000). Identity of tRNA for Yeast Tyrosyl-tRNA Synthetase: Tyrosylation Is More Sensitive to Identity Nucleotides Than to Structural Features†. *Biochemistry* **39:**1725-1733.

Fechter, P., Rudinger-Thirion, J., Tukalo, M. and Giege, R. (2001). Major tyrosine identity determinants in Methanococcus jannaschii and Saccharomyces cerevisiae tRNATyr are conserved but expressed differently. *European Journal of Biochemistry* **268:**761-767.

Feig, L. A. (1999). Tools of the trade: use of dominant-inhibitory mutants of Ras-family GTPases. *Nature cell biology* **1:**E25-E27.

Felgner, P. L., Gadek, T. R., Holm, M., Roman, R., Chan, H. W., Wenz, M., Northrop, J. P. *et al.* (1987). Lipofection: a highly efficient, lipid-mediated DNA-transfection procedure. *Proceedings of the National Academy of Sciences* **84:**7413-7417.

Fenteany, G., Janmey, P. A. and Stossel, T. P. (2000). Signaling pathways and cell mechanics involved in wound closure by epithelial cell sheets. *Current Biology* **10:**831-838.

Firtel, R. A. and Meili, R. (2000). Dictyostelium: a model for regulated cell movement during morphogenesis. *Current Opinion in Genetics & Development* **10:**421-427.

Fleming, S. A. (1995). Chemical reagents in photoaffinity labeling. *Tetrahedron* **51:**12479-12520.

Friedl, P. and Weigelin, B. (2008). Interstitial leukocyte migration and immune function. *Nat Immunol* **9:**960-969.

Frottin, F., Martinez, A., Peynot, P., Mitra, S., Holz, R. C., Giglione, C. and Meinnel, T. (2006). The Proteomics of N-terminal Methionine Cleavage. *Molecular & Cellular Proteomics* **5:**2336-2349.

Fukumoto, Y., Kaibuchi, K., Hori, Y., Fujioka, H., Araki, S., Ueda, T., Kikuchi, A. *et al.* (1990). Molecular cloning and characterization of a novel type of regulatory protein (GDI) for the rho proteins, ras p21-like small GTP-binding proteins. *Oncogene* **5:**1321.

Galli, G., Hofstetter, H. and Birnstiel, M. L. (1981). Two conserved sequence blocks within eukaryotic tRNA genes are major promoter elements.

Gao, Y., Xing, J., Streuli, M., Leto, T. L. and Zheng, Y. (2001). Trp56 of Rac1 specifies interaction with a subset of guanine nucleotide exchange factors. *Journal of Biological Chemistry* **276:**47530-47541.

Gardel, M. L., Schneider, I. C., Aratyn-Schaus, Y. and Waterman, C. M. (2010). Mechanical integration of actin and adhesion dynamics in cell migration. *Annual review of cell and developmental biology* **26:**315-333.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. and Bairoch, A. (2005). *Protein Identification and Analysis Tools on the ExPASy Server*. Humana Press.

Geddes, D. M., Cargill, R. S. and LaPlaca, M. C. (2003). Mechanical stretch to neurons results in a strain rate and magnitude-dependent increase in plasma membrane permeability. *Journal of neurotrauma* **20:**1039-1049.

Gendron, R. P. and Sheppard, D. E. (1974). Mutations in the L-arabinose operon of Escherichia coli B/r that result in hypersensitivity to catabolite repression. *Journal of bacteriology* **117:**417-421.

Giebel, B., Corbeil, D., Beckmann, J., Höhn, J., Freund, D., Giesen, K., Fischer, J. *et al.* (2004). Segregation of lipid raft markers including CD133 in polarized human hematopoietic stem and progenitor cells. *Blood* **104:**2332-2338.

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

Giese, C., Lepthien, S., Metzner, L., Brandsch, M., Budisa, N. and Lilie, H. (2008). Intracellular uptake and inhibitory activity of aromatic fluorinated amino acids in human breast cancer cells. *ChemMedChem* **3:**1449-1456.

Glacy, S. D. (1983). Subcellular distribution of rhodamine-actin microinjected into living fibroblastic cells. *The Journal of cell biology* **97:**1207.

Goodman, H. M., Abelson, J., Landy, A., Brenner, S. and Smith, J. D. (1968). Amber suppression: a nucleotide change in the anticodon of a tyrosine transfer RNA. *Nature* **217:**1019-1024.

Gopalkrishnan, R. V., Christiansen, K. A., Goldstein, N. I., DePinho, R. A. and Fisher, P. B. (1999). Use of the human EF-1α promoter for expression can significantly increase success in establishing stable cell lines with consistent expression: A study using the tetracycline-inducible system in human cancer cells. *Nucleic Acids Research* **27:**4775-4782.

Gorman, C. M., Howard, B. H. and Reeves, R. (1983). Expression of recombinant plasmids in mammalian cells is enhanced by sodium butyrate. *Nucleic acids research* **11:**7631-7648.

Goudarzi, G., Sattari, M., Roudkenar, M. H., Montajabi-Niyat, M., Zavaran-Hosseini, A. and Mosavi-Hosseini, K. (2009). Cloning, expression, purification, and

characterization of recombinant flagellin isolated from Pseudomonas aeruginosa. *Biotechnology letters* **31:**1353-1360.

Graham, F. L., Smiley, J., Russell, W. C. and Nairn, R. (1977). Characteristics of a Human Cell Line Transformed by DNA from Human Adenovirus Type 5. *Journal of General Virology* **36:**59-72.

Greiss, S. and Chin, J. W. (2011). Expanding the Genetic Code of an Animal. *Journal of the American Chemical Society* **133:**14196-14199.

Gritsan, N. and Platz, M. (2006). Kinetics, spectroscopy, and computational chemistry of arylnitrenes. *Chem. Rev* **106:**3844-3867.

Gritsan, N. P. and Pritchina, E. A. (1992). Mechanism of aromatic azides photolysis. *Adv. Khim* **61:**910-939.

Grizot, S., Faure, J., Fieschi, F., Vignais, P. V., Dagher, M. C. and Pebay-Peyroula, E. (2001). Crystal structure of the Rac1-RhoGDI complex involved in nadph oxidase activation. *Biochemistry* **40:**10007-10013.

Guo, F., Debidda, M., Yang, L., Williams, D. A. and Zheng, Y. (2006). Genetic deletion of Rac1 GTPase reveals its critical role in actin stress fiber formation and focal adhesion complex assembly. *Journal of Biological Chemistry* **281:**18652-18659.

Guo, J., Wang, J., Anderson, J. C. and Schultz, P. G. (2008a). Addition of an α-Hydroxy Acid to the Genetic Code of Bacteria. *Angewandte Chemie International Edition* **47:**722-725.

Guo, J., Wang, J., Lee, J. S. and Schultz, P. G. (2008b). Site-Specific Incorporation of Methyl- and Acetyl-Lysine Analogues into Recombinant Proteins. *Angewandte Chemie International Edition* **47:**6399-6401.

Guo, X., Stafford, L. J., Bryan, B., Xia, C., Ma, W., Wu, X., Liu, D. *et al.* (2003). A Rac/Cdc42-specific exchange factor, GEFT, induces cell proliferation, transformation, and migration. *Journal of Biological Chemistry* **278:**13207-13215.

Guzman, L. M., Belin, D., Carson, M. J. and Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* **177:**4121-4130.

Hancock, J. F., Paterson, H. and Marshall, C. J. (1990). A polybasic domain or palmitoylation is required in addition to the CAAX motif to localize p21ras to the plasma membrane. *Cell* **63:**133-139.

Hancock, S. M., Uprety, R., Deiters, A. and Chin, J. W. (2010). Expanding the genetic code of yeast for incorporation of diverse unnatural amino acids via a pyrrolysyl-tRNA synthetase/tRNA pair. *Journal of the American Chemical Society* **132:**14819.

Hao, B., Gong, W., Ferguson, T., James, C., Krzycki, J. and Chan, M. (2002). A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science* **296:**1462.

Hartsock, A. and Nelson, W. J. (2008). Adherens and tight junctions: Structure, function and connections to the actin cytoskeleton. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1778:**660-669.

Himeno, H., Hasegawa, T., Ueda, T., Watanabe, K. and Shimizu, M. (1990). Conversion of aminoacylation specificity from tRNATyr to tRNASer in vitro. *Nucleic Acids Research* **18:**6815-6819.

Hino, N., Okazaki, Y., Kobayashi, T., Hayashi, A., Sakamoto, K. and Yokoyama, S. (2005). Protein photo-cross-linking in mammalian cells by site-specific incorporation of a photoreactive amino acid. *Nature methods* **2:**201-206.

Hino, N., Oyama, M., Sato, A., Mukai, T., Iraha, F., Hayashi, A., Kozuka-Hata, H. *et al.* (2011). Genetic Incorporation of a Photo-Crosslinkable Amino Acid Reveals Novel Protein Complexes with GRB2 in Mammalian Cells. *Journal of Molecular Biology* **406:**343-353.

Hirel, P. H., Schmitter, M. J., Dessen, P., Fayat, G. and Blanquet, S. (1989). Extent of N-terminal methionine excision from Escherichia coli proteins is governed by the side-chain length of the penultimate amino acid. *Proceedings of the National Academy of Sciences* **86:**8247-8251.

Hirshberg, M., Stockley, R. W., Dodson, G. and Webb, M. R. (1997). The crystal structure of human rac1, a member of the rho-family complexed with a GTP analogue. *Nature Structural & Molecular Biology* **4:**147-152.

Hohsaka, T., Kajihara, D., Ashizuka, Y., Murakami, H. and Sisido, M. (1999). Efficient incorporation of nonnatural amino acids with large aromatic groups into streptavidin in in vitro protein synthesizing systems. *Journal of the American Chemical Society* **121:**34-40.

Hohsaka, T. and Sisido, M. (2002). Incorporation of non-natural amino acids into proteins. *Current Opinion in Chemical Biology* **6:**809-815.

Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Marquisee, M., Merrill, S. H., Penswick, J. R. *et al.* (1965). STRUCTURE OF A RIBONUCLEIC ACID. *Science* **147:**1462-1465.

Homma, M., Komeda, Y., Iino, T. and Macnab, R. M. (1987a). The flaFIX gene product of Salmonella typhimurium is a flagellar basal body component with a signal peptide for export. *Journal of Bacteriology* **169:**1493-1498.

Homma, M., Ohnishi, K., Iino, T. and Macnab, R. M. (1987b). Identification of flagellar hook and basal body gene products (FlaFV, FlaFVI, FlaFVII and FlaFVIII) in Salmonella typhimurium. *Journal of Bacteriology* **169:**3617-3624.

Hondal, R. J., Nilsson, B. L. and Raines, R. T. (2001). Selenocysteine in native chemical ligation and expressed protein ligation. *Journal of the American Chemical Society* **123:**5140-5141.

Hou, Y. M. and Schimmel, P. (1989). Modeling with in vitro kinetic parameters for the elaboration of transfer RNA identity in vivo. *Biochemistry* **28:**4942-4947.

Houdebine, L. M. and Attal, J. (1999). Internal ribosome entry sites (IRESs): reality and use. *Transgenic research* **8:**157-177.

Huang, L. Y., Umanah, G., Hauser, M., Son, C., Arshava, B., Naider, F. and Becker, J. M. (2008). Unnatural Amino Acid Replacement in a Yeast G Protein-Coupled Receptor in Its Native Environment†. *Biochemistry* **47:**5638-5648.

Hudziak, R. M., Laski, F. A., Rajbhandary, U. L., Sharp, P. A. and Capecchi, M. R. (1982). Establishment of mammalian cell lines containing multiple nonsense mutations and functional suppressor tRNA genes. *Cell* **31:**137-146.

Huisgen, R. (1955). Aliphatic diazo compounds. *Angew. Chem* **67:**439-463.

Huisgen, R., Vossius, D. and Appl, M. (1958). Die Thermolyse des Phenylazids in primären Aminen; die Konstitution des Dibenzamils. *Chemische Berichte* **91:**1-12.

Hurley, J. H., Baase, W. A. and Matthews, B. W. (1992). Design and structural analysis of alternative hydrophobic core packing arrangements in bacteriophage T4 lysozyme. *Journal of molecular biology* **224:**1143-1159.

Ibba, M. and Söll, D. (2000). Aminoacyl-tRNA synthesis. *Annual Review of Biochemistry* **69:**617-650.

Ibba, M. and Söll, D. (2002). Genetic Code: Introducing Pyrrolysine. *Current Biology* **12:**R464-R466.

Inc., D. (1999). DNAStar.

Inokuchi, H., Yamao, F., Sakano, H. and Ozeki, H. (1979). Identification of transfer RNA suppressors in Escherichia coli: I. Amber suppressor su+2, an anticodon mutant of tRNA2Gln. *Journal of Molecular Biology* **132:**649-662.

Jackson, J. C., Hammill, J. T. and Mehl, R. A. (2007). Site-Specific Incorporation of a 19F-Amino Acid into Proteins as an NMR Probe for Characterizing Protein Structure and Reactivity. *Journal of the American Chemical Society* **129:**1160-1166.

Jakubowski, H. and Goldman, E. (1992). Editing of errors in amino acids selection for protein synthesis. *Microbiol Rev* **56:**412-429.

Jayapal, K. P., Wlaschin, K. F., Hu, W. and Yap, M. G. S. (2007). Recombinant protein therapeutics from CHO cells-20 years and counting. *Chemical Engineering Progress* **103:**40.

Johnson, D. B. F., Xu, J., Shen, Z., Takimoto, J. K., Schultz, M. D., Schmitz, R. J., Xiang, Z. *et al.* (2011). RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. *Nature chemical biology*.

Jones, C. J., Homma, M. and Macnab, R. M. (1989). L-, P-, and M-ring proteins of the flagellar basal body of Salmonella typhimurium: gene sequences and deduced protein sequences. *Journal of bacteriology* **171:**3890-3900.

Joyce, D., Bouzahzah, B., Fu, M., Albanese, C., D'Amico, M., Steer, J., Klein, J. U. *et al.* (1999). Integration of Rac-dependent regulation of cyclin D1 transcription through a nuclear factor-κB-dependent pathway. *Journal of Biological Chemistry* **274:**25245-25249.

Kaplan, S., Stretton, A. O. and Brenner, S. (1965). Amber suppressors: efficiency of chain propagation and suppressor specific amino acids. *Journal of molecular biology* **14:**528-533.

Karney, W. L. and Borden, W. T. (1997). Ab initio study of the ring expansion of phenylnitrene and comparison with the ring expansion of phenylcarbene. *Journal of the American Chemical Society* **119:**1378-1387.

Kent, S. B. H. (1988). Chemical Synthesis of Peptides and Proteins. *Annual Review of Biochemistry* **57:**957-989.

Kiga, D., Sakamoto, K., Kodama, K., Kigawa, T., Matsuda, T., Yabuki, T., Shirouzu, M. *et al.* (2002). An engineered Escherichia coli tyrosyl–tRNA synthetase for site-specific incorporation of an unnatural amino acid into proteins in eukaryotic translation and its application in a wheat germ cell-free system. *Proceedings of the National Academy of Sciences* **99:**9715-9720.

Kim, S. H., Quigley, G. J., Suddath, F. L., McPherson, A., Sneden, D., Kim, J. J., Weinzierl, J. *et al.* (1973). Three-Dimensional Structure of Yeast Phenylalanine Transfer RNA: Folding of the Polynucleotide Chain. *Science* **179:**285-288.

Kim, S. H., Sussman, J. L., Suddath, F. L., Quigley, G. J., McPherson, A., Wang, A. H. J., Seeman, N. C. *et al.* (1974). The general structure of transfer RNA molecules. *Proceedings of the National Academy of Sciences* **71:**4970-4974.

Kirshenbaum, K., Carrico, I. S. and Tirrell, D. A. (2002). Biosynthesis of Proteins Incorporating a Versatile Set of Phenylalanine Analogues. *ChemBioChem* **3:**235-237.

Kiselar, J. G., Mahaffy, R., Pollard, T. D., Almo, S. C. and Chance, M. R. (2007). Visualizing Arp2/3 complex activation mediated by binding of ATP and WASp using structural mass spectrometry. *Proceedings of the National Academy of Sciences* **104:**1552-1557.

Knowles, J. (1972). Photogenerated reagents for biological receptor-site labeling. *Accounts of Chemical Research* **5:**155-160.

Kobayashi, T., Nureki, O., Ishitani, R., Yaremchuk, A., Tukalo, M., Cusack, S., Sakamoto, K. *et al.* (2003). Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion. *Nature Structural & Molecular Biology* **10:**425-432.

Kondoh, H. and Ozeki, H. (1976). Deletion and amber mutants of fla loci in Escherichia coli K-12. *Genetics* **84:**403-421.

Kruh, J. (1981). Effects of sodium butyrate, a new pharmacological agent, on cells in culture. *Molecular and cellular biochemistry* **42:**65-82.

Kuwajima, G. (1988a). Construction of a minimum-size functional flagellin of Escherichia coli. *Journal of Bacteriology* **170:**3305-3309.

Kuwajima, G. (1988b). Flagellin domain that affects H antigenicity of Escherichia coli K-12. *Journal of bacteriology* **170:**485.

Kuwajima, G., Kawagishi, I., Homma, M., Asaka, J., Kondo, E. and Macnab, R. M. (1989). Export of an N-terminal fragment of Escherichia coli flagellin by a flagellum-specific pathway. *Proceedings of the National Academy of Sciences* **86:**4953-4957.

Köhrer, C., Xie, L., Kellerer, S., Varshney, U. and RajBhandary, U. L. (2001). Import of amber and ochre suppressor tRNAs into mammalian cells: A general approach to site-specific insertion of amino acid analogues into proteins. *Proceedings of the National Academy of Sciences* **98:**14310-14315.

Ladner, J. E., Jack, A., Robertus, J. D., Brown, R. S., Rhodes, D., Clark, B. F. and Klug, A. (1975). Structure of yeast phenylalanine transfer RNA at 2.5 A resolution. *Proceedings of the National Academy of Sciences* **72:**4414-4418.

Lapouge, K., Smith, S. J. M., Walker, P. A., Gamblin, S. J., Smerdon, S. J. and Rittinger, K. (2000). Structure of the TPR domain of p67phox in complex with Rac·GTP. *Molecular cell* **6:**899-907.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F. *et al.* (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23:**2947-2948.

Laursen, B. S., Sørensen, H. P., Mortensen, K. K. and Sperling-Petersen, H. U. (2005). Initiation of protein synthesis in bacteria. *Microbiology and Molecular Biology Reviews* **69:**101-123.

Le Moigne, V., Robreau, G. and Mahana, W. (2006). Homologous recombination with linear DNA to insert antigenic protein in the flagellin: improvement of the Th1 immune response. *Microbiology and immunology* **50:**33.

Lee, H. S., Dimla, R. D. and Schultz, P. G. (2009). Protein–DNA photo-crosslinking with a genetically encoded benzophenone-containing amino acid. *Bioorganic & Medicinal Chemistry Letters* **19:**5222-5224.

Lee, N., Francklyn, C. and Hamilton, E. P. (1987). Arabinose-induced binding of AraC protein to araI2 activates the araBAD operon promoter. *Proceedings of the National Academy of Sciences* **84:**8814-8818.

Leyva, E., Platz, M., Persy, G. and Wirz, J. (1986). Photochemistry of phenyl azide: the role of singlet and triplet phenylnitrene as transient intermediates. *Journal of the American Chemical Society* **108:**3783-3790.

Leyva, E., Platz, M. S., Niu, B. and Wirz, J. (1987). Arylaminyl radicals studied by laser flash photolysis of di-tert-butyl peroxide in the presence of arylamines. *The Journal of Physical Chemistry* **91:**2293-2298.

Li, Y., Kirby, J., George, M., Poliakoff, M. and Schuster, G. (1988). 1, 2-Didehydroazepines from the photolysis of substituted aryl azides: analysis of their chemical and physical properties by time-resolved spectroscopic methods. *Journal of the American Chemical Society* **110:**8092-8098.

Lichtenberg, U., Zock, C. and Doerfler, W. (1988). Integration of foreign DNA into mammalian genome can be associated with hypomethylation at site of insertion. *Virus Research* **11:**335-342.

Lim, M., Hamm, P. and Hochstrasser, R. M. (1998). Protein fluctuations are sensed by stimulated infrared echoes of the vibrations of carbon monoxide and azide probes. *Proceedings of the National Academy of Sciences* **95:**15315-15320.

Link, A. J., Vink, M. K. S., Agard, N. J., Prescher, J. A., Bertozzi, C. R. and Tirrell, D. A. (2006). Discovery of aminoacyl-tRNA synthetase activity through cell-surface display of noncanonical amino acids. *Proceedings of the National Academy of Sciences* **103:**10180-10185.

Linstrom, P. J. and Mallard, W. G. (2005). NIST Chemistry WebBook, NIST Standard Reference database Number 69. Gaithersburg MD: National Institute of Standards and Technology (http://webbook.nist.gov).

Liu, C. C., Qi, L., Yanofsky, C. and Arkin, A. P. (2011). Regulation of transcription by unnatural amino acids. *Nature biotechnology* **29:**164-168.

Liu, C. C. and Schultz, P. G. (2006). Recombinant expression of selectively sulfated proteins in Escherichia coli. *Nature biotechnology* **24:**1436-1440.

Liu, D. R., Magliery, T. J., Pastrnak, M. and Schultz, P. G. (1997a). Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo. *Proceedings of the National Academy of Sciences* **94:**10092-10097.

Liu, D. R., Magliery, T. J. and Schultz, P. G. (1997b). Characterization of an 'orthogonal' suppressor tRNA derived from E. coli tRNA2Gln. *Chemistry & Biology* **4:**685-691.

Liu, W., Brock, A., Chen, S. and Schultz, P. G. (2007). Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nature methods* **4:**239-244.

Lu, Z., Murray, K. S., Cleave, V. V., LaVallie, E. R., Stahl, M. L. and McCoy, J. M. (1995). Expression of Thioredoxin Random Peptide Libraries on the Escherichia coli Cell Surface as Functional Fusions to Flagellin: A System Designed for Exploring Protein-Protein Interactions. *Nat Biotech* **13:**366-372.

Luo, D. and Saltzman, W. M. (2000). Synthetic DNA delivery systems. *Nature biotechnology* **18:**33-37.

Machesky, L. M., Mullins, R. D., Higgs, H. N., Kaiser, D. A., Blanchoin, L., May, R. C., Hall, M. E. *et al.* (1999). Scar, a WASp-related protein, activates nucleation of actin filaments by the Arp2/3 complex. *Proceedings of the National Academy of Sciences* **96:**3739-3744.

Macnab, R. (2003). How bacteria assemble flagella.

Malandro, M. S. and Kilberg, M. S. (1996). Molecular biology of mammalian amino acid transporters. *Annual review of biochemistry* **65:**305-336.

Marck, C. and Grosjean, H. (2002). tRNomics: analysis of tRNA genes from 50 genomes of Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific features. *RNA* **8:**1189-1232.

Matsubayashi, Y., Ebisuya, M., Honjoh, S. and Nishida, E. (2004). ERK activation propagates in epithelial cell sheets and regulates their migration during wound healing. *Current biology* **14:**731-735.

Matsumura, M., Becktel, W. J. and Matthews, B. W. (1988). Hydrophobic stabilization in T4 lysozyme determined directly by multiple substitutions of Ile 3. *Nature* **334:**406.

Matthews, B., Dahlquist, F. and Maynard, A. (1973). Crystallographic data for lysozyme from bacteriophage T4. *Journal of Molecular Biology* **78:**575-576.

McCoy, S. and Caughey, W. S. (1970). Infrared studies of azido, cyano, and other derivatives of metmyoglobin, methemoglobin, and hemins. *Biochemistry* **9:**2387-2393.

Merrifield, R. B. (1963). Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *Journal of the American Chemical Society* **85:**2149-2154.

Michaelson, D., Ali, W., Chiu, V. K., Bergo, M., Silletti, J., Wright, L., Young, S. G. *et al.* (2005). Postprenylation CAAX processing is required for proper localization of Ras but not Rho GTPases. *Molecular biology of the cell* **16:**1606.

Michaelson, D., Silletti, J., Murphy, G., D'Eustachio, P., Rush, M. and Philips, M. R. (2001). Differential Localization of Rho Gtpases in Live Cells. *The Journal of cell biology* **152:**111.

Minnihan, E. C., Seyedsayamdost, M. R. and Stubbe, J. (2009). Use of 3-Aminotyrosine To Examine the Pathway Dependence of Radical Propagation in Escherichia coli Ribonucleotide Reductase. *Biochemistry* **48:**12125-12132.

Miyada, C. G., Stoltzfus, L. and Wilcox, G. (1984). Regulation of the araC gene of Escherichia coli: catabolite repression, autoregulation, and effect on araBAD expression. *Proceedings of the National Academy of Sciences* **81:**4120-4124.

Miyake-Stoner, S., Miller, A., Hammill, J., Peeler, J., Hess, K., Mehl, R. and Brewer, S. (2009). Probing protein folding using site-specifically encoded unnatural amino acids as FRET donors with tryptophan. *Biochemistry* **48:**5953-5962.

Miyake-Stoner, S. J., Refakis, C. A., Hammill, J. T., Lusic, H., Hazen, J. L., Deiters, A. and Mehl, R. A. (2010). Generating permissive site-specific unnatural aminoacyl-tRNA synthetases. *Biochemistry* **49:**1667-1677.

Modha, R., Campbell, L. J., Nietlispach, D., Buhecha, H. R., Owen, D. and Mott, H. R. (2008). The Rac1 Polybasic Region Is Required for Interaction with Its Effector PRK1. *Journal of Biological Chemistry* **283:**1492-1500.

Monahan, S. L., Lester, H. A. and Dougherty, D. A. (2003). Site-Specific Incorporation of Unnatural Amino Acids into Receptors Expressed in Mammalian Cells. *Chemistry & Biology* **10:**573-580.

Monypenny, J., Zicha, D., Higashida, C., Oceguera-Yanez, F., Narumiya, S. and Watanabe, N. (2009). Cdc42 and Rac family GTPases regulate mode and speed but not direction of primary fibroblast migration during platelet-derived growth factor-dependent chemotaxis. *Molecular and cellular biology* **29:**2730-2747.

Muir, T. W., Sondhi, D. and Cole, P. A. (1998). Expressed protein ligation: A general method for protein engineering. *Proceedings of the National Academy of Sciences* **95:**6705-6710.

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

Mukai, T., Kobayashi, T., Hino, N., Yanagisawa, T., Sakamoto, K. and Yokoyama, S. (2008). Adding l-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. *Biochemical and biophysical research communications* **371:**818-822.

Mukai, T., Wakiyama, M., Sakamoto, K. and Yokoyama, S. (2010b). Genetic encoding of non-natural amino acids in Drosophila melanogaster Schneider 2 cells. *Protein Science* **19:**440-448.

Muramoto, K., Makishima, S., Aizawa, S.-I. and Macnab, R. M. (1999). Effect of Hook Subunit Concentration on Assembly and Control of Length of the Flagellar Hook ofSalmonella. *Journal of Bacteriology* **181:**5808-5813.

Nakamura, Y. (2007). Codon usage database [Online]. Department of Plant Gene Research, Kazusa: Available at: http://www.kazusa.or.jp/codon/ [Accessed: 10/10/2012].

Nakata, H., Ohtsuki, T., Abe, R., Hohsaka, T. and Sisido, M. (2006). Binding efficiency of elongation factor Tu to tRNAs charged with nonnatural fluorescent amino acids. *Analytical biochemistry* **348:**321.

Narayanan, A., LeClaire, L. L., III, Barber, D. L. and Jacobson, M. P. (2011). Phosphorylation of the Arp2 Subunit Relieves Auto-inhibitory Interactions for Arp2/3 Complex Activation. *PLoS Comput Biol* **7:**e1002226.

Nehring, S., Budisa, N. and Wiltschi, B. (2012). Performance Analysis of Orthogonal Pairs Designed for an Expanded Eukaryotic Genetic Code. *PLoS ONE* **7:**e31992.

Neumann, E., Schaefer-Ridder, M., Wang, Y. and Hofschneider, P. H. (1982). Gene transfer into mouse lyoma cells by electroporation in high electric fields. *The EMBO journal* **1:**841.

Neumann, H., Peak-Chew, S. Y. and Chin, J. W. (2008). Genetically encoding N ε-acetyllysine in recombinant proteins.

Nikolić, D. L., Boettiger, A. N., Bar-Sagi, D., Carbeck, J. D. and Shvartsman, S. Y. (2006). Role of boundary conditions in an experimental model of epithelial wound healing. *American Journal of Physiology-Cell Physiology* **291:**C68-C75.

Nobes, C. D. and Hall, A. (1995). Rho, rac, and cdc42 GTPases regulate the assembly of multimolecular focal complexes associated with actin stress fibers, lamellipodia, and filopodia. *Cell* **81:**53.

Nobes, C. D. and Hall, A. (1999). Rho GTPases control polarity, protrusion, and adhesion during cell movement. *The Journal of cell biology* **144:**1235-1244.

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

Normanly, J., Masson, J. M., Kleina, L. G., Abelson, J. and Miller, J. H. (1986). Construction of two Escherichia coli amber suppressor genes: tRNAPheCUA and tRNACysCUA. *Proceedings of the National Academy of Sciences* **83:**6548-6552.

Nowak, M. W., Kearney, P. C., Sampson, J. R., Saks, M. E., Labarca, C. G., Silverman, S. K., Zhong, W. *et al.* (1995). Nicotinic receptor binding site probed with unnatural amino acid incorporation in intact cells. *Science (New York, N.Y.)* **268:**439-442.

Offord, R. E. (1987). Protein engineering by chemical means? *Protein Engineering* **1:**151-157.

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

Ohtsuki, T., Yamamoto, H., Doi, Y. and Sisido, M. (2010). Use of EF-Tu mutants for determining and improving aminoacylation efficiency and for purifying aminoacyl tRNAs with non-natural amino acids. *Journal of biochemistry* **148:**239-246.

Olson, M. F., Ashworth, A. and Hall, A. (1995). An essential role for Rho, Rac, and Cdc42 GTPases in cell cycle progression through G1. *Science (New York, NY)* **269:**1270.

Omelchenko, T., Vasiliev, J. M., Gelfand, I. M., Feder, H. H. and Bonder, E. M. (2003). Rho-dependent formation of epithelial “leader” cells during wound healing. *Proceedings of the National Academy of Sciences* **100:**10788-10793.

Ozawa, K., Loscha, K. V., Kuppan, K. V., Loh, C. T., Dixon, N. E. and Otting, G. (2012). High-yield cell-free protein synthesis for site-specific incorporation of unnatural amino acids at multiple sites. *Biochemical and Biophysical Research Communications*.

O'Donoghue, P., Prat, L., Heinemann, I. U., Ling, J., Odoi, K., Liu, W. R. and Söll, D. (2012). Near-cognate suppression of amber, opal and quadruplet codons competes with aminoacyl-tRNAPyl for genetic code expansion. *FEBS Letters* **586:**3931-3937.

Pankov, R., Endo, Y., Even-Ram, S., Araki, M., Clark, K., Cukierman, E., Matsumoto, K. *et al.* (2005). A Rac switch regulates random versus directionally persistent cell migration. *The Journal of cell biology* **170:**793-802.

Parrish, A. R., She, X., Xiang, Z., Coin, I., Shen, Z., Briggs, S. P., Dillin, A. *et al.* (2012). Expanding the Genetic Code of Caenorhabditis elegans Using Bacterial Aminoacyl-tRNA Synthetase/tRNA Pairs. *ACS Chemical Biology* **7:**1292-1302.

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:**2277-2286.

Petrie, R. J., Doyle, A. D. and Yamada, K. M. (2009). Random versus directionally persistent cell migration. *Nat Rev Mol Cell Biol* **10:**538-549.

Pollard, T. D. (2003). The cytoskeleton, cellular motility and the reductionist agenda. *Nature* **422:**741-745.

Pollard, T. D. (2007). Regulation of actin filament assembly by Arp2/3 complex and formins. *Annu. Rev. Biophys. Biomol. Struct.* **36:**451-477.

Pollard, T. D. and Borisy, G. G. (2003). Cellular Motility Driven by Assembly and Disassembly of Actin Filaments. *Cell* **112:**453-465.

Polycarpo, C., Ambrogelly, A., Bérubé, A., Winbush, S. M., McCloskey, J. A., Crain, P. F., Wood, J. L. *et al.* (2004). An aminoacyl-tRNA synthetase that specifically activates pyrrolysine. *Proceedings of the National Academy of Sciences of the United States of America* **101:**12450-12454.

Poujade, M., Grasland-Mongrain, E., Hertzog, A., Jouanneau, J., Chavrier, P., Ladoux, B., Buguin, A. *et al.* (2007). Collective migration of an epithelial monolayer in response to a model wound. *Proceedings of the National Academy of Sciences* **104:**15988-15993.

Price, L. S., Leng, J., Schwartz, M. A. and Bokoch, G. M. (1998). Activation of Rac and Cdc42 by integrins mediates cell spreading. *Molecular biology of the cell* **9:**1863-1871.

Quinn, C. L., Tao, N. and Schimmel, P. (1995). Species-specific microhelix aminoacylation by a eukaryotic pathogen tRNA synthetase dependent on a single base pair. *Biochemistry* **34:**12489-12495.

Rackham, O. and Chin, J. W. (2005). A network of orthogonal ribosome· mRNA pairs. *Nature chemical biology* **1:**159-166.

Radhakrishna, H., Al-Awar, O., Khachikian, Z. and Donaldson, J. G. (1999). ARF6 requirement for Rac ruffling suggests a role for membrane trafficking in cortical actin rearrangements. *Journal of Cell Science* **112:**855-866.

Raftopoulou, M. and Hall, A. (2004). Cell migration: Rho GTPases lead the way. *Developmental biology* **265:**23-32.

Railey White, E., Reed, T. M., Ma, Z. and Hartman, M. C. T. (2012). Replacing amino acids in translation: expanding chemical diversity with non-natural variants. *Methods*.

Ramakrishnan, V. (2002). Ribosome structure and the mechanism of translation. *Cell* **108:**557-572.

Rao, S. T. and Rossmann, M. G. (1973). Comparison of super-secondary structures in proteins. *J Mol Biol* **76:**241-256.

Rapley, J., Tybulewicz, V. L. J. and Rittinger, K. (2008). Crucial structural role for the PH and C1 domains of the Vav1 exchange factor. *EMBO reports* **9:**655-661.

Rasband, W. (1997). ImageJ. National Institutes of Health, USA.

Reddington, S. (2013). Thesis PhD, Cardiff University.

Reif, K., Nobes, C. D., Thomas, G., Hall, A. and Cantrell, D. A. (1996). Phosphatidylinositol 3-kinase signals activate a selective subset of Rac/Rho-dependent effector pathways. *Current biology: CB* **6:**1445.

Reiser, A., Bowes, G. and Horne, R. (1966). Photolysis of aromatic azides. Part 1.—Electronic spectra of aromatic nitrenes and their parent azides. *Transactions of the Faraday Society* **62:**3162-3169.

Remington, S., Anderson, W., Owen, J., Eyck, L., Grainger, C. and Matthews, B. (1978). Structure of the lysozyme from bacteriophage T4: An electron density map at 2.4 Å resolution* 1. *Journal of Molecular Biology* **118:**81-98.

Rennert, O. M. and Anker, H. S. (1963). On the Incorporation of 5',5',5'-Trifluoroleucine into Proteins of E. coli*. *Biochemistry* **2:**471-476.

Richmond, M. H. (1959). Incorporation of canavanine by *Staphylococcus aureus*. *Biochem J* **73**.

Richmond, M. H. (1962). THE EFFECT OF AMINO ACID ANALOGUES ON GROWTH AND PROTEIN SYNTHESIS IN MICROORGANISMS. *Bacteriological Reviews* **26:**398-420.

Ridley, A. J. and Hall, A. (1992). The small GTP-binding protein rho regulates the assembly of focal adhesions and actin stress fibers in response to growth factors. *Cell* **70:**389-399.

Ridley, A. J., Paterson, H. F., Johnston, C. L., Diekmann, D. and Hall, A. (1992). The small GTP-binding protein rac regulates growth factor-induced membrane ruffling. *Cell* **70:**401-410.

Robertus, J. D., Ladner, J. E., Finch, J. T., Rhodes, D., Brown, R. S., Clark, B. F. C. and Klug, A. (1974). Correlation between three-dimensional structure and chemical reactivity of transfer RNA. *Nucleic Acids Research* **1:**927-932.

Robinson, R. C., Turbedsky, K., Kaiser, D. A., Marchand, J. B., Higgs, H. N., Choe, S. and Pollard, T. D. (2001). Crystal structure of Arp2/3 complex. *Science* **294:**1679-1684.

Rosqvist, R., Håkansson, S., Forsberg, A. and Wolf-Watz, H. Functional conservation of the secretion and translocation machinery for virulence proteins of yersiniae, salmonellae and shigellae.

Rottner, K. (2010). *Transient transfection of plasmid pGFPRac1.* to: Stephens, P.

Rottner, K. (2011). *Rac1 KO cells express the SV40 large T antigen.* to: Stephens, P.

Rowe, L., Ensor, M., Mehl, R. and Daunert, S. (2010). Modulating the Bioluminescence Emission of Photoproteins by in Vivo Site-Directed Incorporation of Non-Natural Amino Acids. *ACS Chemical Biology* **5:**455-460.

Roy, R. S., Allen, O. and Walsh, C. T. (1999). Expressed protein ligation to probe regiospecificity of heterocyclization in the peptide antibiotic microcin B17. *Chemistry & Biology* **6:**789-799.

Ryu, Y. and Schultz, P. (2006). Efficient incorporation of unnatural amino acids into proteins in Escherichia coli. *Nature methods* **3:**263-265.

Rørth, P. (2009). Collective cell migration. *Annual Review of Cell and Developmental* **25:**407-429.

Sakamoto, K., Hayashi, A., Sakamoto, A., Kiga, D., Nakayama, H., Soma, A., Kobayashi, T. *et al.* (2002). Site specific incorporation of an unnatural amino acid into proteins in mammalian cells. *Nucleic acids research* **30:**4692.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual. Second Edition ed. New York: Cold Spring Harbor Laboratory Press.

Sanders, L. C., Matsumura, F., Bokoch, G. M. and de Lanerolle, P. (1999). Inhibition of myosin light chain kinase by p21-activated kinase. *Science* **283:**2083-2085.

Sankaranarayanan, J., Rajam, S., Hadad, C. M. and Gudmundsdottir, A. D. (2010). The ability of triplet nitrenes to abstract hydrogen atoms. *Journal of Physical Organic Chemistry* **23:**370-375.

Schimmel, P. R. and Soll, D. (1979). Aminoacyl-tRNA synthetases: general features and recognition of transfer RNAs. *Annu Rev Biochem* **48:**601-648.

Schmidt, A. and Hall, A. (2002). Guanine nucleotide exchange factors for Rho GTPases: turning on the switch. *Genes & development* **16:**1587-1609.

Schoenhals, G. and Whitfield, C. (1993). Comparative analysis of flagellin sequences from Escherichia coli strains possessing serologically distinct flagellar filaments with a shared complex surface pattern. *Journal of bacteriology* **175:**5395-5402.

Schrock, A. K. and Schuster, G. B. (1984). Photochemistry of phenyl azide: chemical properties of the transient intermediates. *Journal of the American Chemical Society* **106:**5228-5234.

Schultz, K. C., Supekova, L., Ryu, Y., Xie, J., Perera, R. and Schultz, P. G. (2006). A Genetically Encoded Infrared Probe. *Journal of the American Chemical Society* **128:**13984-13985.

Schuster, G. and Platz, M. (1992). Photochemistry of phenyl azide. *Advances in photochemistry* **17:**69-143.

Sega, G. A. (1984). A review of the genetic effects of ethyl methanesulfonate. *Mutation Research/Reviews in Genetic Toxicology* **134:**113-142.

Severinov, K. and Muir, T. W. (1998). Expressed Protein Ligation, a Novel Method for Studying Protein-Protein Interactions in Transcription. *Journal of Biological Chemistry* **273:**16205-16209.

Sharma, N., Furter, R., Kast, P. and Tirrell, D. A. (2000). Efficient introduction of aryl bromide functionality into proteins in vivo. *FEBS Letters* **467:**37-40.

Shen, B., Xiang, Z., Miller, B., Louie, G., Wang, W., Noel, J. P., Gage, F. H. *et al.* (2011). Genetically Encoding Unnatural Amino Acids in Neural Stem Cells and Optically Reporting Voltage-Sensitive Domain Changes in Differentiated Neurons. *Stem Cells* **29:**1231-1240.

Sherman, J. M., Rogers, M. J. and Söll, D. (1992). Competition of aminoacyl-tRNA synthetases for tRNA ensures the accuracy of aminoacylation. *Nucleic Acids Research* **20:**1547-1552.

Short 3rd, G. F., Golovine, S. Y. and Hecht, S. M. (1999). Effects of release factor 1 on in vitro protein translation and the elaboration of proteins containing unnatural amino acids. *Biochemistry* **38:**8808.

Silverman, M. and Simon, M. (1973). Genetic analysis of flagellar mutants in Escherichia coli. *Journal of Bacteriology* **113:**105.

Smolinsky, G., Wasserman, E. and Yager, W. (1962). The EPR of Ground State Triplet Nitrenes. *Journal of the American Chemical Society* **84:**3220-3221.

Sprinzl, M., Hartmann, T., Meissner, F., Moll, J. and Vorderwülbecke, T. (1987). Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic acids research* **15 Suppl:**r53-188.

Steer, B. and Schimmel, P. (1999). Major anticodon-binding region missing from an archaebacterial tRNA synthetase. *Journal of Biological Chemistry* **274:**35601.

Stephens, P., Grenard, P., Aeschlimann, P., Langley, M., Blain, E., Errington, R., Kipling, D. *et al.* (2004). Crosslinking and G-protein functions of transglutaminase 2 contribute differentially to fibroblast wound healing responses. *Journal of Cell Science* **117:**3389-3403.

Stokes, A. L., Miyake-Stoner, S. J., Peeler, J. C., Nguyen, D. P., Hammer, R. P. and Mehl, R. A. (2009). Enhancing the utility of unnatural amino acid synthetases by manipulating broad substrate specificity. *Mol. Biosyst.* **5:**1032-1038.

Stoll, S. and Schweiger, A. (2006). EasySpin, a comprehensive software package for spectral simulation and analysis in EPR. *Journal of Magnetic Resonance* **178:**42-55.

Studier, F. (2005). Protein production by auto-induction in high-density shaking cultures. *Protein expression and purification* **41:**207-234.

Sugawara, T. and Matsuda, T. (1995). Photochemical protein fixation on polymer surfaces via derivatized phenyl azido group. *Langmuir* **11:**2272-2276.

Sun, C. X., Downey, G. P., Zhu, F., Koh, A. L. Y., Thang, H. and Glogauer, M. (2004). Rac1 is the small GTPase responsible for regulating the neutrophil chemotaxis compass. *Blood* **104:**3758-3765.

Svitkina, T. M. and Borisy, G. G. (1999). Arp2/3 complex and actin depolymerizing factor/cofilin in dendritic organization and treadmilling of actin filament array in lamellipodia. *The Journal of cell biology* **145:**1009.

Taki, M., Hohsaka, T., Murakami, H., Taira, K. and Sisido, M. (2002). Position-Specific Incorporation of a Fluorophore−Quencher Pair into a Single Streptavidin through Orthogonal Four-Base Codon/Anticodon Pairs. *Journal of the American Chemical Society* **124:**14586-14590.

Takimoto, J. K., Adams, K. L., Xiang, Z. and Wang, L. (2009). Improving orthogonal tRNA-synthetase recognition for efficient unnatural amino acid incorporation and application in mammalian cells. *Molecular BioSystems* **5:**931-934.

Takimoto, J. K., Xiang, Z., Kang, J. Y. and Wang, L. (2010). Esterification of an unnatural amino acid structurally deviating from canonical amino acids promotes its uptake and incorporation into proteins in mammalian cells. *Chembiochem* **11:**2268-2272.

Tanaka, Y., Bond, M. and Kohler, J. (2008). Photocrosslinkers illuminate interactions in living cells. *Molecular BioSystems* **4:**473-480.

Thomas, P. and Smart, T. G. (2005). HEK293 cell line: A vehicle for the expression of recombinant proteins. *Journal of Pharmacological and Toxicological Methods* **51:**187-200.

Théobald-Dietrich, A., Giegé, R. and Rudinger-Thirion, J. (2005). Evidence for the existence in mRNAs of a hairpin element responsible for ribosome dependent pyrrolysine insertion into proteins. *Biochimie* **87:**813-817.

Tian, F., Tsao, M.-L. and Schultz, P. G. (2004). A Phage Display System with Unnatural Amino Acids. *Journal of the American Chemical Society* **126:**15962-15963.

Tippmann, E. M., Liu, W., Summerer, D., Mack, A. V. and Schultz, P. G. (2007). A Genetically Encoded Diazirine Photocrosslinker in Escherichia coli. *ChemBioChem* **8:**2210-2214.

Tomioka, H., Ichikawa, N. and Komatsu, K. (1993). Photochemistry of 2-(methoxycarbonyl)phenyl azide studied by matrix-isolation spectroscopy. A new slippery energy surface for phenylnitrene. *Journal of the American Chemical Society* **115:**8621-8626.

Tsao, M.-L., Summerer, D., Ryu, Y. and Schultz, P. G. (2006). The Genetic Incorporation of a Distance Probe into Proteins in Escherichia coli. *Journal of the American Chemical Society* **128:**4572-4573.

Tsugita, A. and Inouye, M. (1968). Complete primary structure of phage lysozyme from Escherichia coli T4. *Journal of molecular biology* **37:**201.

Tsukita, S., Tsukita, S., Nagafuchi, A. and Yonemura, S. (1992). Molecular linkage between cadherins and actin filaments in cell—cell adherens junctions. *Current Opinion in Cell Biology* **4:**834-839.

Turner, L., Ryu, W. S. and Berg, H. C. (2000). Real-time imaging of fluorescent flagellar filaments. *Journal of Bacteriology* **182:**2793-2801.

Van Aelst, L. and D'Souza-Schorey, C. (1997). Rho GTPases and signaling networks. *Genes & Development* **11:**2295-2322.

Vartiainen, M. K. and Machesky, L. M. (2004). The WASP-Arp2/3 pathway: genetic insights. *Current opinion in cell biology* **16:**174-181.

Vaughan, R. B. and Trinkaus, J. P. (1966). Movements of Epithelial Cell Sheets In Vitro. *Journal of Cell Science* **1:**407-413.

Veronese, F. M. and Pasut, G. (2005). PEGylation, successful approach to drug delivery. *Drug Discovery Today* **10:**1451-1458.

Vidal, M., Brachmann, R. K., Fattaey, A., Harlow, E. and Boeke, J. D. (1996). Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions. *Proceedings of the National Academy of Sciences* **93:**10315-10320.

Vidali, L., Chen, F., Cicchetti, G., Ohta, Y. and Kwiatkowski, D. J. (2006). Rac1-null mouse embryonic fibroblasts are motile and respond to platelet-derived growth factor. *Molecular biology of the cell* **17:**2377-2390.

Vitorino, P. and Meyer, T. (2008). Modular control of endothelial sheet migration. *Science Signalling* **22:**3268.

Vonderviszt, F., Aizawa, S. and Namba, K. (1991). Role of the disordered terminal regions of flagellin in filament formation and stability. *Journal of molecular biology* **221:**1461-1474.

Végh, B. M., Gál, P., Dobó, J., Závodszky, P. and Vonderviszt, F. (2006). Localization of the flagellum-specific secretion signal in Salmonella flagellin. *Biochemical and Biophysical Research Communications* **345:**93-98.

Wallace, C. J. A. (1995). Peptide ligation and semisynthesis. *Current Opinion in Biotechnology* **6:**403-410.

Wang, J., Xie, J. and Schultz, P. G. (2006). A Genetically Encoded Fluorescent Amino Acid. *Journal of the American Chemical Society* **128:**8738-8739.

Wang, J., Zhang, W., Song, W., Wang, Y., Yu, Z., Li, J., Wu, M. *et al.* (2010a). A Biosynthetic Route to Photoclick Chemistry on Proteins. *Journal of the American Chemical Society* **132:**14812-14818.

Wang, K., Neumann, H., Peak-Chew, S. Y. and Chin, J. W. (2007a). Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nature biotechnology* **25:**770-777.

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

Wang, L., Magliery, T., Liu, D. and Schultz, P. (2000). A New Functional Suppressor tRNA/Aminoacyl- tRNA Synthetase Pair for the in Vivo Incorporation of Unnatural Amino Acids into Proteins. *J. Am. Chem. Soc* **122:**5010-5011.

Wang, L. and Schultz, P. (2005). Expanding the Genetic Code. *Angewandte Chemie International Edition* **44:**34-66.

Wang, L. and Schultz, P. G. (2001). A general approach for the generation of orthogonal tRNAs. *Chemistry & biology* **8:**883-890.

Wang, L., Xie, J., Deniz, A. A. and Schultz, P. G. (2002). Unnatural Amino Acid Mutagenesis of Green Fluorescent Protein. *The Journal of Organic Chemistry* **68:**174-176.

Wang, L., Zhang, Z., Brock, A. and Schultz, P. G. (2003). Addition of the Keto Functional Group to the Genetic Code of Escherichia coli. *Journal Name: Procurement National Academy of Sciences; Journal Volume: 100; Journal Issue: 1; Other Information: Submitted to the Procurement National Academy of Sciences; Volume 100: No.1; PBD: 7 Jan 2003* Medium: X; Size: vp.

Wang, Q. and Wang, L. (2008). New methods enabling efficient incorporation of unnatural amino acids in yeast. *Journal of the American Chemical Society* **130:**6066-6067.

Wang, W., Takimoto, J. K., Louie, G. V., Baiga, T. J., Noel, J. P., Lee, K. F., Slesinger, P. A. *et al.* (2007b). Genetically encoding unnatural amino acids for cellular and neuronal studies. *Nature neuroscience* **10:**1063-1072.

Wang, X., He, L., Wu, Y. I., Hahn, K. M. and Montell, D. J. (2010b). Light-mediated activation reveals a key role for Rac in collective guidance of cell movement in vivo. *Nature cell biology* **12:**591-597.

Wasserman, E. (1971). Electron spin resonance of nitrenes. *Prog. Phys. Org. Chem* **8:**319–336.

Weaver, L. and Matthews, B. (1987). Structure of bacteriophage T4 lysozyme refined at 1.7 Å resolution* 1. *Journal of molecular biology* **193:**189-199.

Weigert, M. G. and Garen, A. (1965). AMINO ACID SUBSTITUTIONS RESULTING FROM SUPPRESSION OF NONSENSE MUTATIONS. I. SERINE INSERTION BY THE SU-1 SUPPRESSOR GENE. *Journal of molecular biology* **12:**448-455.

Wells, C. M., Walmsley, M., Ooi, S., Tybulewicz, V. and Ridley, A. J. (2004). Rac1-deficient macrophages exhibit defects in cell spreading and membrane ruffling but not migration. *Journal of cell science* **117:**1259-1268.

Westwick, J. K., Lambert, Q. T., Clark, G. J., Symons, M., Van Aelst, L., Pestell, R. G. and Der, C. J. (1997). Rac regulation of transformation, gene expression, and actin organization by multiple, PAK-independent pathways. *Molecular and cellular biology* **17:**1324-1335.

Wheeler, A. P., Wells, C. M., Smith, S. D., Vega, F. M., Henderson, R. B., Tybulewicz, V. L. and Ridley, A. J. (2006). Rac1 and Rac2 regulate macrophage morphology but are not essential for migration. *Journal of cell science* **119:**2749-2757.

Wittmann, T., Bokoch, G. M. and Waterman-Storer, C. M. (2004). Regulation of microtubule destabilizing activity of Op18/stathmin downstream of Rac1. *Journal of Biological Chemistry* **279:**6196-6203.

Worthylake, D. K., Rossman, K. L. and Sondek, J. (2000). Crystal structure of Rac1 in complex with the guanine nucleotide exchange region of Tiam1. *Nature* **408:**682-688.

Wu, N., Deiters, A., Cropp, T. A., King, D. and Schultz, P. G. (2004). A genetically encoded photocaged amino acid. *Journal of the American Chemical Society* **126:**14306-14307.

Wu, Y. I., Frey, D., Lungu, O. I., Jaehrig, A., Schlichting, I., Kuhlman, B. and Hahn, K. M. (2009). A genetically encoded photoactivatable Rac controls the motility of living cells. *Nature* **461:**104-108.

Xie, J., Wang, L., Wu, N., Brock, A., Spraggon, G. and Schultz, P. (2004). The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination. *Nature biotechnology* **22:**1297-1301.

Xie, J. M. and Schultz, P. G. (2006). A chemical toolkit for proteins - an expanded genetic code. *Nature Reviews Molecular Cell Biology* **7:**775-782.

Xu, R., Hanson, S. R., Zhang, Z., Yang, Y.-Y., Schultz, P. G. and Wong, C.-H. (2004). Site-Specific Incorporation of the Mucin-Type N-Acetylgalactosamine-α-O-threonine into Protein in Escherichia coli. *Journal of the American Chemical Society* **126:**15654-15655.

Yang, N., Higuchi, O., Ohashi, K., Nagata, K., Wada, A., Kangawa, K., Nishida, E. *et al.* (1998). Cofilin phosphorylation by LIM-kinase 1 and its role in Rac-mediated actin reorganization. *Nature* **393:**809-811.

Yang, W., Hendrickson, W. A., Crouch, R. J. and Satow, Y. (1990a). Structure of ribonuclease H phased at 2 A resolution by MAD analysis of the selenomethionyl protein. *Science (New York, N.Y.)* **249:**1398-1405.

Yang, W., Hendrickson, W. A., Kalman, E. T. and Crouch, R. J. (1990b). Expression, purification, and crystallization of natural and selenomethionyl recombinant ribonuclease H from Escherichia coli. *Journal of Biological Chemistry* **265:**13553-13559.

Yang, X.-L., Skene, R. J., McRee, D. E. and Schimmel, P. (2002). Crystal structure of a human aminoacyl-tRNA synthetase cytokine. *Proceedings of the National Academy of Sciences* **99:**15369-15374.

Yaremchuk, A., Kriklivyi, I., Tukalo, M. and Cusack, S. (2002). Class I tyrosyl-tRNA synthetase has a class II mode of cognate tRNA recognition. *EMBO J* **21:**3829-3840.

Ye, S., Huber, T., Vogel, R. and Sakmar, T. P. (2009). FTIR analysis of GPCR activation using azido probes. *Nature chemical biology* **5:**397-399.

Ye, S., Köhrer, C., Huber, T., Kazmi, M., Sachdev, P., Yan, E. C. Y., Bhagat, A. *et al.* (2008). Site-specific incorporation of keto amino acids into functional G protein-coupled receptors using unnatural amino acid mutagenesis. *Journal of Biological Chemistry* **283:**1525-1533.

Ye, S., Zaitseva, E., Caltabiano, G., Schertler, G. F. X., Sakmar, T. P., Deupi, X. and Vogel, R. (2010). Tracking G-protein-coupled receptor activation using genetically encoded infrared probes. *Nature* **464:**1386-1389.

Yokoyama, K., Uhlin, U. and Stubbe, J. (2010). Site-specific incorporation of 3-nitrotyrosine as a probe of pKa perturbation of redox-active tyrosines in ribonucleotide reductase. *Journal of the American Chemical Society* **132:**8385-8397.

Yonekura, K., Maki-Yonekura, S. and Namba, K. (2002). Growth mechanism of the bacterial flagellar filament. *Research in microbiology* **153:**191-197.

Yonekura, K., Maki-Yonekura, S. and Namba, K. (2003). Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy. *Nature* **424:**643-650.

Yoshioka, K., Aizawa, S. and Yamaguchi, S. (1995). Flagellar filament structure and cell motility of Salmonella typhimurium mutants lacking part of the outer domain of flagellin. *Journal of bacteriology* **177:**1090-1093.

Young, T. S., Ahmad, I., Yin, J. A. and Schultz, P. G. (2010). An enhanced system for unnatural amino acid mutagenesis in E. coli. *Journal of molecular biology* **395:**361.

Zeng, H., Xie, J. and Schultz, P. G. (2006). Genetic introduction of a diketone-containing amino acid into proteins. *Bioorganic & Medicinal Chemistry Letters* **16:**5356-5359.

Zhang, K., Diehl, M. R. and Tirrell, D. A. (2005a). Artificial polypeptide scaffold for protein immobilization. *Journal of the American Chemical Society* **127:**10136-10137.

Zhang, T., Bertelsen, E., Benvegnu, D. and Alber, T. (1993). Circular permutation of T4 lysozyme. *Biochemistry* **32:**12311-12318.

Zhang, Y., Wang, L., Schultz, P. G. and Wilson, I. A. (2005b). Crystal structures of apo wild-type M. jannaschii tyrosyl-tRNA synthetase (TyrRS) and an engineered TyrRS specific for O-methyl-L-tyrosine. *Protein Science* **14:**1340-1349.

Zhang, Z., Gildersleeve, J., Yang, Y.-Y., Xu, R., Loo, J. A., Uryu, S., Wong, C.-H. *et al.* (2004). A New Strategy for the Synthesis of Glycoproteins. *Science* **303:**371-373.

Zhang, Z., Wang, L., Brock, A. and Schultz, P. G. (2002). The Selective Incorporation of Alkenes into Proteins in Escherichia coli. *Angewandte Chemie International Edition* **41:**2840-2842.

9 Appendix

9.1 A1 Plasmids

Table 18. T4L specific plasmids (Chapter 3)

Plasmid name	Backbone	Resistance	Promoter	Control	Gene	Mutations	Terminator	Ori	Size (Kb)	Use	Provider
pET101T4L153TAG	pET101	Amp	T7	*lacI*	T4 Lysozyme	cys free, 153-->TAG	T7	pBR322		expression of T4L153p-iodo, used with , in BL21(DE3)	Dr Eric Tippmann
pAAT4L153TAG	pAA	CAM	T5 *lpp*	*lac*O	T4 Lysozyme *M j* $tRNA_{CUA}$	cys free, 153-->TAG; Missing ACG codon at T152, codon optimised	T5 *rrn*C	p15A	7.0	expression of T4L153, and cloning	Dr Alicja Antonczak
pT4L153TAG		CAM	T5 *lpp*		T4 Lysozyme *Mj* $tRNA_{CUA}$	cys free, 153-->TAG; no histag	t_o *rrn*C				Dr Eric Tippmann
pBADT4LWT	pBAD	Amp	*ara*BAD	*ara*C	T4 Lysozyme	cys free,	*ara*BAD	pBR322	7	expression of WT T4L	Dr Ryan Mehl
pBADT4L153TAG	pBAD	Amp	*ara*BAD	*ara*C	T4 Lysozyme	cys free, 153-->TAG	*ara*BAD	pBR322	7	expression of T4L153pazF	Dr Ryan Mehl
pDULEazF		Tet	*lpp* *lpp*		azRS $tRNA_{CUA}$		*rrnB* *rrnC*	p15A	6.2	expression of azRS for amber suppression	Dr Jason Chin
pAAazRS	pAA	CAM	*T5* *lpp*	lacO	azRS 2 x *M. j* $tRNA_{CUA}$		*rrn*C *rrn*C	p15A	7.8	express synthetase and tRNA, for expressions of T4L153paz	Dr Matt Edmundson

pBKpIodo	pBK	Kan	*glnS'*	-	*Mj* iodoRS		*glnS'*			expression of iodoRS for amber suppression	Ryu and Schultz 2006
pSUPiodoRS	pSUP	CAM	*glnS'* *proK*		*Mj* iodoFaaRS 2 x (3 x $tRNA_{CUA}$)		*glnS'* *proK*	p15A		expression of T4L153p-iodo, used with , in BL21(DE3)	Dr Eric Tippmann

Table 19. FliC specific plasmids (Chapter 4)

pAAFliC	pAA	CAM	T5 *lpp*	lacO	FliC 2 x *M. j* $tRNA_{CUA}$	-	*rrn*C *rrn*C	p15A	8.0	express synthetase and tRNA and WT FliC, introducing into FliC knockout strains	Josephine Morris
pQEFliC	pQE60	Amp	T5	lacO	FliC	Q76R, D121V, T400A	t_o	ColE1	5	introducing into FliC knockout strains	Josephine Morris
pFD313	pBR322	Amp	FliC natural	-	FliC	missing codons 245 - 301	FliC natural	BR322		introducing into FliC knockout strains, basis for cloning pFDFliC	Berry Group
pFDFliC	pBR322	Amp	FliC natural	-	FliC	Y246R, A247S, A301K, V302L,	FliC natural	BR322		introducing into FliC knockout strains	Josephine Morris
pFDFliC239TAG	pBR322	Amp	FliC natural	-	FliC239TAG	239TAG, Y246R, A247S, A301K, V302L,	FliC natural	BR322		introducing into FliC knockout strains, for amber suppression	Josephine Morris
pFDFliC318TAG	pBR322	Amp	FliC natural	-	FliC318TAG	318TAG, Y246R, A247S, A301K, V302L,	FliC natural	BR322		introducing into FliC knockout strains, for amber suppression	Josephine Morris
pFDFliC54TAG	pBR322	Amp	FliC natural	-	FliC318TAG	54TAG, Y246R, A247S, A301K, V302L,	FliC natural	BR322		introducing into FliC knockout strains, for amber suppression	Josephine Morris
pFDFliC4TAG	pBR322	Amp	FliC natural	-	FliC4TAG	4TAG, Y246R, A247S, A301K, V302L,	FliC natural	BR322		introducing into FliC knockout strains, for amber suppression	Josephine Morris

Table 20. Mammalian Plasmid list (Chapter 5, 6)

Name	Backbone	Bacterial resistance	Promoter	Regulation	Gene of importance	Mutation /notes	Terminator	Ori	Size (Kb)	Use	Created/ supplied by
TPS136_pSVB-Yam	pSVBpUC	Amp			$BstRNA_{CUA}$			pUC9	7.7		Prof. Thomas Sakmar
TPS194_Fluc.Y70am	pCMVTnT	Amp	CMV		Firefly luciferase	Y70→TAG	SV40 late pA		5	5' beta-globin leader sequence	Prof. Thomas Sakmar
TPS192_RS-V1	pcDNA3.1	Amp	CMV		azRS		BGHpA	pBR322	6.7	azRS cloned in to pUAA derived plasmids	Prof. Thomas Sakmar
pSWANGFP37TAG		Amp	CMV	TetO	GFP37TAG-c-Myc,6his 3 x $BstRNA_{CUA}$			pUC		expression of GFP37azF in HEK cells under control of amber suppression technology	Dr Wenshe Liu
peGFP		Kan/Neo	CMV	-	eGFP	-	SV40pA	pUC	4.7	plasmid used to make racGFP construct and as control during transfections	Josephine Morris
pGFPRac		Kan/Neo	CMV	-	GFP-rac fusion	-	SV40pA	pUC	5.3	plasmid used to make racGFP construct and used to restore motility to RacKO fibroblasts	Klemens Rottner
pRacGFPCL		Kan/Neo	CMV	-	rac-GFPCL fusion	CLLL sequence moved to C-term of GFP	SV40pA	pUC	5.3	restore motility to RacKO fibroblasts, preliminary to non-natural amino acid technology implementation	Josephine Morris

pRacGFPRK		Kan/Neo	CMV	-	rac-GFPRK fusion	polybasic and CLLL sequence C-term of GFP	SV40pA	pUC	5.3	restore motility to RacKO fibroblasts, preliminary to non-natural amino acid technology implementation	Josephine Morris
pBluescript-IRES		Amp			IRES	IRES inserted between *Bam*HI and *EcoR*I sites		ColE1	3.6	cloning using IRES	Dr Xiao-Qing Wei
pBluescript-IRES-racGFPCL		Amp	-		IRES-racGFPCL			ColE1	5	cloning azRS-IRES-racGFPCL construct	Josephine Morris
pBluescript-IRES-racGFPRK		Amp	-		IRES-racGFPRK			ColE1	5	cloning azRS-IRES-racGFPRK construct	
pBluescript-azRS-IRES-racGFPCL		Amp	-		azRS-IRES-racGFPCL			ColE1	6.3	cloning azRS-IRES-racGFPCL construct into pcDNA3/zeo	Josephine Morris
pBluescript-azRS-IRES-racGFPRK		Amp	-		azRS-IRES-racGFPRK			ColE1	6.3	cloning azRS-IRES-racGFPRK construct into pcDNA3/zeo	Josephine Morris
pcDNA3.1/zeo RS	pcDNA3.1/zeo	Amp	SV40p CMV		Zeocin EcTyrRS	- added AgeI	SV40pA BGHpA	pUC	6.3	Used in cloning Zeocin selectable plasmid	
pzeoCL	pcDNA3.1/zeo	Amp	SV40p CMV		Zeocin racGFPCL	cloned into NotI & AgeI site of	SV40pA BGHpA	pUC	6.4	Selectable rac-encoding plasmid, used to create stably transfected racKO cells, investigate racGFPCL in stable txns	Josephine Morris

						pzeo					
pzeoGFPCL	pcDNA3.1 /zeo	Amp	SV40 p CMV		Zeocin GFPCL	cloned into NotI & AgeI site of pzeo	SV40pA BGHpA	pUC	5.8	Selectable GFPCL encoding plasmid, used to create stably transfected racKO cells, control for investigation into racGFPCL in stable txns	Josephine Morris
pzeoRK	pcDNA3.1 /zeo	Amp	SV40 p CMV		Zeocin racGFPRK	cloned into XbaI & ApaI site of pzeo	SV40pA BGHpA	pUC	6.4	Selectable plasmid, expressing racGFPRK, not used to due toxicity	Josephine Morris
pUAARK	pcDNA3.1 /zeo	Amp	SV40 p CMV		Zeocin azRS-IRES-racGFPRK	-	SV40pA BGHpA	pUC	8.4	not used in mammalian culture, due to intolerance of racGFPRK construct	Josephine Morris
pUAACL4	pcDNA3.1 /zeo	Amp	SV40 p CMV		Zeocin azRS-IRES-racGFPCL	-	SV40pA BGHpA	pUC	8.4	Selectable plasmid used to create stably transfected racKO cells, investigate function of azRS-IRES-RacGFPCL construct	Josephine Morris
pUAA-rac	pcDNA3.1 /zeo	Amp	SV40 p CMV		Zeocin azRS-IRES-GFPCL	-	SV40pA BGHpA	pUC	8.3	Selectable plasmid used to create stably transfected racKO cells, control for pUAACL	Josephine Morris

pUAAGFP	pcDNA3.1 /zeo	Amp	human Tyr leader sequence SV40p CMV		$BstRNA_{CUA}$ Zeocin azRS-IRES-GFPCL	CUA anticodon -	SV40pA BGHpA	pUC	7.9	control for pUAAY23TAG, no rac	Josephine Morris
pUAAGFP37TAG	pcDNA3.1 /zeo	Amp	human Tyr leader sequence SV40p CMV		$BstRNA_{CUA}$ Zeocin azRS-IRES-GFP37TAGCL	CUA anticodon - GFP 37TAG	SV40pA BGHpA	pUC	7.9	37TAG mutant GFP, in one plasmid system. Test whether azRS, BstRNA constructs work	Josephine Morris
pUAAY23TAG	pcDNA3.1 /zeo	Amp	human Tyr leader sequence SV40p CMV		$BstRNA_{CUA}$ Zeocin azRS-IRES-	CUA anticodon - rac	SV40pA BGHpA	pUC	8.4	Y23TAG mutant of final plasmid, for important switch experiments	Josephine Morris

					racGFPCL	Y23TAG					
			SV40p CMV		Zeocin azRS-IRES-racGFPCL	rac I4TAG	SV40pA BGHpA	pUC			

9.2 A2 Primers

Name	Sequence (5' --> 3')	Nucleotide length that anneals	Tm (Finnzymes; °C)	GC %	Annealing site	Use
Miscellaneous						
T7 promoter	TAATACGACTCACTATAGGG	20	52	40	T7 promoter	sequencing plasmids, e.g. pET, pBluescript-IRES
T7 terminator	GCTAGTTATTGCTCAGCGG	19	61	53	T7 terminator	
JYRS-for	atggacgaatttgaaatgataaagag	26	64	31	anneals to aaRS gene, FP	sequencing synthetase genes
JYRS-rev	ttataatctctttctaattggctc	24	57	29	anneals to aaRS gene, RP	
pBADrev seq 58	tctgatttaatctgtatcaggc	58				
T4Lysozyme						
pAAT4L153TAG + ACG FP	5' - /5Phos/GAA ACG CGT GAT TAC CAC GTA GAA AAC CG - 3'	29	75	48	around codon 153	to try to re-introduce the missing ACG (Threonine) at codon 152, using Phusion site-mutagenesis kit (PCR)
pAAT4L153TAG RP	5' - /5Phos/GCG CGG TTC GGG GTC TGG T - 3'	19	76	74		
JLMt4l003-F	GAAACGCGTGATTACCACGT AGAAAACCGCAC	32	78	50	around codon 153	to try to re-introduce the missing ACG (Threonine) at codon 152, using Quikchange site-mutagenesis kit (PCR)
JLMt4l004-R	GTGCGGTTTTCTACGTGGTAA TCACGCGTTTC	32	78	50		
JLMt4l005-F	CTGGGATGCGTATAAAAACC	26	75	54	around codon 153	to try to re-introduce the missing ACG

	TGGGCG					(Threonine) at codon 152, using Phusion site-mutagenesis kit (PCR)
JLMt4l006-R	GTGCCGGTTTTCTACGTGGTAATCACG	27	74	52		
T4L153TAG opt for	CATATGAACATCTTCGAAATGCTGCG	26	71	42	Anneal pAAT4l153TAG plasmid at optimised T4L sequence beginning and histag at 3' end of gene	colony PCR and sequencing
T4L153TAG opt rev	GCGGCCGCTTATTAATGATGATGATG	26	73	46		
JLMt4l007-F	ATGaatatatttgaaatgttacgtatagatg	31	61	23	primers designed to flank T4L gene within the pT4L153TAG plasmid (no histag),	for colony PCR of expression culture cells
JLMt4l008-R	ttatagatttttatacgcgtccc	23	60	35		
JLMpAA001	CAATTGTGAGCGGATAACAAT	21	62	38	anneals 37bp upstream of XmaI RE site of pAA plasmid	sequence inserts in pAA
JLMpAA002	ATGGAGTTCTGAGGTCATTACTG	23	61	44	anneals 36bp downstream of NotI site of pAA RP	
JLMt4l009	atatatggatccATGAATATATTTGAAATGTTACGTATAGAT	28	59		6extra bases-BamH1site-gene	FP to clone t4l into pLM1
JLMt4l010	atatataagcttTCAATGGTGATGGTGATGA		62		6exctra-HimdIIIsite-end of gene (histag)	RP to clone into pLM1
JLMt4l011	ACCATGAATATATTTGAAATGTTACGTATAG	57	28	21	5'-NcoI recognition site, ATG and start of t4l gene (in pT4L) FP	for cloning into pBAD
JLMt4l012	TTATTAATGATGATGATGATGATGTAGATTTTTATACGCGTCCC	59			3'-end of t4l gene (in pT4L), histag sequence, TAA stop codon, RP	

JLMt4l013	TAGATTTTTATACGCGTCCC	59	20	40	same as jlmt4l012 but no histag or stop	
JLMt4l014	taccatggATGAATATATTTGAAATGTTACGTATAG	57			5' 2bp, NcoI site, t4l gene 3'	
JLMt4l015	atgtttaaacTTATTAATGATGATGATGATGTAGATTTTTATACGCGTCCC	59			5' 2bp, PmeI site, histag, end of t4l 3'	
T4L153 TAG-ttt	AGTCATTACAACGTttAGAACTGGCACTT	27	66	41	anneals around codon 153, fp, 5'phos	mutate TAG153 back to WT Tyr
T4L153 TAG-ttt RP	CACCTAGATTTTTATACGCGTCCC	24	65	46	anneals around codon 153, rp	
FliC						
FliC FP CACC	5'-CACCATGGCACAAGTCATTAATAC-3'	24	64	42	ATG--> of FliC	to clone FliC gene (flagellin) from E.coli's genomic DNA using PCR, various DNA polymerases used (Phire most successful)
FliC RP	5'-ACCCTGCAGCAGAGACAGAAC -3'	21	66	57	end of FliC	
JLMFliC003	CTTACTGGAATTACCCTTTCTACGG	25	64	44	anneals , forward primer	sequencing
JLMFliC004	GGTCATCATCACCATCACCAT	21	65	48	anneals 623-599nts of FliC gene, forward primer	
JLMFliC005	AAGCAGCCACTGATACTGGC	20	65	55	anneals 624-643nts of FliC gene, forward primer	
JLMFliC006	ATGGTGATGGTGATGATGACC	21	65	48	anneals histag at end of FliC, reverse primer	
JLMFliC007	CCGTAGAAAGGGTAATTCCAGTAAG	25	65	44	anneals 623-599nts of FliC gene, reverse primer	

JLMFliC008	aaaaaaCCCGGGATGGCACAAGTCATTAATACCAACAG	26	66	38	anneals first part of FliC gene, + XmaI REsite upstream,	for cloning FliC gene (+TAG already) into pAA, FP
JLMFliC009	aaaaaaCCCGGGATGGCACAAGTCATTAATtagAACAGCC	28	68	39	anneals first part of FliC gene, + XmaI REsite upstream + TAG at codon 7(Tyr),	for cloning FliC gene into pAA (colony30), FP
JLMFliC010	tatatatGCGGCCGCttaACCCTGCAGCAGAGACAGAAC	21	66	57	anneals last part of FliC, +TAA stop codon	to remove V5epitope and histag + NotI REsite, RP
JLMFliC011	GAGCTCGCCCTTATTAACCCTGC	23	70	57	anneals last nts of FliC +TAAT to remove V5 epiope and histag using Phusion polymerase RP	to remove V5 epiope and histag using Phusion polymerase RP
	(GAGCTCGCCCTAACCCTGC)		69			
JLMFliC012	AATTCGAAGCTTGAAGGTAAGCCTATC	27	67		anneals next to JLMFlic011, for use with Phusion	
JLMFliC013	TTAACCCTGCAGCAGAGACAG	21	65	52	RP to clone FliC gene from ASKA plasmid, with TAA stop at end,	clone into pET101
JLMFliC014	atatatccatggATGGCACAAGTCATTAATACCAAC	24	63	38	FP; 6 extra bases + NcoI site + FliC gene	to clone FliC into pQE60
JLMFliC015	atatatagatctTTAACCCTGCAGCAGAGACA	20	64	50	RP ; 6 extra bases + BglII site + end of gene	
Alicja's pQE60 seq for	gagcggataacaatttcacac	21	61		MCS of pQE60	sequencing insertions into pAA
Alicja's pQE60 seq rev	aggagtccaagctcagctaat	21	61			

JLMpQEFliC001	AAATCACCGGTGtagATAACG ATGGGA	24	71	46	FliC middle	mutagenic: for use with Phusion. GTG-->TAG mutation in middle
JLMpQEFliC002	TCGCATAGTAATCATTACCAT TATCAGTATAAACAC	36	67	31		RP for ^
p=with 5' phosphates						
JLMFliC016	TTAACCCTGCAGCAGAGACA G	21	65	52	like FliC RP, but + TAA	
pFD313 bla 01	gttaccaatgcttaatcagtgag	23	60	40	anneals to bla gene	for sequencing pFD313 (in KAF95) & pSWANs
pFD313 bla 02	ctcactgattaagcattggtaac					for sequencing pFD313 (in KAF95) reverse complement of 01
pFD313 1 for	ttctcatgtt tgacagctta	20	57	35	as above, from 1 forward	for sequencing pFD313
pFD313 tet rev	atccgttagcgaggtgc	17	62	59	reverse, end of tetr in pBR322	
pFD313 FliC insertion FP	TACGCAGTAACAGTTGCTAAT GATGGTACAGTGACAATGGC GACTGGAGCAACGGCAAATG CAACTGTAACTGATGCAAATA CTACTAAAGCTACAACTATCA CTTCAGGCGGTACACCTGTTC AGATTGATAATACTGCAGGT TCCGCAACTGCCAACCTTGGT GCTGTTAGCTTAGTAAAACTG CAGGATTCCAAG	27	67	41	27 matched, FliC	5' insertion sequence part of FliC missing from pFD313 FP
pFD313 FliC insertion RP						

pFD313 tat-TAG FP	ACGATGGGAAGTAgAGATCTCCCGGGA	26	74	54	middle of FliC	primer for Phusion mutagenesis, to introduce TAT (tyr) to TAG mutation at codon 244 of FliC within pFD313 plasmid, mutagenic FP
pFD313 tat-TAG RP	TATCACCACCGGTGATTTTCGCATAGTA	28	72	43		primer for Phusion mutagenesis, to introduce TAT (tyr) to TAG mutation at codon 244 of FliC within pFD313 plasmid, RP
pFD313 tat-TAG FP2	ATGGGAAGTAgAGATCTCCCGG	21	65	52	middle of FliC	
pFD313 tat-TAG RP2	CGTTATCACCACCGGTGATTT	21	66	48		
pFD313 tat-TAG FP3	GATACCGATACATAgGCGCTTAAAGATAC	29	64	39	middle of FliC	
pFD313 tat-TAG RP3	ATTACCCTTGGAATCCTGCAGT	22	65	46		
pFD313 - FliC FP	tatataTAGATCTGTAACAGTTGCTAATGA	26	58	31	extra bases, BglII RE site bit of FliC missing from pFD313	cloning missing FliC sequence back into pFD313
pFD313 - FliC RP	atatataagcttACCAAGGTTGGCAGTTG	17	60	53	extra bases, Hind lII RE site bit of FliC missing from pFD313	
pFD313 FliC seq FP	ATTGAGCCGACGGGTG	16	64	62		sequence and screen FliC in pFD313
pFD313 FliC seq RP	TAAGCGCAGCGCATCA	16	65	56		sequence and screen FliC in pFD313
pFD313frag16 TAG FP	AATCACCGGTtagGATAACGATGG	21	66	48	around codon 239 of FliC	mutate codon to TAG for amber suppression
pFD313frag16 TAG RP	TTCGCATAGTAATCATTACCATTATCAGTAT	31	65	29		

FliCPhe54TAG FP	GCGATTGCTAACCGTtagACC TCTAACATTA	29	69	41	mutagenic primers to mutate Phe ttc to TAG, 1st Phe or tyr from codon 1	mutate codon to TAG for amber suppression
FliCPhe54TAG RP	CTGACCCGCTGCGTCAT	17	67	65	primers to mutate Phe ttc to TAG, 1st Phe or tyr from codon 1 RP	
eGFP						
eGFP FP	atggtgagcaagggcga	17	66	59	ATG--> GFP gene	for enhanced GFP in pEGFP-C2 and pSWANGFP
eGFP RP	ttacttgtacagctcgtccatgc	23	65	48	end of GFP gene <--TAA	
eGFP 003	ctgctggagttcgtgacc	18	63	61	anneals just upstream of end of eGFP FP	to sequence pEGFP-C2 plasmid after eGFP gene (Rac1)
eGFP 004	gtccagctcgaccaggat	18	64	61	anneals just downstream of beginning of eGFP RP	to sequence peGFP plasmid before eGFP gene
primer name	**sequence (5' --> 3')**	**nucleotide length that anneals**	**Tm (Finnzymes; °C)**	**GC %**	**Annealing site**	**use**
Rac1						
XhoImRac1 FP	atatatctcgagATGCAGGCCAT CAAGTGTGT	20	66	50	anneals to beginning of mrac1, XhoI site 5' forward primer, 6 spare bps 5'	for cloning mRac1 into TPS194
mRac1NotI RP	atatatgcggccgcTTACAACAGG CATTTTCTCTTCC	23	63	39	anneals to end of mrac1, NotI site 5' forward primer, 6 spare bps 5'	for cloning mRac1 into TPS194
mRac1NotI RP2	atatatgcggccgcTTACAACAGC AGGCATTTTCTCTTC	25	67	40	anneals to end of mrac1, NotI site 5' forward primer, 6 spare bps 5'	for cloning mRac1 into TPS194, missing Leu
mRac1 Y23TAG FP	ctcatcagttaGacgaccaatgcatt	25	67	40	anneals around Y23 of rac1	mutagenic primers mRac1 Tyr23-->TAG for use with Phusion, 5' phos
mRac1 Y23TAG RP	caggcaggttttaccaacagctc	23	68	52	anneals around Y23 of rac1	reverse primer for mRac1 Tyr23-->TAG for use with Phusion, 5' phos
CMV Forward	CGCAAATGGGCGGTAG	16	64	62	Human CMV immediate early promoter	sequence DNA in mammalian plasmids, and after cloning

pCMVTnT seq FP	AGGTGTCCACTCCCAGTTC	19	62	58	anneals just upstream of XhoI site in TPS194 (pCMVTnT) of MCS	for sequencing cloning of mRAC1 into pCMVTnT
pCMVTnT seq RP	GCAATAGCATCACAAATTTCAC	22	61	36	anneals just downstream of NotI site in TPS194 (pCMVTnT) of MCS	
Not1eGFP FP	atatatGCGGCCGCatggtgagcaagggc	15	59	60	extra bases, not1 site, beginning of rac1 gene	clone eGFP into pCMVTNTRac1 c-terminal to rac1 FP
eGFPNot1 RP	atatatGCGGCCGCttacttgtacagctcgtccat	21	59	43	extra bases, not1 site, end of rac1 gene	clone eGFP into pCMVTNTRac1 c-terminal to rac1 RP
rac1eGFP FP	CCTGCTGTTGggcatggtgagca	20	68	55	end of rac1 gene, start of GFP	mutate TAA to GGC (Gly), start of eGFP
rac1eGFP RP	CATTTTCTCTTCCTCTTCTTGACAGGA	27	68	41		RP for mutant FP primer to make Rac1-eGFP fusion protein
	stupid - forgotten NotI site					
Rac1GFP 001	caacagcaggcattttctcttcct	24	69	46	end of rac1	delete taaggc at end of rac1 eGFP cloning,
Rac1GFP 002	CGCatggtgagcaagggc	18	71	67	end of rac1	delete taaggc at end of rac1 eGFP cloning,
Rac1GFP 003	tctgtccccctcctgtcaa	19	66	58	anneals upstream of beginning of rac1 gene, in pCMVTnT, FP	use with eGFP RP to screen pCMVTNTRac1 eGFP cloning, screen for directionality,
racGFP 005	atatatACCGGTatgcaggccatcaagtg	17	62	53	6 extra bases, AgeI site, beginning of rac1 gene FP	for cloning Rac1 into peGFP 5' to eGFP gene
racGFP 004	atatatACCGGTcaacagcaggcatttctc	19	62	47	6 extra bases, AgeI site, end of rac1 gene -TAA. RP	
racGFP 006	atatatACCGGTttttctcttcctcttcttgacag	23	61	39	6 extra bases, AgeI site, end of rac1 gene -CLLL (CAAX) TAG RP	
racGFP 007	atatatACCGGTgacaggaggggacaga	17	62	65	6 extra bases, AgeI site, end of rac1 gene -KKRKRKCLLL (polybasic +	

					CAAX) TAG RP	
racGFP 012	GGCGGGTCCGGCGGCTCGatggtgagcaagggcga	17	66	59	anneals to seq between rac & GFP	add GGSGGS linker sequence between rac1 and GFP in racGFP anneals to beginning of eGFP
racGFP 013	gacaggaggggacagaga	19	65	63	anneals to seq between rac & GFP	RP for racGFP012, racGFP3.1 (-KKRKRKCLLL)
racGFP 014	ttttctcttcctcttcttgacagg	24	65	42	anneals to seq between rac & GFP	RP for racGFP012, racGFP PCR2 (2.5) (-CLLL)
racGFP 015	ttacaacagcaggcacttgtacagctcgtccatgc	20	64	55	anneals to end of eGFP, plus CLLL (from rac1)	mutagenic, RP
racGFP 016	TCCGGCCGGACTCAGA	14	65	71	anneals just downstream of GFP,	for use with racGFP015 and 017, FP
racGFP 017	ttacaacagcaggcattttctcttcctcttcttcttgtacagctcgtccatgc	20	64	55	anneals to end of eGFP, plus polybasic, CLLL (from rac1)	mutagenic, RP
pSWANRS seq	CGTGAAGATCAGGGGATT	18	61	50	anneals mid-EcRS gene to sequence	for sequencing cloning into pBluescript for insertion of IRES
racGFP IRES001	atatatCTCGAGatgcaggccatcaagtg	17	62	53	6bp, XhoI (CTCGAG), start of rac1	for cloning rac1-GFP into pBlue-IRES
racGFP IRES002	ttaattGGTACCttacaacagcaggcacttgtac	22	62	45	6bp, KpnI site (GGTACC), end of GFP, CLL seq	
racGFP IRES003	ttaattGGTACCttacaacagcaggcattttctc	22	63	41	6bp, KpnI site (GGTACC), end of GFP, RKCL seq	
EcTyrRS FP	ATGGCAAGCAGTAACTTGATT	20	59	40	anneals to ATG+ of e coli tyr-RS,	for sequencing and PCR on cDNA from mRNA
EcTyrRS RP	TTACTTGTCGTCATCGTCTTT	21	59	38	anneals to end of EcTyrRS in pUAA for Rt-PCR screen for azRS in pUAA tx't cells	
EcTyrRS RP2	CTTTCCAGCAAATCAGACAGT	21	61	43	anneals to end of EcTyrRS in pUAA for Rt-PCR screen for azRS in pUAA tx't cells	

rac FP	atgcaggccatcaagtg	17	62	53	anneals to ATG+ of mRac1 gene	for use in PCR after RT-PCR of RNA extraction from transfected racKO cells, use with rac FP
rac RP	CTGTCCCCCTCCTGTC	16	60	69	anneals to end of rac gene upstream of polybasic and CLLL,	
racGFP 018	atatattctagaATGCAGGCCATCAAGTGCGT	17	62	53	6 bp, XbaI site, start of rac gene,	for cloning into pSWANRS (pcDNA3.1hygro)
racGFP 019	ttaattgggcccttacaacagcaggcacttgta	21	62	43	6 bp, ApaI site, end of GFPCL gene,	
racGFP 020	ttaattGGGCCCttacaacagcaggcattttctc	22	64	41	6 bp, ApaI site, end of GFPRKCL gene,	
T3 promoter	aattaaccctcactaaaggg	20	57	40		sequence cloning in pBluescript-IRES
IRES 19FP	CCCCTAACGTTACTGGC	17	59	59	anneals 19-35 of IRES	for sequencing
racGFP IRES004	atatatGAATTCatgcaggccatcaagtg	17	62	53	6bp, EcoRI site (GAATTC), start of rac1	for cloning rac1-GFP into pBlue-IRES
racGFP IRES005	ttaattCTCGAGttacaacagcaggcacttgtac	22	62	45	6bp, XhoI site (CTCGAG), end of GFP, CLL seq	
racGFP IRES006	ttaattCTCGAGttacaacagcaggcattttctc	22	63	41	6bp, XhoI site (CTCGAG), end of GFP, RKCL seq	
AzRS IRES001	aattaagcggccgcATGGCAAGCAGTAACTTGA	19	59	42	6bp, NotI site (GCGGCCGC), start of AzRS	for cloning AzRS into pblueIRES-racGFP and into pSWAN
AzRS IRES002	atatattctagaTTACTTGTCGTCATCGTCTT	20	58	40	6bp, XbaI site (TCTAGA), end of AzRS	for cloning AzRS into pblueIRES-racGFP
racGFP IRES007	atatatACCGGTttacaacagcaggcacttgtac	22	62	45	6bp, AgeI site (ACCGGT), end of racGFPCL	for cloning AzRS-IRES-racGFPCL into pSWANRS, for use with AzRS IRES001

racGFP IRES008	atatatACCGGTttacaacagcaggcattttctc	22	63	41	6bp, AgeI site (ACCGGT), end of racGFPRKCL	for cloning AzRS-IRES-racGFPRKCL into pSWANRS
CMV RP	CTACCGCCCATTTGCG	16	64	62	same as FP, but RP	to sequence BstRNA upstream of CMV promoter
pUAA001	ATGGTGAGCAAGGGCGA	17	66	59	5'-P, anneals to eGFP ATG-->	for removing rac from pUAAs
pUAA002	GGCCACAACCATGGAATTC	19	66	53	5'-P, anneals to end of IRES	for removing rac from pUAAs (wrong way I think)
pUAA002-2	GAATTCCATGGTTGTGGCC	19	66	53	5'-P, anneals to end of IRES	for removing rac from pUAAs
pUAA005	ACCATGGAAGTCATGCAGGCC	20	70	55	anneals to seq around ATG of racGFP in pUAACL4	to improve kozak consensus, FP, 5'-P
pUAA006	TGTGGCCATATTATCATCGTGTTTTTC	27	69	37		to improve kozak consensus, RP 5'-P
pUAA007	ACCATGGAAacCATGCAGGCC	20	70	55		to improve kozak consensus, FP, 5'-P (2)
pUAA008	AGTGGCGGCacCATGGCAAGCAGT	22	78	64	anneals to seq around ATG of azRS in pUAACL4,	to improve kozak consensus, FP
pUAA009	GTGCTGGATATCTGCAGAATTCCACC	26	71	46		to improve kozak consensus, RP
pUAA010	TTGCGGCATTTTGCCTTC	18	68	50	anneals to pUAACL (PcDNAz3zeo) after where CMVRP sequences upto,	sequence pUAA plasmids
pUAAA3TAGFP	GAATTCATGCAGtagATCAAGTGCGTG	24	70	46	anneals to end of IRES-rac in pUAACL4	to mutate A3-->TAG
pUAAA3TAGRP	CATGGTTGTGGCCATATTATC	25	70	44		reverse primer for ^

	ATCG					
pUAAI4TAGFP	ATGCAGGCCtagAAGTGCGTG	18	69	61	anneals to end of IRES-rac in pUAACL4	to mutate I4-->TAG
pUAAI4TAGRP	GAATTCCATGGTTGTGGCCATATTA	25	68	40		reverse primer for ^
zeoFP	ATGGCCAAGTTGACCAGTG	19	64	53	anneals to ATG--> of zeocin gene (pcDNA and pUAA)	PCR of cDNA from RNA extraction, and sequencing
zeoRP	TCAGTCCTGCTCCTCGG	17	64	65	anneals to end of Zeocin gene (pcDNA and pUAA)	PCR of cDNA from RNA extraction, and sequencing
pUAA011	CACAACCAgGGAAacCATGCAGGCC	21	70		anneals to kozak consensus and start of rac	mutate consensus of rac (pUAACL4) , and remove previous ATG, need -P (PNK)
pUAA012	GCCATATTATCATCGTGTTTTTCAAAGG	28	68	36	anneals to kozak consensus and start of rac	RP for ^ need -P (PNK)
pUAA013	CACAACCATGGAAacCATGCAGGC	22	72	55	anneals to kozak consensus and start of rac	mutate consensus of rac (pUAACL4) , and remove previous ATG, use with 012 need -P (PNK)
pUAA014	AACCATGGAAacCATGCAGGCCA	21	72	52	anneals to kozak consensus and start of rac	mutate consensus of rac (pUAACL4) , and remove previous ATG, use with 015 need -P (PNK)
pUAA015	GTGGCCATATTATCATCGTGTTTTTC	26	67	39	anneals to kozak consensus and start of rac	RP for ^ need -P (PNK)
β-actin forward	CCACACTGTGCCCATCTACGAGGGGT	26	76.8	62	564 – 589	RT-PCR of β-actin
β-actin reverse	AGGGCAGTGATCTCCTTCTGCATCCT	26	73.6	54	1043 - 1018	RT-PCR of β-actin

9.3 A3 Sequences of NAA-specific synthetases for *E. coli* culture

The engineered aaRSs were provided by Schultz group (Wang et al. 2001) or A. Antonczak (Antonczak 2012 Thesis). The possible sites for mutagenesis were Y32, L65, H70, E107, F108, Q109, D158, I159, L162, A167 and D286. The sequences of RS genes used in the study are compared to that of WT MjTyrRS and are indicated in Figure 124.

```
                    0        10        20        30        40        50        60
                    1234567890123456789012345678901234567890123456789012345678890
MjTyrRS             MDEFEMIKRNTSEIISEEELREVLKKDEKSAYIGFEPSGKIHLGHYLQIKKMIDLQNAGF
iodoRS              MDEFEMIKRNTSEIISEEELREVLKKDEKSALIGFEPSGKIHLGHYLQIKKMIDLQNAGF
azRS                MDEFEMIKRNTSEIISEEELREVLKKDEKSASIGFEPSGKIHLGHYLQIKKMIDLQNAGF
cyanoRS             MDEFEMIKRNTSEIISEEELREVLKKDEKSALIGFEPSGKIHLGHYLQIKKMIDLQNAGF
                    ****************************** ******************************

                            70        80        90       100       110       120
                    1234567890123456789012345678901234567890123456789012345678890
MjTyrRS             DIIILLADLHAYLNQKGELDEIRKIGDYNKKVFEAMGLKAKYVYGSEFQLDKDYTLNVYR
iodoRS              DIIILLADLHAYLNQKGELDEIRKIGDYNKKVFEAMGLKAKYVYGSSFQLDKDYTLNVYR
azRS                DIIILLADLHAYLNQKGELDEIRKIGDYNKKVFEAMGLKAKYVYGSSFQLDKDYTLNVYR
cyanoRS             DIIIVLADLHAYLNQKGELDEIRKIGDYNKKVFEAMGLKAKYVYGSEWMLDKDYTLNVYR
                    ****:***************************************.: ************

                            130       140       150       160       170       180
                    1234567890123456789012345678901234567890123456789012345678890
MjTyrRS             LALKTTLKRARRSMELIAREDENPKVAEVIYPIMQVNDIHYLGVDVAVGGMEQRKIHMLA
iodoRS              LALKTTLKRARRSMELIAREDENPKVAEVIYPIMQVNPLHYEGVDVAVGGMEQRKIHMLA
azRS                LALKTTLKRARRSMELIAREDENPKVAEVIYPIMQVNPLHYEGVDVAVGGMEQRKIHMLA
cyanoRS             LALKTTLKRARRSMELIAREDENPKVAEVIYPIMQVNGAHYLGVDVAVGGMEQRKIHMLA
                    ************************************  ** ******************

                            190       200       210       220       230       240
                    1234567890123456789012345678901234567890123456789012345678890
MjTyrRS             RELLPKKVVCIHNPVLTGLDGEGKMSSSKGNFIAVDDSPEEIRAKIKKAYCPAGVVEGNP
iodoRS              RELLPKKVVCIHNPVLTGLDGEGKMSSSKGNFIAVDDSPEEIRAKIKKAYCPAGVVEGNP
azRS                RELLPKKVVCIHNPVLTGLDGEGKMSSSKGNFIAVDDSPEEIRAKIKKAYCPAGVVEGNP
cyanoRS             RELLPKKVVCIHNPVLTGLDGEGKMSSSKGNFIAVDDSPEEIRAKIKKAYCPAGVVEGNP
                    ************************************************************

                            250       260       270       280       290       300
                    1234567890123456789012345678901234567890123456789012345678890
MjTyrRS             IMEIAKYFLEYPLTIKRPEKFGGDLTVNSYEELESLFKNKELHPMDLKNAVAEELIKILE
iodoRS              IMEIAKYFLEYPLTIKRPEKFGGDLTVNSYEELESLFKNKELHPMDLKNAVAEELIKILE
azRS                IMEIAKYFLEYPLTIKRPEKFGGDLTVNSYEELESLFKNKELHPMDLKNAVAEELIKILE
cyanoRS             IMEIAKYFLEYPLTIKRPEKFGGDLTVNSYEELESLFKNKELHPMDLKNAVAEELIKILE
                    ************************************************************

                       306
                    12345678
MjTyrRS             PIRKRL
iodoRS              PIRKRL
azRS                PIRKRL
cyanoRS             PIRKRL
                    ******
```

Figure 125. Comparison of protein sequences of *Mj*TyrRS synthetases engineered for specific NAAs

9.4 A4 Mass spectrometry spectra

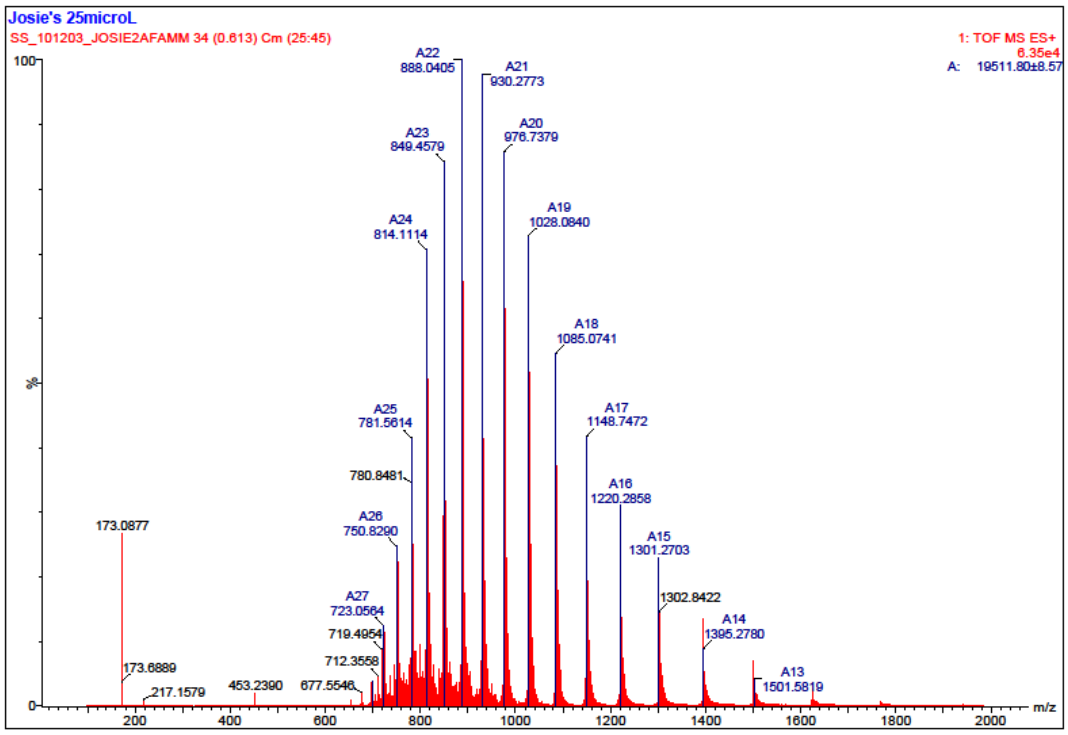


Figure 126 Mass spectrum of WT T4L. MS was obtained using ESI.

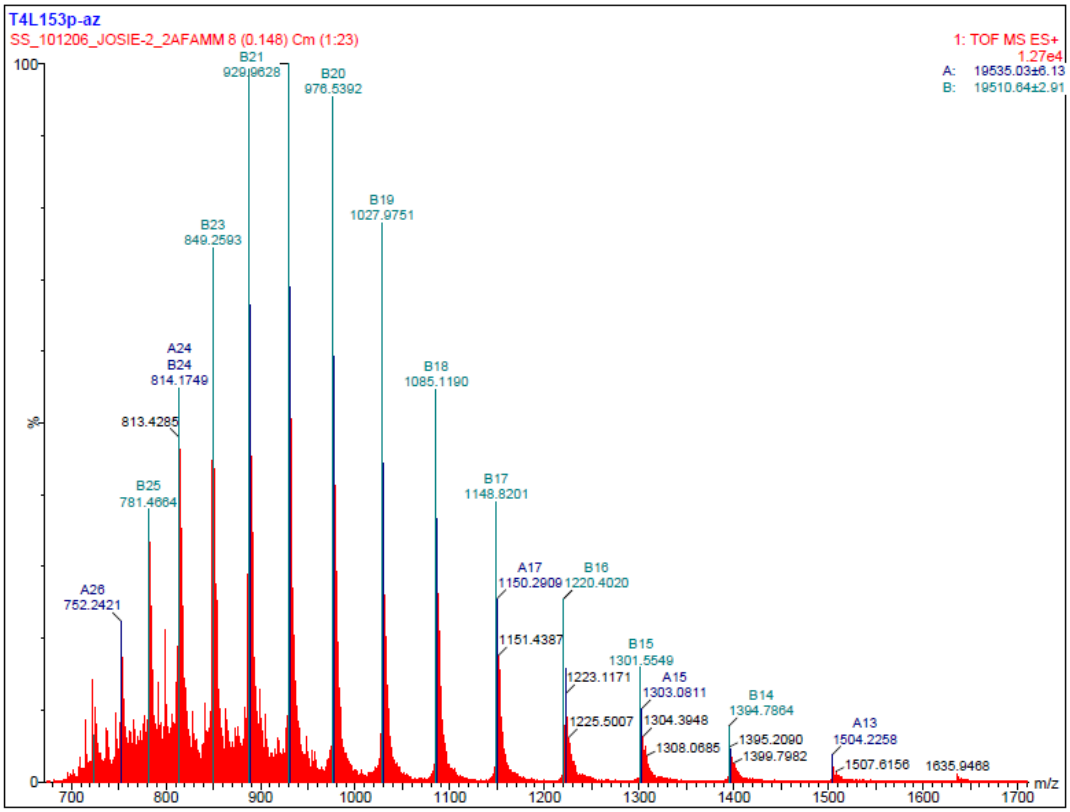


Figure 127. Mass spectrum of T4L153azF. MS was obtained using ESI.

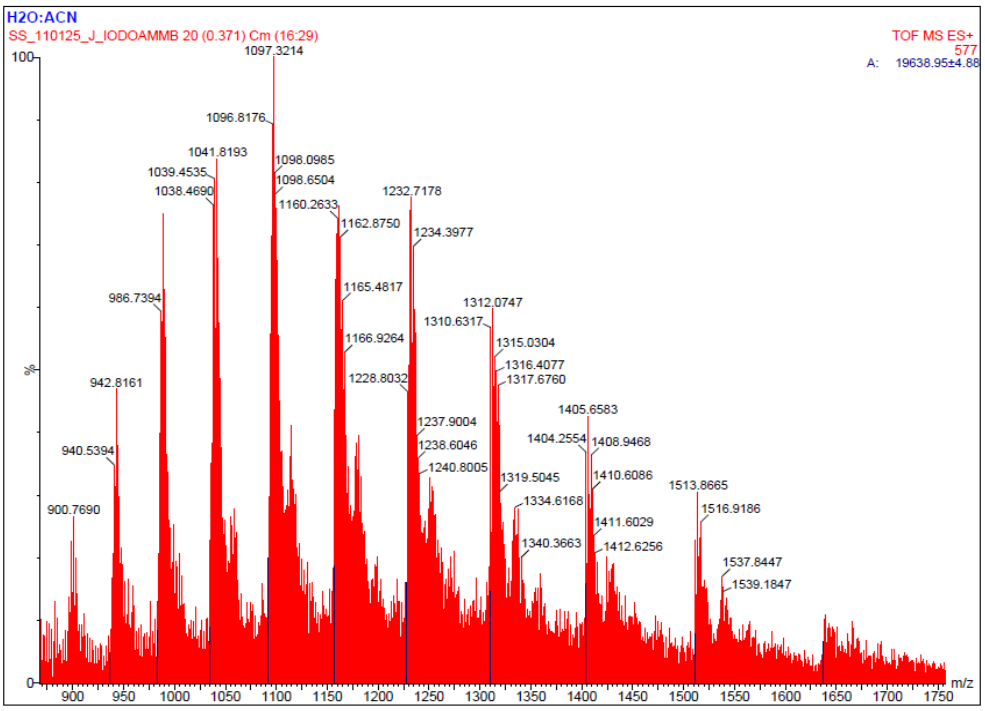


Figure 128. Mass spectrum of T4L153iodoF. MS was obtained using ESI.

9.5 A5 Isolation of KAF95

KAF95 was produced from mutagenesis screens with ethyl methanesulfonate (EMS), which typically results in guanine alkylation, forming O-6-methylguanine. During DNA replication, DNA polymerases frequently place thymine, instead of cytosine, opposite O-6-methylguanine, which, in following subsequent rounds of replication, turns the original G:C pair into an A:T pair.

The *E*. *coli* strain KAF95 was supplied containing the ampicillin resistant plasmid, pFD313. To lose this plasmid, the KAF95 culture was grown sequentially in the absence of antibiotics. Two rounds of 8 – 14 hour liquid culture incubations were performed, a sample from the liquid culture was then spread onto an agar plate. The following day, 10 colonies were picked, resuspended in 300 µl LB and 1 µl samples pipetted onto 5 x agar plates each containing an antibiotic; carbenicillin, tetracycline, kanamycin, chloramphenicol or no antibiotic. This was repeated until a colony was isolated that was susceptible to the four antibiotics.

9.6 A6 Restoration of motility to KAF95 utilising plasmid pFD313

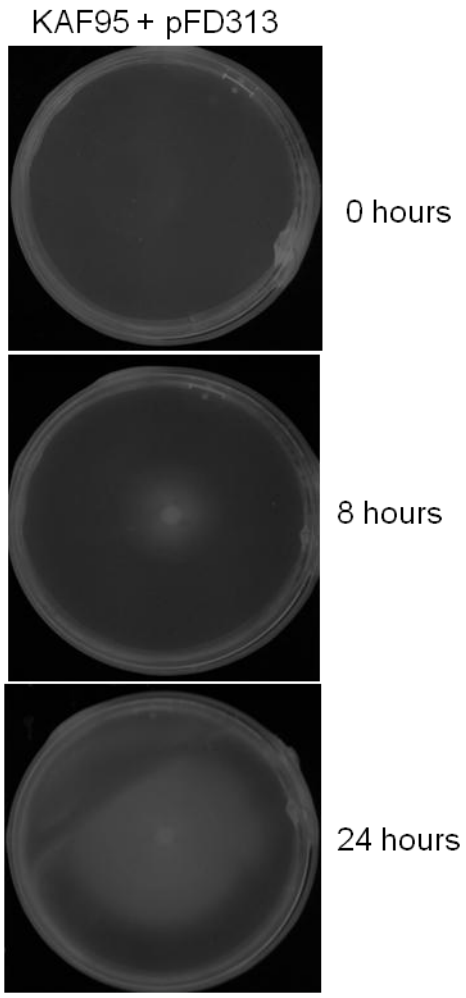


Figure 129. Restoration of motility of KAF95 using plasmid pFD313

9.7 A7 Translated protein sequences of FliC-containing constructs

Figure 130. Comparison of protein sequence of WT Flagellin and that encoded by plasmid pET101FliC. Plasmid DNA was sequenced, the sequence was translated to protein using ExPASy online tool and compared using ClustalW2 tool. Difference from WT FliC are highlighted.

9.8 A8 PCR amplification of β-actin gene from cDNA synthesised from extracted RNA

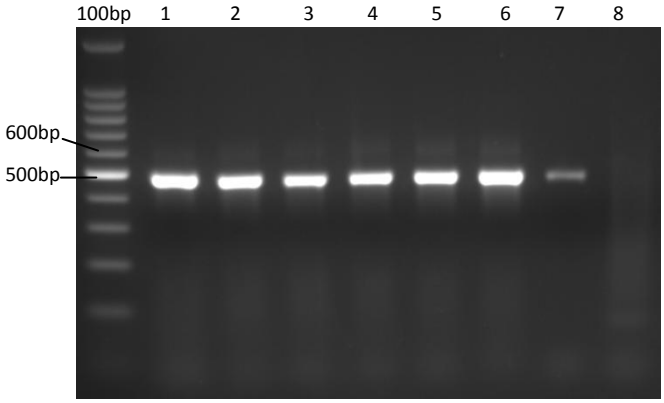


Figure 131. Agarose gel (2%) of PCR to amplify β-actin gene from cDNA synthesised from RNA extracted from control cells (lane 1), racKO cells (lane 2), and racKO cells transfected with peGFP (lane 3), pGFPRac1 (lane 4), pRacGFPCL (lane 5), pRacGFPRK (lane 6), previously isolated cDNA (lane 7) a d DEPC water (lane 8).

9.9 A9 Creation of plasmid pCMVTnTRac1

Plasmid TPS194_FluCYam had previously been created using the commercially available plasmid pCMVTnT™ (Promega) and had been kindly donated by Dr Thomas Sakmar, Rockefeller University (Ye et al. 2009). It encoded a luciferase gene under the control of constitutive CMV promoter, with a 5′ β-globin leader sequence and termination by the late SV40 polyadenylation signal.

The luciferase gene was removed from TPS194 by digesting the plasmid with *Xho*I and *Not*I-HF (NEBuffer 4 + BSA, 37 °C). DNA corresponding to linearised pCMVTnT minus the luciferase gene was excised from a preparative agarose gel and the DNA purified. Rac1 was amplified from plasmid pGFPRac1 using primers that added an *Xho*I and *Not*I site to the 5' and 3' end of the PCR product, respectively. The PCR product and complementarily digested PCR products were ligated and transformed into Top10 *E. coli* and plated onto agar containing carbenicillin. Colony PCR was used to screen colonies for insertion of the Rac1 gene. Positive colonies were confirmed by DNA sequencing. This created plasmid pCMVTnTRac1.

9.10 A10 Sample X-ray diffraction pattern of T4L WT

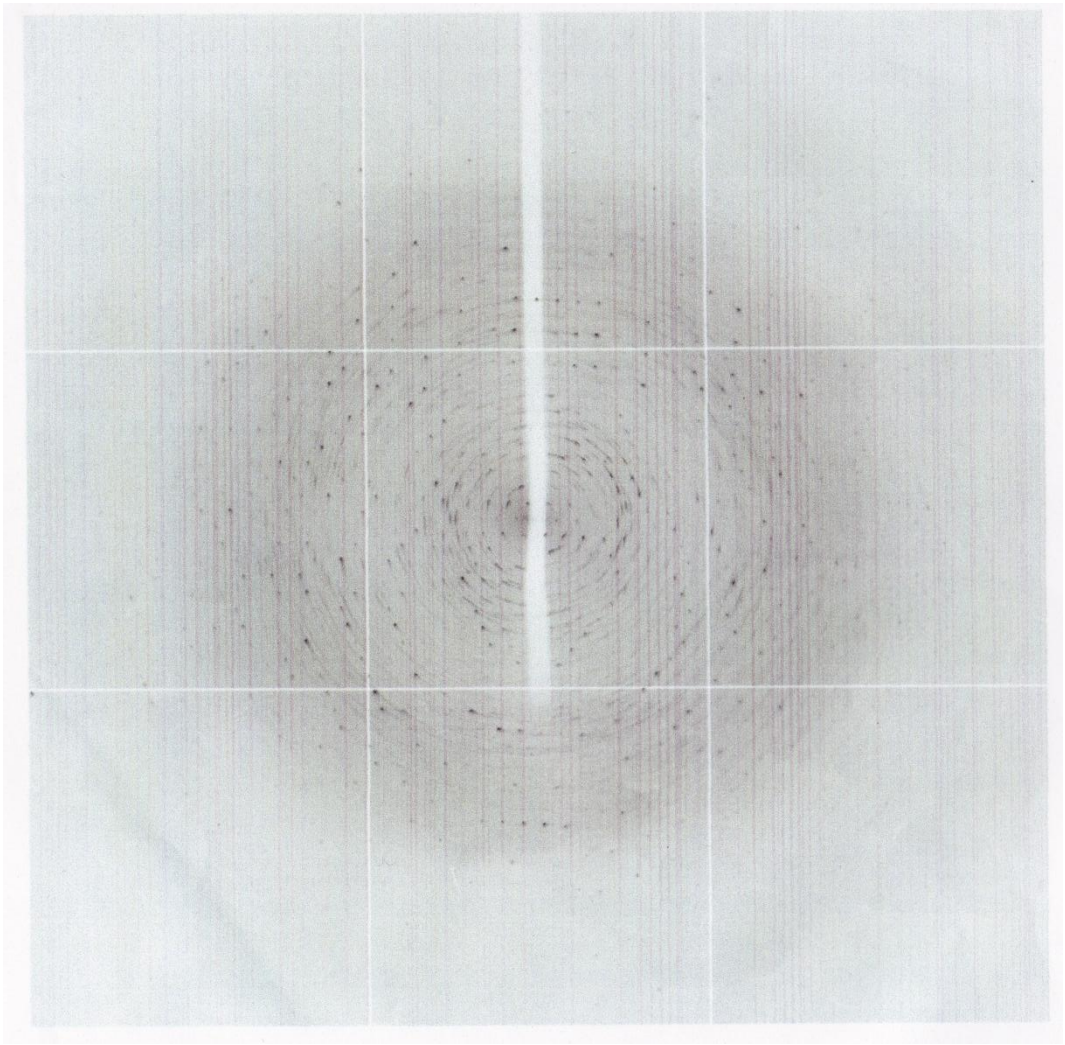


Figure 132. Sample x-ray diffraction of T4L WT. Pattern was collected at Diamond light source, Didcot, by Dr P. Rizkallah.