

Investigating Application of mRNA Elongation Control Models to the Design of Specific Target Region Sequence Variants and the Impact on Protein Synthesis

2023

Louis Darton

A thesis submitted to the University of Kent for the degree
of Doctor of Philosophy

University of Kent

School of Biosciences

Declaration

No part of this Thesis has been submitted in support of an application for any degree or other qualification of the University of Kent or any other University or Institute of learning.

Louis Darton

September 2023

Abstract

During disease, or under cell stress, post-transcriptional programmes are initiated to orchestrate an appropriate cellular response through the regulation and control of protein synthesis. One such avenue for regulation is during translation. Traditionally, much of translational control was thought to occur during the initiation stage, but recent studies suggest the elongation stage is also a central regulatory node. It has been shown that message specific, elongation control is central to disease response mechanisms in both tumorigenesis and neurogenerative disorders. However, it is unclear to what extent factors such as codon usage and the relative abundance of their cognate tRNA impact this regulation. Transcripts can experience ribosomal pausing (stalling) during elongation which may be attributed to heightened dependence on eEF2 translocation, but is also due to tRNA abundance. The university of Kent has developed a computational elongation programme which can model these interactions along a transcripts sequence and can identify regions decoding at a slow rate resulting in ribosomal stalling and potentially a mark of elongation control. Here, we selectively modified these areas for 3 mRNA transcripts by substituting canonical codons with synonymous codons with different decoding rates in order to generate multiple transcripts with differing decoding times. We found that optimising these areas, by using synonymous codons that correspond to abundant cognate tRNA, had a major impact on unoptimised transcripts at the 5' region immediately following the start codon and can be used to tune protein production. Conversely, diosoptimising these regions lead to decreases in protein production compared to native sequences, and in some cases completely prevented translation. slow regions identified further downstream of the start codon were also shown to impact protein production when modified but less reliably. Other codon alterations that weren't fully optimal nor disoptimal lead to a variety of impacts on protein production showcasing the complicated nature of elongation control. Differences were also observed for secreted proteins which may be affected by alteration made at the 5' end of their mRNA transcript more so than non-secretory proteins. Mildly hypothermic conditions also heightened differences in protein production for many sequences compare to protein production of their native sequences. These results assist in confirming the elongation model's accuracy in predicting slow regions involved in elongational control and modifications achieved in this manner could be broadly applied to biotherapeutic production as well as incorporated into medical treatments such as gene therapy or mRNA vaccines.

Acknowledgments

I am extremely grateful to everyone who has lent their support and kept me sane during my PhD experience with an extra special thank you for Liam, Ava, Jack Hat, Jack Nohat, Laura and Phoebe for all the emotional support I could have asked for.

I was extremely lucky to be part of an incredible lab group who somehow made the experience enjoyable and I would like to thank James Budge, Tanya Knight and Emma Mead for sharing so much of their knowledge and expertise with me. JD, who shall remain unnamed, I am also grateful for your small but essential input. An extra big thank you goes to Mark Smales whose patience I certainly tested, but luckily for me, turned out to be boundless. Thank you so much for everything you have done to make this thesis possible.

Finally, I would like to thank my family and in particular my mum and dad, Jenny and Martin, for always supporting me and for helping when I needed it most. Alysia Burrows also gets a special mention for giving me the motivation and encouragement to see this thesis through to completion.

Table of Contents

Abstract	3
Acknowledgments	4
Table of Contents	5
Abbreviations	11
Chapter 1 Introduction	
1.1. mRNA Translation in Eukaryotes; An Overview	14
1.1.1. The Initiation Phase of mRNA Translation	16
1.1.2. The Elongation Phase of mRNA Translation	20
1.1.2.1 <i>tRNA Selection</i>	20
1.1.2.2 <i>Amine and Carboxylic Acid Reaction; Peptide Bond Synthesis</i>	21
1.1.2.3 <i>Translocation</i>	22
1.1.3 Termination Phase	23
1.1.4.1 <i>Recycling</i>	26
1.1.4.2 <i>Rescue</i>	26
1.2. <i>mRNA stability</i>	28
1.3. <i>Impacts on Elongation Rate and its Control on Protein Production and its Regulation</i>	29
1.3.1 <i>Codon Usage</i>	30
1.3.2 <i>Inefficient Peptide Formation</i>	31
1.3.3 <i>mRNA Secondary Structures</i>	33
1.4. <i>Transcript Specific Evasion of Globally Reduced Translation Rates</i>	34

<i>1.5. The Protein Synthesis of Secretory Pathway Proteins</i>	34
<i>1.6. Production of Recombinant Proteins in Mammalian Cell Expression Hosts</i>	35
<i>1.6.1. Cloning Methods</i>	37
<i>1.6.1.1 Gibson Assembly Cloning</i>	39
<i>1.6.1.2 Site Directed Mutagenesis</i>	40
<i>1.7. An Elongation Model of Decoding Speeds and Protein Synthesis Rates of mRNA Transcripts</i>	41
<i>1.8. Model Transcripts Investigated in this Study</i>	42
<i>1.8.1. Cyclin D3 and its Role in Cell Cycle Control</i>	43
<i>1.8.2. Noggin</i>	44
<i>1.8.3. Green Fluorescent Protein</i>	45
<i>1.9. Emerging Applications for RNA Based Therapies</i>	45
<i>1.10. Aims of the Project</i>	46
<i>Chapter 2 Materials and Methods</i>	
<i>2.1 Molecular Biology Methods</i>	47
<i>2.1.1 E. coli Culture</i>	47
<i>2.1.2 Generation of Calcium Competent E. coli Cells</i>	48
<i>2.1.3 RNA extraction from mammalian cells</i>	48
<i>2.1.4 Generation of cDNA from RNA</i>	48
<i>2.1.5 Polymerase Chain Reaction (PCR)</i>	49
<i>Table 2.2: Primers used throughout this project, including their sequence and their relevance in application.</i>	50
<i>2.1.6 Transcript Sequence Design</i>	50
<i>2.1.7 PCR and Gel Clean-up of PCR Products</i>	51

<i>2.1.8 Restriction Enzyme Digests</i>	51
<i>2.1.9 Agarose Gel Electrophoresis Analysis of DNA</i>	51
<i>2.1.10 DNA Ligation</i>	51
<i>2.1.11 Gibson Assembly</i>	52
<i>2.1.12 Site Directed Mutagenesis</i>	52
<i>2.1.13 Transformation of E. coli Cells</i>	52
<i>2.1.14 Mini/Maxi-Prep Isolation of Plasmid DNA</i>	53
<i>2.1.15 DNA Quantification</i>	53
<i>2.1.16 Sequence Alignment Tools and Frame Checker</i>	53
<i>2.1.17 Signal Sequence Checker</i>	53
<i>2.2 Mammalian Cell Culture</i>	53
<i>2.2.1 Adherent Cell Culture Maintenance</i>	53
<i>2.2.2 Cryopreservation and Revival of HEK293 Cells</i>	54
<i>2.2.3 Cell Counting</i>	54
<i>2.2.4 Calculating Cell Index on the xCELLigence System</i>	54
<i>2.2.5 Transfection of DNA in Mammalian Cells by Lipofectamine</i>	55
<i>2.2.6 Proteasome Inhibition Experiments</i>	56
<i>2.3 Protein Analysis</i>	56
<i>2.3.1. Protein Sample Collection and Generation</i>	56
<i>2.3.1.1. Intracellular Sample Collection</i>	56
<i>2.3.1.2. Extracellular Sample Collection</i>	56

<i>2.3.2. Quantification of Protein Lysates using the Bradford Assay</i>	<i>57</i>
<i>2.3.3. Protein Analysis by Sodium Dodecyl Polyacrylamide Gel Electrophoresis (SDS-PAGE)</i>	<i>57</i>
<i>2.3.4. Western Blot Analysis</i>	<i>57</i>
<i>2.3.5. Densitometry Analysis</i>	<i>58</i>
<i>2.3.6. Immunofluorescence Imaging</i>	<i>58</i>
<i>2.3.6.1. Poly-L-Lysine Cell Adhering</i>	<i>58</i>
<i>2.3.6.2. Cell Fixing and Permeabilisation</i>	<i>58</i>
<i>2.3.6.3. Antibody Exposure</i>	<i>59</i>
<i>2.3.6.4. Confocal Microscope Imaging</i>	<i>59</i>
<i>2.3.7. Antibodies Used within this Study</i>	<i>60</i>
<i>Chapter 3 Design, development and generation of model sequences and constructs with different predicted mRNA decoding speeds for evaluation in mammalian cells</i>	
<i>3.1 Introduction</i>	<i>61</i>
<i>3.2 Design and Generation of Cyclin D3 (CCND3) Transcript Sequences with Alternative Codon Usage.</i>	<i>63</i>
<i>3.2.1 Sequence Design</i>	<i>63</i>
<i>3.2.2 CCND3 Cloning Strategy</i>	<i>64</i>
<i>3.2.3 Insertion of Tagged and Untagged Native CCND3 into the pET23b(+) Plasmid</i>	<i>65</i>
<i>3.2.4 Insertion of Codon Modified Regions of Interest into the Native CCND3 Sequence</i>	<i>66</i>
<i>3.2.5 Re-insertion of CCND3 Sequences with ROI Codon Modifications into the pcDNA3.1 Vector</i>	<i>67</i>
<i>3.2.6 Assessing Expression of CCND3 Constructs in HEK293 Cells</i>	<i>68</i>
<i>3.2.6.1 Western Blot Analysis of Transiently Transfected HEK293 cells for CCND3 Expression</i>	<i>68</i>
<i>3.2.6.2 Evaluating Poor Transient Exogenous CCND3 Expression</i>	<i>70</i>

<i>3.2.7 Site Directed Mutagenesis to Remove the CT Insertion in Modified CCND3 ROI and Correct the ORF</i>	77
<i>3.2.7.1 Primer design for site directed mutagenesis experiments</i>	78
<i>3.2.7.2 Site directed Mutagenesis Results for 775G>T and 507_508delCT and In-Frame Confirmation</i>	79
<i>3.2.8 Preparation and Restriction Digest Confirmation of the Final CCND3-pcDNA3.1 Constructs</i>	82
<i>3.3 Generation of Open Reading Frame Codon Variant Noggin Sequences</i>	83
<i>3.3.1 Cloning strategy for generating codon transcript variants of Noggin</i>	83
<i>3.3.2 Cloning of Alternative Noggin ROIs into the Noggin Sequence</i>	86
<i>3.3.3 Insertion of the Noggin Native Sequence into the pcDNA3.1 Vector</i>	87
<i>3.4 Generation of eGFP Alternative Codon Transcript Sequences</i>	88
<i>3.4.1 Design of Alternative eGFP Sequences and Cloning Strategy</i>	88
<i>3.4.2 Assessment of the Generation of eGFP Codon Variant Transcript Sequences</i>	90
<i>3.5. Alterations to mRNA Predicted Secondary Structures Caused by Codon Alterations</i>	91
Chapter 4 Determining the Impact of Codon Manipulation in Target Areas of the Open Reading Frame of Cyclin on Subsequent Recombinant, Exogenous CCND3 Protein Production in HEK293 Cells	
<i>4.1 Introduction</i>	94
<i>4.2 Western Blot Analysis of Temperature Dependent Production of Cold Inducible RNA-Binding Protein (CIRP), Eukaryotic Elongation Factor 2 (eEF2), and Phosphorylation of eEF2 (eEF2-P)</i>	97
<i>4.3 Preliminary Confirmation of CCND3 Protein Production from all Codon Designed Transcripts Through Western Blot Analysis</i>	98
<i>4.3.1 Qualitative Western Blot Analysis for CCND3 Protein Production from the Different Codon Variant Transcripts</i>	99
<i>4.3.2 Assessing Time Dependent Transient Expression Profiles of CCND3 at 32°C and 37°C for Gv constructs by Western Blot</i>	101
<i>4.4 Investigation into the Impact of CCND3 Transcript Codon Sequence Modifications on Transient CCND3 Protein Production at 32°C and 37°C 24 and 48 Hours Post Transfection</i>	103
<i>4.4.1 Expression Analysis of CCND3 for Gs Codon Transcript Variant Modifications at 32°C and 37°C</i>	103
<i>4.4.1.1 Comparison of CCND3 Transient Protein Production at 24 and 48 Hour Post Transfection from Codon Variant Transcripts</i>	103
<i>4.4.1.2 Western Blot Analysis of Experimental Replicates of Protein Production from each CCND3 Transcript Variant Sequence</i>	106
<i>4.4.2 Protein Production Analysis of CCND3 From Gv Transcript Variant Modifications</i>	111
<i>4.4.2.1 Comparison of CCND3 Protein Production at 24 and 48 h Post Transient Transfection from Different Transcript Variants</i>	111

4.4.2.2 Analysis of Experimental Replicates of each CCND3 Sequence for Statistical Analysis.	113
4.4.3 Protein Production Analysis of CCND3 from Ts Transcript Variant Modifications	118
4.4.3.1 Comparison of Transient CCND3 Protein Production at 24 and 48 Hour Post Transfection from Different Transcript Variants	118
4.4.4 Analysis of CCND3 Protein Production from Ts Transcript Variant Modifications	127
4.4.4.1 Comparison of Transient Protein Production of CCND3 at 24 and 48 hour Post Transfection	127
4.4.4.2 Western Blot Analysis of Protein Production from Experimental Replicates of Each CCND3 Transcript Variant Sequence	130
4.5 Determining any Impact of Stable CCND3 Transcript Variant, and Subsequent Exogenous CCND3 Protein Production, on the Growth Profiles of HEK293 Using the XCELLigence System	133
Chapter 5 Investigating Codon Transcript Variant Impact on GFP and Noggin Protein Production	
5.1 Introduction	137
5.1 eGFP Protein Production from Different Transcript Variants in HEK293 Cells Incubated at 37°C or 32°C at 24 and 48 hour Post Transient Transfection	138
5.3 Secretory eGFP Protein Production from Different Transcript Variants in HEK293 Cells Incubated at 37°C or 32°C at 24 and 48 hour Post Transient Transfection	143
5.4 Assessment of the Impact of GFP Transcript Variant Expression on HEK293 Growth using the xCELLigence Instrument	151
5.5 Noggin Expression from 5' Open Reading Frame Codon Transcript Variants Encoding for the Same Protein	153
5.5.1 Comparison of Noggin Protein Production from Different Transcript Variants in HEK293 Cells Cultured at 37°C or 32°C, 24 and 48 hour Post Transient Transfection	153
5.5.2 Assessment of the Impact of Noggin Transcript Variant Expression on HEK293 Growth using the xCELLigence Instrument	161
Chapter 6 Discussion	
6.1. Comparing codon modifications and transcript variants of all transcripts tested and the impact on protein synthesis	163
6.2. The Potential Implications and Applications of Region-Specific Codon Optimisation	170
6.3. Further Supplementary Studies to Potentially Improve the Ability to Design Transcript Variant Sequences to Tune Protein Production to Desired Amounts	171
6.4. Limitations of this Study	174
6.5. Final Conclusions	175
References	177
Appendices	200

Abbreviations

aa-tRNA	Aminoacyl-tRNA
ABCE1	ATP-binding cassette family E member 1
Arg	Arginine
ATP	Adenosine triphosphate
BSA	Bovine serum albumin
Ca²⁺	Calcium ion
CCND3	Cyclin D3
CD	D-type cyclin
CDK	Cyclin-dependent kinase
CHO	Chinese hamster ovary
CIRP	Cold inducible RNA-binding protein
CMV	Cytomegalovirus
Cryo-EM	Cryogenic electron microscopy
eEF	Eukaryotic elongation factor
EF	Elongation factor
eGFP	Enhanced green fluorescent protein
eIF	Eukaryotic initiation factor
ER	Endoplasmic reticulum
eRF	Eukaryotic release factor
Fe-S	Iron-sulphur
FITC	Fluorescein isothiocyanate

GDP	Guanosine diphosphate
GFP	Green fluorescent protein
GGQ	Glycine-glycine-glutamine
GTP	Guanosine triphosphate
HEK	Human embryonic kidney
IRES	Internal ribosome entry site
Leu	Leucine
M7G	7-methylguaylate
Met-tRNAⁱMet	Initiator methionine transfer RNA
mRNA	Messenger ribonucleic acid
NaAc	Sodium Acetate
NMD	Nonsense mediated decay
NOG	Noggin
ORF	Open reading frame
PABP	Poly(A)-binding protein
PBS	Phosphate buffered saline
PNPase	Polynucleotide phosphorylase
Pro	Proline
PTC	Peptidyl transfer centre
Rb	Retinoblastoma
Rbm3	RNA binding motif protein 3
RNA	Ribonucleic acid
rRNA	Ribosomal ribonucleic acid

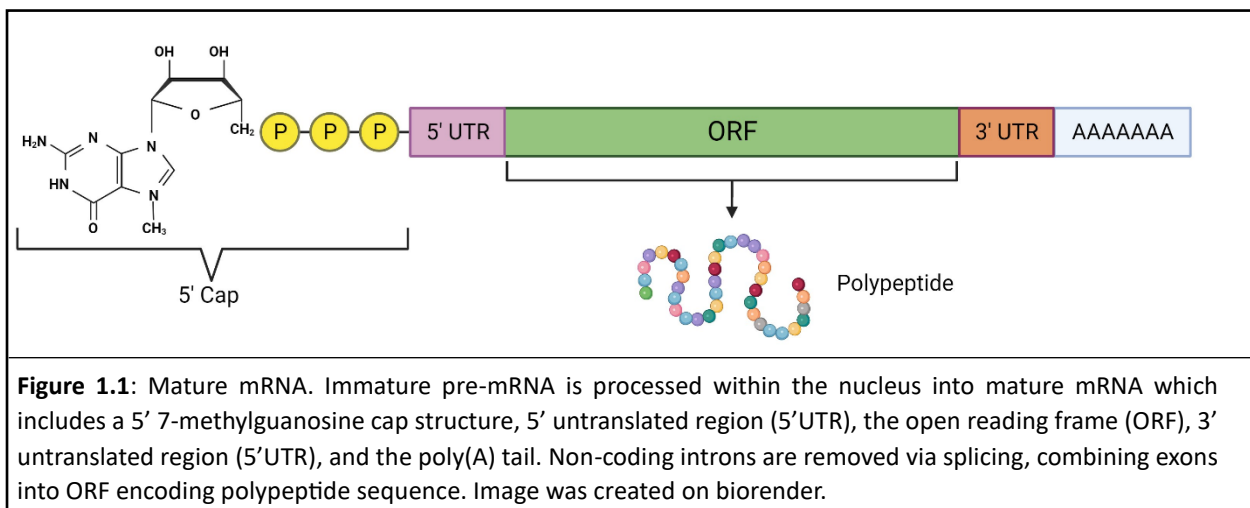
RTN3	Reticulon 3
sGFP	Secretory-green fluorescent protein
siRNA	Small interfering ribonucleic acid
SRP	Signal recognition particle
SRPR	Signal recognition particle receptor
TGF-β	Transforming growth factor beta
tRNA	Transfer ribonucleic acid
UTR	Untranslated reading frame

CHAPTER 1

Introduction

1.1. mRNA Translation in Eukaryotes; An Overview

The production of proteins is integral for all life processes and they account for a large fraction of biological macromolecules within our body- comprising approximately 20% to 30% of cellular content (G. C. Brown, 1991; Wiśniewski et al., 2014). In eukaryotic cells, this process primarily originates within the nucleus of cells where messenger-ribonucleic acid (mRNA) transcripts encoding specific genes are transcribed from template strands of deoxyribonucleic acid (DNA) in a complex process involving RNA polymerase, which synthesises singled stranded mRNA (Sainsbury et al., 2015). mRNA is initially transcribed in the nucleus as an immature pre-mRNA primary transcript which is then further processed to generate the mature mRNA before it is transported from the nucleus for translation. A mature mRNA broadly consists of a 5' Cap structure, a 5' untranslated region (5'UTR), the open reading frame (ORF) that encodes for the polypeptide sequence, a 3' UTR and a polyadenylation region at the 3' end (Figure 1.1). The processing of the immature pre-mRNA includes the splicing out of introns (non-coding regions) and joining of the exon (coding regions) to form the required ORF (Pan et al., 2008), the addition of a 5' 7-methylguanosine cap (m^7G cap) that is a modified guanine nucleotide important for recruitment of the translational machinery and stability (Mukherjee et al., 2014; Ramanathan et al., 2016), and the addition of a polyadenylation sequence at the 3' end of the message to form a poly(A) tail (Millevoi & Vagner, 2010). The poly(A) tail is usually 100-200 nucleotides in length and is important in protecting the mRNA from degradation by exonucleases, transport out of the nucleus whereby specific proteins bind to it, and binding of the poly(A) binding protein (PABP) that helps stabilise the transcript (Katahira, 2015; Passmore & Collier, 2022). PABP interacts with proteins bound at the 5' end of the mRNA to form a circularised mRNA that is thought to be more efficiently translated (Passmore & Collier, 2022; Vicens et al., 2018).



Once transcribed and processed, mRNA may exit the nucleus through a protein complex known as a nuclear pore and enters the cytoplasm, freely floating within the cytosol where protein synthesis known as mRNA translation commences. Mature mRNAs in the nucleus are recognised by the completion of the modifications described above and by binding to specific proteins that aid in transport (Katahira, 2015). The half-life of mRNA can be relatively long (Raghavan, 2002) in the context of rapid changes in the proteome required in response to environmental stresses and thus rapid regulation of protein production rates is achieved by controlling mRNA translation, and thus rates of protein synthesis from mRNAs, as well as protein degradation rates (Bastide et al., 2017; Knight et al., 2020; Spriggs et al., 2010). In general, the rate of protein synthesis is proportional to the concentration and translation efficiency of its mRNA transcript. Translation is also a highly conserved process evolutionarily, with many studies in bacteria still able to provide insight into the eukaryotic process. Most studies covered in this introduction focus on eukaryotes unless specified otherwise (Rodnina & Wintermeyer, 2009).

Ribosomes are the chief machinery involved in translation and transfer ribonucleic acids (tRNA) are responsible for decoding mRNA nucleotide sequence into an amino acid sequence for newly formed polypeptides. The ribosome is a large macromolecular ribonucleoprotein complex comprised of many ribosomal proteins (r-proteins) and lengths of ribosomal ribonucleic acids (rRNAs) which are organised into 2 subunits; in eukaryotes these are a larger 60S, and a smaller 40S, subunit named due to their ultracentrifugation sedimentation coefficients expressed in Svedberg, with a fully intact ribosome being 80S in size (Figure 1.1) (Ben-Shem et al., 2011). These subunits interact primarily via inter-subunit bridges at the predominantly rRNA-based subunit interface and contain many features vital to their function (Ben-Shem et al., 2011; Weisser & Ban, 2019). Some of note are the peptidyl transferase centre (PTC), the polypeptide exit tunnel, L1 stalk, P1/P2 stalk, and the mRNA decoding centre. The PTC resides within the 60S subunit and is responsible for catalysing chemical reactions involving the polypeptide for elongation and eventual release (Weisser & Ban, 2019). The L1 stalk coordinates the ejection of E-site tRNAs from the ribosome whereas the P1/P2 stalk and sarcin-ricin loop assist the recruitment and activation of ribosome-dependent guanosine triphosphatases (GTP) (Trabuco et al., 2010; Wahl & Moller, 2002). Lastly, verification of base pairing between mRNA codon and tRNA anticodon is achieved by rRNA and r-protein interactions within the 40S A site in a region known as the mRNA decoding centre (Blanchard et al., 2004; Geggier et al., 2010). The ribozyme nature of ribosomes likely indicates that their evolution origin dates back to an early RNA-based world.

Full sized 80S ribosomes are assembled on mRNA transcripts to begin translation (Figure 1.1). Once initiated (the process of loading and assembling ribosomes on a transcript) at the start AUG codon, the ribosomes move along the transcript sequence creating an ever-elongating nascent polypeptide by utilising charged tRNA (tRNA with appropriate amino acid linked for delivery to the ribosome) until the amino acid sequence is completed; signalled by the presence of a stop codon in the mRNA sequence. There are several factors and machinery residing within the cytosol that facilitate this process which can be broken into 4 main phases: initiation, elongation, termination and ribosomal recycling (Figure 1.2).

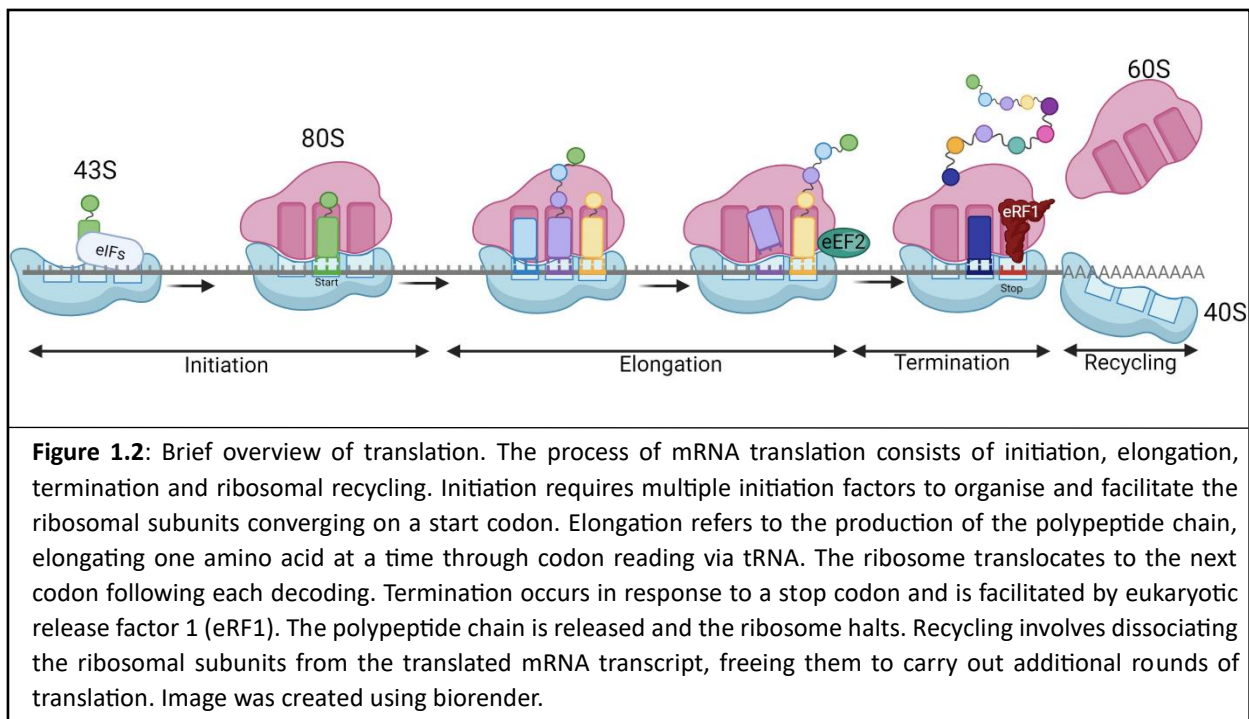


Figure 1.2: Brief overview of translation. The process of mRNA translation consists of initiation, elongation, termination and ribosomal recycling. Initiation requires multiple initiation factors to organise and facilitate the ribosomal subunits converging on a start codon. Elongation refers to the production of the polypeptide chain, elongating one amino acid at a time through codon reading via tRNA. The ribosome translocates to the next codon following each decoding. Termination occurs in response to a stop codon and is facilitated by eukaryotic release factor 1 (eRF1). The polypeptide chain is released and the ribosome halts. Recycling involves dissociating the ribosomal subunits from the translated mRNA transcript, freeing them to carry out additional rounds of translation. Image was created using biorender.

1.1.1. The Initiation Phase of mRNA Translation

Initiation is a major regulatory node for mRNA translation and traditionally thought to be the predominant method of translational control (Jung et al., 2014). It is a highly complex process requiring at least 12 additional initiation factors which are also targets of regulatory pathways, notably the de/phosphorylation of key translation factors that in turn attenuates or promotes translation initiation (Hinnebusch & Lorsch, 2012). Evidence is still emerging that not all factors have been identified, highlighted by the implicated involvement of DEAD-box helicase DDX3 and DHX29 helicase in the process (Hinnebusch & Lorsch, 2012; Lai et al., 2008). Initiation occurs predominantly in either a cap-dependent (relying on the 5' cap) or independent manner, with cap dependant being the most common in eukaryotes (Figure 1.3).

As outlined above, eukaryotic mRNA is co-transcriptionally modified at the 5' end to include a N⁷-methylguanosine (m⁷G) cap (Wang et al., 2002; Wurm & Sprangers, 2019). For cap-dependent initiation, the mRNA is thought to be organised into a closed loop structure by multiple interactions involving the eIF4F complex (where eIF is an abbreviation for eukaryotic initiation factor) and the poly(A)-binding protein (PABP) (Vicens et al., 2018; Wells et al., 1998). eIF4F is made from a number of different translation factors; eIF4A, eIF4E and eIF4G components, each with a role to play. eIF4A is an ATP-dependent RNA helicase which resolves secondary mRNA structures important for future 43S preinitiation complex scanning (DHx29 also facilitates this) (Hinnebusch, 2011, 2014). eIF4E binds to the 5' cap and eIF4G acts as the framework that directly associates with other eIF4s as well as PABP (Shatkin, 1976; Wells et al., 1998). eIF4 and PABP interactions create the mRNA closed loop structure as PABP coats the poly(A) tail of mRNA- another common modification for eukaryotic mRNA found at the 3' end as described above, so their interactions effectively lead to circular mRNA. It is poorly understood why this is advantageous to translation efficiency but has been suggested that it aids in stabilisation and facilitating the utilisation or recycling of 40S ribosomes, however recent reviews still question its function (Vicens et al., 2018).

Away from the mRNA strand, the 43S preinitiation complex forms comprised of: the 40S ribosome subunit, a ternary complex of methionyl-tRNA (Met-tRNA^{Met}), eIF2 and GTP, and eIFs 1, 1A, 3 and 5. These attach to the 5' end of mRNA and move toward the 3' end in an ATP driven scanning process until a start AUG codon is reached. This is a complex process summarised in figure 1.3. Initiation factors either directly assist in association with the mRNA, scanning or stabilisation of the complex (Benne & Hershey, 1978; Hinnebusch, 2014; Jackson et al., 2010; Pestova et al., 1998). eIF3 is a large complex of 13 individual subunits that interacts with the eIF4F complex and prevents premature binding of the 60S subunit (Kolupaeva et al., 2005). eIF2 in the ternary complex is responsible for directing the initiator Met-tRNA^{Met} to the P site of the 40S ribosome subunit (Hinnebusch, 2014; Schreier & Staehelin, 1973). Once an AUG start codon is located, eIF2 hydrolyses GTP signalling for the dissociation of several initiation factors from the 40S ribosome subunit (Adomavicius et al., 2019). GTP hydrolysis also requires GTPase-activating protein eIF5 and allows the 60S subunit to associate, completing the 80S ribosome complex capable of elongation (Hinnebusch, 2011; Unbehaun et al., 2004).

There are multiple methods of regulation surrounding this process but initiation is not the main focus of this study so only a select few will be covered here. eIF2 function can be regulated via phosphorylation of a conserve serine residue in the α -subunit (eIF2 α). GTP hydrolysis signalled by encountering an AUG codon converts eIF2-GTP to eIF2-GDP and while in this form, eIF2 cannot bind to another Met-tRNA^{Met}. eIF2 functionality is rescued by eIF2B acting as a guanine nucleotide exchange factor, successfully reintroducing GTP to eIF2 . Phosphorylation of the specific serine residue (Ser51) within the eIF2 α subunit interferes with this process, thereby preventing the ternary complex from reforming and causes global reductions in protein synthesis (Adomavicius et al., 2019; Singh et al., 2006). This form of regulation can occur in response to multiple factors including amino acid starvation, protein misfolding or stress in the ER and viral infection of the cell. Likewise, the eIF4F complex can also be inhibited by 4EBP binding to eIF4E. This inhibits interactions between itself and eIF4G (Rong et al., 2008). Growth factors can prevent this loss of initiation and encourage protein synthesis by phosphorylating 4EBP which in turn reduces its affinity for eIF4E.

Cap independent initiation commonly occurs for viral genomes but does also occur in mammalian cells, although to a lesser extent (Miller et al., 2007). This process forgoes 5' end scanning and the translational machinery assembles close to or directly at the start AUG codon. In these cases, the area of assembly is referred to as an internal ribosome entry segment (IRES) (Jackson, 2013). This method of initiation in eukaryotes commonly occurs for specific transcripts in response to cellular stresses which generally results in translation reduction or attenuation, perhaps as a means of evading this affect (Godet et al., 2019).

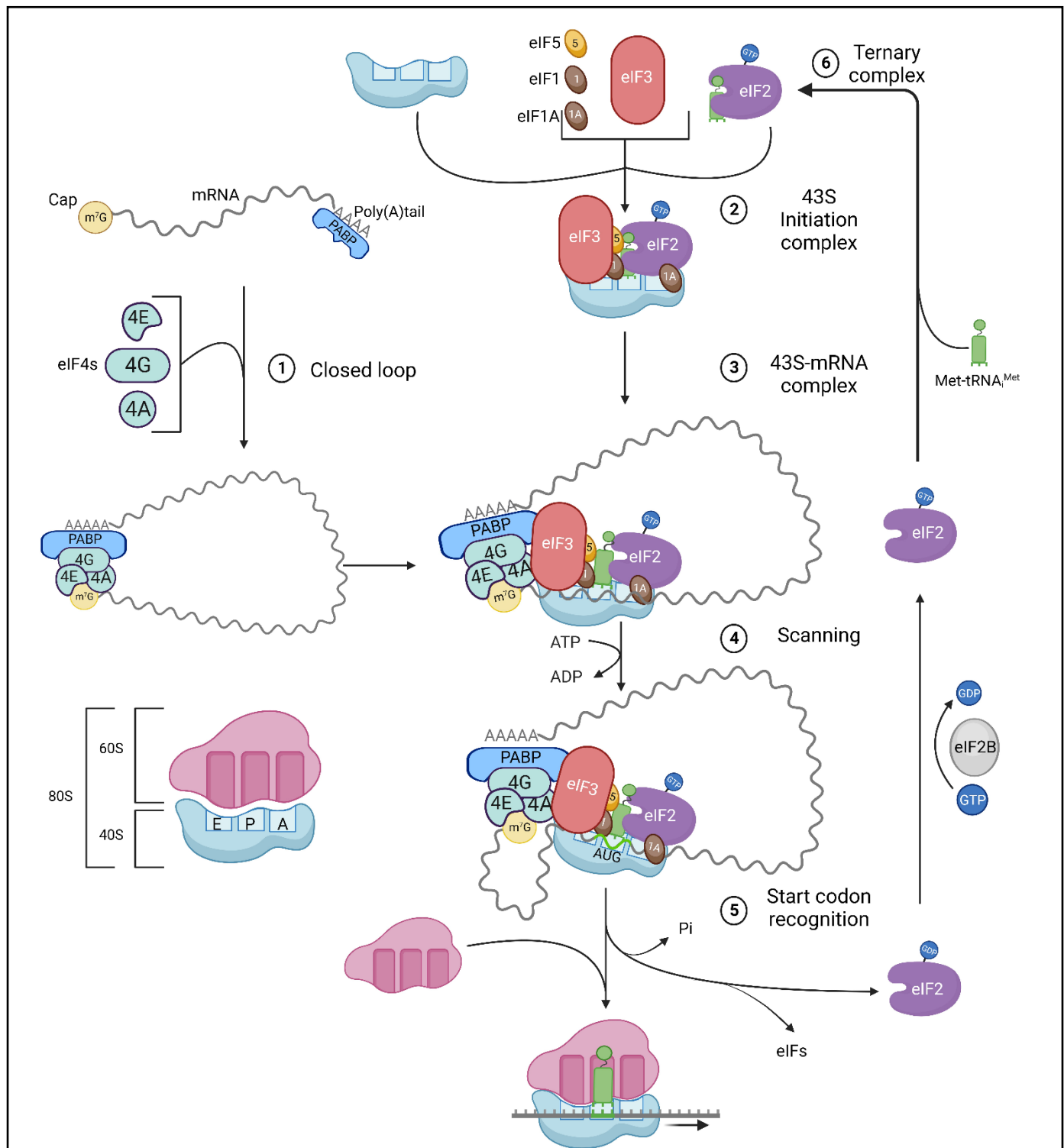


Figure 1.3: Cap-dependent initiation. 1) eIF4A, eIF4E and eIF4G come together to form the eIF4F complex which associates with 5' cap of mRNA as well as PABP bound to the 3' poly(A) tail. This creates a closed loop mRNA structure. 2) The 43S initiation complex is formed from several constituents including the ternary complex, the 40S ribosome subunit and additional initiation factors. eIF1 and eIF1A assists in stabilising the initiation complex, eIF3 interacts with eIF4F and inhibits premature 60S ribosome subunits from binding, whilst eIF5 promotes GTP hydrolysis once start codon is identified. The ternary complex is comprised of GTP, eIF2 and Met-tRNA^{Met}, which is delivered to the P site within the 40S subunit. 3) The 43S initiation complex attaches to the 5' end of the closed loop structure mediated by eIF3- eIF4F interactions. 4) Initiation complex scans mRNA sequence in an ATP dependent manner. 5) Once start codon is located GTP associated with eIF2 is hydrolysed releasing phosphate, triggering release of initiation factors and the recruitment of the 60S ribosome subunit in preparation for elongation. 6) eIF2 ability to associate with Met-tRNA^{Met} is restored through eIF2B activity replenishing GTP. Image was created using biorender.

1.1.2. The Elongation Phase of mRNA Translation

The next step of translation after completion of initiation and assembly of the ribosome on the AUG start codon, is the elongation phase and is highly relevant to the studies carried out in this thesis. The elongation phase is strongly conserved between eukaryotes and even into prokaryotes (Dever et al., 2018; Dever & Green, 2012). This phase begins with the subsequent codon after the initiator AUG codon directly following initiation. During elongation, the polypeptide chain is formed by linking specific amino acids sequentially directly correlating to the triplicate nucleotide codons in the mRNA open reading frame code. This is repeated until each codon has been successfully decoded up to the stop/termination codon, incorporating amino acids until a complete polypeptide is synthesised. This phase can itself be viewed as consisting of 3 major stages: transcript decoding via tRNA selection, peptide bond synthesis, and translocation of the mRNA-tRNA complex in relation to the ribosome. Ribosomes can encounter several hindrances during this process which are covered in 1.3

1.1.2.1 tRNA Selection

The 80S ribosome contains 3 sites which can be occupied by tRNAs, the aminoacyl site (A), the peptidyl site (P) and the exit site (E) (Figure 1.3). The P site houses the Met-tRNA^{Met} or the tRNA with the growing polypeptide chain attached and the A site is where the next aminoacyl-tRNA (aa-tRNA) are selected for subsequent amino acid incorporation (Figure 1.4) (Weisser & Ban, 2019). Each site is centralised around a single triplicate nucleotide codon. aa-tRNAs are transported to ribosomal A sites by the GTPase eEF1A (elongation factor Tu in bacteria) in a ternary complex with GTP, somewhat similar to Met-tRNA^{Met} transfer seen in initiation (Carvalho et al., 1984; Fischer et al., 2015). This continues stochastically until a cognate tRNA to the codon within the A site is identified. Once these codon-anticodon interactions occur by classic Watson-Crick interactions, the ribosome acts as a GTPase activating factor to stimulate eEF1A hydrolyses of GTP, releasing eEF1A from the ribosome and enabling the cognate aa-tRNA to occupy the A site (Maracci et al., 2014). eEF1A is recycled by eEF1B catalysing guanine nucleotide exchange; swapping GDP for GTP (Pittman et al., 2006; Sasikumar et al., 2012). Though eEF1A and EF-Tu in bacteria show high homology, this aspect of elongation differs distinctly in bacteria and may provide greater elongation control in Eukaryotes (Sivan et al., 2011).

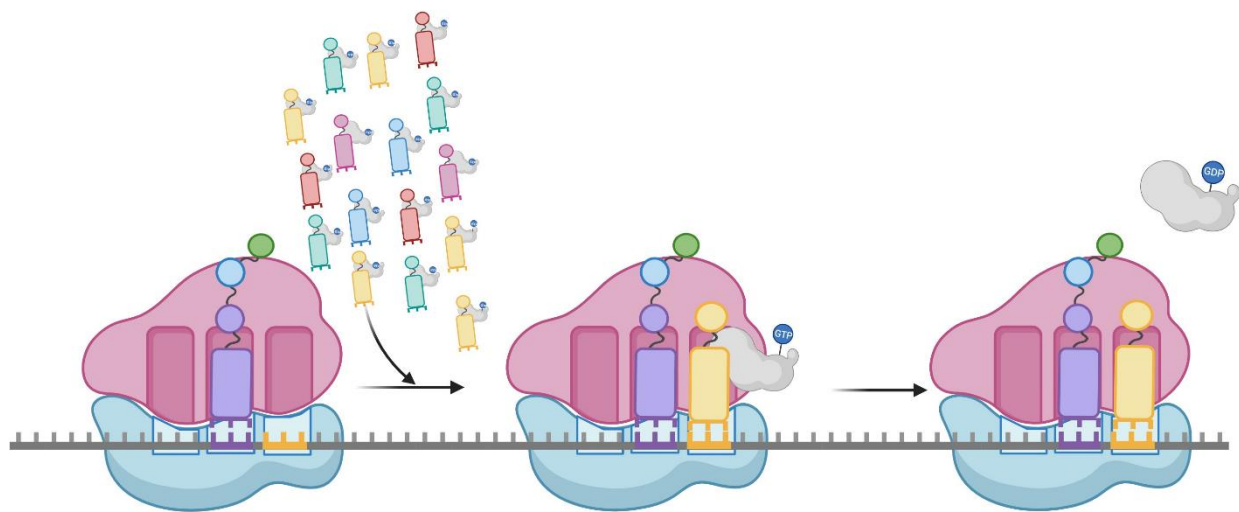
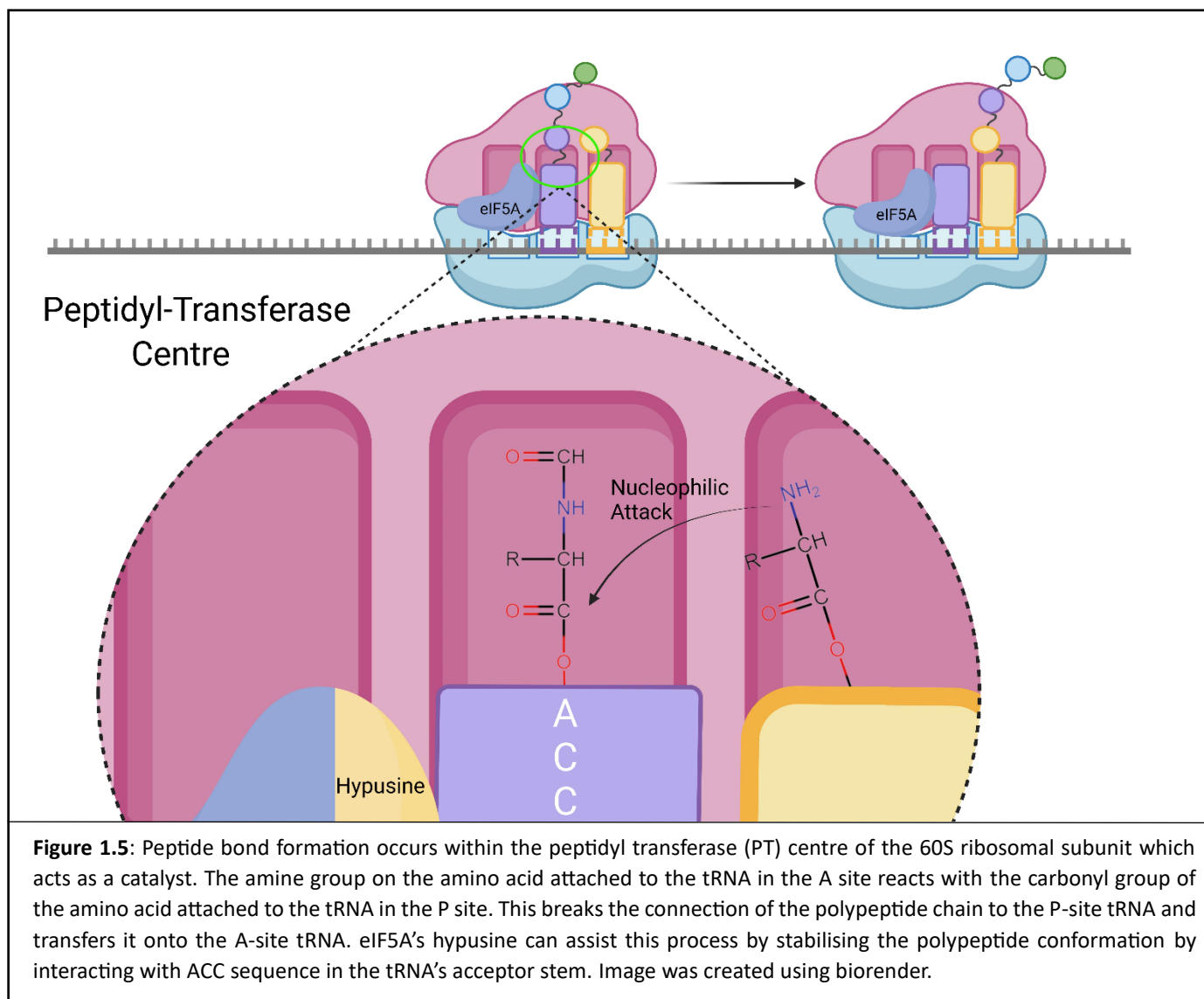


Figure 1.4: The tRNA selection process is governed by stochastic interactions between tRNA populating the cytosol and the mRNA codon in the ribosomal A site. tRNAs are delivered by eEF1A (grey) and once cognate pairing is identified, eEF1A hydrolyses associated GTP which causing eEF1A to dissociate and allows the cognate aa-tRNA to occupy the A site. Image was created using biorender.

1.1.2.2 Amine and Carboxylic Acid Reaction; Peptide Bond Synthesis

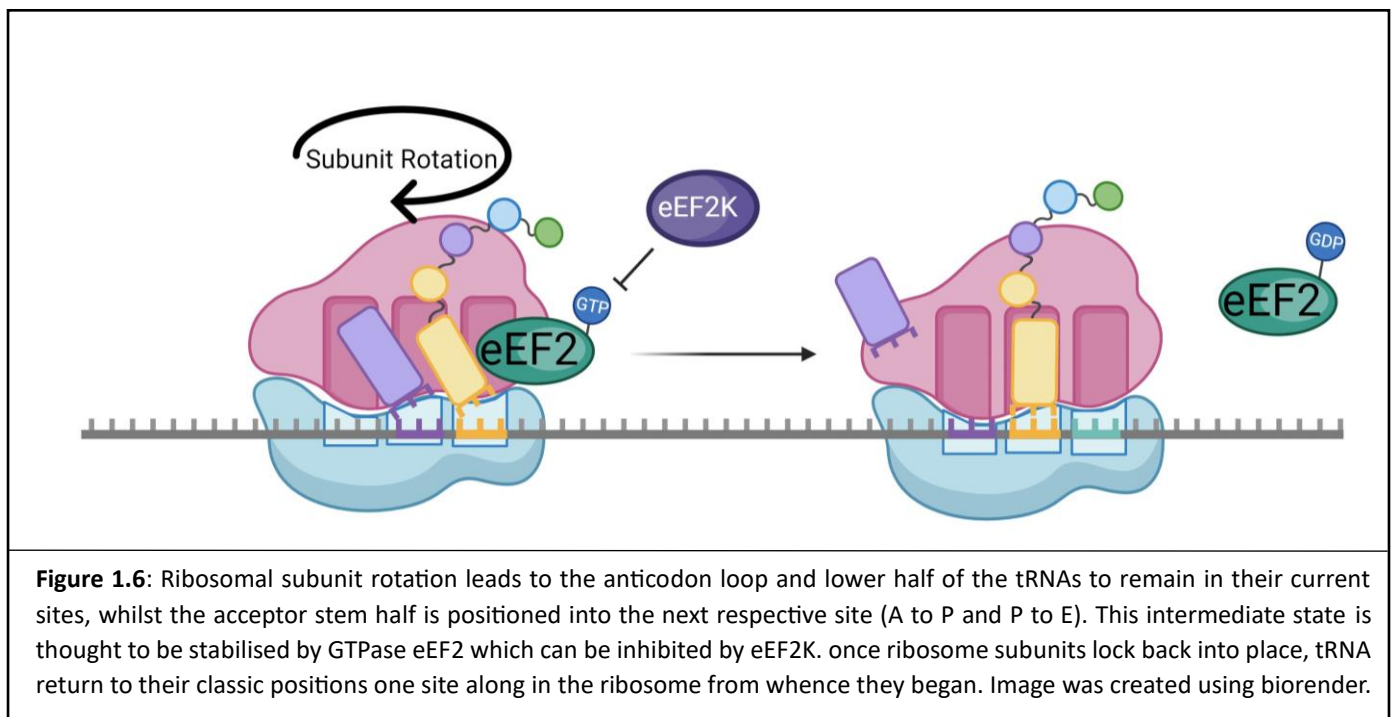
Amino acids contain both a basic amine group and an acidic carboxylic acid group. This allows two to become linked by formation of an amide bond: a peptide bond, by way of nucleophilic attack (Figure 1.5) (Beringer et al., 2005). The α -amino group of the aa-tRNA in the A site attacks the electrophilic carbonyl group present in the ester linkage in the peptidyl-tRNA in the P site. This condensation reaction is facilitated by the PT centre of the 60S ribosome subunit employing entropic catalysis (Beringer & Rodnina, 2007; Wohlgemuth et al., 2006). The effect is that the growing nascent polypeptide peptide is transferred to the A site tRNA leaving the P site tRNA uncharged. eIF5A can assist in stimulating this reaction by occupying the E site if it remains empty for a prolonged period of time (Choi & Puglisi, 2017). Specific examples of this function are outlined in section 1.3 but it has been shown to assist peptide bond formation for multiple tRNA combinations and may carry out this role throughout elongation (Pelechano & Alepuz, 2017; Saini et al., 2009).



1.1.2.3 Translocation

Following successful polypeptide transfer, the ribosomal subunits need the next codon to be available in the A site and the current uncharged tRNA and peptidyl-tRNA to translocate into the next site. Thus, movement is required which is achieved by the 2 subunits rotating in relation to each other (Figure 1.6)(Ratje et al., 2010). This causes the anticodon loop and lower half of the tRNAs to remain in their current sites, whilst the acceptor stem half is positioned into the next respective site (A to P and P to E) (Behrmann et al., 2015; Moazed & Noller, 1989). This intermediate state is thought to be stabilised by GTPase eEF2 which is essential for translocating the mRNA-tRNA complex relative to the ribosome (Taylor

et al., 2007). Rapid GTP hydrolysis brings about conformational changes in the eEF2 structure which are thought to 'unlock' ribosomal subunits allowing mRNA-tRNA complex movement before 'relocking' subunits in a post-translational state (Shoji et al., 2009; Taylor et al., 2007). This returns tRNAs to their classical positions before eEF2 dissociates with the need to replace GDP with GTP for further translocation (Spahn et al., 2004). In this fashion, eEF2 is also thought to prevent backwards movement of tRNAs during the 'unlocked' state of ribosomal subunits (Dever & Green, 2012; Gao et al., 2009). The phosphorylation of eEF2 by Ca^{2+} activated protein kinase eEF2K has been shown to inhibit translation potentially by impairing its binding affinity to the ribosome (Carlberg et al., 1990). A summary of this process is shown in Figure 1.6.



1.1.3. Termination Phase

Once all codons in the ORF have been translated the ribosome will progress to a stop codon. In eukaryotes these codons are limited to UAA, UAG and UGA. At this point, the complete polypeptide must be released from the final cognate tRNA (Figure 1.7). Similar to elongation, a nucleophilic attack on the peptidyl-tRNA is executed, except this time not from another aa-tRNAs amino group, but rather a water molecule is utilised facilitated by release factor 1 (eRF1) (Song et al., 2000). Given its ability to enter the A site to fulfil

its role, it is unsurprising that eRF1 is similar in shape to tRNA, but where tRNA are cognate to one codon, eRF1 is responsible for recognition of all stop codons with high specificity through highly conserved Gly-Thr-Ser and YxCxxF motifs in the eRF1 protein (A. Brown et al., 2015; Schuller & Green, 2018; Song et al., 2000). eRF1 contains other highly conserved elements including an Asn-Ile-Lys-Ser (NIKS) tetrapeptide located at the N-terminal domain, and a Gly-Gly-Gln (GGQ) motif. NIKS is postulated to also be integral in stop codon recognition and ribosome binding. When altered through site directed mutagenesis, alterations caused a decrease in termination promotion sensitive to stop codon involved and mutation achieved (L. Frolova et al., 2002). GGQ is positioned at the exposed tip of eRF1's domain 2, which coincides with the acceptor loop of tRNA and allows GGQ to extend into PTC and promote nascent peptide release (Song et al., 2000). When GGQ was mutated via mutagenesis of both Gly residues, eRF1 lost all functionality as a release factor for all 3 stop codons (L. Y. Frolova et al., 1999).

eRF1 also requires eRF3, a GTPase, for proper function. eRF3 stimulates termination in a GTPase-dependent manner and delivers eRF1 to the A site to catalyse peptidyl hydrolysis at the PTC (L. Frolova et al., 1996; Salas-Marco & Bedwell, 2004). This has also been shown to be stimulated by eIF5A presence in the E site. Once the eRF1-eRF3 complex reaches the A site with a stop codon present, eRF3 dissociates following GTP hydrolysis and the GGQ motif of eRF1 is able to direct H₂O hydrolysis in the PTC to release nascent polypeptide from the P site tRNA (des Georges et al., 2014). ATP binding cassette sub-family E member 1 (ABCE1) plays a role in termination, though it is primarily involved in ribosome recycling and rescue. Here, it stabilises active conformation of eRF1 thereby exciting catalytic capabilities (Figure 1.7) (Pisarev et al., 2010; Preis et al., 2014).

Readthrough is a potential issue in termination where the ribosome continues to translate past a stop codon into the 3' UTR and poly(A) tail. This has been shown to be caused by some components of eIF3 (Beznosková et al., 2013, 2015). This is suggested to be indicative of reduced termination efficiency. Stop codon usage may also contribute to readthrough probability and potentially the nucleotide immediately following the stop codon in a 4+ nucleotide model. For example, studies have identified that UGAC results in higher readthrough tendencies compared to UAAG sequence and that stop codon +C in general may lead to increased readthrough (Beznosková et al., 2016; Bonetti et al., 1995; A. Brown et al., 2015). These factors underpin the complexity in fully understanding translation and its control.

As mentioned above, eIF5A has been implicated in assisting termination as well as other areas of translation. Termination as a process occurs at a slower rate than peptidyl transfer implying that the E site is empty for a prolonged period of time as seen when awaiting a rare cognate tRNA (Gnirke et al., 1989; McClary et al., 2017). eIF5A's involvement in termination lends further evidence that the E site availability acts as a sensor for translation rates and if assistance is required.

PABP may also impact termination rates. It can assist in eRF1-eRF3 complex recruitment, promote eRF1 catalysis function and may be involved in mRNA mediated decay protection (Ivanov et al., 2016). Nonsense mediated decay (NMD) quality control pathways prevent the production of truncated proteins from mRNA with premature stop codons (Brognia & Wen, 2009). These proteins could attribute to proteotoxicity and therefore their prevention is integral to healthy cell homeostasis. PABP may prevent erroneous degradation of functional mRNA transcripts by its proximity to legitimate stop codons (Peltz et al., 1993). Once artificially added in the vicinity of premature stop codons, NMD has been shown to decrease hinting at PABP's regulatory purpose (Amrani et al., 2004).

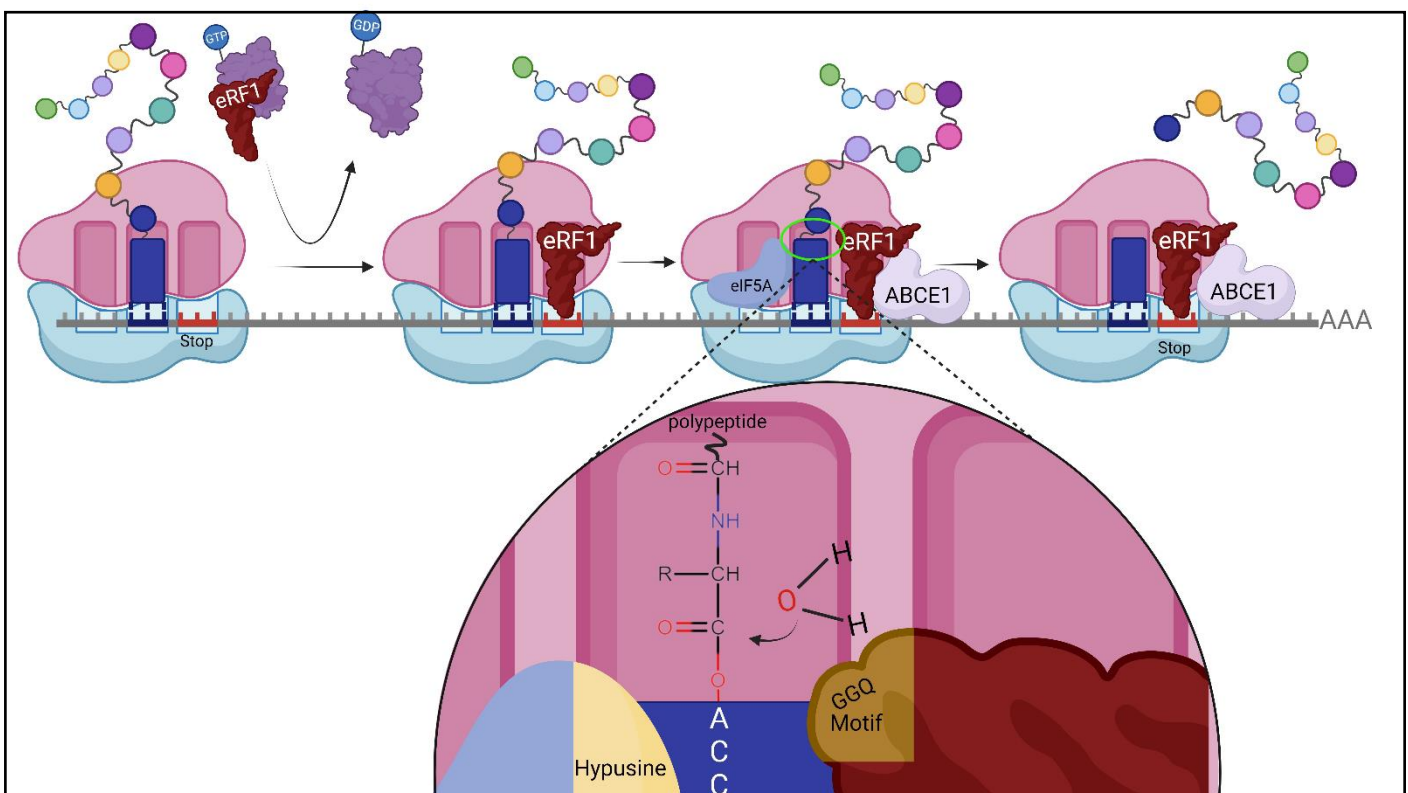


Figure 1.7: The termination phase of elongation. Stop codon in the ribosome A site is identified by eRF1 which is delivered by eRF3. eRF1 facilitates a nucleophilic attack of a water molecule on the carbonyl group from the amino acid attached to the P-site tRNA. This releases the polypeptide from the tRNA for further processing. ABCE1 and eIF5A assist by stabilising the conformational shape of eRF1 and tRNA (thereby the polypeptide chain) respectively. Image was created using biorender.

1.1.4. Recycling & Rescue

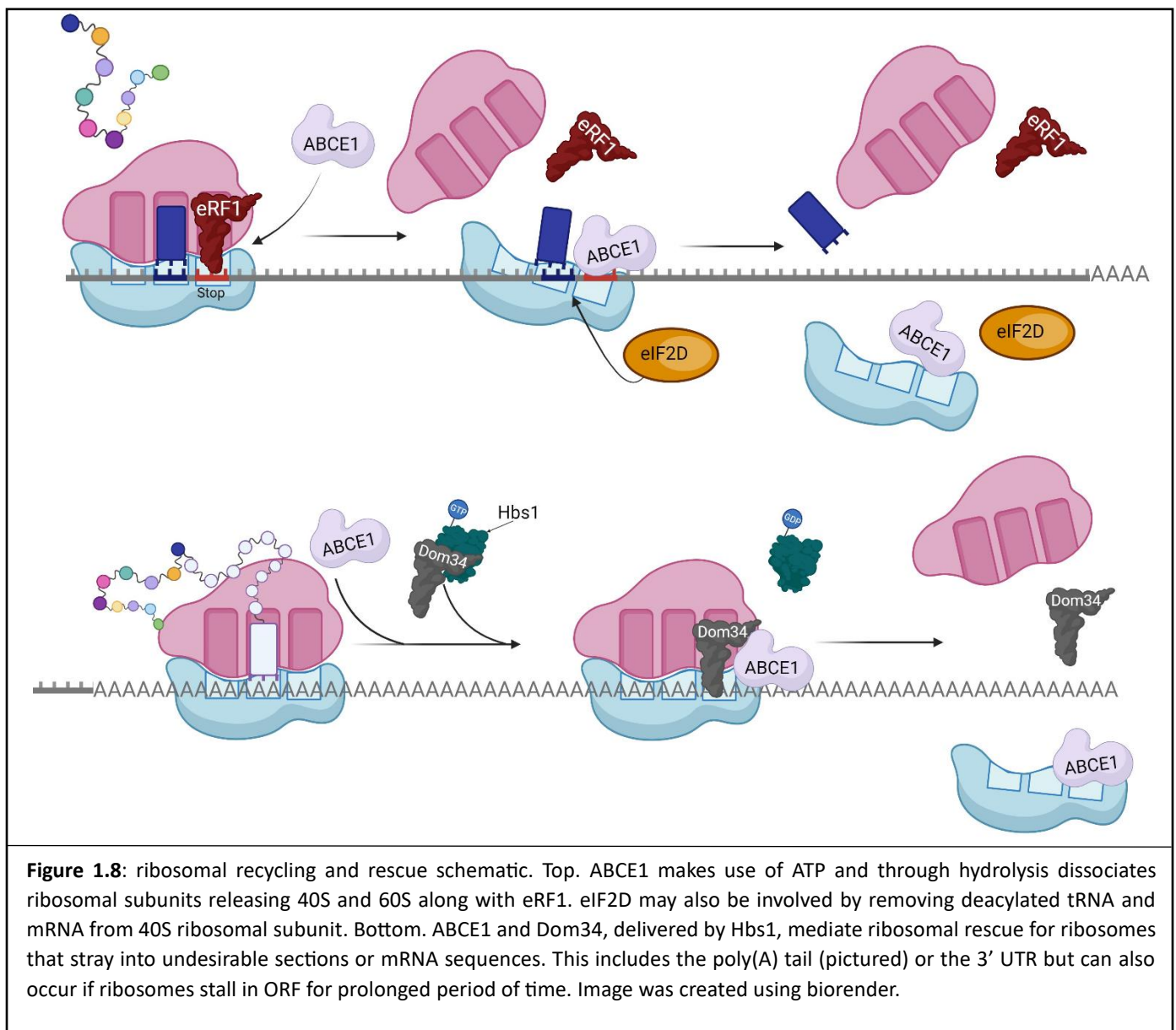
1.1.4.1 Recycling

The fourth and final step in mRNA translation is the recycling of the ribosomal subunits so that they can begin the process anew. The 80S complex along with deacylated tRNA-mRNA must be disassembled into individual 40S and 60S subunits and release the mRNA transcript. ABCE1 is essential in this process for all eukaryotes (Barthelme et al., 2011; Young et al., 2015). It contains an iron-sulphur (Fe-S) cluster thought to be necessary for this process (Barthelme et al., 2011). As the name implies, ABCE1 makes use of ATP and through hydrolysis dissociates ribosomal subunits releasing 40S and 60S along with eRF1, linking termination and recycling (Barthelme et al., 2011; Khoshnevis et al., 2010). ABCE1 not only links recycling to termination but also to initiation due to evidence of co-immunoprecipitation with several initiation factors (Dong et al., 2004; Pisarev et al., 2007). This may be due to the cyclical nature of translation and shows recycling at work. It is unclear how or if ABCE1 plays a role in initiation or the recruitment of initiation factors. Cryo-EM analysis of ABCE1 bound to either 80S or 40S highlighted a change in Fe-S cluster conformation pre- and post-split that suggests this movement forces eRF1 further into ribosome inter-subunit space, leading to dissociation (Heuer et al., 2017). Furthermore, post-split conformation also prevents the 60S subunit from rejoining, ensuring separation is final. It is noted that eIF2D may also be involved by removing deacylated tRNA and mRNA from 40S ribosomal subunit, and thus completing the process of returning the ribosomal subunits to their pre-translation state, now available for initiation factors to prepare for further initiation (Skabkin et al., 2010).

1.1.4.2 Rescue

Ribosomal rescue is required when stalling occurs along an mRNA transcript and involves removing the ribosome and aborting translation (Figure 1.8). Most stalls within the ORF of accurate mRNAs are resolved without such drastic actions being required and are covered in section 1.3 which, for brevity, will not be repeated here. However, some occur where ribosomes move to inappropriate areas which will be covered in this section. In many instances, the response is orchestrated by multiple factors and machinery whose dual purpose is not only to rescue stranded ribosomes, but also to degrade nascent polypeptide and faulty mRNA (Joazeiro, 2017).

Dom34 alongside ABCE1 aid in rescuing stalled ribosomes who find themselves within 3'UTRs, 3' ends of truncated mRNAs, and ones translating in the poly(A) tail region (Figure 1.8) (Guydosh & Green, 2014, 2017). Dom34 can selectively cause the degradation of mRNA sequence features that inhibit translation or lead to complete stalling such as secondary structures like stem loops, stretches of non-optimal codons or truncated mRNAs (Tsuboi et al., 2012). Structurally, Dom34 shows high homology to eRF1 although there are notable differences. The absence of NIKs and GGQ motifs are logical due to difference in function and Dom34 is chaperoned by a different GTPase: Hbs1 (Doma & Parker, 2006; Graille et al., 2008; van den Elzen et al., 2014).



1.2. mRNA stability

mRNA turnover within the cell is another central axis for post-transcriptional regulation of gene expression (Harigaya & Parker, 2016). Mechanisms leading to mRNA decay are plentiful but use similar machinery to achieve degradation. Decapping complexes act on the 5' end of transcripts to remove 5' cap structures that protect against rapid exonucleolytic degradation. Enzymatic activity of Dcp2, a component of the complex, is responsible for removal by hydrolysing the m⁷G cap generating a 5' monophosphate mRNA and m⁷GDP (Wang et al., 2002; Wurm & Sprangers, 2019). Once removed, exonucleases such as Xrn1 proceed with degradation (Łabno et al., 2016). mRNA decay can also occur in the 3'-5' direction by exoribonucleolytic exosome complexes (Januszyk & Lima, 2014; Kilchert et al., 2016). These are involved in general 3' processing of stable RNA and aid in the removal of aberrant mRNA achieved by several co-factors. Enzymes essential for the latter are from three enzyme classes and catalyse: processive hydrolytic RNA decay (Rrp44), distributive hydrolytic RNA decay (Rrp6) and processive phosphorolytic exoribonuclease activity (PNPase) (Januszyk & Lima, 2014). P-bodies are also used to 'store' mRNAs and primarily consist of translationally repressed mRNAs and proteins related to mRNA decay (Luo et al., 2018).

The nonsense mediated decay (NMD) pathway checks for premature stop codons which can arise due to issues such as improper splicing following transcription and mutations (Brognia & Wen, 2009; Schneider-Poetsch et al., 2010). Once erroneous stop codons are identified by NMD proteins, decapping processes are triggered alongside removal of the poly(A) tail and mRNA decay proceeds. The loss of the poly(A) tail heavily impacts mRNA stability, which as mentioned in previous sections, assists in creating the closed loop mRNA model which may aid in prevention of both decapping and exosome complex activity (Mangus et al., 2003; Muhlrads et al., 1994; Passmore & Collier, 2022). Some elements within the mRNA sequence can promote this. Arginine and uridine rich regions found in the 3' UTRs of mRNA and can be bound by multiple proteins, some of which stimulate poly(A) tail removal such as AUF1 (Gratacós & Brewer, 2010; Ross, 1995). Small interfering RNA are thought to be targeted by the innate immune response to defend against viral RNA and are also used experimentally to silence specific genes (Kaushal, 2023). They achieve this by forming RNA induced silencing complexes over mRNA areas complementary to siRNA sequence. exonuclease component of complex cleaves mRNA rendering it functionless (Pratt & MacRae, 2009).

Single stranded mRNA can also form secondary structures via complimentary base pair interactions between its own nucleotides. Examples of such structures include stem loops (similar to tRNA arm loops) and pseudoknots. Their affects through elongation stalling can impact mRNA stability but they also have a direct if mRNA becomes too tangled for translation to occur in any capacity (Namy et al., 2006; Tholstrup et al., 2012).

It has also been shown that translation strongly affects mRNA stability in a codon dependent manner; not just based on nucleotide sequence (Hanson & Collier, 2017). In a study where optimal codons within a transcript were substituted with synonymous non-optimal codons and vice versa, the authors observed that these changes impacted mRNA dramatically, with non-optimal codon sequences leading to destabilised transcripts. The elongation phase was attributed to play a role in this result (Presnyak et al., 2015). This tendency for some codons to promote mRNA stability and others to lead to destabilisation has been observed in multiple other studies and there is a suggestion that tRNA levels also contribute to this, although it has been observed that these factors have less impact on mRNA stability when global translational rates are reduced (Hanson & Collier, 2017; Narula et al., 2019; Wu & Bazzini, 2018). One reported study used viral infection to instigate translation reduction/attenuation and noted that codon usage had less impact on mRNA stability, with optimal and non-optimal codon transcripts showing similar degradation rates (Wu et al., 2019). Ribosomal occupancy has been shown to improve stability of active mRNA transcripts, presumably by protecting the mRNA from decay. The accumulation of these results heavily implicates elongation as a key factor when considering mRNA stability. Some initiation factors have also been attributed to blocking of the decapping processing. eIF4E and eIF4G have been shown to block decapping and PABP may also assist in blocking the exosome complex (Gratacós & Brewer, 2010).

1.3. Impacts on Elongation Rate and its Control on Protein Production and its Regulation

mRNA concentration is a central regulator for protein synthesis and can be affected by transcription rates of individual transcripts and their stability within the cytosol (Elkon et al., 2010; Herzog et al., 2017; Ross, 1995). During disease, or under cell stress, predominantly post-transcriptional programmes are initiated to orchestrate an appropriate cellular response through the regulation and control of protein synthesis (Bastide et al., 2017; Spriggs et al., 2010). Traditionally, much of translational control was thought to occur during the initiation stage as described in section 1.1, but recent studies suggest that the elongation stage is also a central regulatory node. It has been shown that message specific, elongation control is central to

disease mechanisms in both tumorigenesis and neurogenerative disorders (Bastide et al., 2017; Peretti et al., 2015; Spriggs et al., 2010). However, it is unclear to what extent factors such as cognate tRNA abundance and modifications impact this regulation. Ribosomal stalling is believed to play a major role by tuning translation rates due to, and resolved by, multiple factors.

1.3.1 Codon Usage

Codon bias describes the unequal use of synonymous codons used in an organism's protein coding genes and how this can relate to the concentration of tRNAs in the pool within the host's cells. Thereby gene copy number and the frequency of codons used within the genome emerged as a commonly used process for optimising codons in a transcript to increase production rates of biotherapeutic recombinant proteins. This is a well-documented process.

Codon usage has been shown to have a notable effect on the rate of elongation. Decoding speed of each codon in the ORF correlates with the frequency with which it is used within the translome and cognate tRNA availability (Renana & Tamir, 2014; Sharp & Li, 1987). Codons whose cognate tRNA are lowly expressed within the population of tRNA typically show slower translational elongation rates under homeostatic conditions (Plotkin & Kudla, 2011; Quax et al., 2015). This will cause ribosomal stalling at specific points along the transcript whilst the ribosome awaits cognate tRNA presentation (Figure 1.9a). The inclusion of these naturally occurring slow regions may assist in proper co-translational protein folding of certain regions of polypeptides. Issues from poor aminoacylation of tRNA have the same effect whereby the ribosome stalls waiting for peptidyl transfer of nascent polypeptide from the P site tRNA to A site tRNA (Figure 1.9b) (Elf et al., 2003).

Another odd example for codon usage impacting elongation was highlighted in a screening study (Gamble et al., 2016). A *GFP* transcript containing randomised codons were screened and multiple troublesome codon pairs were highlighted. Some of these were attributed to inefficient wobble decoding or rarer cognate tRNA abundance, but some were thought to be solely caused by codon order (Figure 1.9c). This implicated interactions between ribosomal sites and further complexities brought about by codon usage. This study highlighted CUC-CCG (Leu-Pro) and CGA-CCG (Arg-Pro) pairs to be particularly inhibitory to

elongation, both involving proline which by itself provides additional ribosomal stalling tendencies (Gamble et al., 2016).

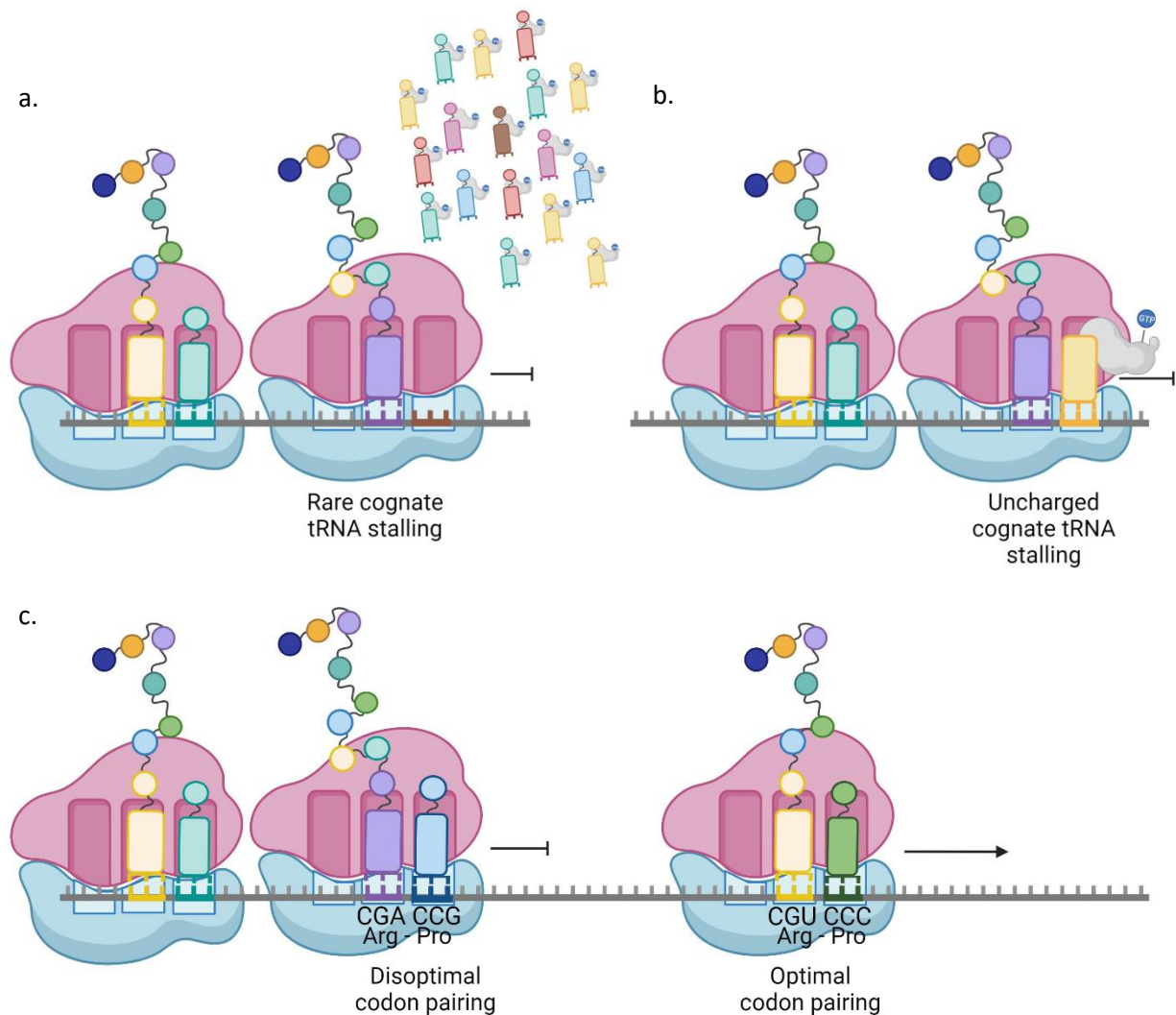
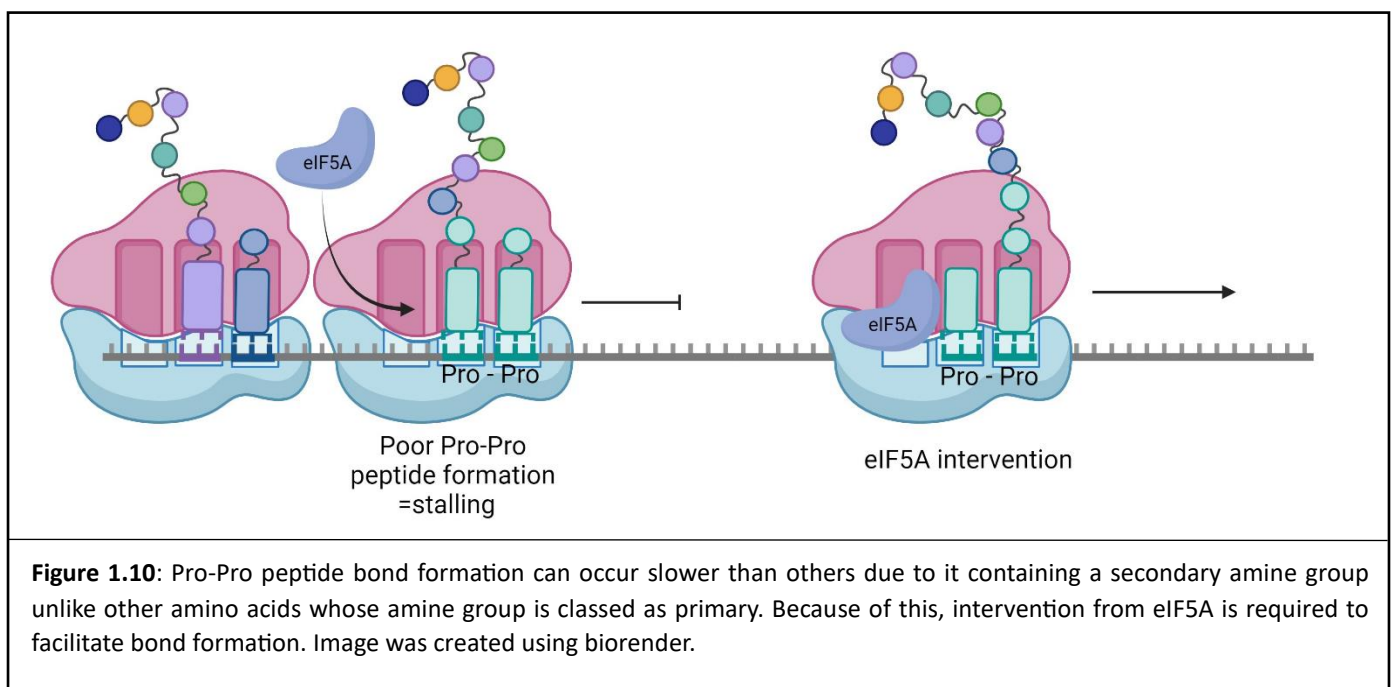


Figure 1.9: Different forms of ribosome stalling brought up through codon usage within mRNA. a) ribosomes can stall whilst waiting for a cognate tRNA to enter the ribosomal A site. This can be prolonged if the cognate tRNA is less abundant in the cell's pool of tRNA. b) Poor aminoacylation can also cause stalling whereby the cognate tRNA enters the A site, but peptide bond synthesis cannot occur as this tRNA lacks its respective amino acid. c) some codon pair combinations can also cause stalling with CGA-CCG suboptimal examples for Arg-Pro whilst CGU-CCC shows improved elongation rates. Image was created using biorender.

1.3.2 Inefficient Peptide Formation

Cognate tRNA abundance does not appear to be the only factor that can impact elongation rates. Unlike the majority of catalytic constructs that excel at facilitating a single reaction between the same substrates,

ribosomes are responsible for generating peptide bonds between any two of the 20 amino acids resulting in 400 possible combinations. Though all are achieved, disparities in effectiveness are observed. Proline is notable due its amine group being contained with a unique ring feature. This downgrades the amine group from primary to secondary and greatly reduces its ability to carry out the nucleophilic attack on the P site peptidyl-tRNA. This issue only compounds with each subsequent consecutive proline inclusion as Pro-Pro peptide bonds are slower than X-Pro formations (Pavlov et al., 2009; Wohlgemuth et al., 2008). This affect can also be brought about by inhibitory conformations of the nascent polypeptide within the exit tunnel (Lucent et al., 2010; Wohlgemuth et al., 2008). Fortunately for all mRNA with proline sequences, eIF5A can alleviate this form of ribosomal stalling (Figure 1.10) (Gutierrez et al., 2013). As previously stated, there is evidence to suggest that a prolonged empty E site acts an indicator of ribosomal stalling, and eIF5A enters this site to assist (Greggio et al., 2009; Pelechano & Alepuz, 2017; Schuller et al., 2017). eIF5A contains a proteome wide unique post-translational lysine modification, with the addition of the 4-aminobutyl moiety from the polyamine spermidine to create hypusine (M. H. Park et al., 1981, 2010). This feature has been shown to be vital in eIF5As ability to interact with ribosomal machinery and impact elongation. Studies where this amino acid is mutated abolished any evidence of interaction (Jao & Chen, 2006). It has also been shown that rather than catalysing peptide formation directly, hypusine involvement focusses on stabilising the P site tRNA through 3' end interactions which may aid in successful nucleophilic attack of substrate in A site (Schmidt et al., 2016). It is likely that this same mechanism aids eRF1 in termination and has also been shown to assist multiple amino acid combinations, not only proline centric (Pelechano & Alepuz, 2017).



1.3.3 mRNA Secondary Structures

Lastly, secondary structures along the length of mRNA can result in inhibited elongation, and initiation. Stem loop formation and more complex structures such as pseudoknots occur by mRNA folding, allowing for secondary and potentially tertiary nucleotide interactions. The longer the length of these structures, the more stable they tend to be and thus, require more energy for the ribosome to overcome (Kozak, 1986). Some of these weaker structures will resolve in a matter of time but some require ribosomal rescue, aborting translation altogether (Schuller & Green, 2018). In some instances, these structures bring about programmed ribosomal frameshift whereby the ribosome shifts a single nucleotide backwards into an overlapping reading frame (Gurvich et al., 2005). This alleviates the strength of the 3' obstruction allowing for elongation to commence. This is only achievable due to another element present in the mRNA sequence immediately before the obstruction coined as a 'slippery sequence' (Figure 1.11) (Temperley et al., 2010). This form of elongation control is predominantly seen in viral pathogens maximising their small genome (Dinman, 2012; Jacks et al., 1988; Wilson et al., 1988).

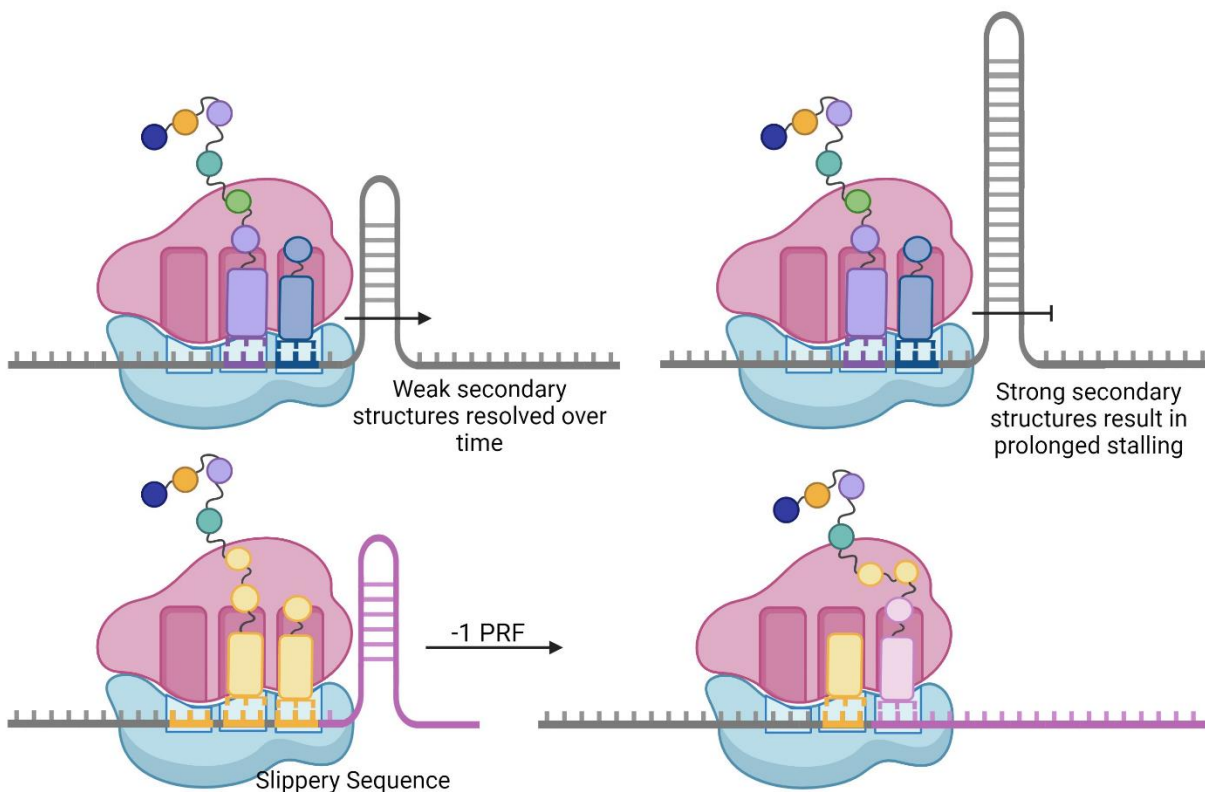


Figure 1.11: Ribosome stalling caused by mRNA secondary structures. a) Secondary structures that have weaker binding strength can pause ribosomes temporarily and are typical resolved without any further intervention required. More elaborate secondary structures including pseudoknots can cause indefinite stalling whereby rescue machinery is required to free the ribosome, abandoning the translation of the mRNA. b) Programmed ribosomal frameshift (PRF) can resolve secondary structures stalling the ribosome whereby the ribosome moves back one nucleotide in length, adjusting the reading frame in the process. This can only be achieved when the secondary structure is preceded by a slippery sequence. Image was created using biorender.

1.4. Transcript Specific Evasion of Globally Reduced Translation Rates

The mammalian cells cold shock response brought about by hypothermic conditions (around 32°C), is highly relevant to this study and causes many controlling mechanisms to be activated which focus on both transcriptional and post-transcriptional control axis to enact genome wide changes (Knight et al., 2015). Some of these changes have been shown to be neuroprotective in nature (Bastide et al., 2017; Peretti et al., 2015). In two HEK293 studies, the cells translatoome was heavily impacted by being cultured at 32°C resulting in global reductions in protein synthesis. It was found that both eIF2 α and eEF2 phosphorylation levels increased, likely a driver for this global reduction (Knight et al., 2015). Through the use of polysome profiling and computational models, subsets of mRNA were identified that were able to evade global reduction rates including RTN3 and noggin (Bastide et al., 2017). A 5' sequence in the open reading frame of both mRNAs encoding these proteins contained codons cognate to less abundant tRNAs and were highlighted as regions that could escape repression of elongation. The increased Rtn3 protein production was also driven by neuroprotective rbm3 protein binding to RTN3-mRNA, thereby mediating rbm3 induced neuroprotection (Bastide et al., 2017). Other studies have shown that cold-inducible protein (CIRP) is another well-known protein to be upregulated during the cold shock response and is thought to act as an RNA chaperone (Al-Fageeh & Smales, 2009; Zhong & Huang, 2017).

As mentioned above, phosphorylation of eEF2 by Ca²⁺ activated protein kinase eEF2K inhibits translation potentially by impairing binding affinity to the ribosome (Carlberg et al., 1990). Interesting however, the effect is reversed for specific transcripts in some tissue types. In neurons phosphorylation of eEF2 has been linked to enhanced localised translation of activity-regulated cytoskeleton-associated protein Arc/Arg3.1 (S. Park et al., 2008). Another example of increased expression in the presence of global repression is the study using cycloheximide- a translation inhibitor. A subset of mRNA showed enhance expression despite reductions globally (Walden & Thach, 1986). It is possible that general inhibition enhances translation of normally uncompetitive mRNA by freeing up machinery for their translation (Gebauer et al., 2012).

1.5. The Protein Synthesis of Secretory Pathway Proteins

As mentioned above, translation is the general process by which all proteins are synthesised, but variations and localisation does occur for many proteins to maintain order and organisation within the cell. This is the case for all proteins destined to be transported out of the expressing cell, often co-translationally, or

those that are destined for the ER and other organelles where proteins are directed to from the ER. The nascent polypeptide in these cases begin their synthesis within the cytosol much like the majority, but are quickly transported to the rough endoplasmic reticulum to complete their generation and for further processing. This can be undertaken post-translationally, but for most proteins is undertaken co-translationally.

mRNA transcripts are freely dispersed within the cytosol awaiting translation initiation as described in section 1.1.1 and once the ribosomal machinery has converged on the start codon of the mRNA, elongation commences. For secreted (and membrane) proteins, the first sequence to be translated typically relates to a feature distinct from the final protein's function. This sequence usually codes for 10-30 amino acids and functions to signal that the nascent peptide is bound for the secretory pathway (von Heijne, 1985). The signal recognition particle (SRP), another cytosolic protein, recognises the N-terminal signal peptide and arrests elongation requiring activity of the SRP9/14 subunit to do so (Mason et al., 2000; Walter & Blobel, 1981). Elongation arrest is thought to increase the timeframe for SRP-nascent peptide-ribosome complex to migrate to the rough ER membrane (Mason et al., 2000; Walter & Blobel, 1981). SRP is recognised by the α subunit of the docking protein: signal recognition particle receptor (SRPR) (Akopian et al., 2013; Legate, 2000). This interaction allows delivery of the nascent polypeptide and ribosome to the Sec61 translocon, another ER membrane bound protein, which binds to the ribosome and facilitates translocation of the polypeptide into rough ER lumen (Linxweiler et al., 2017). At this point, SRP dissociates from the complex coupled with GTP hydrolysis and translation resumes (Legate, 2000). The elongating polypeptide is funnelled into the rough ER lumen through the translocon where a signal peptidase cleaves the signal peptide (Linxweiler et al., 2017). The ribosome subunits are recycled and migrate away from the ER to complete further rounds of translation.

1.6. Production of Recombinant Proteins in Mammalian Cell Expression Hosts

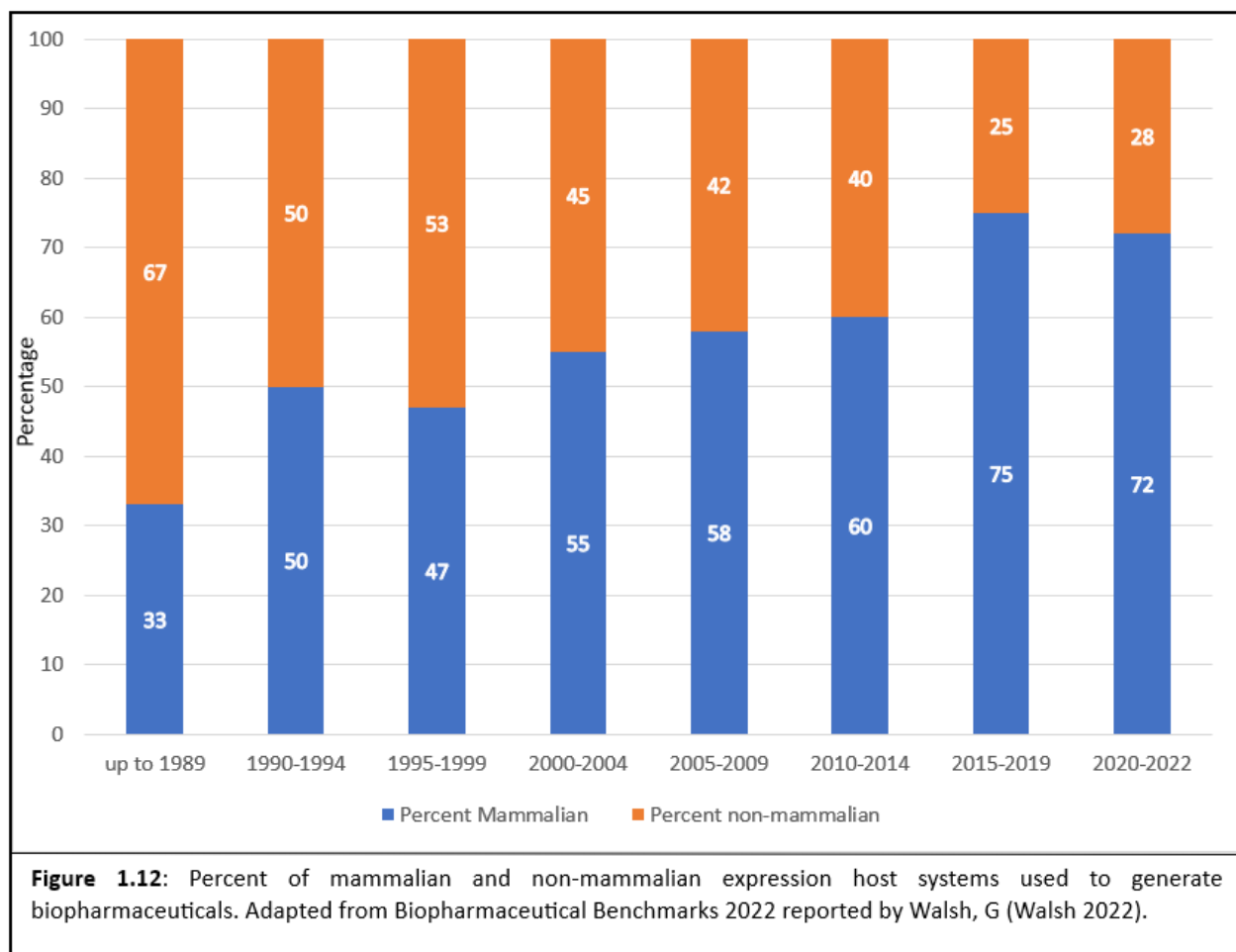
The use of mammalian cell expression host systems as a means of producing therapeutic recombinant proteins has been steadily increasing for the past 30 years and are desirable for multiple reasons (Figure 1.12) (Walsh and Walsh, 2022). Other expression systems fail to generate post-translational modifications such as glycosylation and protein folding in a manner comparable with human characteristics. This poses increased risk of eliciting an immune response towards therapeutic drugs the body would identify as

foreign as well as other stability issues. Mammalian cell expression systems mitigate these issues generating human-like alterations (Barnes et al., 2003; Jefferis, 2009). Human embryonic kidney (HEK) and Chinese hamster ovary (CHO) cells are predominantly used, but several other expression systems have been used or explored including HeLa and HepG2. Table 1.1 reports the top selling recombinant biotherapeutics in 2021 and demonstrates the scale of the biotherapeutic industry (Walsh, 2022).

Nonetheless, improvements in protein production are always desirable. Sequence optimisation techniques have been widely implemented to tailor foreign DNA encoding recombinant proteins to the tRNA abundance levels and gene copy number of codons to specific expression hosts to maximise cell exploitation (Mauro, 2018). These typically focus on the entirety of the gene rather than a selective section so it is unclear what impact modifying specific sections of the open reading frame codon usage would have on protein production. Temperature shift culturing approaches are also frequent, especially with CHO cell expression systems (McHugh et al., 2020; Xu et al., 2019). This involves reducing the bioreactor incubation temperature to mildly hypothermic conditions (Al-Fageeh & Smales, 2006). This can arrest cell proliferation and lead to sustained high viable cell densities over longer incubation periods, if timed correctly, ultimately leading to increased product titre (Ahn et al., 2008; Kaufmann et al., 1999). It has also been shown to reduce aggregation and improve ease of the downstream purification process for some monoclonal antibodies (Goey et al., 2017; Tharmalingam et al., 2008).

Table 1.1 Total reported 2021 biopharmaceuticals global sales values. Adapted from Biopharmaceutical Benchmarks 2022 reported by Walsh, G (Walsh 2022)

Biopharmaceutical Category	Reported sales value (\$ billions)
Originator recombinant proteins: mABs	217.3
Originator recombinant proteins: non-mABs	53.6
Covid vaccines (Comirnaty and Spikevax)	54.5
Biosimilars	11.1
Nucleic acid and engineered cell based	6.8
Total Value	343.3



1.6.1. Cloning Methods

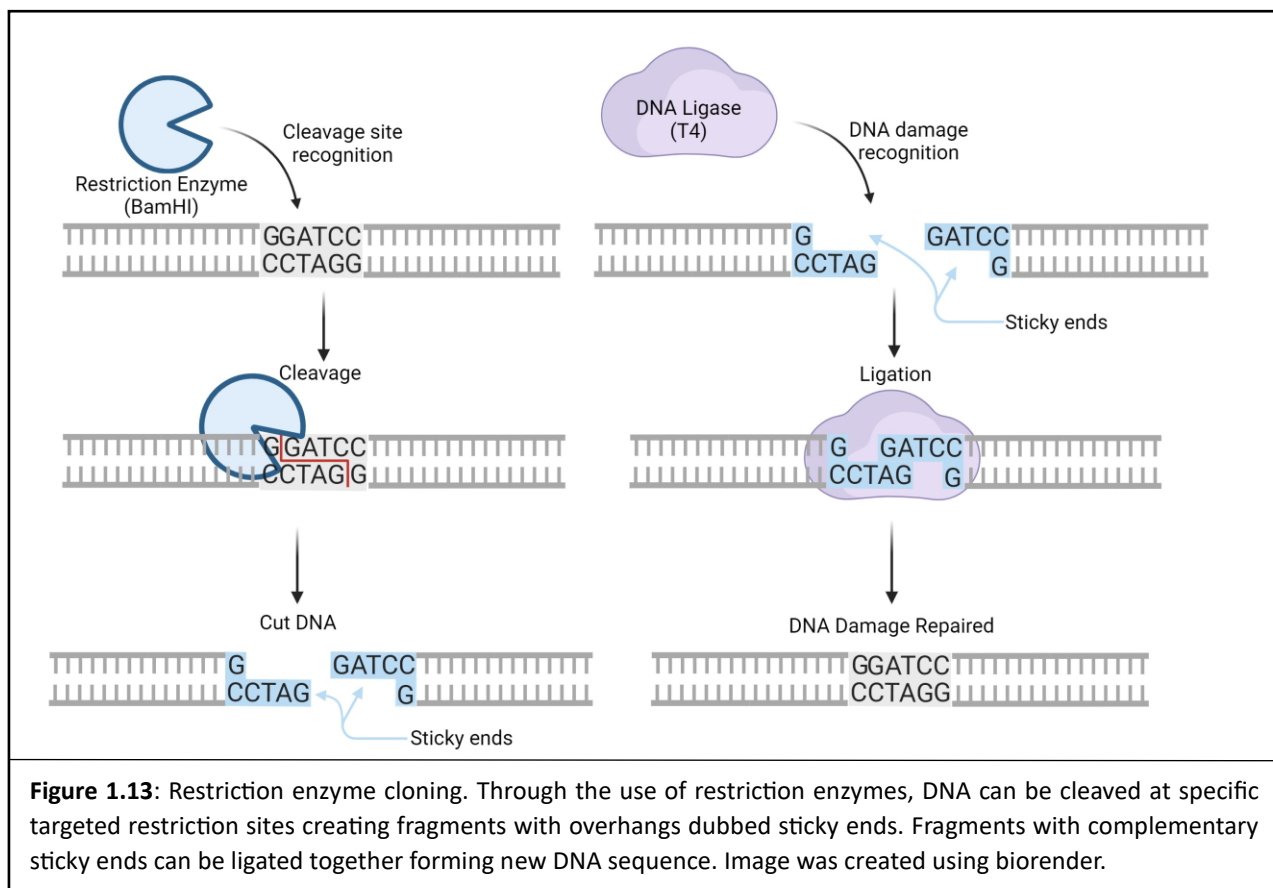
In order to utilise expression hosts of any origin, genetic information for a respective biotherapeutic must be introduced into the host cell. Transfection is the process by which this is typically achieved and can be accomplished stably or transiently; either being incorporated directly into the host cells genome or, usually, delivered transiently in a plasmid form to the nucleus for transcript but where integration does not occur and thus expression is lost over time. Large scale biotherapeutic production typically deploys stable transfection as this creates cell lines that can express the recombinant gene over many cell divisions. This is in direct contrast to cells containing transiently transfected genetic material which is diluted within

the population with each round of cell division, losing expression effectiveness over time. Transient transfection still has its uses in short term investigation of gene expression and is a much faster process than generating stably transfected cell lines, rapidly expressing recombinant DNA. Regardless of method, introduction of genetic material is essential and their preparation relies upon molecular cloning techniques.

In the early 1950s, Mary Human, Salvador Luria, Joe Bertani and Jean Weigle discovered that certain bacterial strains were more resistant to bacteriophage infection than others (Luria & Human, 1952). This observation was built upon in the 1960s when Werner Arber observed that bacteriophage DNA was heavily degraded after invading these resistant strains (Roberts, 2005). He suggested that these strains must be producing an enzyme capable of degrading DNA and was proven correct in the 1970s when Hamilton Smith successfully purified a restriction enzyme (Smith & Welcox, 1970). Kathleen Danna and Daniel Nathans demonstrated for the first time that this restriction enzyme, dubbed endonuclease R, could produce specific DNA fragments from simian virus 40 (Danna & Nathans, 1971). The restriction enzyme digests we know today stemmed from this research and are now an indispensable tool for DNA manipulation. It is through this tool that most genes are introduced into plasmid vectors in preparation for transfection and expression. Restriction enzyme cloning is achieved by exploiting the site-specific nature of restriction enzymes. Each only identifies and cuts specific nucleotide sequences known as restriction sites, allowing for accurate modifications of genes and their incorporation into plasmids to be achieved (Figure 1.13). Treating a plasmid and desirable insert with the same restriction enzymes creates complimentary sticky ends which can be ligated through the use of a ligase enzyme. Blunt end restriction enzymes can also be used in a similar manner.

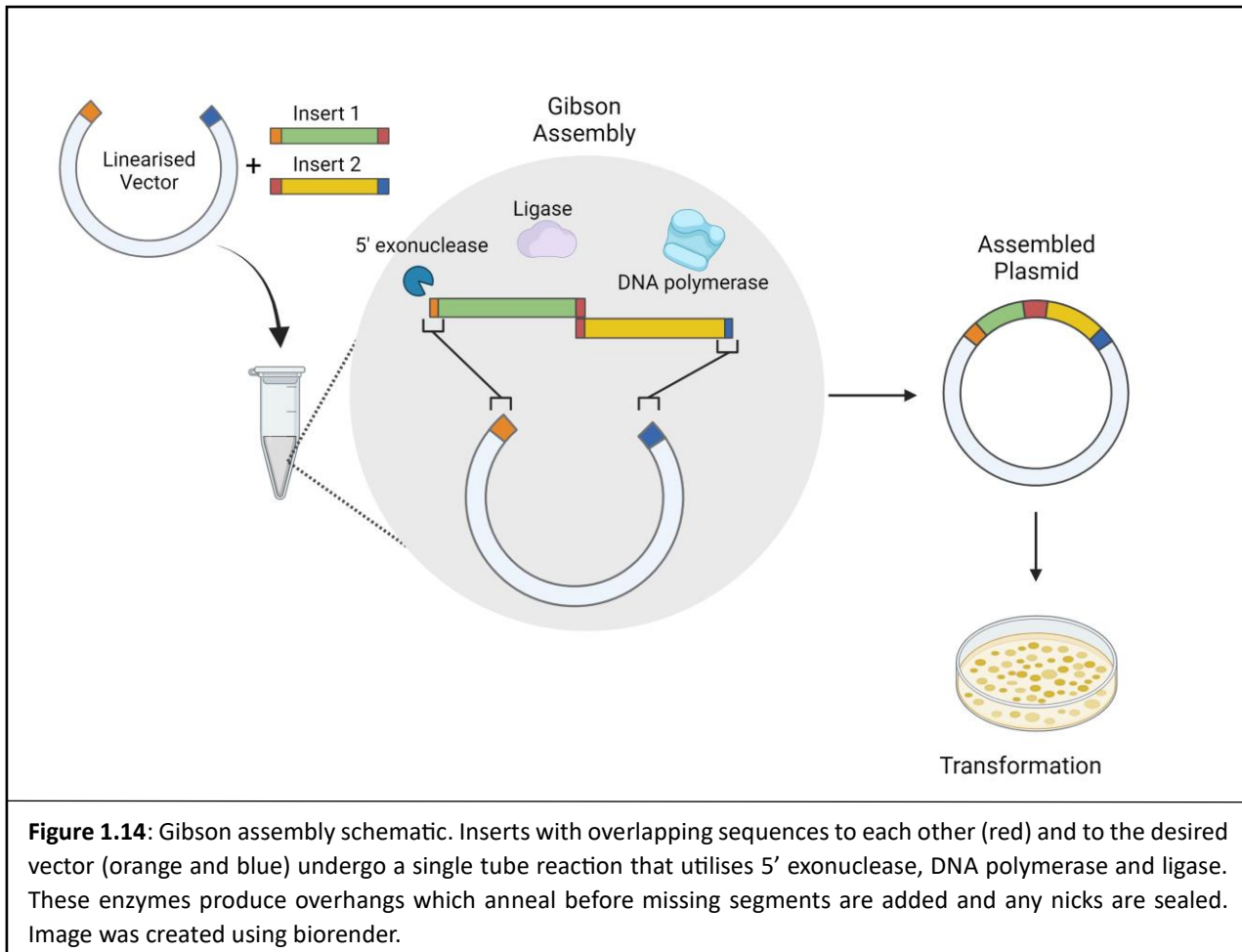
Recombinant plasmid DNA often contains additional genes and motifs to facilitate expression or the transfection process. An expression cassette is the site for gene inclusion. Promoters such as CMV, SV40 and T7 promoters, are located upstream of the expression cassette and describes the combination of the RNA polymerase binding site and response elements that allow for increased expression of the included gene. Within the cassette are usually multiple restriction sites for ease of gene incorporation. A poly(A) tail is also often found at the 3' end of the binding cassette. Plasmids also often contain genes providing resistance to various selection pressures such as the ones used in this study; hygromycin, ampicillin and

kanamycin resistant genes. Applying these selection pressures to transfect or transformed cultures allows for effective screening for successfully introduced genetic information.



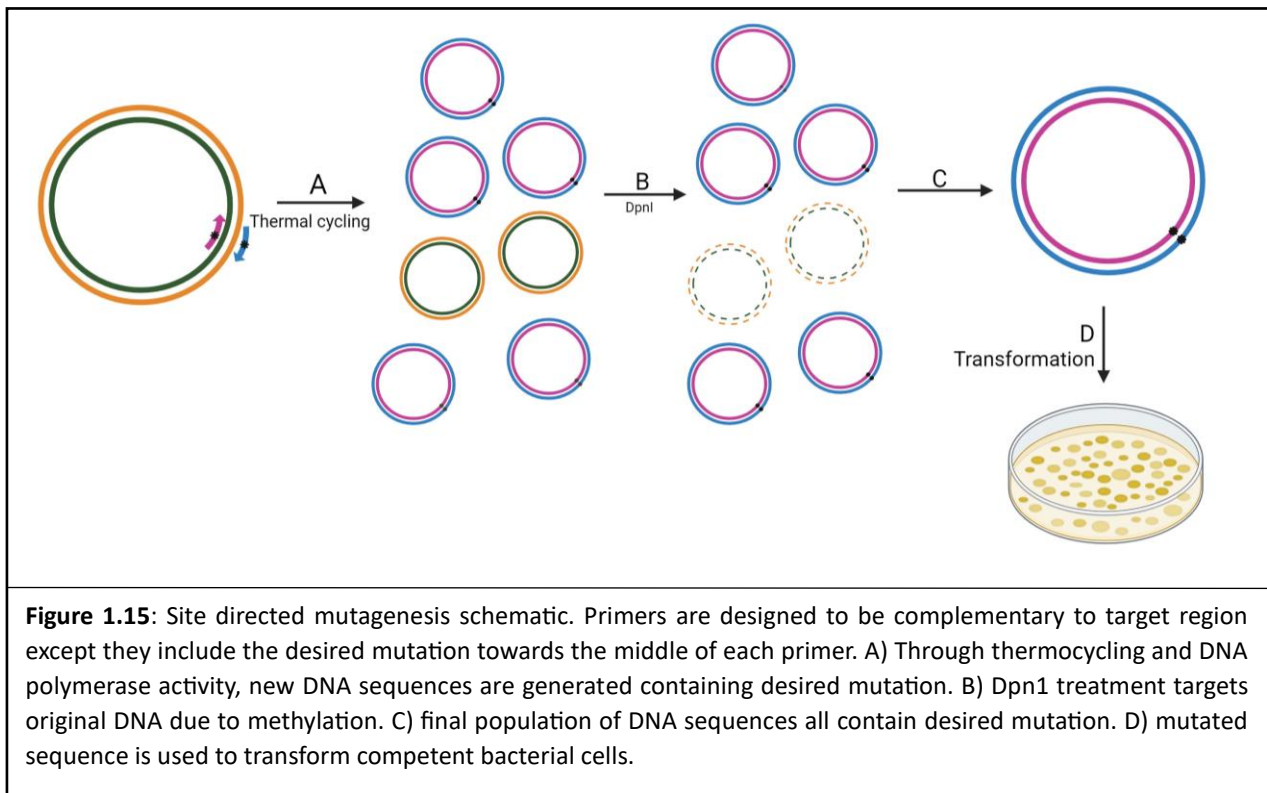
1.6.1.1 Gibson Assembly Cloning

Unfortunately, suitable restriction sites are not always available and, in these cases, other molecular cloning techniques are needed. Daniel Gibson described a robust alternative utilising an exonuclease-based method of DNA assembly (Gibson et al., 2009, 2010). The process involves three enzymatic activities: 5' exonuclease that generates long overhangs, a polymerase to extend annealed strands, and a DNA ligase to seal remaining nicks post polymerase activity. The process is schematically described in Figure 1.14 below.



1.6.1.2 Site Directed Mutagenesis

Site directed mutagenesis is another powerful cloning tool and allows for highly targeted alterations to be made to a DNA sequence (DeCero et al., 2020). These mutations include substitutions, insertions or deletion of nucleotides or a string of nucleotides by including the desired mutation(s) within primers otherwise complimentary to the target region (Figure 15). This has a wide range of uses and can be implemented to cause changes as silent nucleotide mutations (e.g., change one codon for an amino acid for another codon for the same amino acid) or to have an impact at the amino acid level. It can also amend sequences if errors have been incorporated in the gene cloning process. Through thermocycling and DNA polymerase activity, copies of a gene with the desired mutation are generated, though unmutated template DNA will also be present. This is typically digested by treatment with restriction enzymes. In the case of Dpn1, template DNA is selected due to methylation, whilst newly synthesised strands lack methylation (Mierzejewska et al., 2014).



1.7. An Elongation Model of Decoding Speeds and Protein Synthesis Rates of mRNA Transcripts

Although there are multiple studies linking codon usage and its effects on protein production, very few translational control models are known linking the two. A model created by University of Kent collaborations has attempted to fill this hole in our knowledge by modelling a novel translational control model that responds to the rate of ribosomal translocation across an mRNA transcript (Chu et al., 2012). This model does not take into account transcript-transcript competition for ribosomes but does allow for system-wide and dynamic representation of translation and works by taking into account ribosomal collisions, tRNA abundance, charged/uncharged tRNA states and other factors outlined in table 1.2 (Chu et al., 2012; Chu, Thompson, et al., 2014). Initial applications of this model highlighted that the initial 5' sequence of the ORF immediately following the start AUG codon have a substantial impact on overall protein production rates and may be integral to elongational control (Chu et al., 2012; Mead et al., 2014). This is the same model that identified the aforementioned *RTN3* and *NOG* 5' ORF sequences that may assist in evading translation repression in response to hypothermic conditions (Bastide et al., 2017). It was identified that optimal codon usage within this 5' ORF sequence allowed for effective ribosome clearing and thus, liberated the start codon for additional ribosomes to begin initiation (Chu et al., 2012; Chu,

Kazana, et al., 2014; Chu, Thompson, et al., 2014). This model is not only capable of identifying regions of elongation control towards the 5' end of the mRNA, but also identifying predicted slow decoding sequences throughout the entire transcript. As such, multiple mRNA sequences were analysed by colleagues at Kent University and some of note were chosen as model transcripts for the further studies carried out and described within this thesis. These model transcripts each displayed potentially slowly decoded regions within their sequence at differing locations.

Table 1.2: Outline of discrete events driving elongation simulation that are scheduled using Gillespie's algorithm. Table adapted from information in 'A novel and versatile computational tool to model translation' by von der Haar et al (Chu et al., 2012).

	Simulation Drivers
1	Binding of free ribosomes to an initial binding site on a transcript.
2	Binding of ribosomes located on an initial binding site to the first codon (AUG site).
3	Binding of near/non-cognate aa-tRNAs to unoccupied bound ribosomes.
4	Unbinding of near/non-cognate tRNAs from a bound ribosome.
5	Translocation of ribosomes to the next codon.
6	Aminoacylation of uncharged tRNA.
7	Re-initiation/dissociation of ribosome once the ribosome has reached the end of the ORF.

1.8. Model Transcripts Investigated in this Study

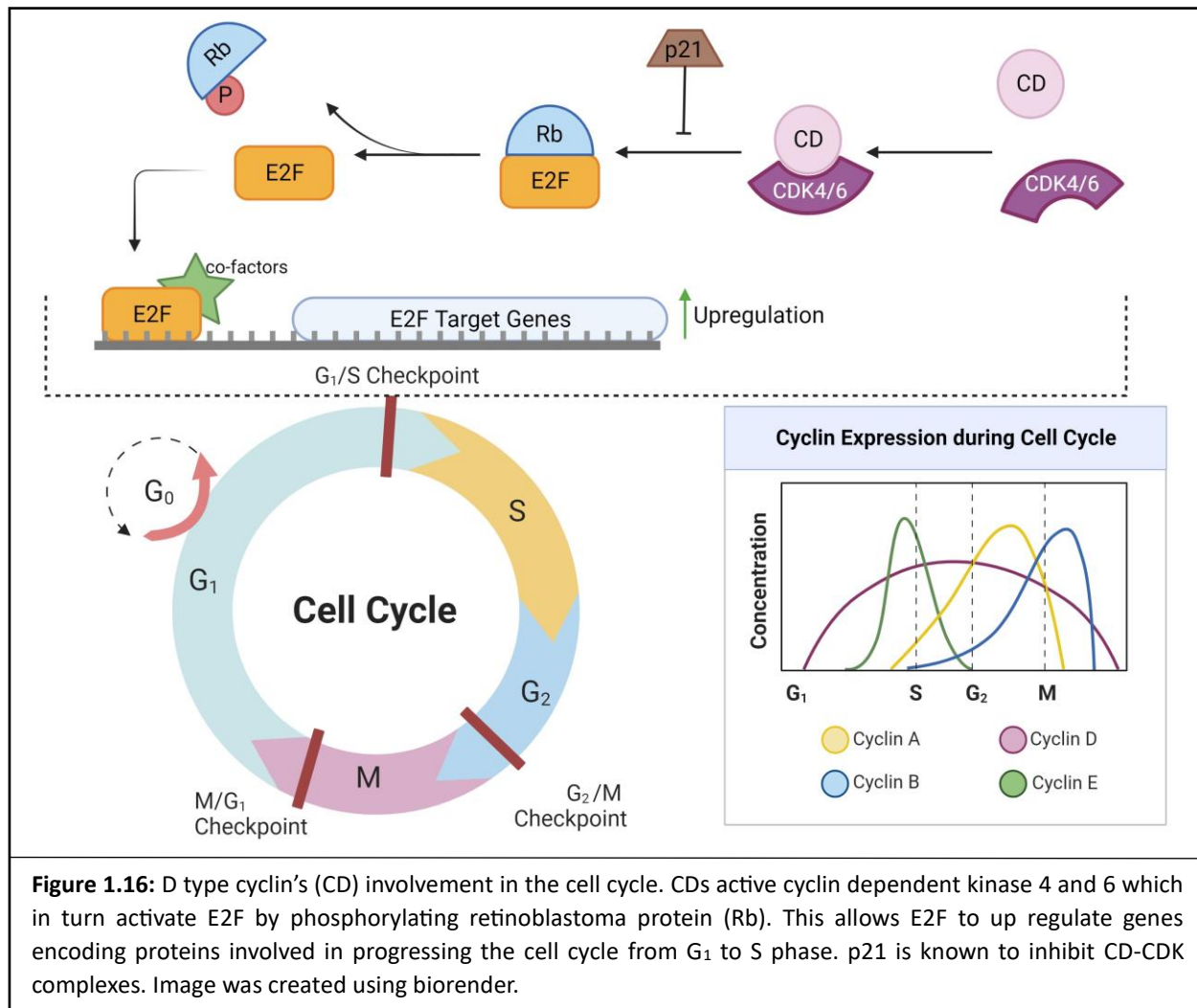
Using predicted translation rates from the elongation model explained in section 1.7, mRNA transcripts for three proteins were identified to contain areas relevant to the investigation of mRNA sequence on elongation control. These model proteins are used throughout this study to explore the effects synonymous codon usage has on elongation rates.

1.8.1 Cyclin D3 and its Role in Cell Cycle Control

Cyclin D3 (CCND3) is a member of the cyclin family of proteins whose role is to assist in the regulation of the mammalian cell cycle. D type cyclins (CD) are predominantly involved in progressing the cell cycle by promoting G1/S transition. They act as regulators for cyclin dependent kinase 4 and 6 (CDK4, CDK6), binding to them creating an active complex (Figure 1.16) (Sherr, 1995; Sherr & Roberts, 2004). These complexes are thought to enable progression functions through catalysing phosphorylation of retinoblastoma tumour suppressor (Rb). Rb binds to E2F family members which in turn, are responsible for recruiting transcriptional co-factors to gene promoters upregulating specific genes important to G1/S phase progression (Vélez-Cruz & Johnson, 2017). Rb interactions with E2F block this recruitment process thereby repressing transcription and preventing G1/S progression (Vélez-Cruz & Johnson, 2017). Once phosphorylated by CD-CDK4/6 complexes, Rb loses its affinity to bind E2F and dissociates. This alleviates downregulation of targeted genes and thus, promotes the transition from G1 to S phase (Johnson, 1998).

Cyclin-dependent kinase inhibitors such as p21 and p27 work antagonistically against this process. Promoted by p53, these inhibitors bind and inhibit CD-CDK complexes thereby causing cell cycle arrest (Engeland, 2022). This has been seen in response to DNA damage and is thought to safeguard cells by arresting cell division until DNA repair can be achieved (Engeland, 2022). This is not the only form of regulation surrounding cyclin activity. Throughout the cell cycle, the concentration of cyclin proteins changes with the majority fluctuating at transitional states (Hochegger et al., 2008). However, D type cyclin production appears cyclical, steadily rising and falling across the cell cycle. These variations in cyclin levels are due to fluctuating gene expression and ubiquitin mediated proteasome degradation and without this pattern, control over the cell cycle would be lost (Dang et al., 2021; Hu et al., 2002).

The transcript elongation translation model described above identified a region of elongation control towards the middle of the *CCND3* ORF which is notable as the majority of identified regions are at the 5' end of genes immediately following the start codon. This may prove useful in determining the impact codon usage has on elongation rates without influencing initiation and therefore this transcript and region of interest was selected as one for further study.



1.8.2 Noggin

Noggin (NOG) is a secreted polypeptide that is important for regulating signalling pathways at the developmental stages of early life, particularly bone development, but it has also been shown to assist neural tissue and muscle development (Costamagna et al., 2016; Zimmerman et al., 1996). In the case of bone growth, noggin acts antagonistically against bone morphogenic proteins (BMPs), signalling to proteins from the transforming growth factor- β superfamily (TGF- β) (Brunet et al., 1998). BMPs cause a phosphorylation cascade within cells leading to activation of transcription factors by binding to cell surface receptors which allow for cell differentiation during embryogenesis (Zakin & De Robertis, 2010). This can cause the transformation of stem cells into immature osteoblasts (Cheng et al., 2003). Noggin regulates this process and its absence has dire consequences on joint development with *NOG* mutations leading to

conditions such as proximal symphalangism (Marcelino et al., 2001). Its relevance to this study is that a region highlighted by the elongational control model at the 5' end of its mRNA transcript was identified as a region of elongational control with a stretch of predicted slowly decoded codons. This is of note as this region coincides with its signal sequence marking the nascent peptide for secretion. The transcript was therefore selected as a model transcript to assess the effects codon alterations in this region have on secretion efficiency. This transcript has also been identified in additional studies to evade global translation reductions in response to hypothermic conditions and is relevant to neuroprotective studies (Bastide et al., 2017; Díaz-Moreno et al., 2018).

1.8.3 Green Fluorescent Protein

Green fluorescent protein (GFP) was isolated from jellyfish *Aequorea victoria* in the 1960s (Shimomura et al., 1962), and since then has become another indispensable tool used throughout biological research as a biomarker for gene expression as well as many other applications. GFPs role in jellyfish is to act as an energy-transfer acceptor, transducing blue chemiluminescence from aequorin into green fluorescent light (Prasher et al., 1992). Lab utilised GFP has been engineered to produce several mutants, secretory versions, and brighter fluorescence. Enhanced GFP (eGFP) which includes 64Phe->Leu and 65Ser-> Thr mutations is a commonly used variant due to its increased fluorescence (Cormack et al., 1996; Ou et al., 2014). GFP is often codon optimised based on codon gene copy numbers of the host cells genome (Guohong, 1996; Rouwendal* et al., 1997). GFP was another protein highlighted to have a 5' end sequence relevant to elongation control by the novel elongation control model and as such was another protein incorporated within this study. Due to its highly optimised nature, it was interesting to investigate whether additional improvements to expression could be achieved through codon manipulation.

1.9 Emerging Applications for RNA Based Therapies

Emerging nucleotide base therapeutics have increased in recent years and their potential in combatting disease states is being explored. Gene therapy involving the introduction of functional genes into patients to combat disease states have been slow to reach patients but products have been approved (Walsh & Walsh, 2022). The cost of treatment may be beyond many which in part is due to production costs. mRNA vaccines are also of interest, especially in the advent of the Covid-19 pandemic where mRNA vaccines gained notoriety from large scale roll out of Pfizer and Moderna mRNA vaccines. This was possible in part

due to high yields for in vitro reactions easing the burden of production scalability (Pardi et al., 2018). Recent improvements in this field have led to increased translation, stability, reduced immunogenicity, and improved delivery of in vitro transcribed (IVT) mRNAs (Guan & Rosenecker, 2017; Karikó et al., 2008; Kauffman et al., 2016; Thess et al., 2015). However, protein production from IVT mRNAs remains far from optimal with the potential for enhancement (Pardi et al., 2018). If further codon optimisation of IVT mRNAs and gene therapies can be achieved by utilising elongation model tools, this could aid in furthering improvements seen in expression from both gene therapy and mRNA vaccination.

1.10 Aims of the Project

Codon optimisation is common practice for transcripts that encode for biotherapeutic proteins, tailoring recombinant DNA sequences to the gene copy number of codons within the expression host cell's genome. Here we test a method of codon usage choice based upon another rationale modelled by a novel elongation tool. Whilst doing so, only small regions of the ORFs for 3 different transcripts will be modified, leaving the remaining sequence untouched. This will help elucidate whether these regions do impact elongational control as anticipated and predicted by the elongation control model. Temperature has also shown to be of note both during biotherapeutic production and investigated diseased states such as neurodegeneration. To this end, codon usage effects at varying temperatures was also to be analysed.

In summary, the specific aims of this project were to:

- Assess the impact on protein production of modifying codon usage in the 5' region of the ORF on mRNA transcripts based upon the predicted decoding speeds for these areas of interest by an elongation speed computation model.
- Determine if hypothermic conditions (32°C) alter gene expression profiles from modified transcript variants.
- Determine whether transcript variants with modified regions away from the AUG start codon impacted mRNA translation as determined by monitoring protein production from such transcripts.
- To evaluate the elongation model of von der Haar and colleagues efficacy at identifying regions relevant to elongational control.

Chapter 2

Materials and Methods

General protocols used in the work are outlined in this chapter. Experiment specific details are included within relevant result sections if alterations were made. For the entire duration of this study, all reagents used were of analytical grade or greater.

2.1 Molecular Biology Methods

2.1.1 *E. coli* Culture

The following growth media were used when culturing *E. coli* (DH5 α strain).

Lysogeny Broth (LB)

LB media was prepared containing 10 g tryptone, 5 g yeast, and 10 g sodium chloride (NaCl) per litre of MilliQ water and autoclaved to sterilise before use.

Super Optimal Broth (SOB)

SOB was prepared containing 20 g tryptone, 5 g yeast, 0.5 g NaCl, 10 mL 0.25 M potassium chloride (KCl) per litre of milliQ water. This was then autoclaved to sterilise and 5 mL of 2 M magnesium chloride (MgCl₂) was then added per litre.

Super Optimal Broth with Catabolite Repression (SOC) Media

Filter sterilised 10 mL of 50% glucose solution (w/v) was added to SOB media per litre.

Agar Plates

Agar plates were prepared by adding 15 g of agar per litre of LB media and autoclaving. Once cooled, but prior to setting, either ampicillin or kanamycin antibiotic was added where appropriate. After mixing, medium was poured into dishes and left to finish setting.

Preparation of Ampicillin and Kanamycin Solutions

Both antibiotics were used to select for successfully transformed *E. coli* cells in agar plates and LB media. Ampicillin was used at 100 µg/mL working concentration whilst kanamycin was used at 50 µg/mL.

2.1.2 Generation of Calcium Competent *E. coli* Cells

50 mL of SOB media was inoculated with 1 mL of overnight *E. coli* culture and incubated at 37°C, 200 rpm. Growth was monitored at 600 nm using a spectrophotometer. Once the OD₆₀₀ had reached 0.4, cultures were chilled and centrifuged at 2500 x g for 15 minutes at 4°C (Megafuge 16R, rotor 75003181). The cell pellets were resuspended in 10 mL of filter sterilised 100 mM calcium chloride solution (CaCl₂) and incubated on ice for 30 minutes before a final centrifugation, also at 2500 x g for 15 minutes at 4°C. Cell pellets were then gently resuspended in 4 mL of ice cold 100 mM CaCl₂ with 1 mL 80% (v/v) glycerol. 100 µL per cryotube of the cell solution were aliquoted and flash frozen on dry ice before -80°C long term storage.

2.1.3 RNA extraction form mammalian cells

The original native human transcript sequence for *CCND3* originated from HEK293 cells. RNA was extracted from cells using the commercially available RNeasy kit (Qiagen). Lysates were homogenised using a QIAshredder spin column and manufacturer's protocol was then followed for isolation of total RNA from the cells.

2.1.4 Generation of cDNA from RNA

Isolated RNA was used to generate complementary DNA (cDNA) to allow amplification of *CCND3* gene by PCR. QIAGEN RNeasy Minikit (QIAGEN) was used to extract RNA from confluent HEK293 culture in a 6 well plate, following the manufacturer's instructions. RNase Zap and ethanol were used to clear workspace of any potential RNA contaminants. M-MLV Reverse Transcriptase pack (Promega) was used as well as RNasein Ribonuclease Inhibitor (Promega), Oligo(dt)₁₅ Primer (Promega), RNase-free water and nuclease free water to generate cDNA. 4 µg RNA, 4 µl of Oligo(dt)₁₅ Primer and RNase free water were mixed (15 µL total) and incubated at 70°C for 5 minutes, then transferred to ice for 2 minutes. Whilst on ice, 10 µL MMLV RT 5X buffer, 2.6 10 mM deoxynucleotide triphosphates (dNTPs), 1.4 µL RNasein, 2 µL M-MLV reverse transcriptase, and nuclease-free water (50µ total) were mixed. Solution was then incubated for 1 hour at

42°C and the resultant cDNA was purified using Promega's Wizard® SV gel and PCR clean up system (section 2.18) and stored at -80°C for future use.

2.1.5 Polymerase Chain Reaction (PCR)

Sequence amplification was undertaken using NEB Phusion™ polymerase protocol. *CCND3* genetic sequence was amplified using cDNA generated from HEK293 isolated RNA (section 2.1.4). PCR thermocycling steps are outlined in table 2.1. Primers were designed to amplify target *CCND3* gene, whose sequences can be found in table 2.2, and introduced restriction sites for subsequent cloning into expression vectors. Amplicons were synthesised by mixing 100 ng cDNA, 20 µL 5 X HF buffer, 4 pmol forward primer, 4 pmol reverse primer, 0.2mM dNTPs and 1 µL Phusion™ DNA Polymerase (100 µL total) in 0.2 mL PCR tubes and conditions were applied based on primer melting temperatures and target strand length.

Table 2.1: A breakdown of the PCR amplification process indicating the temperature and time taken at each step. a: Annealing temperature required is determined melting temperatures of primers involved, using the lowest melting temperature of the pair. b: The extension time is determined by the length of sequence destined for amplification and increases by 30 seconds for every 1 kb of amplicon.

Step	Temperature	Time	35 cycles
Initial Denaturation	98°C	30 seconds	
Denaturation	98°C	10 seconds	
Annealing	a (primer T _m) °C	30 seconds	
Extension	72°C	b (30 seconds per kb DNA)	
Final Extension	72°C	7 minutes	
Hold	4°C	∞	

Table 2.2: Primers used throughout this project, including their sequence and their relevance in application.

Application	Primer Name	Sequence 5'-3'
<i>CCND3</i> gene amplification	CCND3 Native Forward	CCTCCTACTTCCAGTGCGTG
	CCND3 Native Reverse	AGACAGGTAGCGATCCAGGT
<i>CCND3</i> ROI sequence confirmation	CCND3 ROI Forward1	GACTGGGAGGTGCTGGTCCTA
	CCND3 ROI Reverse1	GTCCACTTCAGTGCCAGTGATCC
	CCND3 ROI Forward2	GCCAGACCAGCACTCCTACA
	CCND3 ROI Reverse2	GATGGCTGGCAACTAGAAGG
<i>CCND3</i> site directed mutagenesis G->T	SDM GT Forward	GAGCTGGTCTGAGAGGCTTCCCTGAGG
	SDM GT Reverse	CCTCAGGGAAGCCTCTCAGACCAGCTC
<i>CCND3</i> site directed mutagenesis CT deletion	Opt/Fast SDM CT Forward	GCACCGGCTCTCTCTGCCTCGTGACC
	Opt/Fast SDM CT Reverse	GGTCACGAGGCAGAGAGAGCCGGTGC
	Dis SDM CT Forward	CACCGGCTCTCTCTCCCCGCGAT
	Dis SDM CT Reverse	ATCGCGGGGAGAGAGAGCCGGTG
	Mid SDM CT Forward	CACCGGCTCTCTCTGCCGCGTGAC
	Mid SDM CT Reverse	GTCACGCGGCAGAGAGAGCCGGTG
	Swap SDM CT Forward	GCACCGGCTCTCTCTCCCTCGCGATC
	Swap SDM CT Reverse	GATCGCGAGGGAGAGAGAGCCGGTGC
<i>GFP</i> & <i>Noggin</i> ROI sequence confirmation	T7	TAATACGACTCACTATAGGG
	BGHR	TAGAAGGCACAGTCGAGG

2.1.6 Transcript Sequence Design

Sequence modifications described in Chapter 3 utilised an elongation model created at the University of Kent to calculate individual codon decoding times based upon tRNA abundance rather than gene copy number. This tool was used to identify regions within transcripts that displayed potential elongation control characteristics, that is to say, contained a run of codons that were predicted by the model to be decoded slowly due to the low abundance of the tRNAs for these codons. Modified regions of interests were designed and ordered commercially from either Genewiz or GeneArt services.

2.1.7 PCR and Gel Clean-up of PCR Products

PCR reactions and DNA bands in gel slices were purified using commercially available reagents and kits. Promega's Wizard® SV gel and PCR clean up system was used to purify DNA bands from gels and PCR products respectively, following the manufacturer's protocol.

2.1.8 Restriction Enzyme Digests

All restriction enzymes used in the cloning process were FastDigest restriction enzymes obtained from Thermo Scientific. Specific restriction enzymes used in each instance are included and described in the relevant results sections. Unless stated otherwise in specific result sections, a total volume reaction of 20 μL was used, including 2 μL 10x FastDigest green buffer, 1 μg DNA, 1 μL of each relevant restriction enzyme, and nuclease free water to increase total reaction volume to 20 μL . The reaction solutions were incubated at 37 °C for 20 minutes.

2.1.9 Agarose Gel Electrophoresis Analysis of DNA

DNA was analysed by agarose gel electrophoresis to either isolate DNA fragments for purification or for visual confirmation of successful fragment insertion into vector backbone following ligation. This is achieved by passing an electrical current through the gel causing the fragments to migrate towards the positive electrode at different speeds based on the size of the DNA fragment being analysed. 8 well gels and 40 well gels were used at 1% or 2% (w/v) agarose concentration, as appropriate. Agarose was added to TAE buffer to achieve these percentages and heated in a microwave. Once cooled, 3 μL of ethidium bromide was added for 8 well gels and 9 μL was added for 40 well gels, before transferring the solution to a gel cassette to set. A set gel was placed in a gel rig (Scotlab Electrophoresis Rig) and 1 X TAE buffer was added to the rig. DNA samples were loaded alongside DNA ladders of 1 kb or 100 bp size (Promega).

2.1.10 DNA Ligation

T4 ligase (Promega) was used to ligate digested DNA vector backbone with digested inset. These two DNA fragments were previously digested with the same restriction enzymes creating complimentary ends and providing insertion orientation. Ligation reactions consisted of 50 ng of digested backbone and digested

inserts at ratios of 1:0 and 1:1, 1 µL ligase buffer and nuclease free water to bring the solution volume to 10 µL. Reactions were incubated at 4°C overnight as recommended in the manufacturer's instructions.

The formula below was used to calculate the volume of insert to add based on ligation ratios:

$$\text{Vol of Insert (}\mu\text{L)} = \frac{50 \text{ (ng)} * \text{Insert Size (kb)}}{\text{Clean up Conc (ng}\mu\text{L}^{-1}) * \text{Vector Size (kb)}} * \text{Insert:Vector}$$

2.1.11 Gibson Assembly

GFP sequence did not contain suitable restriction sites and thus Gibson Assembly® was implemented (New England BioLabs). The manufacturers guidelines were followed and assembled plasmid was transformed into NEB® 5-alpha Competent *E. coli*.

2.1.12 Site Directed Mutagenesis

Manufacturer's guidelines were followed and thermal cycling parameters are outlined in table 2.3. Subsequent transformation utilised XL1-Blue competent cells.

Table 2.3: Thermal cycling parameters used during site directed mutagenesis. 12 cycles for the second stage were used for G-> T point mutations reactions and 16 was used for CT deletion.

Cycles	Temperature	Time
1	95	30 seconds
12/16	95	30seconds
	55	1 minute
	68	1 minute per kb of plasmid length

2.1.13 Transformation of *E. coli* Cells

50 - 100 µL of ice-chilled calcium competent *E. coli* cells were used per transformation. The DNA plasmid to be transformed was added, not exceeding 10% of the volume containing the cells. Vials were placed on ice for 30 minutes followed by brief heat shock at 42°C for 90 seconds before being returned onto ice for 2 minutes. 1 mL of sterile LB was then added and the culture was then transferred to a shaking incubator

at 37°C, 200 rpm for one hour. 50-100 µL of culture was then spread upon an agar plate containing the appropriate antibiotic selection pressure and incubated overnight at 37°C.

2.1.14 Mini/Maxi-Prep Isolation of Plasmid DNA

Plasmid DNA was isolated and purified following the manufacturer's instructions. QIAprep spin mini kit (Qiagen) were used for 5 mL *E. coli* cultures and the Qiagen Plasmid Maxi Kit (Qiagen) was used for 250 mL cultures.

2.1.15 DNA Quantification

1 -2 µL of DNA-containing solution was loaded onto the stage of a NanoDrop Spectrometer ND-1000 (Thermo Scientific) set to 260nm, which was used to determine the DNA concentration within the solution.

2.1.16 Sequence Alignment Tools and Frame Checker

Confirmation of DNA sequences was achieved by using commercially available Genewiz sequencing which was then compared to the expected DNA sequence on the sequence analysis service from EMBL-EBI. The reading frame was also confirmed post cloning using the Expasy Translate Tool webservice.

2.1.17 Signal Sequence Checker

SignalP- 5.0 (DTU Health Tech) was used to analyse gene sequences for presence of signal peptide encoding regions.

2.2 Mammalian Cell Culture

2.2.1 Adherent Cell Culture Maintenance

HEK293 adherent cell cultures were maintained using Dulbecco's modified Eagle's medium (DMEM, Gibco) supplemented with 50 mL fetal bovine serum (FBS, Sigma Aldrich) per 500 mL total medium. Cultures were typically maintained in T25 (Starstedt) flasks in 5 mL of DMEM incubated in a static incubator at 37°C with a 5% CO₂ atmosphere. Turbidity and media colouration were consistently observed to identify any

potential contamination and images obtained during cell counting via the ViCell were also scrutinised to ensure cultures were contamination free.

Cultures were passaged every 3-4 days as needed (once 80% to 90% confluency was achieved) whereby medium was aspirated and cells were washed with Phosphate buffered saline (PBS) solution preheated to 37°C. 500 µL 0.05% trypsin-EDTA solution (Gibco) was applied and T25 flasks were returned to the same static incubator for approximately 5 minutes. HEK293 cells would detach from the flask's surface during this time and 2.5 mL of DMEM-FBS was added to inactivate trypsin. Subsequently, 0.2 mL of cell solution was transferred to a fresh T25 flask supplemented with 4.8 mL DMEM-FBS medium. The passaged cells were then returned to the same incubator.

2.2.2 Cryopreservation and Revival of HEK293 Cells

HEK293 cells were stored long term via cryopreservation. A target cell density at 4×10^6 cells/mL was achieved and was resuspended in DMEM with additional 7% dimethyl sulfoxide (DMSO) following centrifugation at 800 x g (Megafuge 16R, rotor 75003181) for 5 minutes. 1 mL aliquots of cell containing solution was transferred per cryo-vial (1.8 mL, Nunc) which were placed into cryo-freezing container (NALGENE) for gradual freezing, stored at -80°C overnight. The following day the cryo-vials were transferred to cryostat filled with liquid nitrogen.

2.2.3 Cell Counting

A ViCell (Beckman Coulter) instrument was used to measure viable cells and total cell counts of HEK293 cultures using 1 mL from cell cultures, utilising the trypan blue staining exclusion method.

2.2.4 Calculating Cell Index on the xCELLigence System

The xCELLigence (ACEA Biosciences) equipment and software was used to calculate cell index profiles of HEK293 cell cultures following the instructions in the manufacturer's manual. 3,000 cells were taken from cultures experiencing log growth phase to seed 18-well E-plates (Amgen). Cell index values were automatically taken every 30 minutes for 7 or 10 days by electronically sweeping the E-plates. Cells were cultured under 5% CO₂ conditions at either 37°C or 32°C in a static incubator.

2.2.5 Transfection of DNA in Mammalian Cells by Lipofectamine

Lipofectamine® 2000 (Invitrogen) was used to transfect HEK293 cells transiently, or to create stably transfected cell pools, following the instructions from the manufacturer.

Transient Transfection of HEK293 Cells

HEK293 cells were seeded in 6 well plates and left overnight to adhere. Following 70-90% confluency, 4 µg of DNA was mixed with 250 µL Opti-MEM™ and 12 µL lipofectamine® 2000 was mixed with 250 µL Opti-MEM™ in a separated microcentrifuge tube per transfection. These mixtures were left to incubate at room temperature for 5 minutes before being combined in a fresh microcentrifuge tube and mixed thoroughly. Combined mixture was incubated at room temperature for 20 minutes. Medium was aspirated from wells and refreshed alongside addition of lipofectamine® 2000-DNA-Opti-MEM™ mixture. The 6 well plate was then returned to static incubator at appropriate temperature at 5% CO₂.

Generation of Stable Cell Pools

DNA plasmids were linearised before transfection for the purpose of generating stable cell pools to facilitate incorporation/integration into the host cell DNA genome. Restriction enzyme FspI was used to digest plasmids within the ampicillin resistance gene. 15 µg of DNA was digested by 2 µL of FspI in 50 µL nuclease free water and 10% enzyme buffer overnight at 37°C per transfection. DNA clean-up was achieved by adding 150 µL TE buffer, 20 µL 3 M sodium acetate (NaAc), and 550 µL 100% Ethanol to the solution and mixing thoroughly before incubating at -20°C for an hour. The solution was then centrifuged at 11,500 x g for 30 minutes (Eppendorf Minispin® Plus, rotor F-45-12-11) and the ethanol was aspirated off. The DNA was then washed with 1.5 mL 70% ice cold ethanol and centrifuged again at 11,500 x g for 5 minutes. Ethanol was aspirated again and tubes were left open until all residual ethanol had evaporated. Once dry, DNA was resuspended in 100 µL sterile TE buffer.

Stable transfections were carried out on HEK293 cultures grown in T25 flasks to 70-90% confluency. The same mixing strategy as for transient transfection was implemented using 10 µg DNA in 625 µL Opti-MEM and 30 µL lipofectamine 2000 in 620 µL Opti-MEM. Once mixed, culture medium was replaced and mixture was added. Following a 24 hour incubation at 37°C in a 5% CO₂ static incubator, hygromycin B (Gibco) was added for a final concentration of 200 µg/mL, and was added with each subsequent passage as a selection pressure.

2.2.6 Proteasome Inhibition Experiments

MG132 (Sigma-Alrich) was used to inhibit proteasome activity of stably transfected HEK29 cell lines. Cultures were grown to 70-90% confluency in T25 flasks and treated with either 0 M (control), 62.5 nM, or 125nM MG132. Protein samples were extracted 24 hours later by method explained in section 2.3.1.

2.3 Protein Analysis

2.3.1. Protein Sample Collection and Generation

For both lysate and supernatant sample collection, HEK293 cultures were moved from an incubator to a heat pad set to the appropriate temperature where samples were extracted (either 37 or 32°C). Extraction was undertaken as quickly as possible due to the temperature dependant nature of this study and samples were immediately flash frozen on dry ice once extraction was completed. Samples were stored at -80°C until examined. Sample were stored in lysis buffer which consisted of 0 mM HEPES-NaOH, pH 7.2, containing 100 mM NaCl, 1% (w/v) Triton-X100, protease inhibitors (10 mg/ml leupeptin, 2 mg/ml pepstatin, 0.2 mM phenylmethylsulphonyl fluoride) and protein phosphatase inhibitors (50 mM NaF and 1 mM activated Na₃VO₄).

2.3.1.1. Intracellular Sample Collection

Media was siphoned off cells using a glass syringe attached to a pump. Cells were then washed with sterile phosphate buffered saline before lysed by cell scrapers. Lysate was collected in lysate buffer for liberation and solubilisation of the cellular proteins before storage at -20°C until analysed.

2.3.1.2. Extracellular Sample Collection

1 mL of the media supplementing HEK293 cells was aliquoted from each 6 well plate well, centrifuged at 800 x g (Megafuge 16R, rotor 75003181) for 5 minutes at 4 °C, and then the supernatant removed and stored in a microcentrifuge tube at -20°C until required for analysis.

2.3.2. Quantification of Protein Lysates using the Bradford Assay

Protein content of cell lysates was quantified using Bradford assays in 96 well plate format. 1 μ L of centrifuged, debris-free lysate diluted in 19 μ L of double-distilled water were aliquoted into wells and mixed with 200 μ L of Bradford reagent. Known concentrations of bovine serum albumin (BSA) were also treated with Bradford reagent. Following a 10 minute incubation, the absorption of the solution at A600 nm using a FLUOstar® OPTIMA microplate reader (BMG Labtech) was measured and protein concentration was calculated using a standard curve generated from the BSA samples.

2.3.3. Protein Analysis by Sodium Dodecyl Polyacrylamide Gel Electrophoresis (SDS-PAGE)

SDS-PAGE was implemented to resolve proteins in lysate and supernatant samples by their molecular weight. Gels 8% or 12% polyacrylamide content depending upon molecular weight of the primary protein of interest were prepared. 30% polyacrylamide (Biorad), 1 M Tris (pH 6.8) 1.5 M Tris (pH 8.8), 10% SDS, 10% ammonium persulphate and tetramethylethylenediamine were used to achieve this. Protein samples, unless stated otherwise, contained 20 μ g of total protein as determined using a Bradford assay and were prepared in 5x reducing Laemmli sample buffer which contained 0.02% bromophenol blue, 1% β -mercaptoethanol, 2% SDS, 10% glycerol, 12.5 mM EDTA and 50 mM Tris-HCl (pH 6.8) and were heated at 95°C for 5 minutes in a heat block.

Gels were placed into gel tanks also filled with 1X SDS running buffer made from 10X stock (1% SDS, 200 mM Glycine 250 mM Tris) and samples were loaded along with a PageRuler™ Plus Prestained Protein Ladder (Thermo Scientific) for protein size approximations. SDS-PAGE were run for 30 minutes at 100 V followed by 1-1.5 hours at 150 V to allow for clear protein separation until the loading dye front reached the bottom of the gel.

2.3.4. Western Blot Analysis

Protein profiles resolved on SDS-PAGE gels were transferred to 0.45 μ m pore-size nitrocellulose membrane (Amersham™ ProTran™) by means of TE22 Mini Tank Blotting Unit (Hoefer™) run at 0.75 A for 1 hour at 4°C. The nitrocellulose membrane was blocked with 5% (w/v) milk/TBS solution (150 mM NaCl, 50 mM Tris, 2 mM KCl) and washed with TBST solution (TBS+ 0.1% Tween20). The relevant primary antibody, diluted in 3% BSA-TBST was introduced and the membrane was left on a rocking platform overnight in a

4°C cold-room. Following 3 additional TBST washes, the appropriate secondary antibody-peroxidase conjugate was introduced for 1 hour at room temperature. All antibodies and their dilutions are outlined in table 2.4. Chemiluminescence exposed targeted polypeptides using ECL Western Blotting Substrate (Thermo Scientific™) and signal was captured on Hyperfilm ECL (GE healthcare).

2.3.5. Densitometry Analysis

Densitometry analysis was performed on western blot data using the freeware ImageJ software to quantify band intensities. Each sample's target polypeptide's Chemiluminescent signal was normalised to its housekeeping protein signal (β -tubulin) unless stated otherwise in specific result sections. These normalised values were made relative to a sample condition for accurate comparisons which are specified in each relevant results section. Error bars were plotted using the standard error of the mean for densitometry graphs containing multiple repeats per experiment condition. No overlap of error bars between samples was determined to be an indicator that results were different in a statistically significant way. P- values were also calculated and can be found in the appendix section.

2.3.6. Immunofluorescence Imaging

Immunofluorescence imaging was implemented for *In vivo* protein localisation study as described in Budge et al. (Budge et al., 2020a, 2020b)

2.3.6.1. Poly-L-Lysine Cell Adhering

To ensure strong adherence of HEK293 cells to coverslips for imaging, coverslips were incubated for 15 minutes at room temperature while submerged in poly-L-lysine. Coverslips were then allowed to dry before being placed in 24 well plate.

2.3.6.2. Cell Fixing and Permeabilisation

HEK293 culture medium was aspirated and cells were washed with 1 mL PBS before incubated on ice for 5 minutes with 500 μ L ice chilled methanol added to each well. Methanol was aspirated and three 1 mL PBS washes ensured methanol's complete removal.

2.3.6.3. Antibody Exposure

Cells were then blocked with 250 μ L 3% BSA in PBS per well incubated for 20 minutes at room temperature. Following this incubation, BSA-PBS solution was aspirated and coverslips were removed from 24 well plate and washed by being placed facedown upon 100 μ L droplet of PBS. Coverslips were dried and placed upon 25 μ L droplet of V5 antibody (1:500 dilution) and left to incubate overnight at 4°C. all antibodies used throughout study are outlined in table 2.4.

The coverslips were washed in 100 μ L 0.1% Tween PBS solution for 5 minutes four times. They were then placed upon 25 μ L Anti-mouse fluorescein isothiocyanate (FITC) conjugate antibody (1:100 dilution) in a darkroom and left to incubate for 24 hours at 4°C. Coverslips were washed twice with 0.1% Tween PBS before they were treated with 50 μ L 4',6-diamidino-2-phenylindole (DAPI) at 10 mg/mL for 1 minute, after which two addition washes were performed. Finally, coverslips were dabbed to remove excess liquid and cells were mounted on cell slides using a drop of mounting reagent and left overnight.

2.3.6.4. Confocal Microscope Imaging

Cell images were taken using LSM 880 Elyra, AxioObserver confocal microscope (Zeiss) utilising 2 channels. Channel 1 used 405nm laser at 0.2% power detecting a 456nm emission (blue), and channel 2 used the Airyscan detector array, 488nm argon laser at 0.4% power (green). No bandpass filters were used and Airyscan images were deconvoluted after acquisition.

2.3.7. Antibodies Used within this Study

Table 2.4: Specifics on antibodies used through this study to facilitate protein production analysis. Company of origin, the organism used to produce them and their dilutions are included. Anti V5 antibody was used in both western blotting and immunofluorescent techniques and therefore was used at two dilutions. All antibodies are whole IgG molecules.

Antibody	Company/ Reference Number	Species	Dilution
Anti-V5	<i>Sigma V8012</i>	Mouse	Western Blot- 1:5000 Immunofluorescence- 1:500
Anti-CCND3	<i>Invitrogen MAS-12717</i>	Mouse	1:1000
Anti-eGFP	<i>Abcam ab184601</i>	Mouse	1:1000
Anti-eEF2	<i>Cell Signalling #2332S</i>	Rabbit	1:1000
Anti-eEF2-P	<i>Cell Signalling #2331S</i>	Rabbit	1:1000
Anti-CIRP	<i>ProteinTech 10209-2-AP</i>	Rabbit	1:1000
Anti-Tubulin (TAT)	<i>In House*</i>	Mouse	1:1000
Anti- β -actin	<i>Sigma A5441</i>	Mouse	1:5000
Anti-Mouse FITC conjugate	<i>Sigma F0257</i>	Goat	1:100

* TAT monoclonal antibody was gifted by Professor Keith Gull from the University of Oxford and was originally used as part of a set of antibodies used to define structural components of the cytoskeleton. The antibody was generated by cloning hybridomas produced from mice injected with proteins from the cytoskeleton or its extracts (Woods et al., 1989).

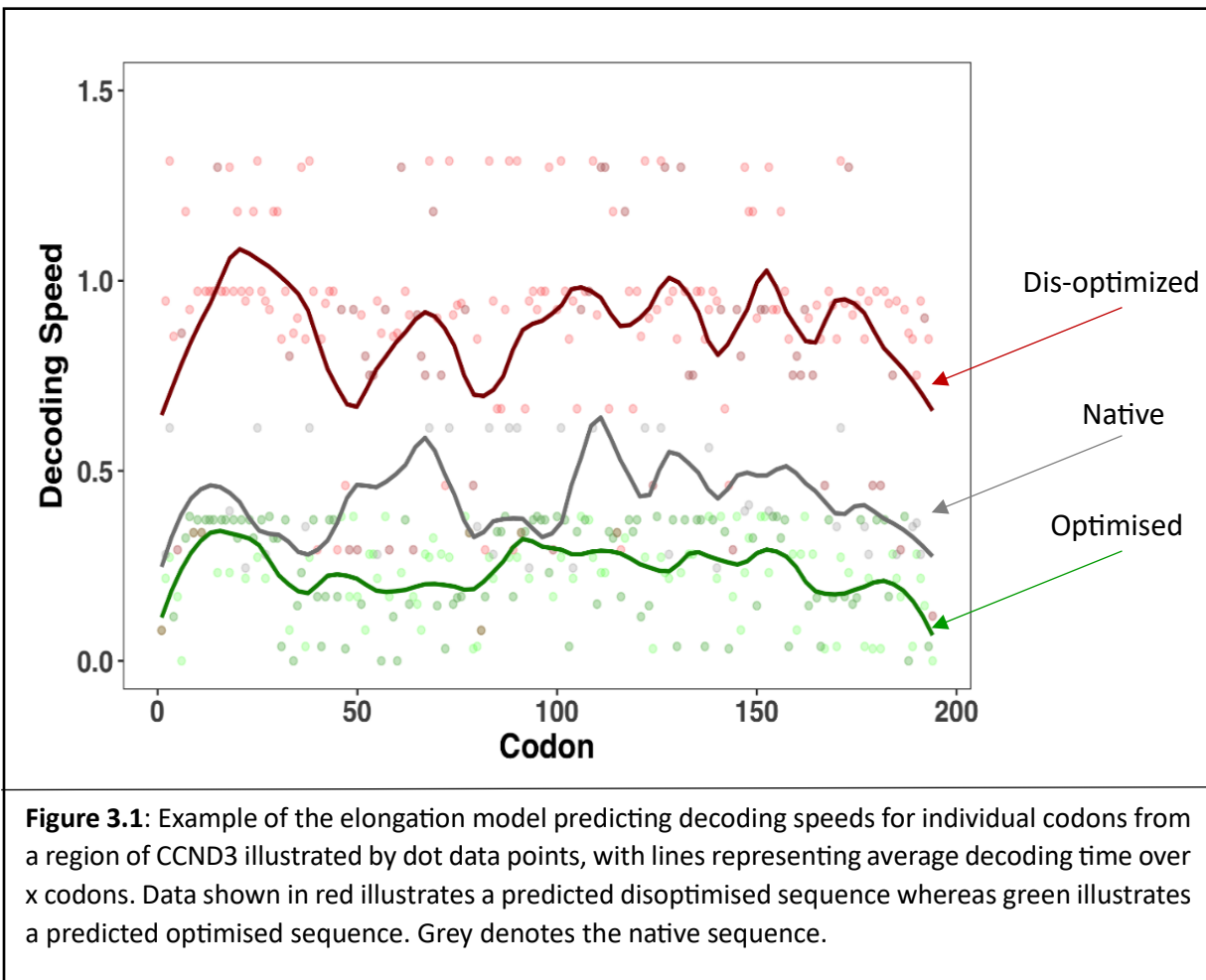
Chapter 3

Design, development and generation of model sequences and constructs with different predicted mRNA decoding speeds for evaluation in mammalian cells

3.1 Introduction

As discussed in detail in the main Introduction chapter of this thesis (Chapter 1), multiple factors can impact the three basic steps of translation elongation during protein synthesis which include: codon decoding/tRNA selection, peptide bond formation, and translocation of mRNA-tRNA complex. These variables supply cells with a level of elongation control, tuning protein production often in response to cellular stresses (Bastide et al., 2017; Godet et al., 2019; Spriggs et al., 2010). The primary focus of our investigations was to explore the codon decoding/tRNA selection axis of this control. With regards to this, ribosomes can experience pausing/stalling whilst waiting for a correct cognate tRNA to enter the A site for peptide bond synthesis to occur (Plotkin & Kudla, 2011; Quax et al., 2015). Codons that have corresponding abundant tRNAs enhance elongation speed whereas codons whose tRNA counterparts are at low or limited abundance reduce predicted and experimental elongation speed, resulting in ribosomal stalling. This stalling not only impacts elongation rates, but can hinder translation initiation as well (Chu, Kazana, et al., 2014; Plotkin & Kudla, 2011).

A novel elongation model created by University of Kent collaborations was designed to simulate translational control incorporating tRNA dynamics and the impact they cause on ribosomal collisions and thereby elongation rates across an entire mRNA transcript (Figure 3.1) (Chu et al., 2012). This model does not account for competition for ribosomes caused by different transcripts, but does allow for system-wide representation of translation and account for tRNA abundance, charged states, near cognate tRNAs and translocation, as well as being driven by initial binding of ribosomes to mRNA transcripts. This algorithm-based model was used to identify mRNA transcripts which contained regions of slow decoding speeds relevant to elongational control. Applications of this model have highlighted that the initial 5' sequence of the ORF immediately following the start AUG codon have a substantial impact on overall protein production rates and may be integral to elongational control. To assess the impact these regions have on elongational control, three of the identified mRNA transcripts: *CCND3*, *noggin*, and *eGFP*, were chosen and modifications were made and their protein production was observed.



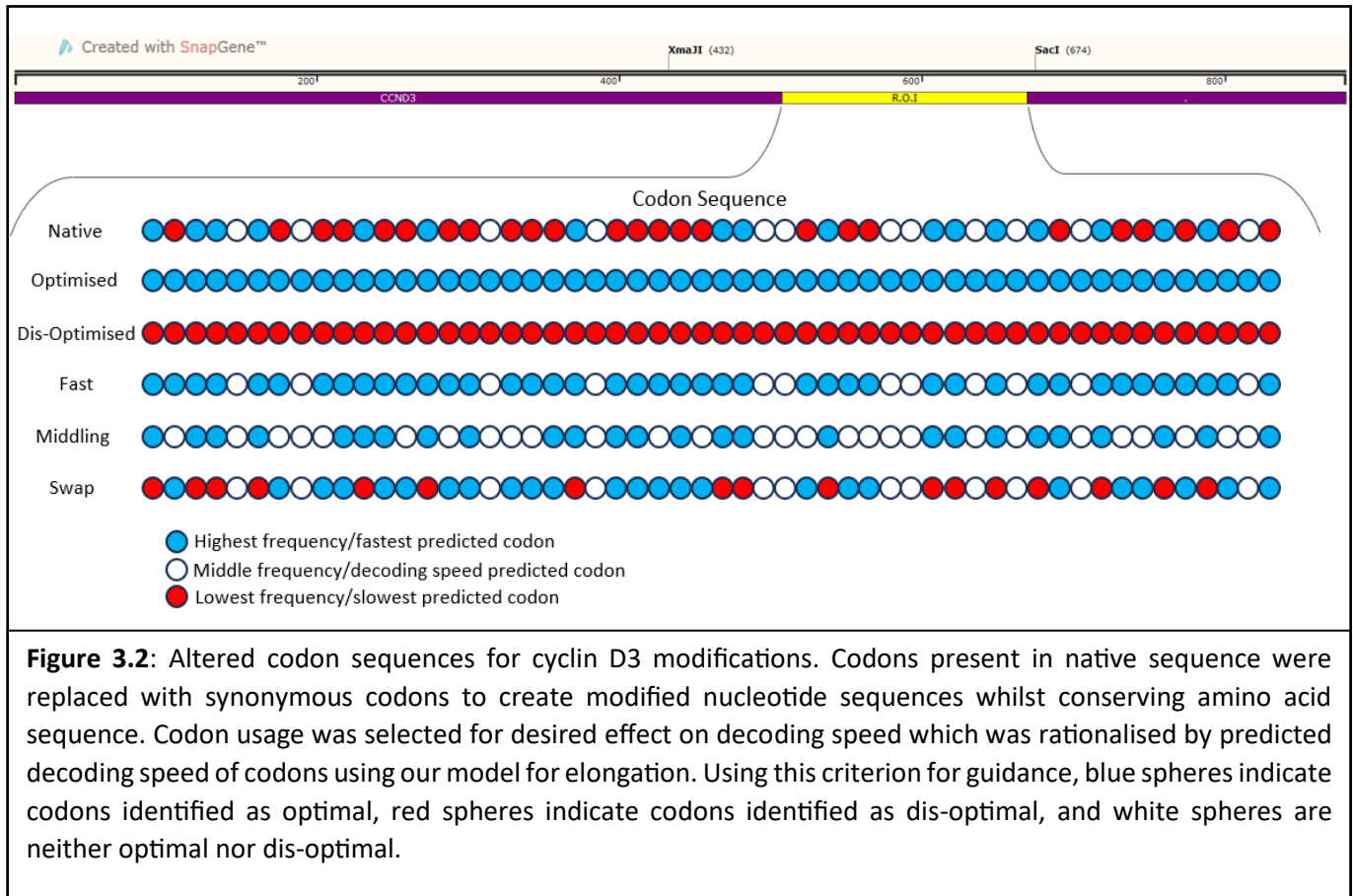
Modifications were made only within the slowly decoded regions identified and not the entire transcript, and these newly designed sequences were incorporated into their respective gene via the cloning methods outlined in methods chapter of this thesis (chapter 2). *Noggin* and *eGFP* sequences were typical of many sequences identified as their regions highlighted to have a substantial impact on elongation was their initial 5' sequences of the ORF immediately following their start AUG, but the location for the *CCND3* sequence was towards the middle of the ORF, which is of note as this is predicted to have less impact on initiation rates (Chu et al., 2012). Design of these region modifications focused on creating optimised (fastest decoded codons) and disoptimised (slowest decoded codons) sequences without altering the resultant protein's amino acid sequence by using synonymous codons with different decoding speeds. Other combinations of codons were also implemented to design other sequences to enhance our understanding of how codon usage impact elongation. These sequences including swap, middling and fast sequences and are described in each relevant section of this chapter. Each protein's gene required different approaches to incorporate these modified sequences which are described in this chapter.

3.2 Design and Generation of Cyclin D3 (*CCND3*) Transcript Sequences with Alternative Codon Usage.

3.2.1 Sequence Design

The initial work involved the design of sequences containing different codons within the open reading frame (ORF) transcript of *CCND3* that all code for the same amino acid sequence. *CCND3* is a primary control protein/driver of the cell cycle and previous work had identified a section of codons that were predicted to be decoded slowly, referred to as the region of interest in this case. Here, region of interest (ROI) codon modifications were focused on assessing the effects on overall *CCND3* protein production levels, by altering codon usage with codons with different predicted decoding times within this region. While fully optimising transcript decoding speeds to use only those codons with the fastest decoding times may intuitively lead to the fastest overall transcript decoding times and hence the highest increase in protein production compared to non-codon optimised sequences, other factors such as ribosome collisions may hinder production rates (Meydan & Guydosh, 2021). Conversely, the presence of slower decoded codons may lead to more effective concurrent elongation of multiple ribosomes on a single mRNA/transcript at once. With this in mind, multiple sequences with different ORF codon usage for the ROI in *CCND3* were designed as described in figure 3.2. ‘Optimised’ and ‘dis-optimised’ sequences were designed to utilise the fastest or slowest codon for their respective amino acid. The ‘fast’ sequence was designed to replace any slow codons from the sequence with the fastest codons without altering any other codons used, whereas the ‘middling’ sequence replaced any slow codons with ‘averagely’ decoded speed codons where possible, or the fastest codon where only two codons for that amino acid exist. The ‘swap’ sequence replaced each fastest codon with the slowest decoded codon for that amino acid and vice versa.

Two sets of codon manipulated sequences for the region of interest in *CCND3* were generated, with and without the stop codon, to create a C-terminal V5/His tagged (without stop) and untagged version. These were generated to allow the distinction between endogenous *CCND3* and exogenous *CCND3* expression containing a V5 tag, and the combination (or total) endogenous and exogenous *CCND3* expression using a *CCND3* specific antibody. Both antibodies for the V5 tag and *CCND3* protein are detailed in materials Chapter 2.



3.2.2 CCND3 Cloning Strategy

The original *CCND3* ORF 'native' or endogenous sequence /main transcript was cloned from HEK293 cells by isolating mRNA and amplifying *CCND3* mRNA sequences, using PCR with *CCND3* specific primers and restriction sites BamHI & XhoI included at 5' and 3' primers respectively. Two 3' primers were used, one with stop codon removed to generate V5/His tagged version of *CCND3* as this was in the plasmid and the gene was placed in frame with this. Primers outlined in methods chapter (2.1.5). Modified sequences were designed using the rationale explained in 3.2.1 and the 5 designed sequences were commercially synthesised using GeneArt, outlined in figure 3.2. The expression vector used for expression of the different transcripts with different codon usage was a modified version of commercial vector pcDNA3.1 (see appendix).

A challenge with modifying regions of encoding ORF sequences is that if these are to be introduced using traditional restriction enzyme cloning, as in the work described here, only pre-existing restriction sites found in the native sequence were viable for use. Fortunately, two such sites exist flanking the *CCND3* ROI:

XmaJI & SacI. Unfortunately, as seen in figure 3.3, these sites were also present in the vector of choice. Therefore, an additional cloning step was necessary wherein *CCND3* was first subcloned into pET23b(+) vector to insert the different ROI codon sections before the new modified sequence was then subcloned back into the pcDNA3.1 expression vector for use in expression studies.

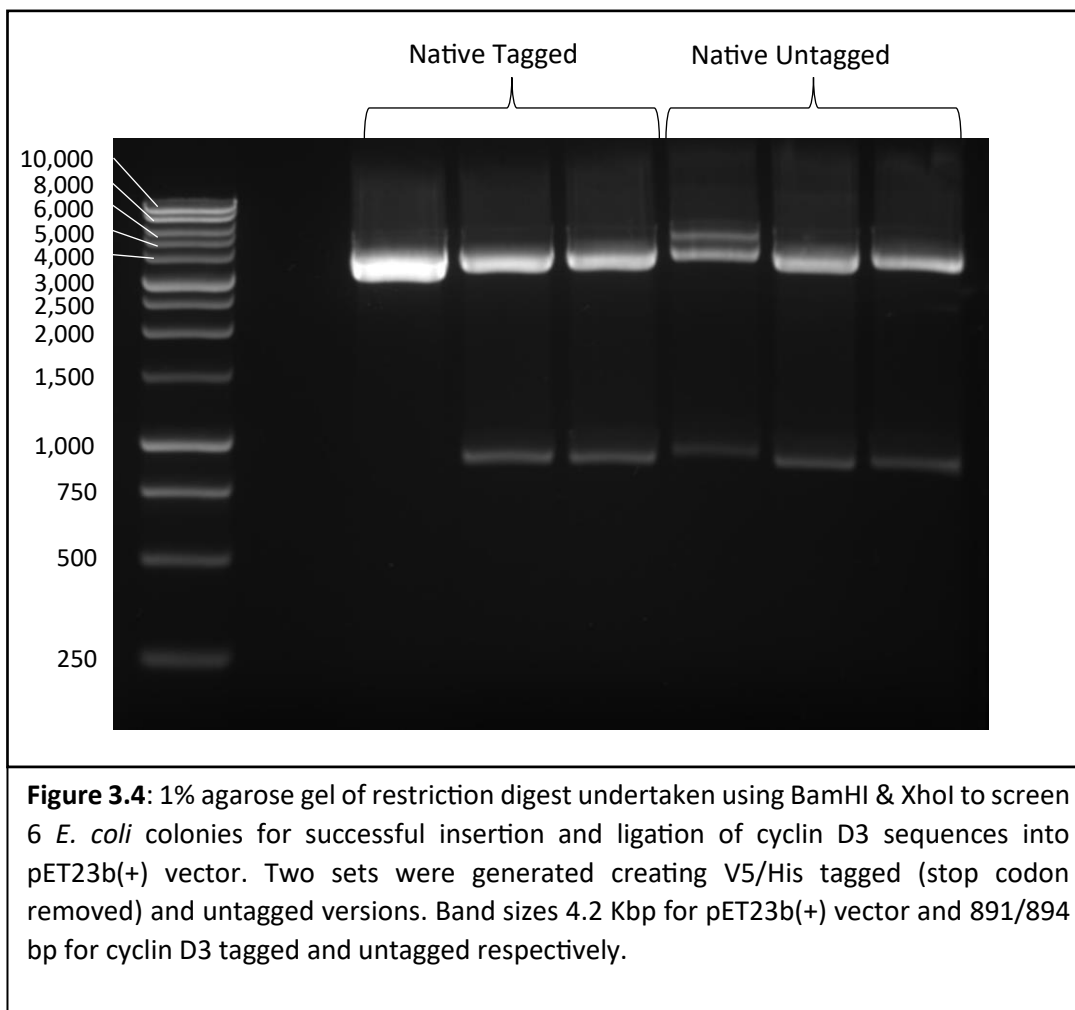


Figure 3.3: Schematic of *CCND3* sequences. Top *CCND3* sequence with highlighted region of interest in yellow. Bottom: *CCND3* sequence in pcDNA3.1 hygromycin vector and pET23b(+) vector. Relevant restriction sites are included.

3.2.3 Insertion of Tagged and Untagged Native *CCND3* into the pET23b(+) Plasmid

The initial cloning step was achieved by digesting both pcDNA3.1 vector containing the HEK293 derived *CCND3* native/endogenous sequence and pET23b(+) with XhoI & BamHI restriction enzymes. The DNA segments were then separated by electrophoresis using a 1% agarose gel. The commercially available Promega Wizard clean up system was then used to purify digested *CCND3* and pET23b(+) segments before ligation and subsequent transformation into calcium competent DH5α *E. coli* cells as described in the methods section. Colonies were grown on agar plates containing ampicillin, cultured overnight and minipreped to isolate plasmid DNA which was used to confirm successful pET23b(+)/*CCND3* ligation. This was carried out for both V5/His-tagged and untagged versions. Figure 3.4 displays minipreped plasmid

ligation confirmation by restriction enzyme mapping with two colonies of each version displaying the correct band sizes at 4.2 kbp for the pET23b(+) vector backbone and 891/894 bp for *CCND3* tagged and untagged genes respectively. Protocols for these processes are outlined in methods 2.1.7-2.1.10.



3.2.4 Insertion of Codon Modified Regions of Interest into the Native *CCND3* Sequence

The cloning step to insert the codon modified ROI utilised restriction enzymes XmaI & SacI to digest GeneArt modified sequences along with pET23b(+)/*CCND3* tagged and untagged versions. DNA segments were similarly separated, purified, ligated, and transformed into competent DH5α *E. coli* cells. Subsequent colonies were screened for modified sequence insertions as seen in figure 3.5, on a 2% agarose gel. Out of these, the majority displayed what was thought to be bands at the correct sizes of 242 bp and 4.2 kbp for the modified ROI and remaining vector respectively, denoting successful ligations. The exceptions were

dis-optimised colony 3 and 4, and fast colony 4 which did not show bands at 242 base pairs, suggesting the modified sequence failed to be incorporated into the *CCND3* backbone. Colony 1, successful for each modification, was chosen to take forward for the third and final round of cloning that was preformed to reintroduce *CCND3* sequences into the final destination pcDNA3.1 vector.

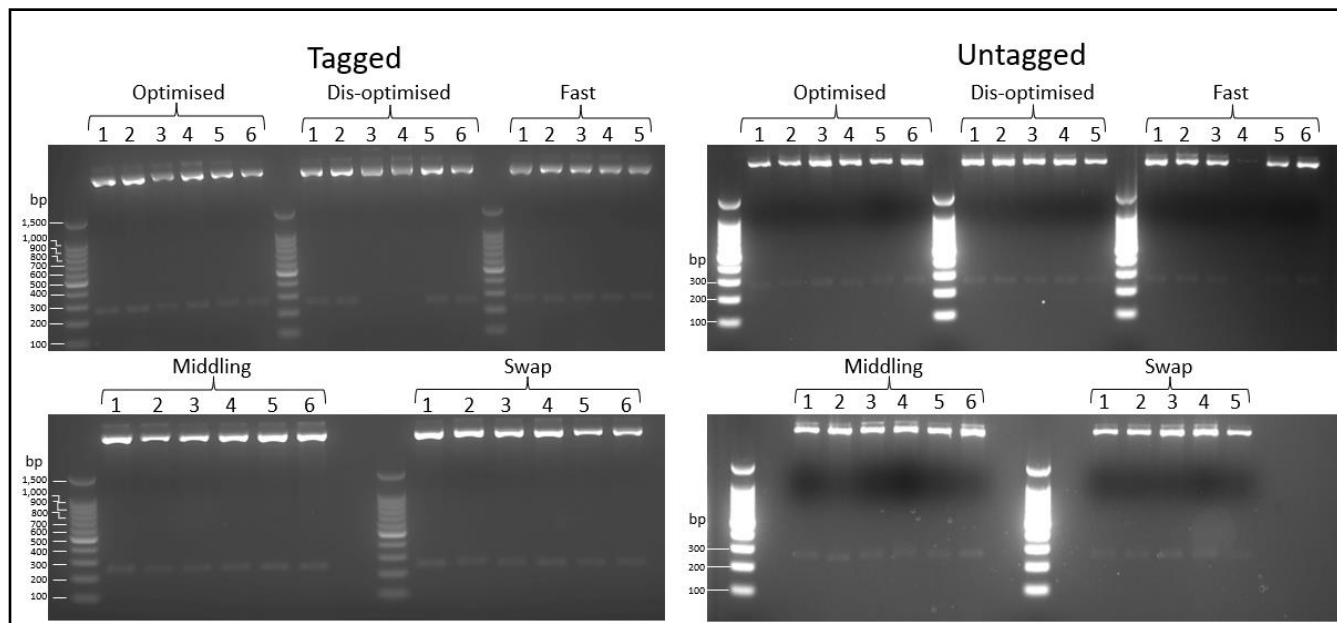


Figure 3.5: 2% agarose gel of restriction digest undertaken using XmaI & SacI to screen *E. coli* colonies for successful insertion and ligation of modified Cyclin D3 sequences in Pet23b(+) vector. Two sets were generated creating V5/His tagged (stop codon removed) and untagged versions. Band size at 242bp was expected for modified insertions and vector was too large for scale (4.2kp).

3.2.5 Re-insertion of *CCND3* Sequences with ROI Codon Modifications into the pcDNA3.1 Vector

To move the modified *CCND3* from the pET23 vector and into the final pcDNA3.1 destination vector, BamHI & XhoI were used to excise *CCND3* sequences from pET23b(+). *CCND3* sequences were isolated on a 1% agarose gel and purified for subsequent ligation into BamHI & XhoI digested pcDNA3.1 destination vector. Ligation solutions were used to transform DH5α *E. coli* cells, producing colonies grown on ampicillin containing agar plates. Overnight cultures of these colonies were miniprepmed and the resultant purified plasmid DNA were again digested with BamHI & XhoI to screen for successful ligation products. Figure 3.6 shows the restriction digest screen for all tagged and untagged *CCND3* sequences. Bands were expected at 5.6 kbp for vector and 891/894 bp for *CCND3* tagged and untagged drop out bands respectively. Each

modified sequence had colonies exhibiting the correct band profiles and colony 1 from each was used to produce overnight cultures in preparation for maxiprep. Correct sequence for each construct were confirmed through commercial Genewiz sequencing services using primers outlined in methods table 2.2. The open reading frame translation frame for each sequence for both the tagged and untagged versions was confirmed using the Expsy translate tool.

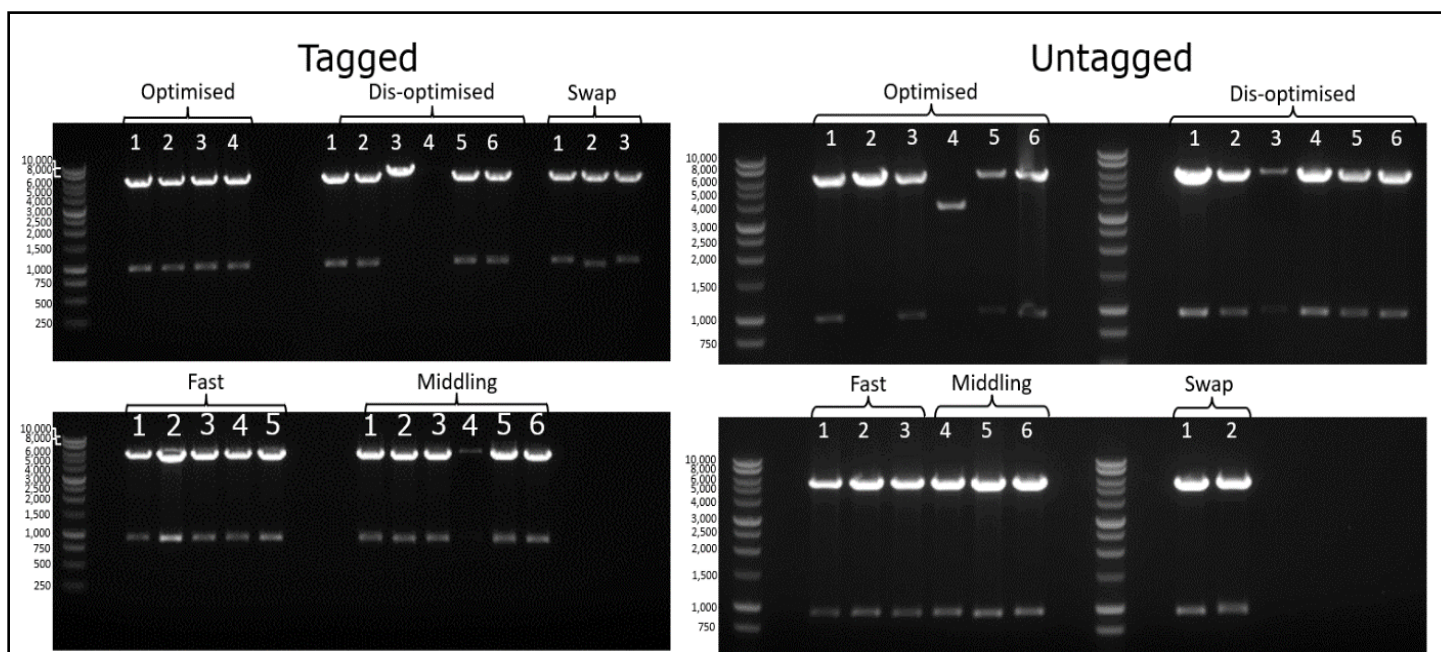


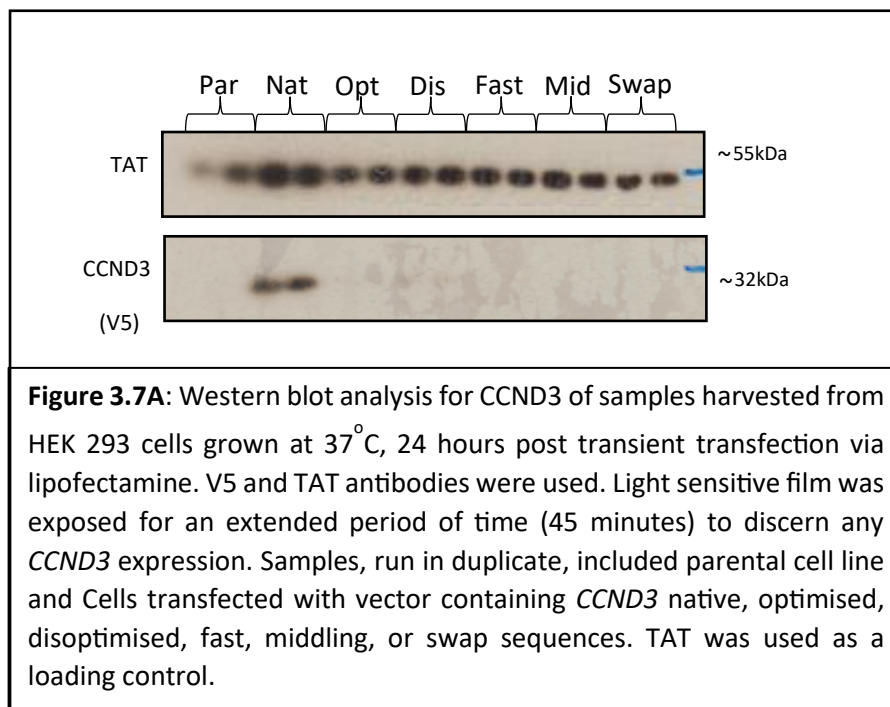
Figure 3.6: Restriction digest undertaken using BamHI & XhoI to screen transformed DH5 α *E. coli* colonies for successful reinstatement of modified cyclin D3 sequences into pcDNA3.1Hygro(+) V5His vector. Two sets were generated creating V5&His tagged (stop codon removed) and untagged versions. Bands were expected at 5.6kbp for vector and 891/894bp for cyclin tagged and untagged respectively.

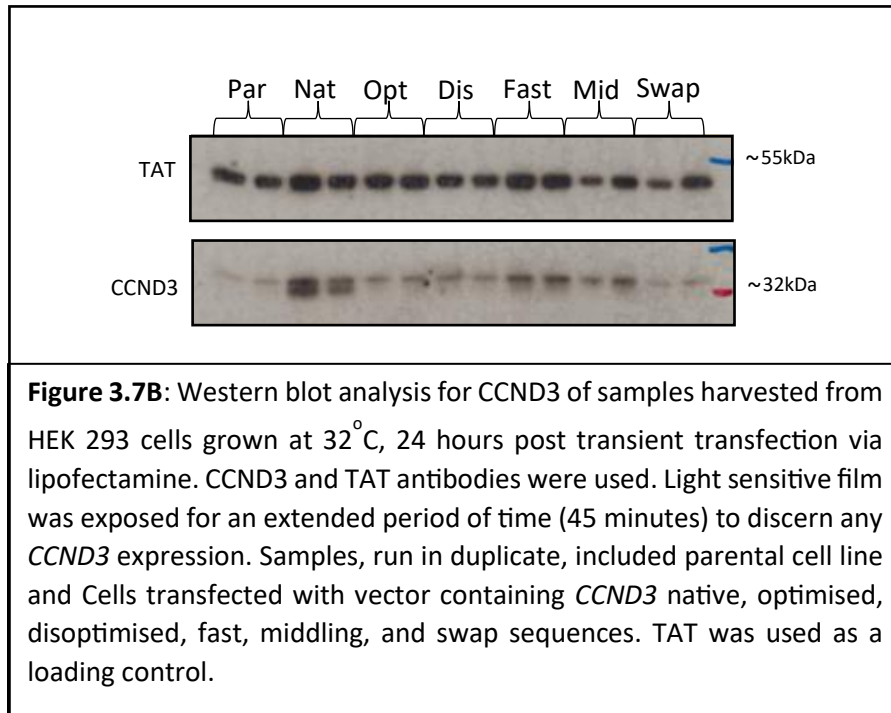
3.2.6 Assessing Expression of *CCND3* Constructs in HEK293 Cells

3.2.6.1 Western Blot Analysis of Transiently Transfected HEK293 cells for *CCND3* Expression

Cells were grown at 37°C in 6 well plates seeded with HEK293 cells at 1.5×10^5 cells/mL 24 hours prior to transient transfection of *CCND3* constructs in duplicate. The Thermofisher lipofectamine 2000 and Opti-MEM protocol was used for transfection. Cell lysis samples were taken 24 hours post transfection and stored in lysis buffer. Cell debris was removed and total protein concentrations were determined by the Bradford assay. Samples containing 10 μ g protein in reducing buffer were subjected to SDS-PAGE analysis/separation followed by western blotting, once transferred to nitrocellulose blotting paper.

Anti-V5 and anti-CCND3 antibodies were used to attempt to visualise CCND3 production for tagged and untagged sequences respectively, and tubulin (TAT) production was used as a loading control. As seen in figure 3.7A, all V5 tagged modified *CCND3* sequences failed to show any signs of expression, even after extended exposure times of 45 minutes to an hour. V5 tagged native sequence did show some evidence of expression after prolonged exposure times. Figure 3.7B also required prolonged exposure times, however *CCND3* expression was visible across all samples with the native *CCND3* sequence yielding the strongest signal. Notably, CCND3 specific antibody as oppose to V5 specific antibody was used for this blot, consequently endogenous and exogenous expression cannot be distinguished. Native samples appear to exhibit two similarly sized bands which may relate to exogenous and endogenous proteins.



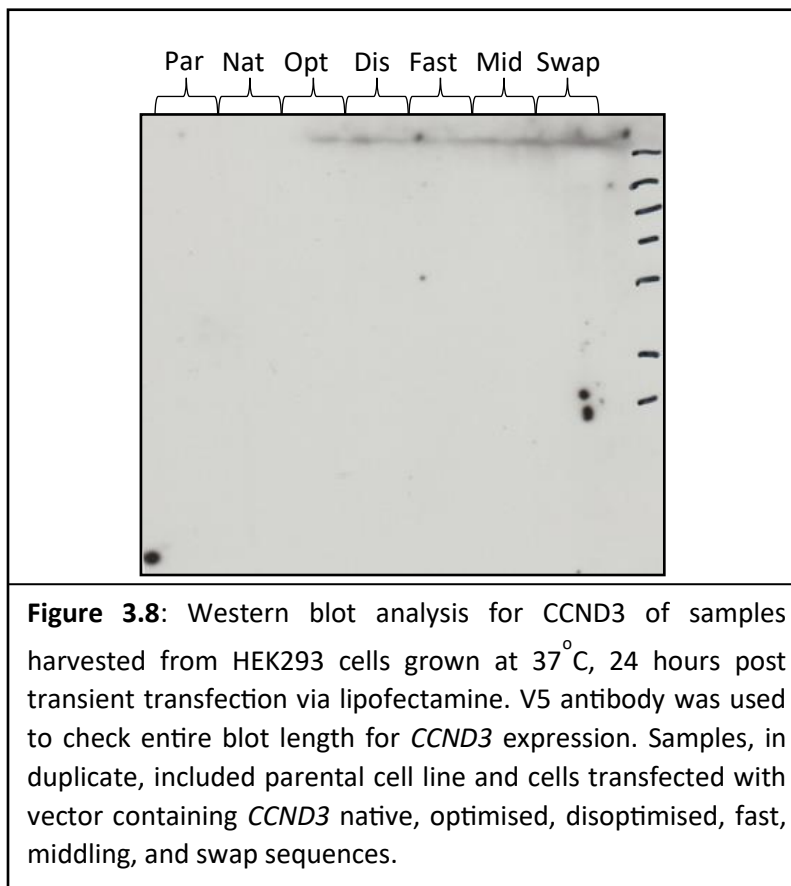


Results from these blots were confirmed and highlighted expression issues with the generated CCND3 constructs, an issue that impacted modified ROI constructs more dramatically than native.

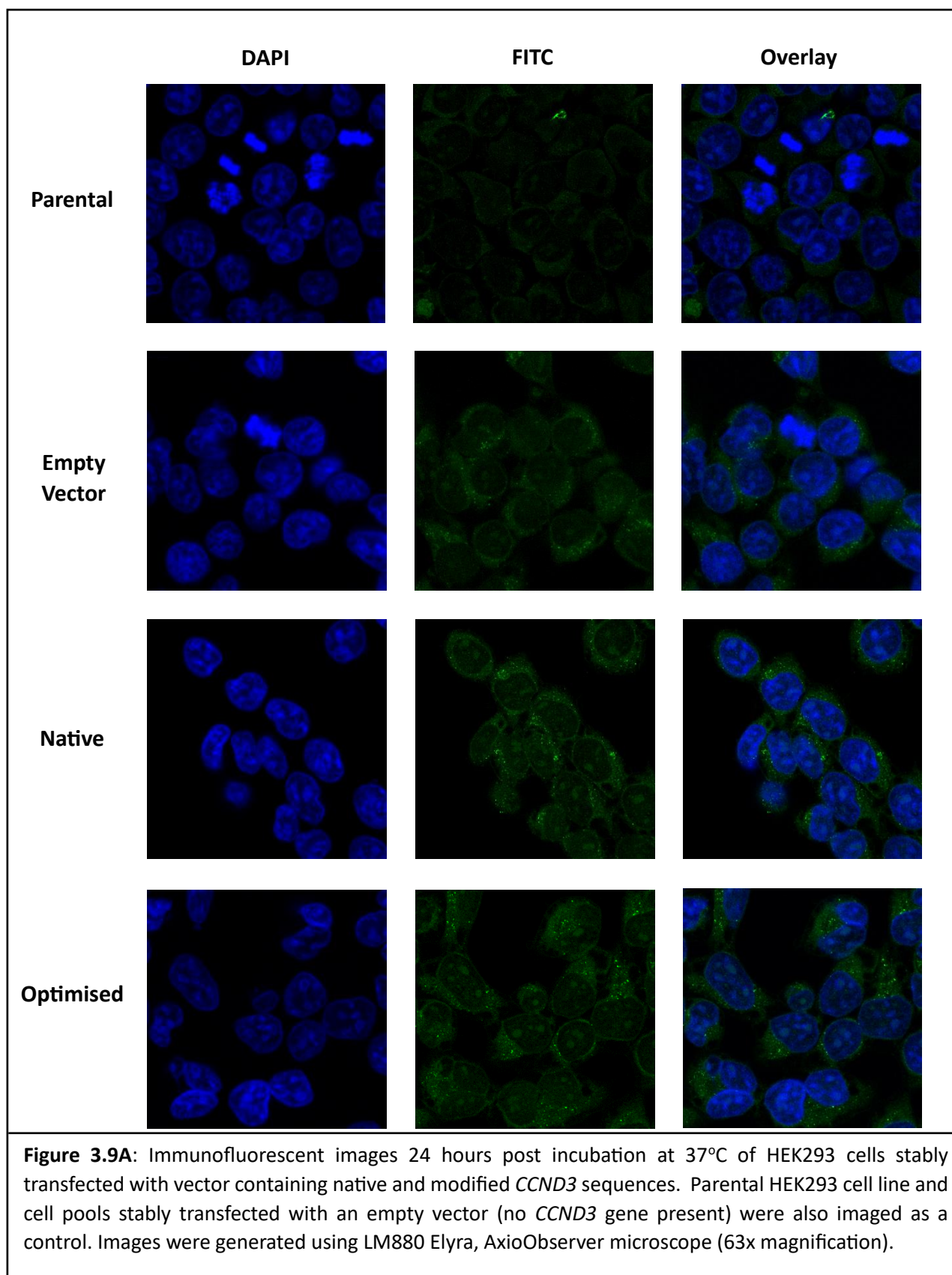
3.2.6.2 Evaluating Poor Transient Exogenous CCND3 Expression

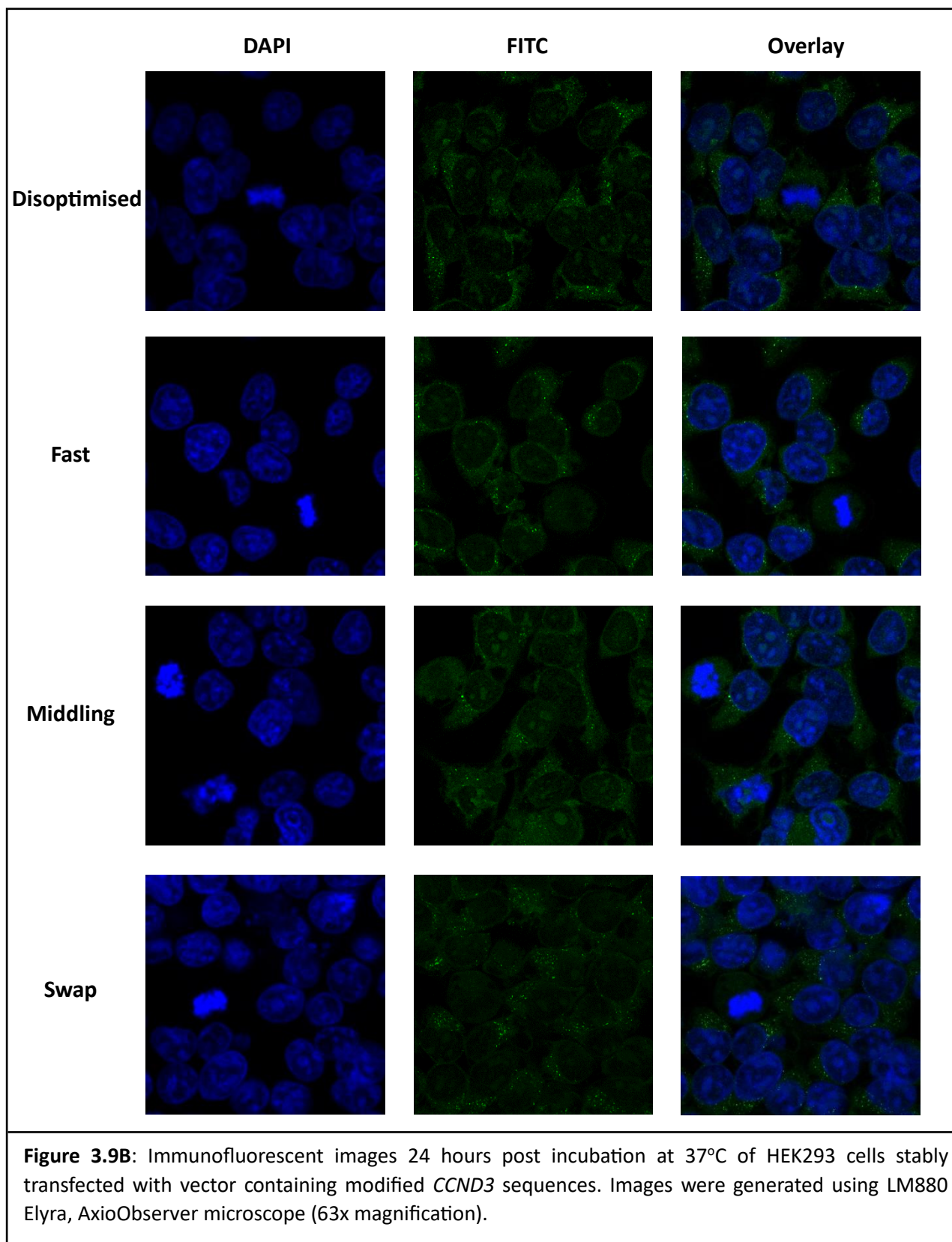
To investigate the CCND3 expression issues exhibited in figure 3.7A and figure 3.7B, multiple experiments were proposed and some explored before the issue was identified.

To confirm CCND3 protein size had not been drastically impacted by the cloning process, a full nitrocellulose membrane blot of the entire SDS-PAGE was treated with V5 antibody to pick up V5 tagged protein of any size. Figure 3.8 shows that this further confirmed no protein production was being observed.

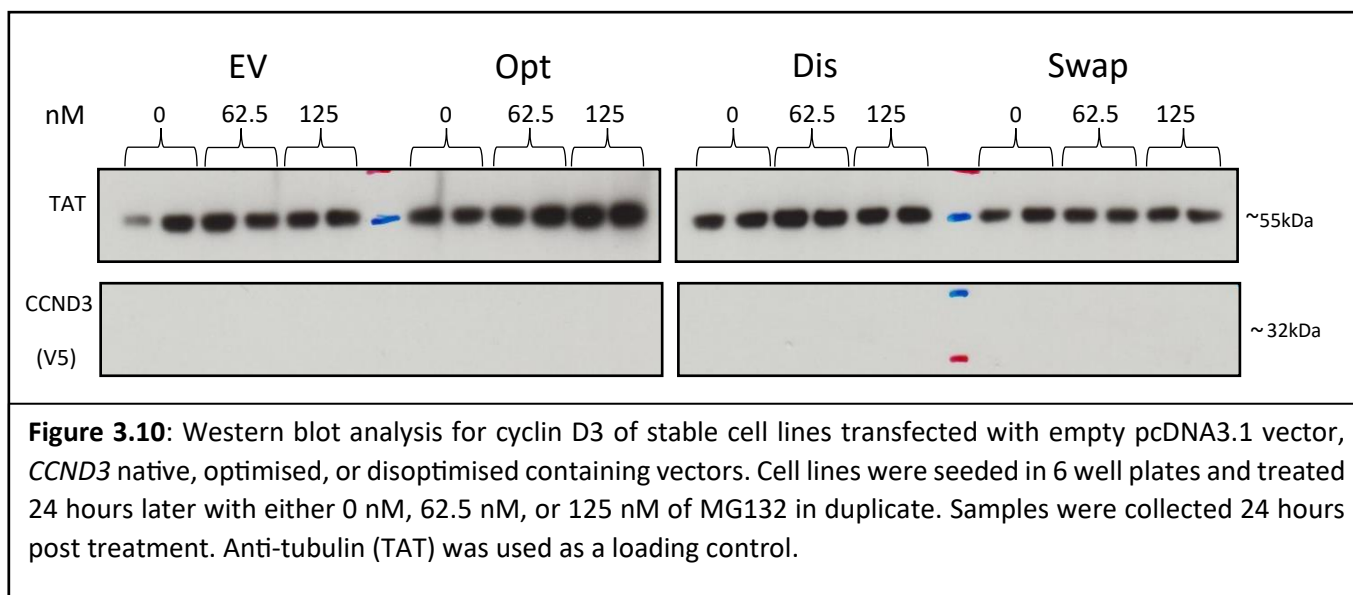


Immunostaining stably transfected cell lines using DAPI for dsDNA staining and FITC for exogenous CCND3 staining was carried out using confocal microscope and Zen software for image processing as seen in figure 3.9A and figure 3.9B. cell cultures were incubated at 37°C and samples were taken after 24 hours. The antibody dilutions are outlined in table 2.4. These images again confirmed a lack of expression, although some expression did appear to be present at low levels and localised within the cell's cytoplasm. This localisation, coupled with low expression levels and CCND3's highly regulated nature within mammalian cells (Dang et al., 2021; Hu et al., 2002) suggested the possibility that CCND3 was being targeted for degradation rapidly; localised hot spots relating to proteasomes. Signals are so weak that they could also be due to cross reactive background noise within cell.





To explore this, HEK293 cells stably transfected with empty pcDNA3.1 vector, optimised, dis-optimised or swap *CCND3* sequences were treated with varying concentrations of proteasome inhibitor MG132 intending to prevent ubiquitin mediated degradation of *CCND3* (Dang et al., 2021). Western blot analysis of subsequent samples in figure 3.10 shows that exogenous *CCND3* expression from these cell lines when treated with proteasome inhibitors did not improve compared to previous transient studies.



Subsequently, a closer evaluation of the sequences generated during the insertion of the modified ROI into *CCND3* was undertaken. Analysis as described in Figure 3.11, shows that the lack of expression originated in the design of the initial modified sequence from GeneArt. The correct nucleotide length between and including restriction enzyme sites *Xma*I & *Sac*I spans 244 bps, however a sequence of 246 was erroneously ordered for ROI generation and cloning. The *CCND3* fast sequence is shown as an example of the systemic error seen throughout all modified sequences, both tagged and untagged, due to a CT insertion between position 507 and 508 nucleotides at the beginning of the ROI. The cascading effect of this double insertion resulted in all subsequent codons to be frame shifted, drastically altering the amino acid sequence and truncating the protein. This also prevented translation of the V5 tag for all tagged modified models. Truncated proteins produced can be seen in figure 3.12. ExPASy online translate tool was used to discern the affect.

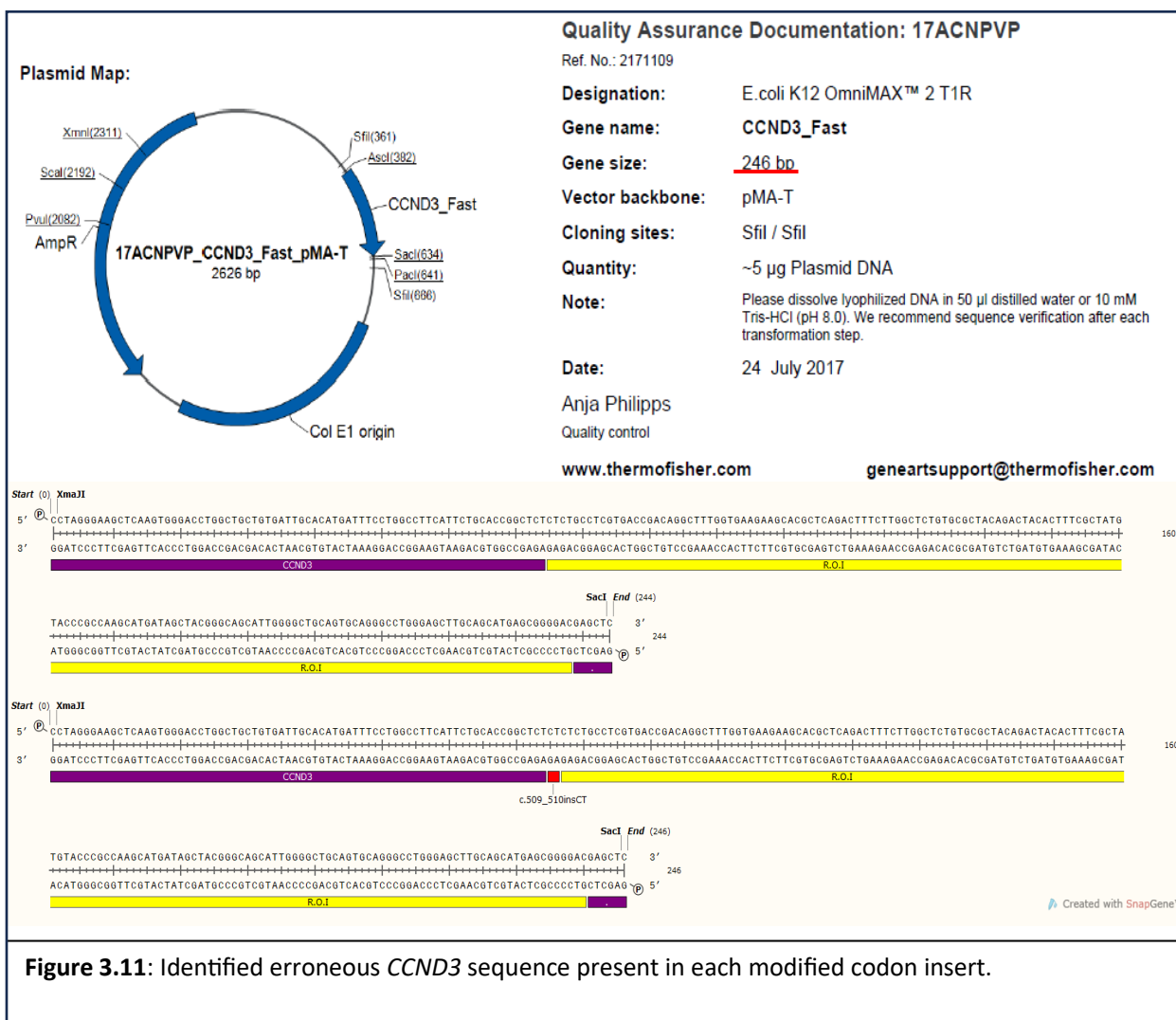


Figure 3.11: Identified erroneous *CCND3* sequence present in each modified codon insert.

Native -317

MELLCCEGTRHAPRAGPDPRLDQORVLQSLRLLEERYVPRASYFQCVQREIKPHMRKMLAYWM
LEVCEEQRCEEEVFPLAMNYLDRYLSCVPTRKAQLQLLGAVCMLLASKLRETTPLTIEKLCIYT
DHAIVSPQRLRDWEVLVLGKLKWDLAAVIAHDFLAFLHRLSLPRDRQALVKKHAQTFLALCATD
YTFAMYPPSMIATGSIGAAVQGLGACMSGDELTELLAGITGTEVDCLRACQEQIEAALRESLR
EAAQTSSSPAPKAPRGSSSQGPSQTSTPTDVTAIHLLEGKPIPNPLLGLDSTRTGHHHHHH

Optimised- 178

MELLCCEGTRHAPRAGPDPRLDQORVLQSLRLLEERYVPRASYFQCVQREIKPHMRKMLAYWM
LEVCEEQRCEEEVFPLAMNYLDRYLSCVPTRKAQLQLLGAVCMLLASKLRETTPLTIEKLCIYT
DHAIVSPQRLRDWEVLVLGKLKWDLAAVIAHDFLAFLHRLSLCLVTVRLW-RSTLRLSWLCALL
TTLSLCTLLA-LLAALELLCRDWELAA-AETSSQSCWQGSALAKWTACGPVRSRSLHSGRAS
KPLRPAPAQRPKPPGAPAAKGPAPALLQMSQPYTCSRVSLSLTLSSVSILRVPVIITITI

Dis-Optimised- 202

MELLCCEGTRHAPRAGPDPRLDQORVLQSLRLLEERYVPRASYFQCVQREIKPHMRKMLAYWM
LEVCEEQRCEEEVFPLAMNYLDRYLSCVPTRKAQLQLLGAVCMLLASKLRETTPLTIEKLCIYT
DHAIVSPQRLRDWEVLVLGKLKWDLAAVIAHDFLAFLHRLSLSPAIAKPSKSNMPKPFSPSPVP
IIPLPICIPPP-SPPVPSVPPSKVSVVPCPVMSQSCWQGSALAKWTACGPVRSRSLHSGRAS
GKPLRPAPAQRPKPPGAPAAKGPAPALLQMSQPYTCSRVSLSLTLSSVSILRVPVIITITI

Fast- 178

MELLCCEGTRHAPRAGPDPRLDQORVLQSLRLLEERYVPRASYFQCVQREIKPHMRKMLAYWM
LEVCEEQRCEEEVFPLAMNYLDRYLSCVPTRKAQLQLLGAVCMLLASKLRETTPLTIEKLCIYT
DHAIVSPQRLRDWEVLVLGKLKWDLAAVIAHDFLAFLHRLSLCLVTDRLW-RSTLRLSWLCAQ
TTLSLCTRQA--LRAALGLQCRAWELAA-AGTSSQSCWQGSALAKWTACGPVRSRSLHSGRAS
GKPLRPAPAQRPKPPGAPAAKGPAPALLQMSQPYTCSRVSLSLTLSSVSILRVPVIITITI

Middling- 178

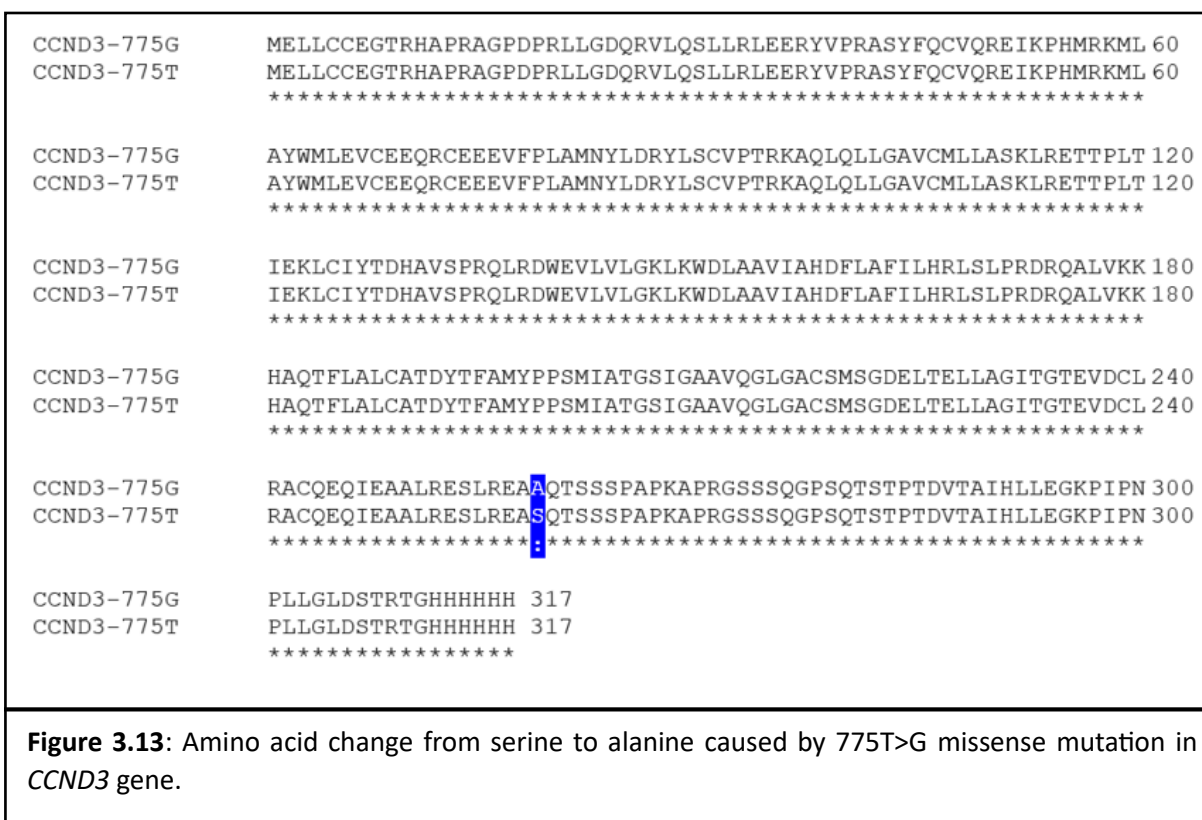
MELLCCEGTRHAPRAGPDPRLDQORVLQSLRLLEERYVPRASYFQCVQREIKPHMRKMLAYWM
LEVCEEQRCEEEVFPLAMNYLDRYLSCVPTRKAQLQLLGAVCMLLASKLRETTPLTIEKLCIYT
DHAIVSPQRLRDWEVLVLGKLKWDLAAVIAHDFLAFLHRLSLCRVTDRLW-RSTHRRSWRYALQ
TTHSRCTRHR-LRRAALGLQCRAWARALCLGTSSQSCWQGSALAKWTACGPVRSRSLHSGRAS
GKPLRPAPAQRPKPPGAPAAKGPAPALLQMSQPYTCSRVSLSLTLSSVSILRVPVIITITI

Swap- 178

MELLCCEGTRHAPRAGPDPRLDQORVLQSLRLLEERYVPRASYFQCVQREIKPHMRKMLAYWM
LEVCEEQRCEEEVFPLAMNYLDRYLSCVPTRKAQLQLLGAVCMLLASKLRETTPLTIEKLCIYT
DHAIVSPQRLRDWEVLVLGKLKWDLAAVIAHDFLAFLHRLSLSLAIDKLW-RNTLKLWLCAQ
TTLSLCIRQA-LRAPSGPQSRASELVA-AGTSSQSCWQGSALAKWTACGPVRSRSLHSGRAS
GKPLRPAPAQRPKPPGAPAAKGPAPALLQMSQPYTCSRVSLSLTLSSVSILRVPVIITITI

Figure 3.12: Effect of 507_508insCT on primary protein sequence for modified *CCND3* variants. Native sequence contains 317 amino acids, including the V5/His tags whilst others are truncated. Highlighted sequences are amino acid incorporated until a stop codon is encountered during translation.

This issue was recognised by aligning modified sequences to the current canonical ENSEMBL *CCND3* sequence rather than the sequence historically used for the project. During alignment another variation was identified that had been added to the database since the inception of the project: a single base pair substitution at position 775 from G to T within all *CCND3* sequences. At the time of synthesis outlined in section 3.2.1, excluding erroneous CT insertion, the historically used sequence was considered the accepted human mRNA sequence and matched the sequence cloned from HEK293 cells, which was corroborated by cloning the same sequence from T-rex HEK293 cells and HEK293 cells sourced from a different university laboratory (University of Manchester). The G variation at this position is now listed as a known mis-sense mutation leading to 259Ser>Ala amino acid change (Figure 3.13) (Ensembl, n.d.).



3.2.7 Site Directed Mutagenesis to Remove the CT Insertion in Modified *CCND3* ROI and Correct the ORF

As outlined above, discrepancies in the modified *CCND3* sequences were identified and therefore site directed mutagenesis was used for CT deletion and G>T substitution in the sequences previously generated to obtain the final, correct and in frame sequences. As the 775G variation originated from HEK293 cells used in this study and 775T is the canonical sequence, both were used and tagged and

untagged version of these sequences were made. Table 3.1 describes the sequences required and the abbreviations used to describe them throughout the remainder of this thesis. 507_508delCT was required for all modified samples and 775G>T substitution was carried out on *GsN* and *GvO* to generate T variant back ends of *CCND3* for tagged and untagged sets.

Table 3.1: The four sets of *CCND3* sequences generated and their abbreviations. V and s denote V5 tagged and stop codon respectively.

<i>CCND3</i> Variants	Modified Sequence Abbreviations					
	Native	Optimised	Disoptimised	Fast	Middling	Swap
G Tagged	<i>GvN</i>	<i>GvO</i>	<i>GvD</i>	<i>GvF</i>	<i>GvM</i>	<i>GvS</i>
G Untagged	<i>GsN</i>	<i>GsO</i>	<i>GsD</i>	<i>GsF</i>	<i>GsM</i>	<i>GsS</i>
T tagged	<i>TvN</i>	<i>TvO</i>	<i>TvD</i>	<i>TvF</i>	<i>TvM</i>	<i>TvS</i>
T Untagged	<i>TsN</i>	<i>TsO</i>	<i>TsD</i>	<i>TsF</i>	<i>TsM</i>	<i>TsS</i>

3.2.7.1 Primer design for site directed mutagenesis experiments

Primers for both G>T substitution and the CT deletion were designed using Agilent Primer Design webservice and are described in table 3.2 with imperfect primer-template duplexes shown in table 3.3. Imperfect primer-template duplex formations such as these are characterised as having a “higher free energy than fully matching duplexes” and “the difference between these two values is described as the energy cost” (Agilent, n.d.). Primers are designed to minimise this cost by ensuring the mismatched nucleotide bases are centred around the middle of each primer flanked by template-complementary strands to form a more stable duplex.

Table 3.2: Primers designed for site directed mutagenesis.

Required Mutagenesis	Sequence	Primer Name	Primer Sequence
G>T Substitution	Native & Opt	775G>T	5'-GAGCTGGTCTGAGAGGCTTCCCTGAGG-3'
		775G>Tantisense	5'-CCTCAGGGAAGCCTCTCAGACCAGCTC-3'
CT deletion	Opt	Optdel509_510	5'-GCACCGGCTCTCTCTGCCTCGTGACC-3'
		Optdel509_510antisense	5'-GGTCACGAGGCAGAGAGAGCCGGTGC-3'
	Dis	Disdel509_510	5'-CACCGGCTCTCTCTCCCCGCGAT-3'
		Disdel509_510antisense	5'-ATCGCGGGGAGAGAGAGCCGGTG-3'
	Fast	Fastdel509_510	5'-GCACCGGCTCTCTCTGCCTCGTGACC-3'
		Fastdel509_510antisense	5'-GGTCACGAGGCAGAGAGAGCCGGTGC-3'
	Mid	Middel509_510	5'-CACCGGCTCTCTCTGCCGCGTGAC-3'
		Middel509_510antisense	5'-GTCACGCGGCAGAGAGAGCCGGTG-3'
	Swap	Swapdel509_510	5'-GCACCGGCTCTCTCTCCCTCGCGATC-3'
		Swapdel509_510antisense	5'-GATCGCGAGGGAGAGAGAGCCGGTGC-3'

Table 3.3: Duplexes formed during site directed mutagenesis between primers used and template strand DNA. The template strand originates from *CCND3* sequences containing G at position 775 and insertion 507_508insCT.

Primer-Template Duplex	G>T Substitution	gagcctcaggggaagccgctcagaccagctccag 3'-ggagtcaccttcggagagctctggtcgag-5' 5'-cctcaggggaagcctctcagaccagctc-3' ctcggagtcaccttcggcgagctctggtcgaggtc
	CT Deletion	5'-gcaccggctctct--ctgcctcgtgacc-3' agacgtggccgagagagagacggagcactggcag tctgcaccggctctctctctgcctcgtgaccgctc 3'-cgtggccgagaga--gacggagcactgg-5'

3.2.7.2 Site directed Mutagenesis Results for 775G>T and 507_508delCT and In-Frame Confirmation

The commercially available QuikChange II site-directed mediated mutagenesis protocol was followed using the primers described in table 2.2. The PCR product was used in transformation of XL1-blue supercompetent cells (provided with the QuikChange II kit) and colonies were subsequently observed on ampicillin agar plates which were used to grow overnight cultures. Miniprep samples were then digested with restriction enzymes BamHI & XhoI to screen for correct band profiles of 5.6 kbp for pcDNA3.1 vector and 891/894 bp for *CCND3* tagged and untagged genes respectively, on a 1% agarose gel (figure 3.14). Electrophoresis lacks the precision to determine successful mutagenesis and it was necessary to confirmed this through commercial Genewiz sequencing service using primers in table 2.2. Figure 3.15 shows the subsequent sequencing around the site of mutagenesis and the flanking regions taken from the sequencing data. All sequences were also confirmed to be in-frame and thus the correct and final *CCND3* codon variant sequences had been generated.

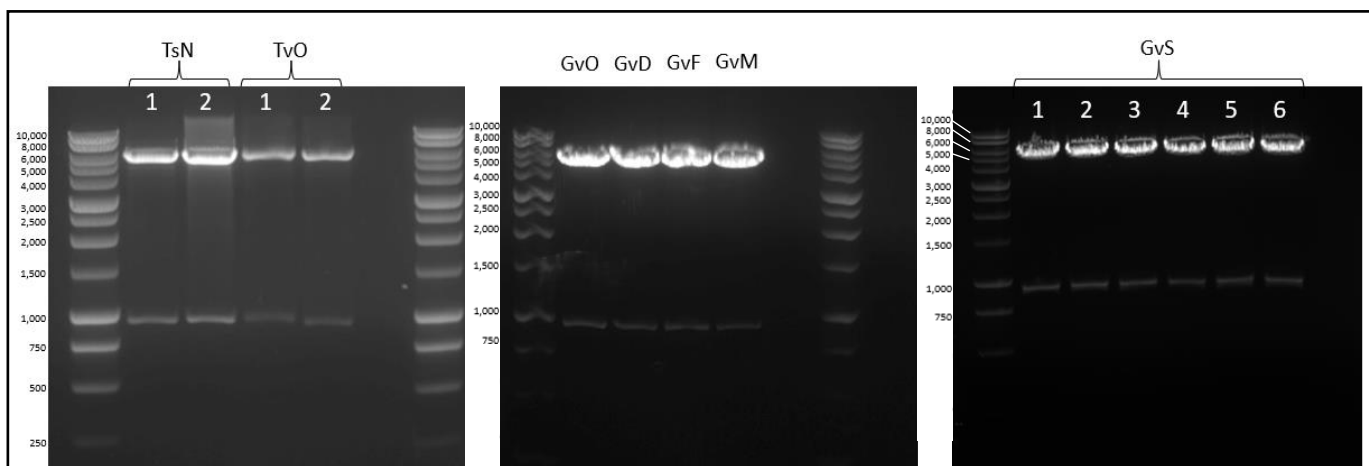


Figure 3.14: Restriction digest undertaken using BamHI and XhoI to screen transformed DH5 α E. coli colonies for cyclin D3 containing plasmids following site directed mutagenesis (SDM). Left: Untagged native and tagged optimised samples underwent SDM for 775G>T substitution to create tagged and untagged transcripts containing T at respective site. These were used to generate of tagged and untagged T variants through further cloning. Middle & Right: Tagged optimised, dis-optimised, fast, middling, and swap sequences underwent SDM to remove 505_506insCT. These ROI were inserted into *CCND3* backbone to generate the other *CCND3* modification sets. Bands were expected at 5.6kp for pcDNA3.1hygro (+) V5/His vector and 891/894bp for cyclin D3 tagged and untagged respectively.

i.

CCND3Nat-G	TGCACTCAGGGAGAGCCTCAGGGAAGCCCTCAGACCAGCTCCAGCCCAGCGCCCAAAGC	60
F1CCND3Nat-T	TGCACTCAGGGAGAGCCTCAGGGAAGCCCTCAGACCAGCTCCAGCCCAGCGCCCAAAGC	60

RevCCND3Nat-G	GGGCGCTGGGCTGGAGCTGGTCTGAGCGGCTTCCCTGAGGCTCTCCCTGAGTGCAGCTTC	60
R4CCND3Nat-T	GGGCGCTGGGCTGGAGCTGGTCTGAGCGGCTTCCCTGAGGCTCTCCCTGAGTGCAGCTTC	60

CCND3Opt-G	ACTCAGGGAGAGCCTCAGGGAAGCCCTCAGACCAGCTCCAGCCCAGCGCCCAAAGCCCC	60
F1CCND3Opt-T	ACTCAGGGAGAGCCTCAGGGAAGCCCTCAGACCAGCTCCAGCCCAGCGCCCAAAGCCCC	60

RevCCND3Opt-G	GCTTTGGGCGCTGGGCTGGAGCTGGTCTGAGGCTTCCCTGAGGCTCTCCCTGAGTGCA	60
R4CCND3Opt-T	GCTTTGGGCGCTGGGCTGGAGCTGGTCTGAGGCTTCCCTGAGGCTCTCCCTGAGTGCA	60

ii.

CCND3Opt-CT	TCCTGGCCTTCATTCTGCACCGGCTCTCTCTGCTCGTGACCGTCAGGCTCTGGTGAA	60
F1CCND3Opt	TCCTGGCCTTCATTCTGCACCGGCTCTCTCTGCTCGTGACCGTCAGGCTCTGGTGAA	58

RevCCND3Opt-CT	TTCACCAGAGCCTGACGGTCACGAGGCAGAGAGAGAGCCGGTGCAATGAAGGCCAGGA	60
R4CCND3Opt	TTCACCAGAGCCTGACGGTCACGAGGCAGAGAGAGAGCCGGTGCAATGAAGGCCAGGA	58

CCND3Dis-CT	TCCTGGCCTTCATTCTGCACCGGCTCTCTCTCTCCCCGCGATCGCCAAGCCCTCGTCAA	60
F1CCND3Dis	TCCTGGCCTTCATTCTGCACCGGCTCTCTCTCTCCCCGCGATCGCCAAGCCCTCGTCAA	58

RevCCND3Dis-CT	GTTTTTTGACGAGGGCTTGGCGATCGCGGGGAGAGAGAGAGCCGGTGCAATGAAGGC	60
R4CCND3Dis	GTTTTTTGACGAGGGCTTGGCGATCGCGGGGAGAGAGAGAGCCGGTGCAATGAAGGC	58

CCND3Fast-CT	TTTCTGGCCTTCATTCTGCACCGGCTCTCTCTGCTCGTGACCGACAGGCTTTGGTG	60
F1CCND3Fast	TTTCTGGCCTTCATTCTGCACCGGCTCTCTCTGCTCGTGACCGACAGGCTTTGGTG	58

RevCCND3Fast-CT	ACCAAAGCCTGTCGGTCACGAGGCAGAGAGAGAGCCGGTGCAATGAAGGCCAGGAAAT	60
R4CCND3Fast	ACCAAAGCCTGTCGGTCACGAGGCAGAGAGAGAGCCGGTGCAATGAAGGCCAGGAAAT	58

CCND3Mid-CT	TCCTGGCCTTCATTCTGCACCGGCTCTCTCTGCGCGTGACCGACAGGCATTGGTAAA	60
F1CCND3Mid	TCCTGGCCTTCATTCTGCACCGGCTCTCTCTGCGCGTGACCGACAGGCATTGGTAAA	58

RevCCND3Mid-CT	ATGCCTGTGCGTCACGCGGCAGAGAGAGAGCCGGTGCAATGAAGGCCAGGAAATCATG	60
R4CCND3Mid	ATGCCTGTGCGTCACGCGGCAGAGAGAGAGCCGGTGCAATGAAGGCCAGGAAATCATG	58

CCND3Swap-CT	TTCTGGCCTTCATTCTGCACCGGCTCTCTCTCCCTCGCGATCGACAAGCTTTGGTGA	60
F1CCND3Swap	TTCTGGCCTTCATTCTGCACCGGCTCTCTCTCCCTCGCGATCGACAAGCTTTGGTGA	58

RevCCND3Swap-CT	AAGCTTGTGATCGCGAGGGAGAGAGAGAGCCGGTGCAATGAAGGCCAGGAAATCATG	60
R4CCND3Swap	AAGCTTGTGATCGCGAGGGAGAGAGAGAGCCGGTGCAATGAAGGCCAGGAAATCATG	58

Figure 3.15: Relevant section of Genewiz Sanger sequencing to confirm successful site directed mutagenesis. i. 775G>T substitution. ii. 507_509delCT.

3.2.8 Preparation and Restriction Digest Confirmation of the Final *CCND3*-pcDNA3.1 Constructs

Successfully site directed mutated sequences were then subjected to the same cloning strategy outlined in section 3.2.2 for Ts, Tv, and Gs variants; *CCND3* sequences were cloned into pET23b(+), modified ROIs were added and cloned back into pcDNA3.1 vector. *TvO* and *TsN* were used to provide T variant *CCND3* backbone for tagged and untagged versions. Maxiprepped pcDNA3.1-*CCND3* plasmids were screened for correct band profiles on 1% agarose gel using BamHI and XhoI restriction enzyme digests to yield bands at 5.6 kbp for the pcDNA3.1 vector backbone and 891/894 bp for *CCND3* tagged and untagged respectively (Figure 3.16). GeneArt sequencing confirmed correct *CCND3* sequences and Expasy webservice confirmed the correct translation frame, concluding generation of *CCND3* sequences.

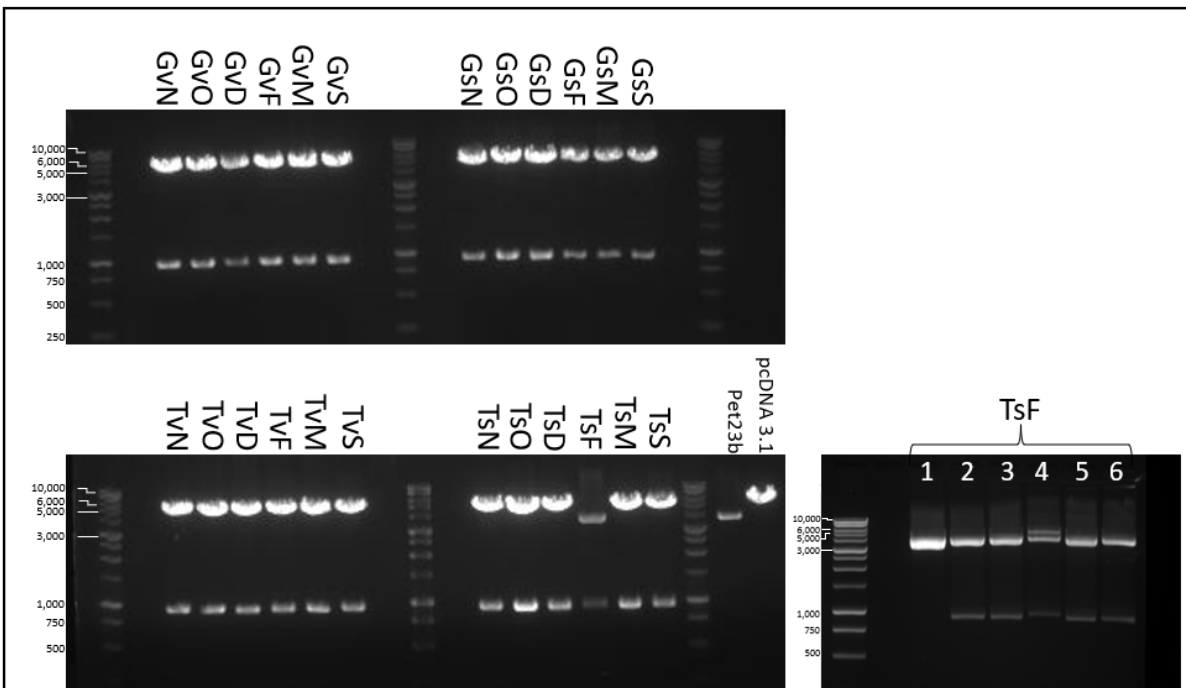


Figure 3.16: Restriction digest undertaken using BamHI&XhoI to confirm maxiprepped modified Cyclin D3 sequences. Modified sequences include optimised, dis-optimised, fast, middling, and swap codon changes. 4 sets of these were generated: G at position 775 variants, tagged (top left) and untagged (top right), and T at position 775 variants, tagged (bottom left) and untagged (bottom right). Empty pcDNA3.1 *hygro* (+) V5/His vector and Pet23b(+) were also included. An additional colony screen for TsF samples was repeated separately. Bands were expected at 5.6kbp for pcDNA3.1 *hygro*(+) V5/his vector, 3.6kbp for Pet23b(+) vector, and 891/894bp for cyclin dropout, tagged and untagged respectively.

3.3 Generation of Open Reading Frame Codon Variant *Noggin* Sequences

3.3.1 Cloning strategy for generating codon transcript variants of *Noggin*

As for the *CCND3* transcript, previous work in the laboratory at Kent had identified a region of the *Noggin* transcript (encoded by the *NOG*) that was predicted to contain stretches of slowly decoded codons and hence would result in slow elongation. Unlike the *CCND3* gene, the region of interest (ROI) in *Noggin* was at the 5' end of the sequence. Once again, design of ROI codon modifications focused on assessing the affects that individual codon decoding times of each codon in the ROI had on overall expression levels. While fully optimising decoding speed may intuitively lead to the highest increase in expression as previously described, other factors such as ribosomal collision may hinder expression rates (Meydan & Guydosh, 2021), conversely the presence of slower codons may lead to more effective concurrent elongation of multiple ribosomes on mRNA at once. With this in mind, multiple sequences were designed as described in Figure 3.17. Further, as the modified sequence in this case was at the 5' end of the ORF, it was adjacent to the 5'UTR and translation initiation sites and thus was hypothesised to potentially have a greater impact on protein synthesis than the *CCND3* modifications which were located centrally in the ORF.

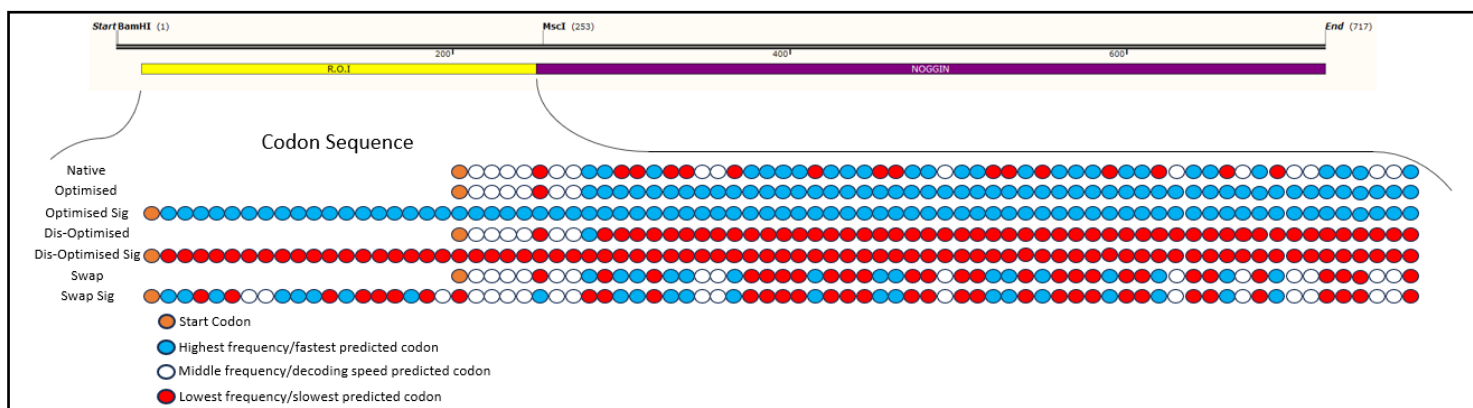


Figure 3.17: Altered codon sequences for *Noggin* modifications. The identified potential signal sequence (leading 28 codons) was either modified correspondingly to the rest of the sequence removed; 'Sig' sequences have modified signal sequences. Codons present in native sequence were replaced with synonymous codons to create modified nucleotide sequences whilst conserving amino acid sequence. Codon usage was selected for desired effect on decoding speed which was rationalised by predicted decoding speed of codons using our model for elongation. Using these criteria for guidance, blue spheres indicate codons identified as optimal, red spheres indicate codons identified as dis-optimal, and white spheres are neither optimal nor dis-optimal. Orange spheres denote start codon.

Optimised, disoptimised, and swap versions of *Noggin* were designed in the ROI using alternative codons as described. As for the *CCND3* transcript designs, optimised and disoptimised sequences utilised the fastest or slowest codon for their respective amino acid. The swap sequence replaced each fastest codon with the slowest codon for that amino acid and vice versa. Using SignalP-5.0 webservice, a potential signal sequence was identified (figure 3.18) at the 5' end of the ORF and versions of *Noggin* with this sequence modified or removed were created for optimised, dis-optimised and swap sequences resulting in two sets of modified *Noggin* sequences. The natural stop codon at the end of the transcript was removed allowing for reading into the V5 and His tags to be incorporated in final protein sequences. The sequences designed and generated, and their abbreviations are outline in table 3.4.

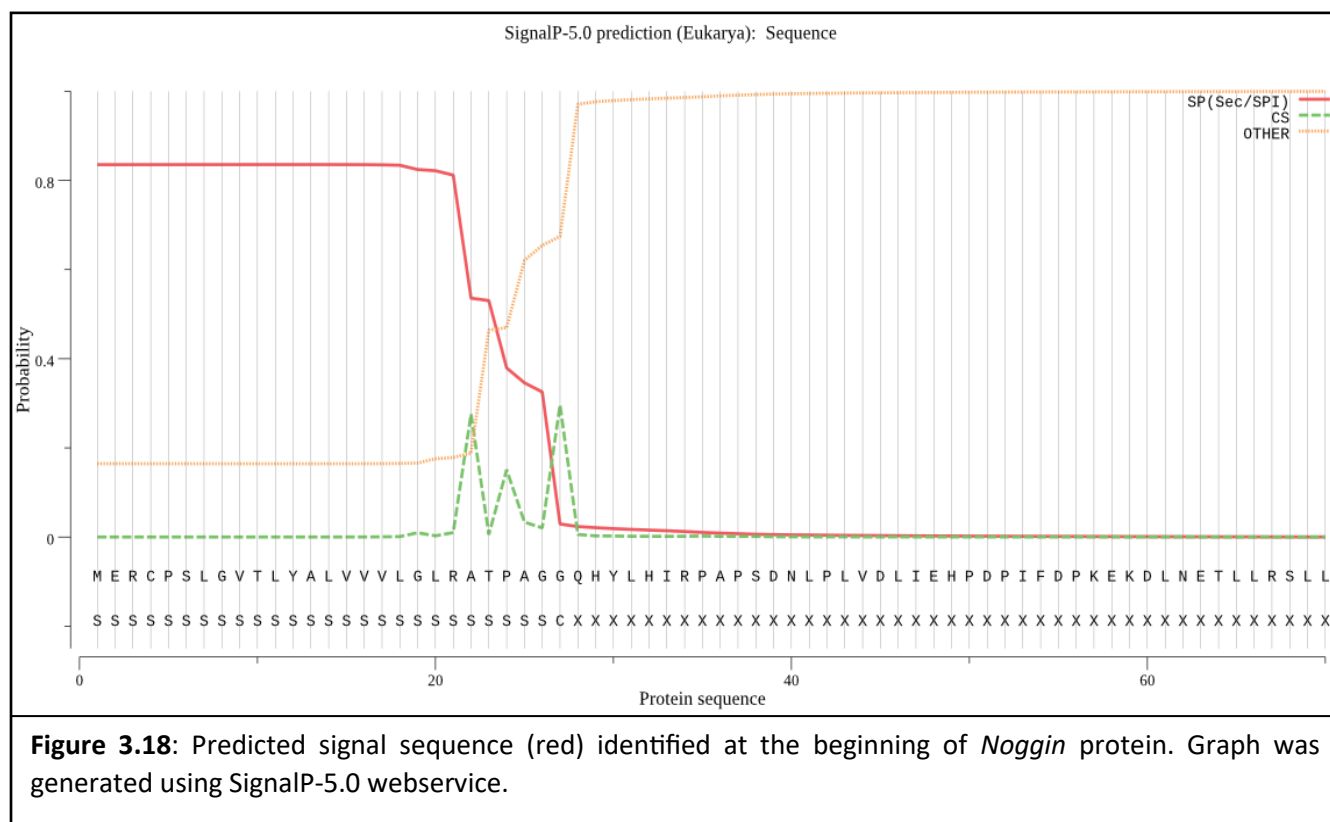
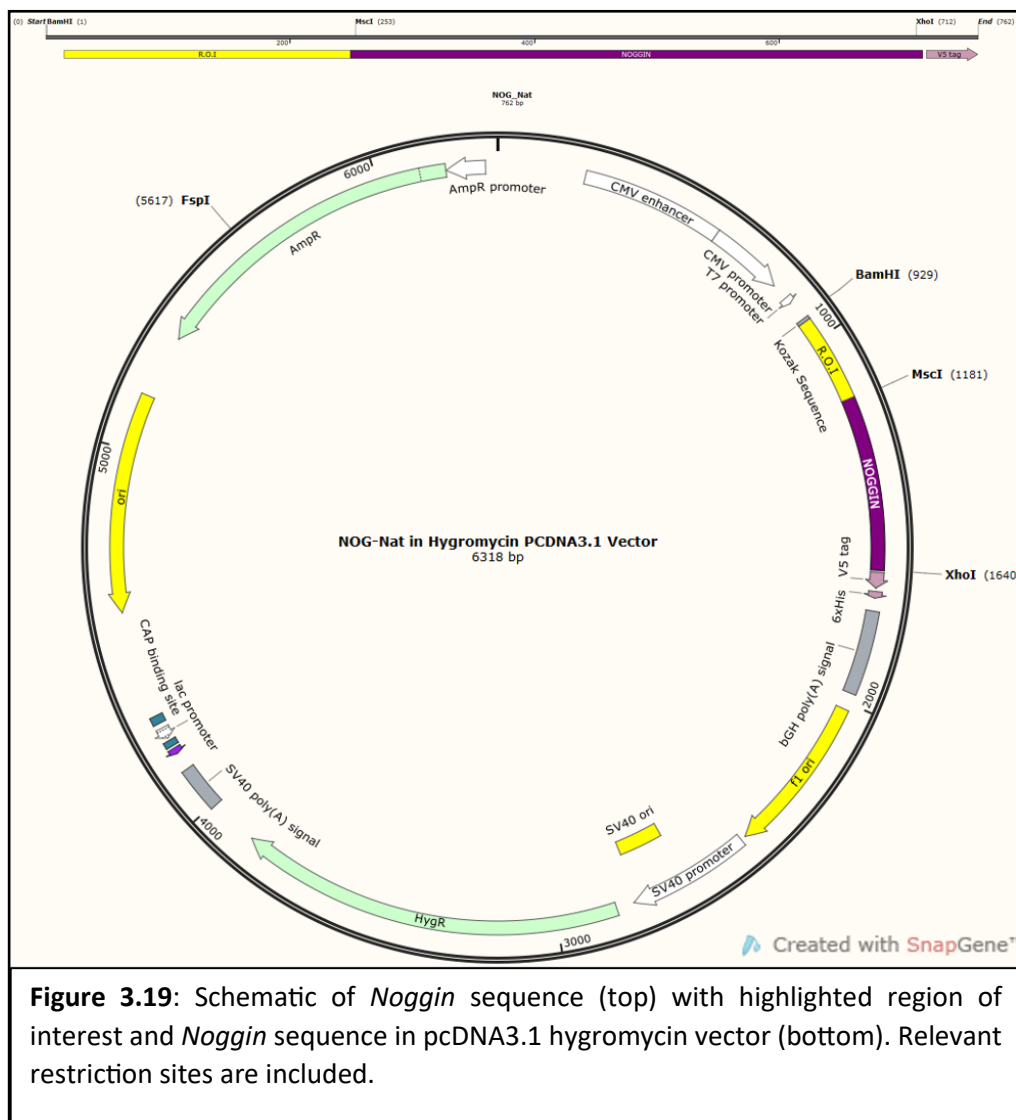


Table 3.4: *Noggin* sequences and their abbreviations.

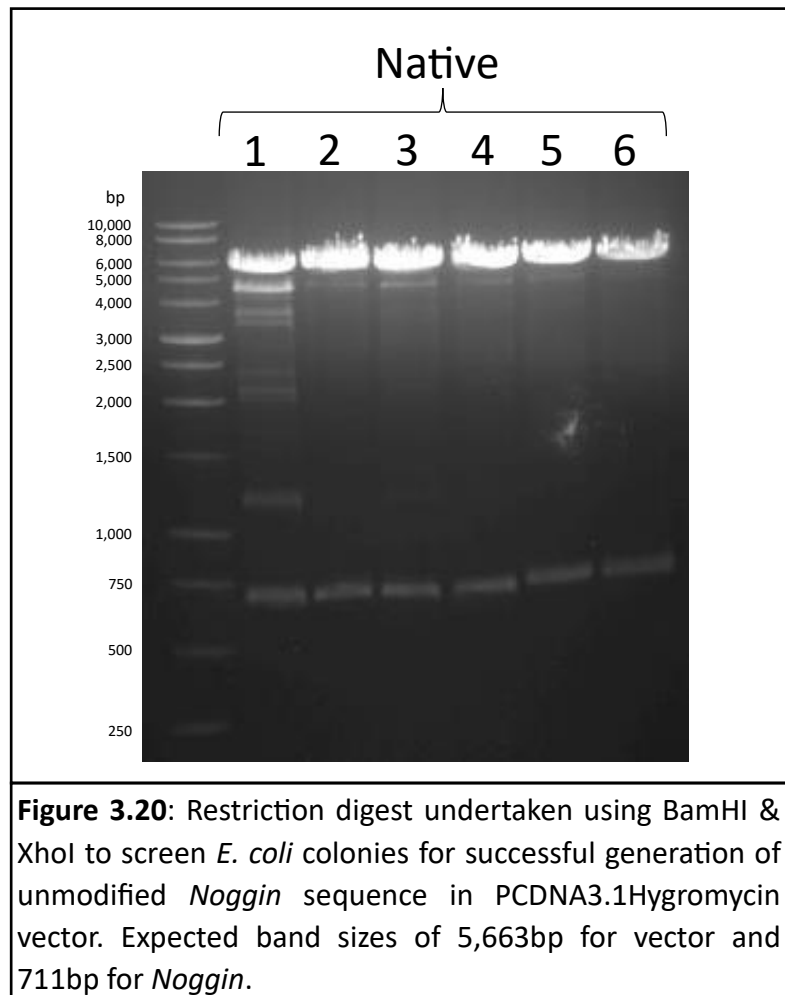
Noggin variants	Modified Sequence Abbreviations			
	Native	Optimised	Dis- optimised	Swap
Modified Signal Sequence		<i>SNogO</i>	<i>SNogD</i>	<i>SNogS</i>
Removed Signal Sequence	<i>NogN</i>	<i>NogO</i>	<i>NogD</i>	<i>NogS</i>

To incorporate the modified sequence into the ROI, restriction site BamHI present in pcDNA3.1 vector and internal restriction site MscI in the *Noggin* sequence were used as seen in figure 3.19.



3.3.2 Cloning of Alternative *Noggin* ROIs into the *Noggin* Sequence

The *Noggin* native sequence, along with the optimised, sig-optimised, dis-optimised, sig-dis-optimised, swap, and sig-swap insert sequences were ordered commercially from TWIST. The *Noggin* native sequence in the supplied TWIST vector, as well as pcDNA3.1 vector, were then subjected to restriction digestion using BamHI & XhoI restriction enzymes. Digest reactions were then run on 1% agarose gels to separate the digested fragments. The *Noggin* native sequence and pcDNA3.1 digested vector were then ligated and transformed into calcium competent DH5 α *E. coli* cells grown on ampicillin agar plates. Colonies were then screened by restriction digest screening as shown in Figure 3.20 below.



3.3.3 Insertion of the *Noggin* Native Sequence into the pcDNA3.1 Vector

Cultures from successful colonies in section 3.3.2 were minipreped to purify the plasmid DNA and used to insert the modified ROIs into. Restriction enzymes BamHI & MscI were used to clone and introduce the modified ROI sequences into the *Noggin* backbone. Following minipreped overnight cultures from successful transformed colonies, BamHI & MscI restriction digests were again used to screen for successful ligations with correct band profiles displayed in figure 3.21. Colonies with the expected bands at 252 bp for ROI and 6.1 kbp for the pcDNA3.1 backbone and remainder of the *Noggin* gene, were commercially sequenced by Genewiz for further confirmation of successful modified ROI incorporation. The Expasy translate tool was used to confirm the correct sequence was present and the sequence was in frame. Figure 3.21 shows the agarose gel analysis of the screened colonies after digestion with colonies 1 being chosen for optimised, sig-optimised, sig-dis-optimised, swap, and sig-swap samples. Colony 2 was chosen for the disoptimised sequence.

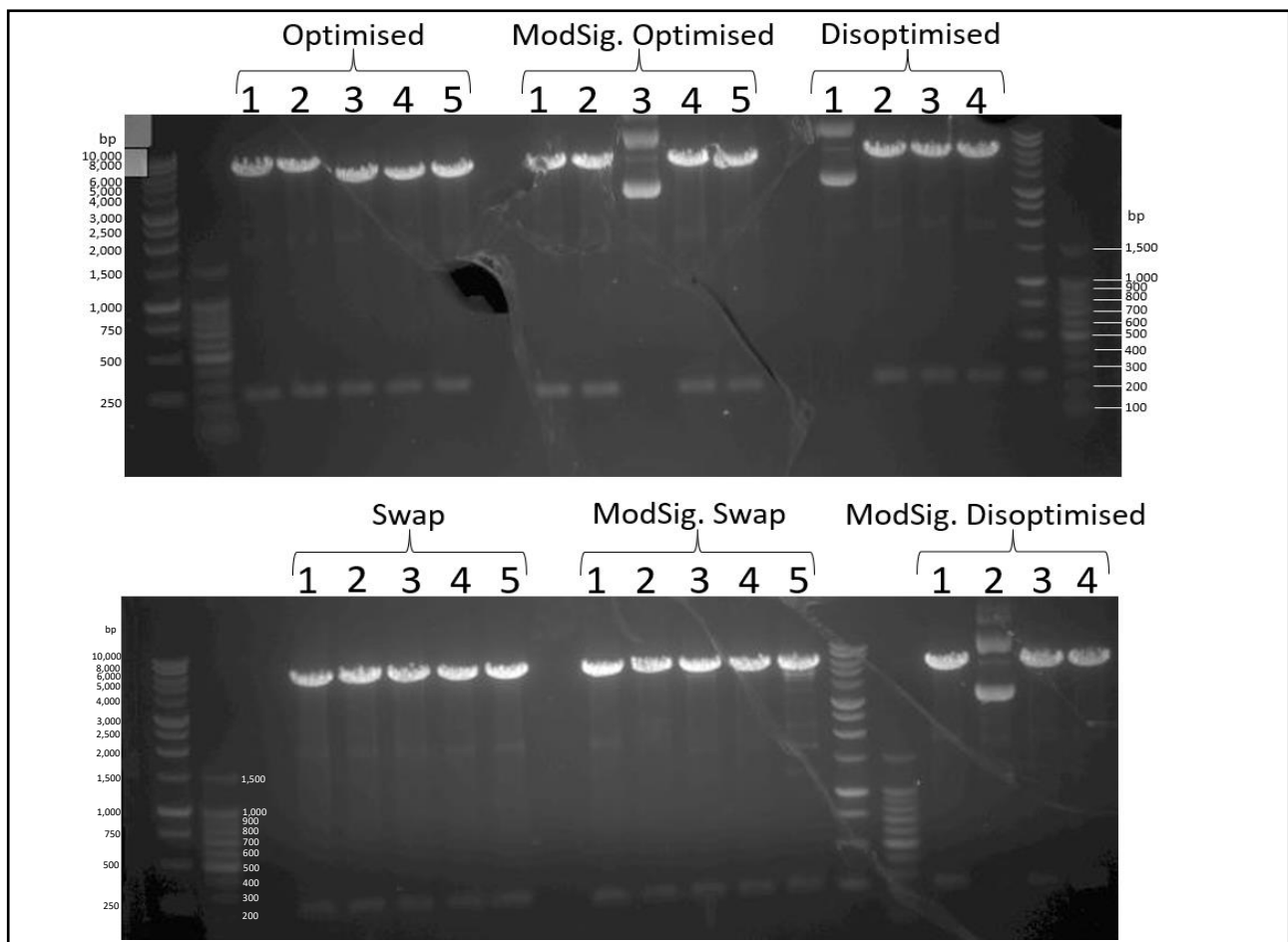
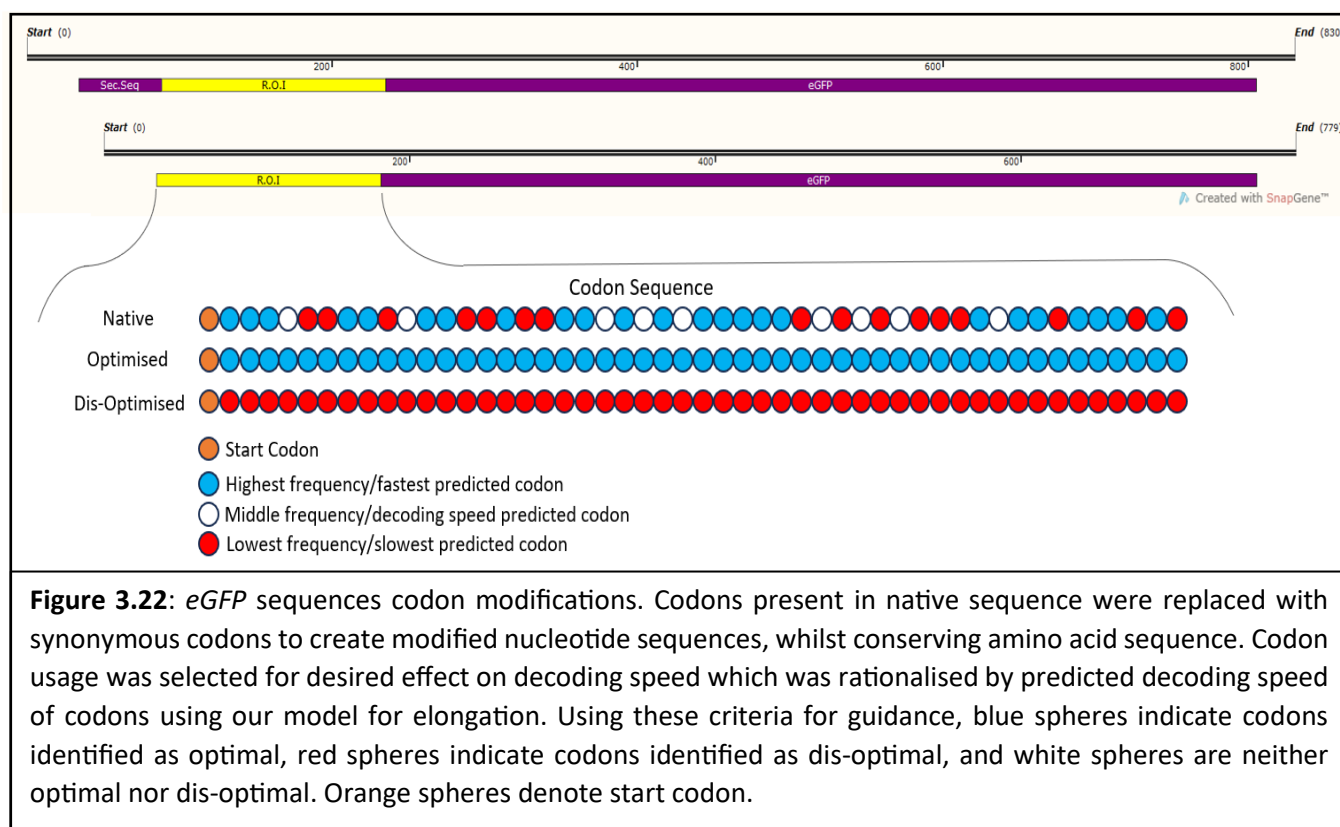


Figure 3.21: Restriction digest undertaken using MscI & XhoI to screen *E. coli* colonies for successful generation of modified *Noggin* sequences with and without signal sequences modified: optimised, disoptimised, and swap in PCDNA3.1Hygromycin vector. Expected band sizes of 6,066bp for plasmid and 252bp for modified *Noggin* section.

3.4 Generation of eGFP Alternative Codon Transcript Sequences

3.4.1 Design of Alternative eGFP Sequences and Cloning Strategy

In addition to *CCND3* and *Noggin*, *eGFP* ORF codon variant transcripts were designed to help further explore the impact of codon manipulation and predicted decoding speeds on protein production. In this case, ROI modifications focused on assessing the effects on overall expression levels of altering codon decoding times within this region. Once again, multiple *eGFP* ROI sequences were designed as described in figure 3.22. Optimised and dis-optimised sequences again utilised the fastest or slowest decoded codon for their respective amino acid respectively. Secretory *eGFP* was also designed (with an ER secretory signal peptide at the 5' end) and used the same ROI modifications as described in figure 3.22, creating two sets of *eGFP* constructs which are outlined in table 3.5. This would allow further investigation of the impact on protein synthesis of modifying the decoding speed of the signal peptide on proteins.



The *eGFP* sequence did not contain any suitable internal restriction sites close to the end of the ROI, therefore Gibson assembly was used to construct sequences. As described in subsection 1.6.1.1, Gibson assembly does not rely on restriction enzymes but rather complementary base pairing of overhanging

segments generated through 5' exonuclease activity. Strings of DNA were ordered commercially from TWIST consisting of the unaltered 3' half of *eGFP*, which was used for all sequences, as well as the different 5' halves for *eGFP* native, optimised and dis-optimised designed sequences, with and without secretory sequences. As seen in figure 3.23, these strings were designed to contain overlapping sequences at the ends, aligning to the pcDNA3.1 vector digested with BamHI and XhoI (pink and orange sections), and also to each other to complete the *eGFP* sequence (blue sections). All DNA reagents were treated with 5' exonuclease, 3' extension activity of a DNA polymerase and DNA ligase activity as per the NEB Gibson assembly protocol. This created complimentary overhanging ends (exonuclease activity) which were annealed, extended (polymerase activity), and any nicks ligated (ligase activity). This created the *eGFP* containing pcDNA3.1 plasmids which were transformed into calcium competent DH5α *E. coli* cells.

Table 3.5: *eGFP* transcript variant sequences designed and investigated in this chapter.

<i>eGFP</i> variants	Modified Sequence Abbreviations		
	Native	Optimised	Dis-optimised
Secretory- <i>eGFP</i>	<i>sGFPN</i>	<i>sGFPO</i>	<i>sGFPD</i>
<i>eGFP</i>	<i>GFPN</i>	<i>GFPO</i>	<i>GFPD</i>

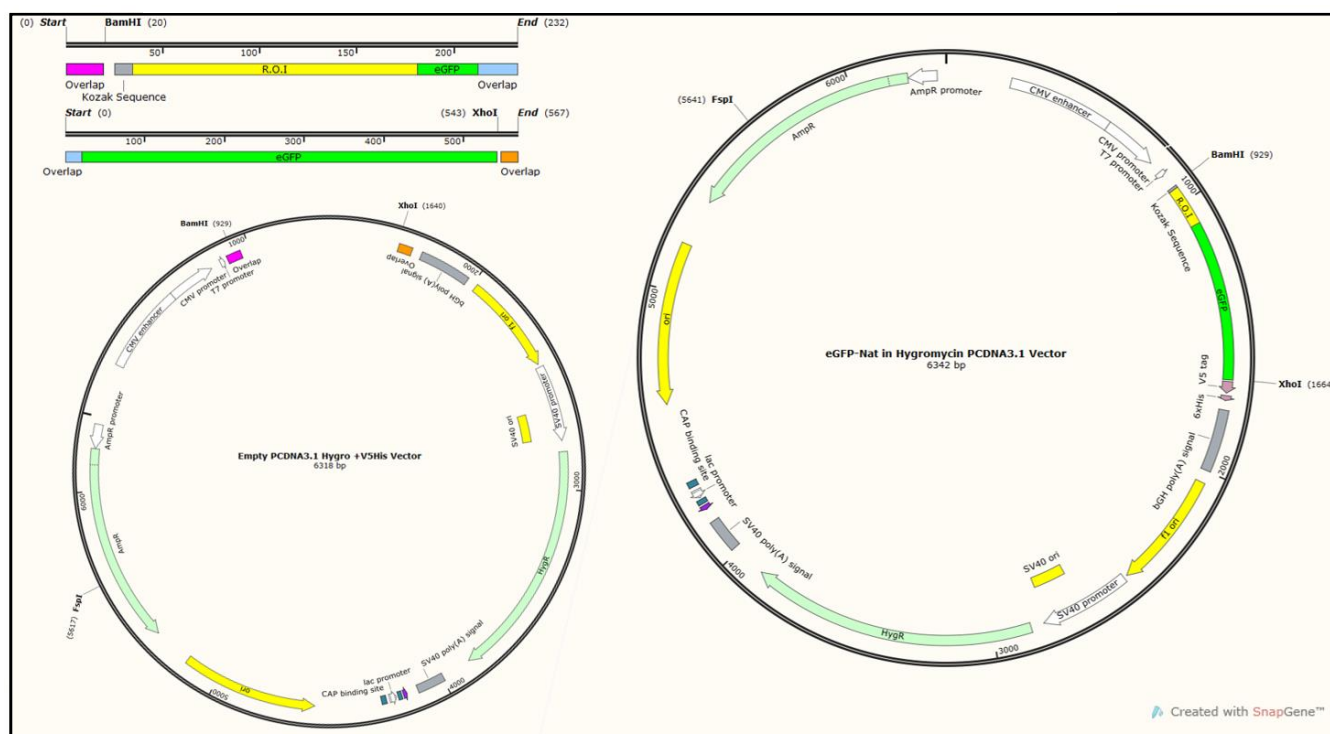
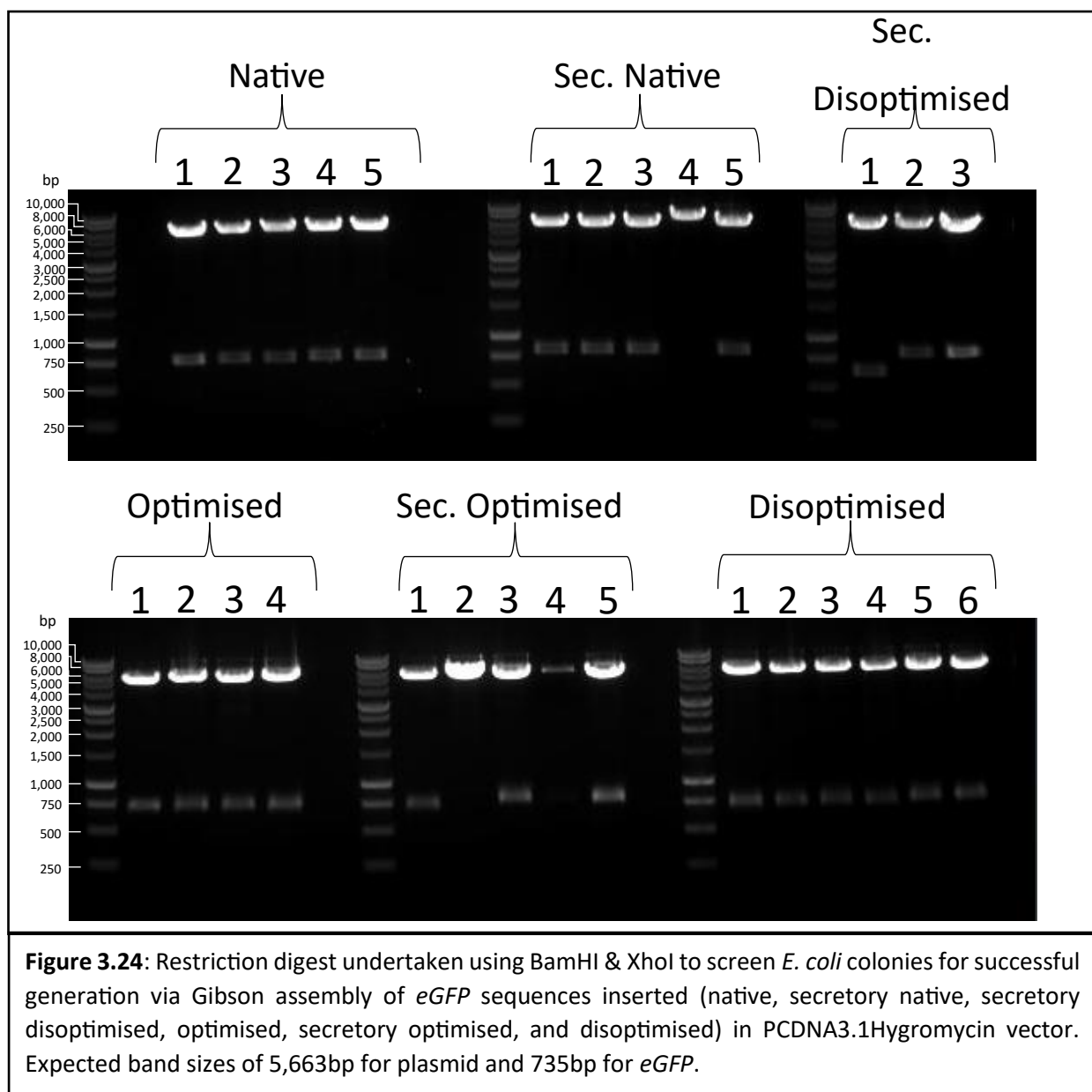


Figure 3.23: Cloning strategy for *eGFP*. Left: BamHI and XhoI digested vector and *eGFP* strings with overlapping regions in matching colours. Pink and blue flanked string contains region of interest differing between sequences whereas blue and orange flanked string is consistent throughout all sequences. Gibson assembly protocol was used to assemble. Right: complete plasmid post Gibson assembly.

3.4.2 Assessment of the Generation of eGFP Codon Variant Transcript Sequences

As described for the generation of the *CCND3* and *Noggin* constructs, plasmid DNA was purified from overnight cultures grown from Gibson assembly colonies and digested with the restriction enzymes BamHI and XhoI to confirm the correct/expected band profiles were observed. Bands were expected at 5.6 kbp for pcDNA3.1 vector and 735 bp/771 bp for eGFP and secretory eGFP respectively (Figure 3.24). Sequences were then confirmed by commercial Genewiz Sanger Sequencing using their T7 and BGHR primers and the presence of the correct translation open reading frame confirmed by Expasy translation tool.



3.5. Alterations to mRNA Predicted Secondary Structures Caused by Codon Alterations

As mentioned in detail in the introduction chapter (chapter 1), mRNA stability is heavily influenced by its nucleotide sequence which were directly altered for our model transcripts by cloning techniques (Hanson & Coller, 2017; Wu et al., 2019). These alterations may impact the mRNA's ability to form secondary structures through hydrogen bond interactions between its own nucleotides. A lack of secondary structures can render transcripts unstable reducing their persistence within cells and thereby preventing their translation. Conversely secondary structures can also impact elongation efficiency as they can cause ribosomes to stall during elongation, reducing translation rates and thereby protein production (Tholstrup et al., 2012). In extreme cases however, secondary structures cannot be resolved and translation must be abandoned with ribosome rescue machinery mediating this process (Doma & Parker, 2006). ViennaRNA web services were used to generate predicted secondary structures for each transcript used in this study to provide insight into the severity of changes caused by codon modifications (Tables 3.6). mRNA modelling made it clear that all codon modifications caused changes in mRNA secondary structures when compared to the native transcripts and is further discussed in chapter 6.

Table 3.6A: Predicted secondary mRNA structures for *eGFP* and secretory *GFP* generated using Vienna RNAseq.

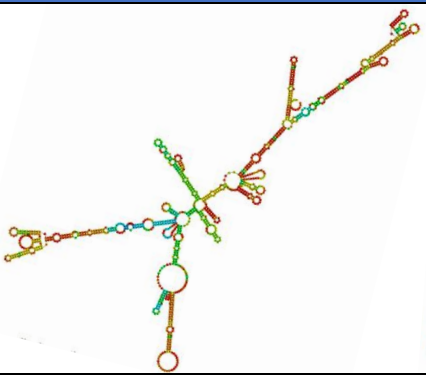
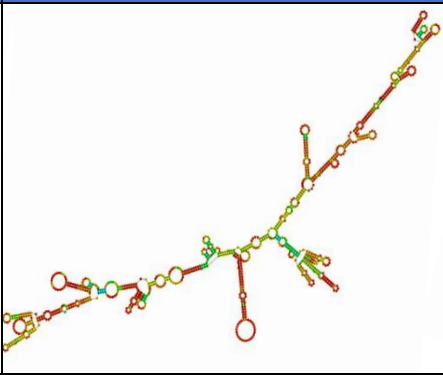
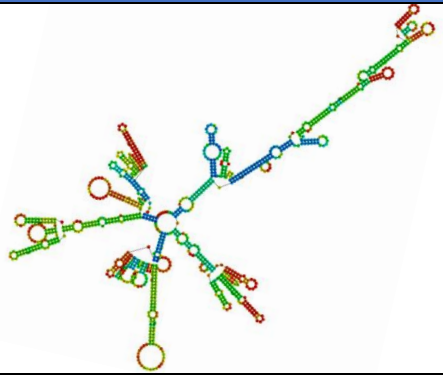
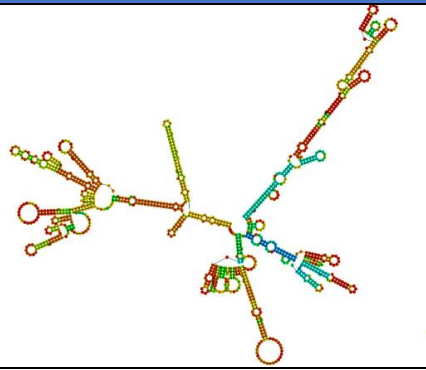
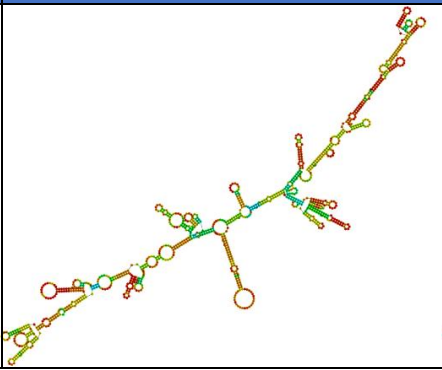
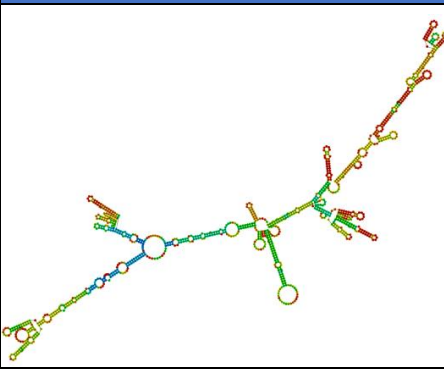
Native	Optimised	Disoptimised
		
Secretory Native	Secretory Optimised	Secretory Disoptimised
		

Table 3.6B: Predicted secondary mRNA structures for Tv variant *CNND3* generated using Vienna RNAseq.

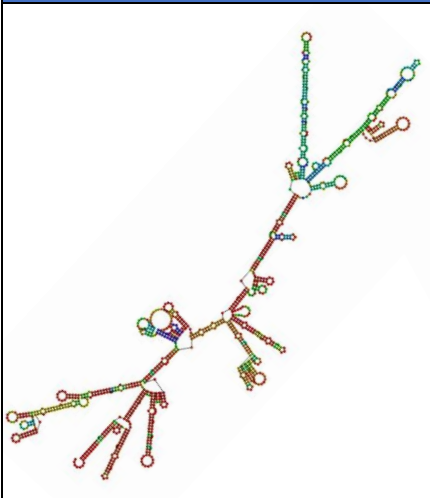
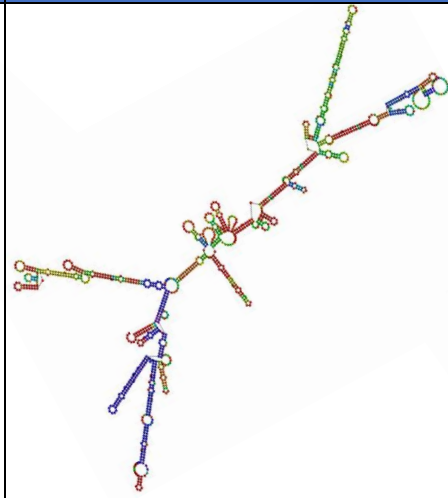
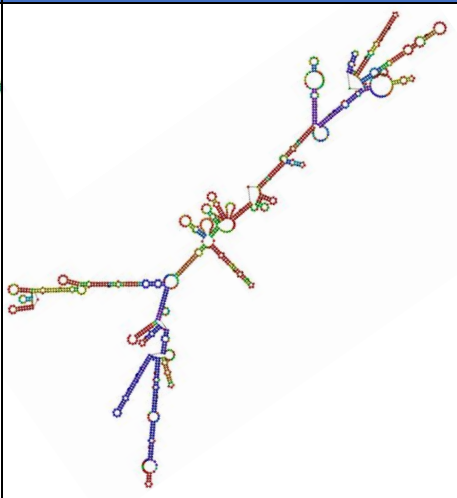
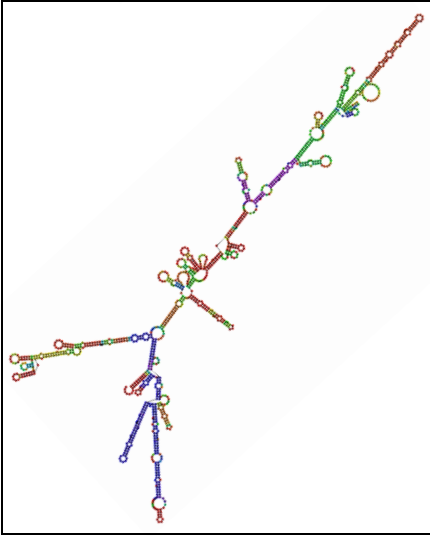
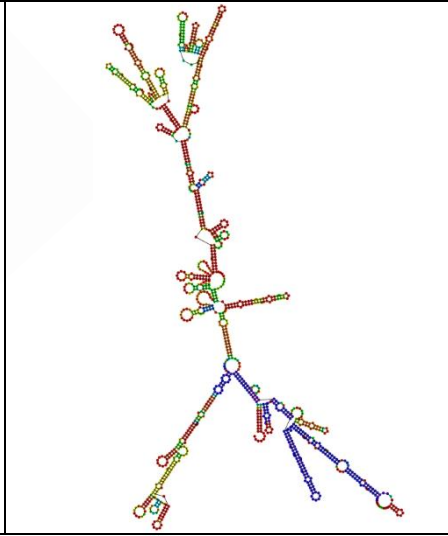
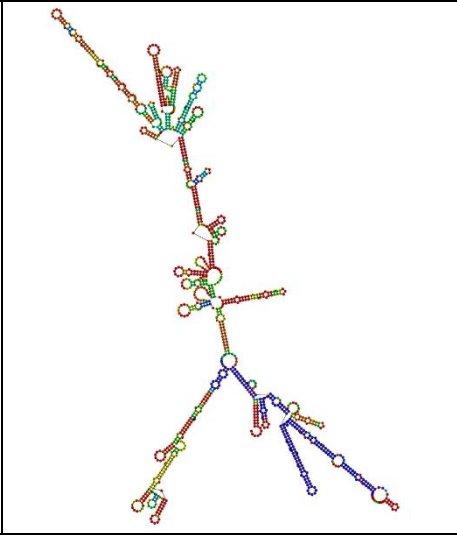
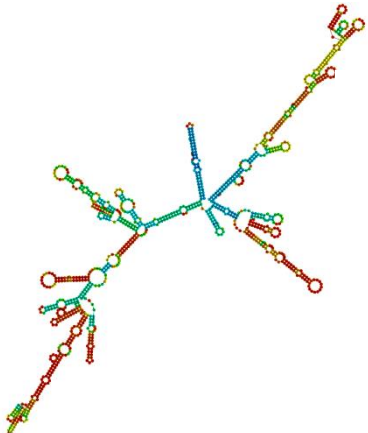
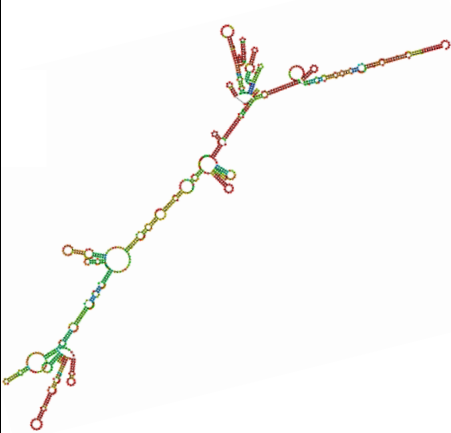
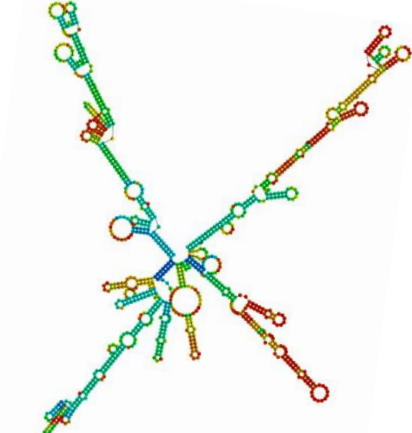
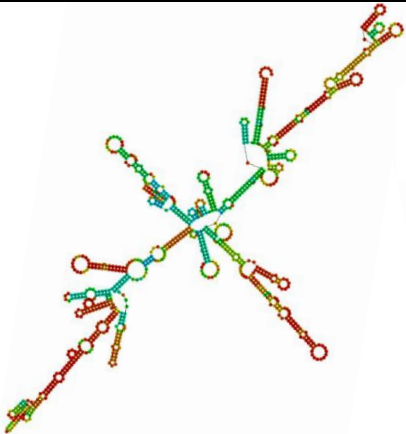
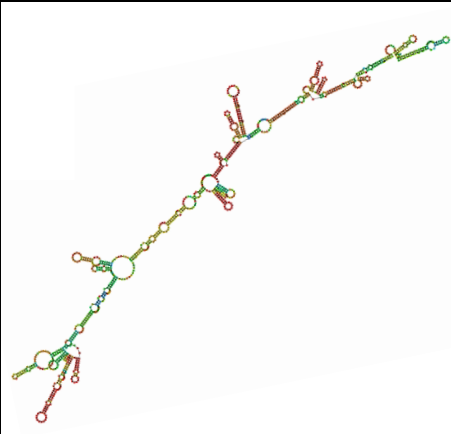
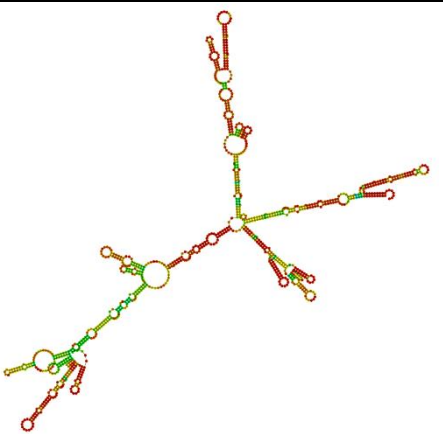
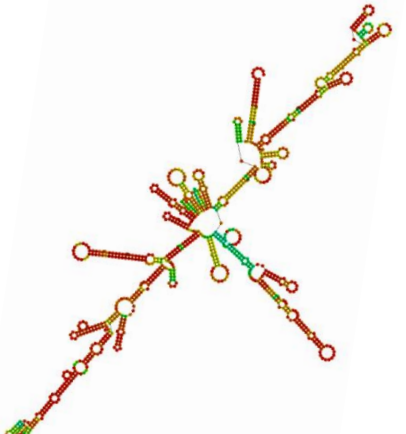
<i>TvN</i>	<i>TvO</i>	<i>TvD</i>
		
<i>TvF</i>	<i>TvM</i>	<i>TvS</i>
		

Table 3.6C: Predicted secondary mRNA structures for *noggin* and secretory *noggin* generated using Vienna RNAseq.

<i>sNOG-O</i>	<i>sNOG-D</i>	<i>sNOG-S</i>
		
<i>NOG-O</i>	<i>NOG-D</i>	<i>NOG-S</i>
		
<i>NOG-N</i>		
		

Chapter 4

Determining the Impact of Codon Manipulation in Target Areas of the Open Reading Frame of Cyclin on Subsequent Recombinant, Exogenous CCND3 Protein Production in HEK293 Cells

4.1 Introduction

As discussed in more depth in the main Introduction chapter of this thesis (Chapter 1), multiple factors can impact the three basic steps of translation elongation during protein synthesis which include: codon decoding/tRNA selection, peptide bond formation, and translocation of mRNA-tRNA complex. These factors include slow peptidyl-transfer kinetics between specific amino acids such as Pro-Pro peptide formation, downstream secondary mRNA structures inhibiting ribosomal progression, availability of associated translocation machinery, amongst others (Carlberg et al., 1990; Pavlov et al., 2009; Tholstrup et al., 2012). As outlined in previous chapters, this study has a particular focus on how codon decoding times due to tRNA abundance and near-cognate tRNA abundance impact protein synthesis rates for a given transcript. With regard to this, ribosomes can experience pausing/stalling whilst waiting for a correct tRNA to enter the A site. Evidence suggests that the initial open reading frame codons proceeding the start codon have a greater impact on elongation rates in some mRNA transcripts (Chu et al., 2012). Codons in this region that have corresponding abundant tRNAs enhance elongation speed whereas codons whose tRNA counterparts are at low or limited abundance reduce predicted and experimental elongation speed, resulting in ribosomal stalling. This stalling not only impacts elongation rates, but can hinder translation initiation as well (Chu et al., 2012; Chu, Kazana, et al., 2014).

Out of the three model transcripts investigated in this study, *CCND3* is of particular note as the identified ROI, hypothesised to be elongation control relevant, is towards the middle of the mRNA open reading frame sequence. Thus, alterations made to the codons across this area, whilst maintaining the same coded for amino acid sequence, should have little or no impact on the initiation step of mRNA translation. mRNA translation initiation is a well-documented regulatory node of translation that can control both transcript specific and global translation rates and hence protein synthesis, mostly via the availability and de/phosphorylation of key translation initiation factors but is not the focus of this study. Modifying codons,

and hence codon usage in a given transcript encoding for the same amino acid sequence, immediately following the AUG start codon provides insight into the impact of codon usage upon elongation control with the added caveat that manipulation of elongation decoding times in this section of sequence can either increase availability of the start codon and potentially increase initiation, or conversely decrease the availability of the start codon and hence potentially reduce initiation rates. Modifications made further downstream from the AUG start codon are likely to mitigate this effect. Thus, although initiation and elongation are often thought of as independent processes, the control of each is linked; however precisely the impact of each process upon the other is complex and difficult to untangle but should be considered in interpreting the findings of codon ‘optimisation’ strategies to increase elongation speed and hence protein synthesis.

To address the impact of different codon usage, predicted decoding speed and hence protein synthesis in the ROI of *CCND3*, the ROI codon usage was modified in the ROI to generate transcript sets encoding the same amino acid sequence while utilising different/alternative codons as described in Chapter 3. Codon usage was modified based upon tRNA abundance and decoding times used in the elongation model to generate transcripts with different predicted decoding speeds. Multiple sets of modified *CCND3* sequences were designed, generated and studied to account for disparities between the canonical *CCND3* sequence on Ensembl database and the sequence amplified from the HEK293 cell line directly used in this study. Both sequences had C-terminal V5 tagged and untagged versions generated to allow comparison of total and exogenous *CCND3* protein production upon transfection of the different constructs into HEK293 cells. Both canonical and innate sequences were included to generate relevant data beyond the HEK293 cell line (canonical), and to also evaluate the same *CCND3* sequence as that endogenously produced (innate). Table 4.1 reports the sequences generated and used, and their abbreviations, in this Chapter on *CCND3*.

Table 4.1: Table of four sets of *CCND3* sequences and their abbreviations. “v” and “s” denote V5 tagged and stop codon present respectively.

CCND3 Variants	Modified Sequence Abbreviations					
	Native	Optimised	Disoptimised	Fast	Middling	Swap
G Tagged	GvN	GvO	GvD	GvF	GvM	GvS
G Untagged	GsN	GsO	GsD	GsF	GsM	GsS
T tagged	TvN	TvO	TvD	TvF	TvM	TvS
T Untagged	TsN	TsO	TsD	TsF	TsM	TsS

In addition to codon usage, gene expression within cells can alter greatly in response to various stresses perceived by the cell and occur swiftly, with mechanisms of control often operating at the post-transcriptional level. Synthesis of proteins involved in a specific stress response typically show mRNA translation upregulation to increase protein synthesis and assist in reducing potential harmful effects caused by the stress, safeguarding the cell. One such stress condition relevant to this study is cold stress and the mammalian cold shock response, which in part hinges upon elongation control and codon usage in mRNA transcripts. During mild hypothermic conditions, global protein synthesis reductions occur throughout the cell, with notable exceptions. Specific proteins such as CIRP and RBM3 are upregulated in their production upon cold-shock perception in mammalian cells (Al-fageeh, 2008; Peretti et al., 2015). These proteins are believed to improve targeted mRNA stability and aid in translation of other transcripts whose expression is presumably required by the cell in response to the perceived cold-stress. Phosphorylation of the mRNA translation elongation factor eEF2 on Thr⁵⁶ also increases upon cold-stress in mammalian cells, which reduces global protein synthesis (Knight et al., 2015). Phosphorylation of eEF2 reduces/prevents its ability to enter the A site, reducing the rate of ribosomal translocation and elongation as a whole, thus globally reducing protein synthesis rates. Thus, CIRP and eEF2-/eEF2-P ratios were also analysed in this study to confirm that reduction in incubation temperature from 37°C to 32°C was appropriate to initiate a cold shock response without causing cell death. This culture condition (32°C) would allow the investigation of whether the codon modified sequences impacts elongation control at different temperatures and if changes in expression at the protein level can subsequently be tuned and observed in cells cultured at 32°C as oppose to 37°C.

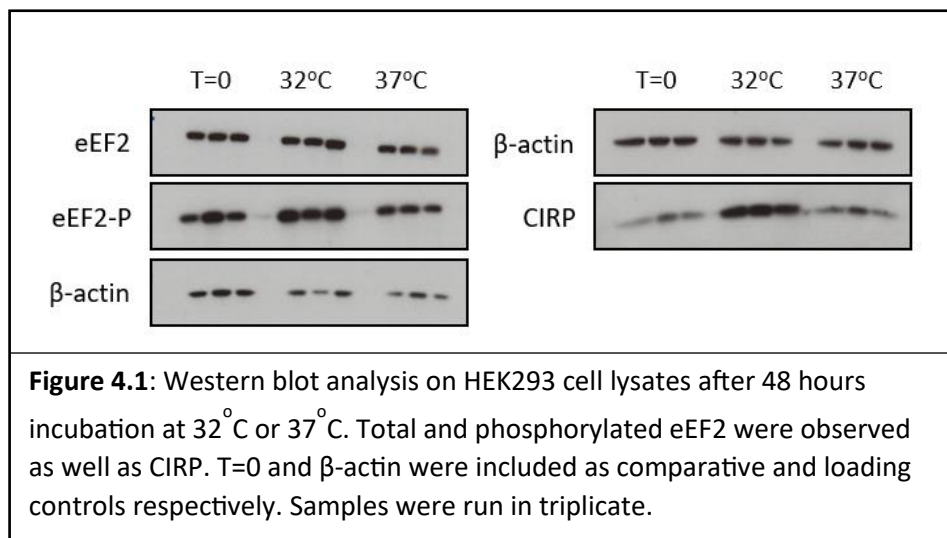
Preliminary studies of CCND3 protein production also showed differences in gene expression dependent on culture time, therefore multiple time points were also included in the study to monitor any change in expression trends between modified sequences over time and throughout culture. The impacts of codon manipulation on cell growth were a general consideration at any temperature due to CCND3s involvement in cell cycle regulation. D type cyclins are a core component in cell cycle machinery whose expression drives cell proliferation, aiding in G1/S transition through associated kinase activity (Sherr, 1995). As such, unregulated expression of *CCND3* is associated with tumorigenesis including leukaemia and bladder cancer (Johnson, 1998; Sawai et al., 2012). Conscious of this, growth of *CCND3* expressing stably transfected cell lines was also monitored in the study using an xCELLigence system as outlined below.

4.2 Western Blot Analysis of Temperature Dependent Production of Cold Inducible RNA-Binding Protein (CIRP), Eukaryotic Elongation Factor 2 (eEF2), and Phosphorylation of eEF2 (eEF2-P)

As discussed in section 4.1, specific transcripts evade reduction in global translation and hence protein production in response to temperature decreases. Many of these are involved in the cellular cold shock response and there is evidence that the mechanism by which they achieve this evasion involves mRNA translation elongation control. To confirm that cold-shock of the model HEK293 cells used in this project is sensed at 32°C as determined by changes in production and phosphorylation of key proteins central to global translation and cold-shock responses, proteins whose production or phosphorylation is known to change upon cold-shock perception, CIRP, eEF2 and eEF2-P were analysed from HEK293 cells grown at 37°C and 32°C from samples collected 48 hours post incubation. *CIRP* is known to increase in expression in response to mild hypothermic conditions (32°C) (Al-fageeh, 2008) and the ratio of non-phosphorylated eEF2 to phosphorylated eEF2-P ratio is modulated if a cold shock response is activated. Phosphorylated eEF2 (eEF2-P) is inactive in terms of translation elongation. To assess this, HEK293 cell samples were seeded at 1×10^6 cells/mL for T=-24 h, and then either continued to be incubated at 37°C or cold-treated samples were then transferred to 32°C and this time designated as T=0 h.

Total *eEF2* expression was more-or-less consistent across both temperatures investigated, however the amount of phosphorylated protein showed a marked increase at 32°C compared to 37°C samples after 48 h (Figure 4.1). Inactivation of eEF2 through phosphorylation at position Thr⁵⁶ prevents peptidyl-tRNA translocation from ribosomal A to P site, thereby suppressing global protein production as part of the cold shock response (Knight et al., 2015). eEF2-P is considered a marker of cold-shock and confirmed 32°C as an appropriate temperature to achieve desired conditions within the cell for evaluating expression of our model transcripts, with a readout of protein production, of CCND3, GFP and Noggin. Contrary to the general response of decreased mRNA translation and protein synthesis, *CIRP* expression increased at 32°C compared to 37°C; within the cytosol its function is to increase mRNA stability during cold shock and other stresses (Qiang et al., 2013). This increase in expression shows that transcript specific control can allow increased expression of transcripts at reduced temperature compared to more physiologically relevant temperature. It has been postulated that this form of upregulation may be exploitable to increase recombinant protein production at 32°C compared to that at 37°C. The combination of the increased eEF2-P and CIRP production supplies precedent for examining expression of our model transcripts and proteins at 32°C as well as 37°C. Although general growth and protein synthesis may be reduced at 32°C, cold shock

control mechanisms may be exploitable to increase endogenous and exogenous recombinant protein production if engineered correctly.



4.3 Preliminary Confirmation of CCND3 Protein Production from all Codon Designed Transcripts Through Western Blot Analysis

Due to the initial difficulties with *CCND3* expression outlined in Chapter 3, production of protein from the different *CCND3* codon designed transcripts was confirmed for each set of modifications using transient transfection experiments before undertaking more extensive analysis. To that end, HEK293 cells were transiently transfected with *CCDN3* Tv, Ts, Gv, and Gs constructs and western blotting of cell lysates was used to confirm and determine expression. The purpose of the experiment and resulting western blot analysis shown in Figure 4.2 was not to compare expression levels from the different codon variants but rather to qualitatively confirm presence of recombinant *CCND3* protein from the constructs generated. As such, the maximum volume of cell lysate that could be loaded onto the gel were used to produce reduced samples with less emphasis on the precise protein masses loaded. Exposure times required to observe *CCND3* were considerably shorter when compared to the blots shown in Section 3.2.6. This was true for all subsequent *CCND3* blots shown in this chapter where samples were adjusted to contain equal total protein masses and the same antibodies were used. These data confirmed and demonstrated exogenous *CCND3* expression in all cases from the codon variant constructs and that the constructs were suitable for further experiments to investigate the impact of manipulation of the ROI codon usage on subsequent *CCND3* expression in HEK293 cells.

4.3.1 Qualitative Western Blot Analysis for CCND3 Protein Production from the Different Codon Variant Transcripts

Expression for all *CCND3* codon variant constructs was verified by western blot analysis of samples from HEK293 cells 24 hours post transient transfection with the relevant constructs. Each construct successfully produced CCND3 at the expected band size with the exception of the *TsF* modification, where transfected cell samples showed no evidence of any expression. This necessitated further cloning and sequence confirmation before expression was confirmed through an additional transfection and subsequent blot analysis (Figure 4.2E). There was a small but clear difference in size between V5/His tagged and untagged CCND3 protein, with tagged versions appearing heavier (running slower and less distance into the gel) than untagged samples as expected. This discrepancy is due to the tag adding to the mass of the protein and was consistent throughout all tagged samples. An additional difference between tagged and untagged CCND3 samples was a double-banding profile observed in the tagged samples, best observed in Gv vs Tv western blot (Figure 4.2C) and again in Figure 4.3 and Figure 4.4. This double band observation is not uncommon when blotting for CCND3 and is most likely due to phosphorylation differences within the population resulting in two distinct bands close in size. Untagged samples were observed as only one band; CCND3 is more homogenous and uniform in size for these samples suggesting that the presence of the tag may impact phosphorylation or some other post-translational modification, or alternatively that phosphorylated and non-phosphorylated CCND3 is indistinguishable without the presence of the tag.

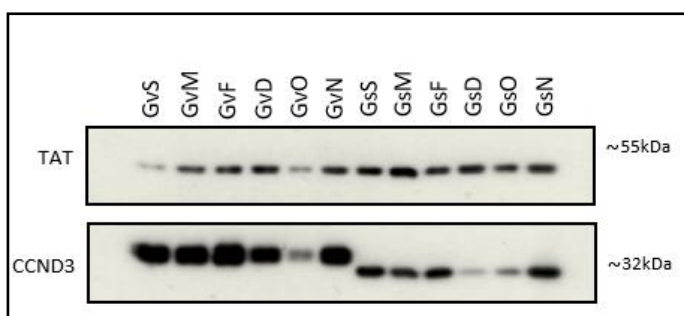


Figure 4.2A: Western blot analysis for Gv and Gs *CCND3* variant samples harvested from HEK 293 cells grown at 37°C, 24 hours post transient transfection via lipofectamine. Samples included *CCND3* native, optimised, fast, middling, or swap sequences. TAT was used as a loading control.

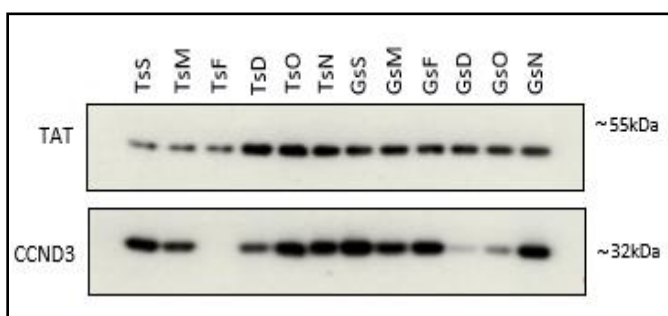


Figure 4.2B: Western blot analysis for Ts and Gs *CCND3* variant samples harvested from HEK 293 cells grown at 37°C, 24 hours post transient transfection via lipofectamine. Samples included *CCND3* native, optimised, fast, middling, or swap sequences. TAT was used as a loading control.

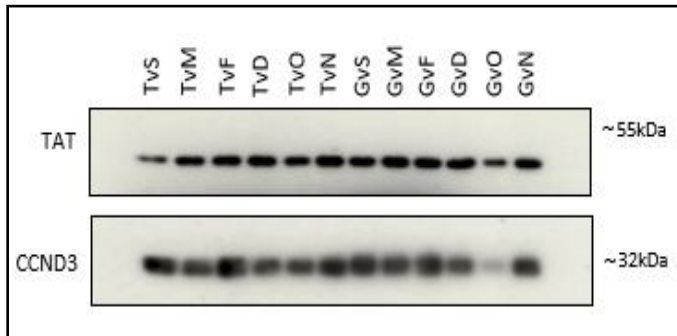


Figure 4.2C: Western blot analysis for *CCND3* Gv and Tv variant samples harvested from HEK 293 cells grown at 37°C, 24 hours post transient transfection via lipofectamine. Samples included *CCND3* native, optimised, fast, middling, or swap sequences. TAT was used as a loading control.

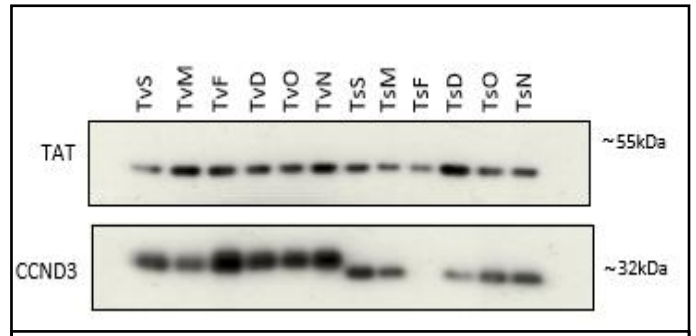


Figure 4.2D: Western blot analysis for *CCND3* Ts and Tv variants harvested from HEK 293 cells grown at 37°C, 24 hours post transient transfection via lipofectamine. Samples included *CCND3* native, optimised, fast, middling, or swap sequences. TAT was used as a loading control.

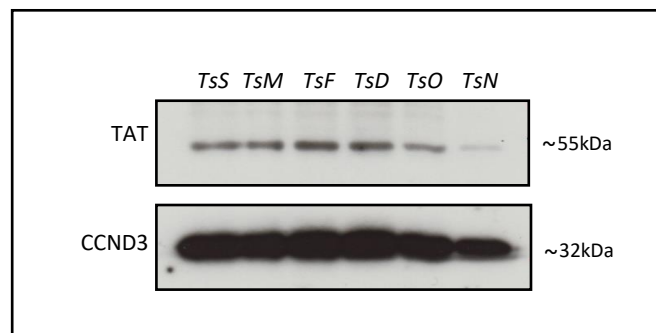
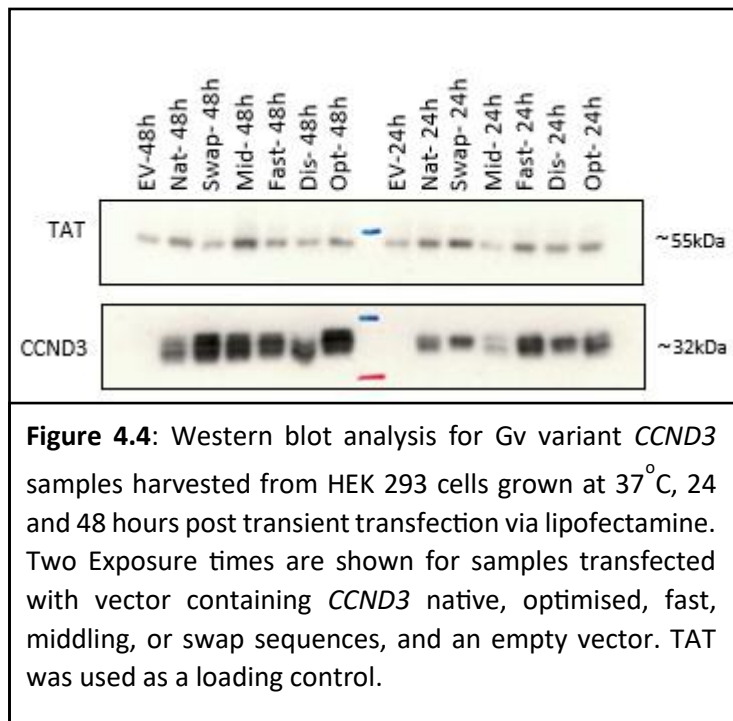
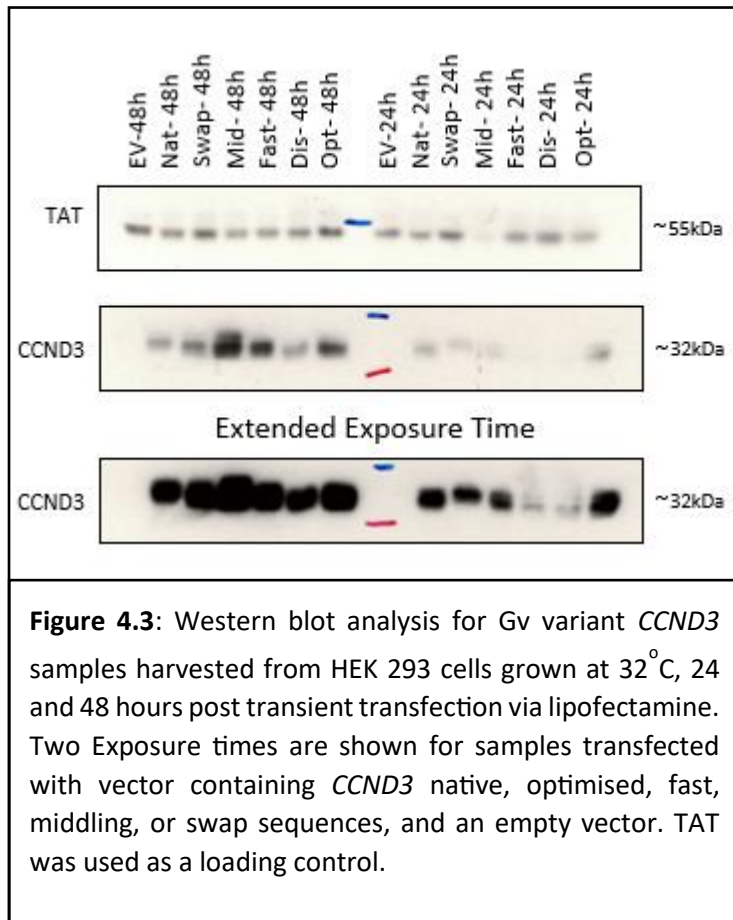


Figure 4.2E: Western blot analysis for Ts *CCND3* variants harvested from HEK 293 cells grown at 37°C, 24 hours post transient transfection via lipofectamine. TAT was used as a loading control.

4.3.2 Assessing Time Dependent Transient Expression Profiles of CCND3 at 32°C and 37°C for Gv constructs by Western Blot

To allow comparison of protein production across different samples, equal total protein loading was achieved by first confirming protein content of cell lysates through using a Bradford assay and sample amounts were adjusted accordingly to give the same concentration of total protein loaded. The double band profile described in section 4.3.1 was present and observed for all tagged samples. Empty vector (EV) samples had no detectable CCND3 signal suggesting endogenous material was not detectable and all 32 kDa bands were due to the presence of exogenous, recombinant CCND3. At both temperatures investigated (32°C and 37°C) different expression trends were observed between the two time points. At 32°C, detectable 24 hour CCND3 protein production for *GvF* and *GvD* samples was low and required increased exposure time to ensure an observable and distinct expression signal on film (10 minutes opposed to 10 seconds). This was not reflected at 48 hours with *GvF* omitting stronger signals than *GvN* and *GvS*. *GvD* production yielded similar, if slightly reduced, signals compared to *GvN* by 48 hours. *GvM* had the highest production 48 hours post transfection but was produced less than *GvN* at 24 hours post-temperature shift. 37°C incubation resulted in less drastic differences in protein production between samples as multiple exposures were not necessary to visualise each band. At 24 hours, *GvM* gave the weakest signal, *GvF* the strongest with the others being comparable to *GvN*. By 48 hours however, *GvM* expression was stronger than *GvN*, whilst *GvO* and *GvS* had more intense bands than *GvF* and was the highest expressing construct/variant. All samples had increased expression of CCND3 by 48 hours compared to 24, suggesting CCND3 concentration within the cell was increasing/accumulating over the 48 h time period. This is of note as endogenous CCND3 is heavily regulated through proteasome mediated degradation, but the exogenous expression appeared to continually increase CCND3 in amount. The different expression observed of modified sequences at various time points suggested that multiple time points were required to accurately assess the impact codon sequence modifications have on CCND3 protein production within the HEK293 cell. As such, 24 hour and 48 hour time points were assessed for CCND3 production in subsequent experiments.



4.4 Investigation into the Impact of *CCND3* Transcript Codon Sequence Modifications on Transient *CCND3* Protein Production at 32°C and 37°C 24 and 48 Hours Post Transfection

The results described in sections 4.2 and 4.3 guided further experiment rationale for assessing any effects of transcript sequence modifications on protein production levels. In this section, each set of modified *CCND3* transcript sequences and their subsequent impact on protein production was examined at 37°C and 32°C culture temperatures 24 and 48 hours post transfection. For this, 6 well plates were seeded with HEK293 cells which were transfected with plasmids containing *CCND3* sequence or 'empty' for transfection control in duplicate. Un-transfected 'parental' cell line was also included as an additional control. As multiple data points for each sample were desirable for comparison, multiple blots were performed; single wells for each sample were used to compare 24 and 48 h *CCND3* protein production directly and blots containing replicates were used to focus on one time point per temperature. Issues were experienced with the tubulin housekeeping TAT antibody meaning some results were difficult to interpret with confidence as equal loading amount could not be confirmed. However, equal total protein was loaded based upon a Bradford assay. Error bars were plotted using the standard error of the mean for densitometry graphs containing multiple repeats per experiment condition. No overlap of error bars between samples was determined to be an indicator that results were different in a statistically significant way. P- values were also calculated and can be found in the appendix section.

4.4.1 Expression Analysis of *CCND3* for Gs Codon Transcript Variant Modifications at 32°C and 37°C

4.4.1.1 Comparison of *CCND3* Transient Protein Production at 24 and 48 Hour Post Transfection from Codon Variant Transcripts

When samples from cells transfected with the *CCND3* transcript variants were cultured at 37°C, the resulting western blot analysis (Figure 4.5A) of the Optimised and Native bands were comparable visually after 24 hours, however once normalised for the loading control, the Optimised transcript was shown to result in increased protein expression compared to the Native (Figure 4.5B). This increase was also observed after 48 hours, both visually on the western blot and in the densitometry analysis where there was a 33% increase in the band intensity compared to that from the Native construct after 48 hours. Other transcript variant sequence samples gave reduced protein production compared to the native transcript sequence at both time points. The Swap and Fast transcripts gave *CCND3* protein production at 24 hours that was higher than that from the Disoptimised and Middling transcript variants, with the disoptimised transcript resulting in the lowest protein production as confirmed by densitometry and in line with the model predictions of decoding speeds. This trend was not mirrored in the 48 hour samples however.

Resolution of the Disoptimised, Fast, Middling, and Swap CCND3 derived protein signals at 48 hours were poor but unexpectedly the Disoptimised signal was the most intense of these. Protein derived from the Swap transcript variant gave the lowest protein production at 48 hours as shown by densitometry analysis of the bands, followed closely by the Middling transcript derived protein band.

For 32°C transcript variant sample western blot analysis (Figure 4.6A), the tubulin TAT controls showed higher than expected variation resulting in the western blot visualisation and densitometry results appearing contradictory to each another. Blots were repeated with fresh samples following further Bradford assays but tubulin loading variation did not improve. It is possible that tubulin expression was not constant between HEK293 cultures even though this is usually considered a ‘good’ housekeeping protein, or that unequal amounts of protein were loaded between samples despite Bradford assaying samples. This particular issue with tubulin visualisation is unlikely to be caused by the antibody difficulties seen in blots displayed later on in this thesis as that would affect all samples present on the blot, not only individual samples. Loading control issues are discussed further in section 6.3. When focussing on CCND3 protein production as analysed by western blot, the transcript optimised sequences yielded the strongest signal at both 24 and 48 hour time points as might be expected. However, this was not reflected in the densitometry analysis as fast transcripts appeared to give the highest protein production at 24 h and native transcripts yield the highest protein production at 48 h relative to the tubulin TAT loading controls (Figure 4.6B).

Some cell lysate samples showed high variance in tubulin (TAT) signals. For example, it is unlikely that the tubulin (TAT) protein present in the native transcript transfected cell 48 hour sample would be as low as visualised when compared to other tubulin bands from transcript variant transfected samples at 48 hours. This was despite this sample, and preparations of this, being analysed multiple times. Tubulin expression in Fast transcript variant 24 hour samples was also low compared to other transcript variant samples at 24 hour samples. These results potentially skew the CCND3 densitometry analysis when normalising to tubulin as a house keeper. At both 24 and 48 h time points, Optimised sequences give the most intense CCND3 signal, as predicted from the decoding speed. By 48 hours, Fast and Swap transcript variant sequences also gave higher protein production levels than the native transcript. However, the tubulin normalisation meant that the densitometry analysis suggests that the Fast transcript at 24 h performs best at 24 hour with no modifications out performing Native at 48 hours, skewed by the tubulin expression as described above.

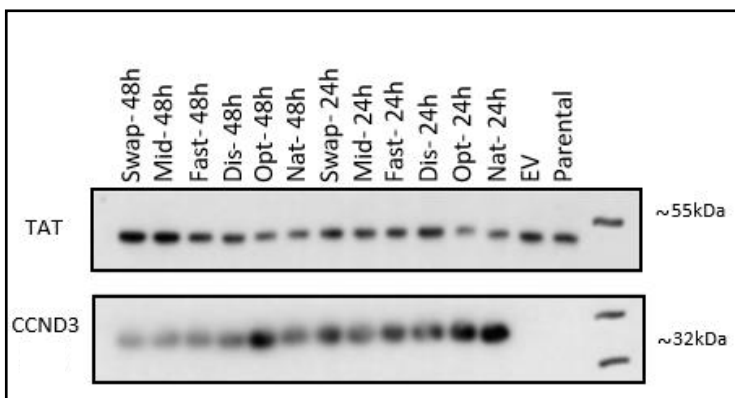


Figure 4.5A: Western blot analysis for *CCND3* of Gs variant samples harvested from HEK 293 cells grown at 37°C, 24 and 48 hours post transient transfection via lipofectamine. Samples included parental cell line and cells transfected with vector containing *CCND3* native, optimised, disoptimised, fast, middling, or swap sequences, and an empty vector. TAT was used as a loading control.

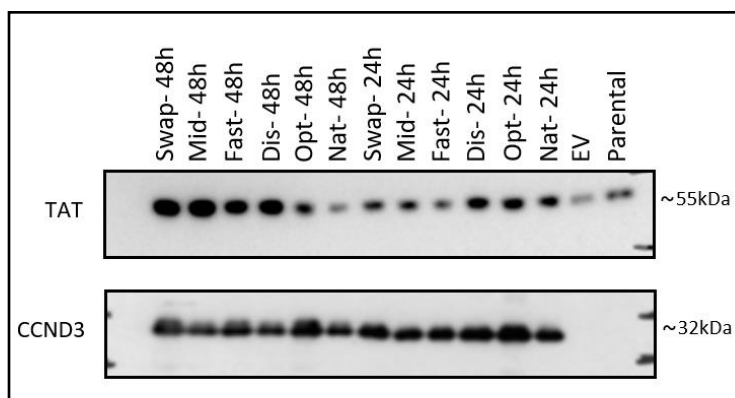
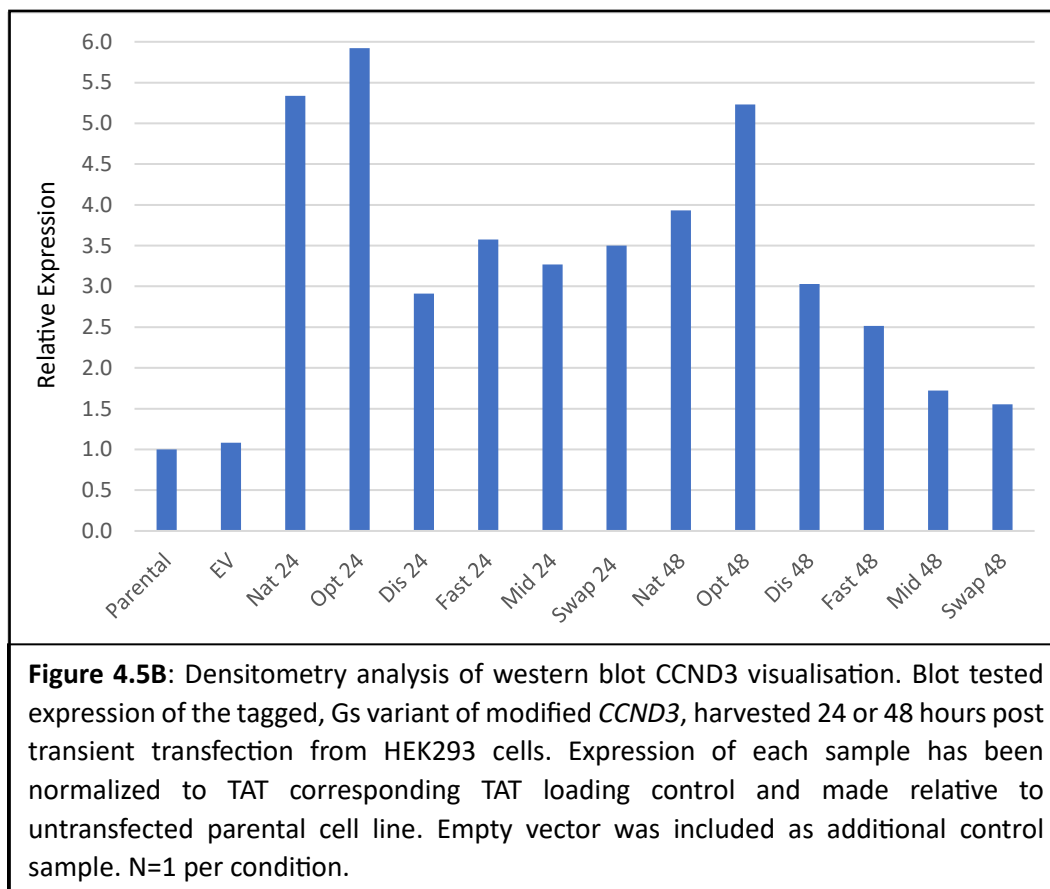
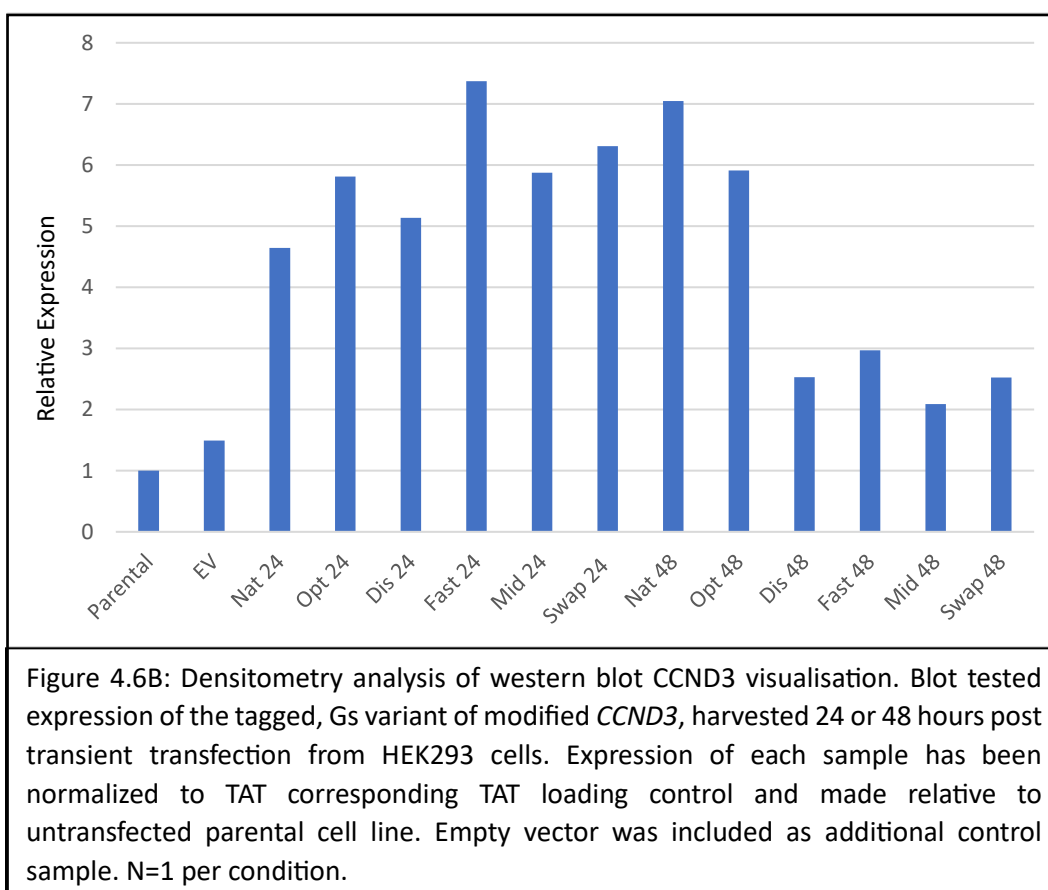


Figure 4.6A: Western blot analysis for *CCND3* of Gs variant samples harvested from HEK 293 cells grown at 32°C, 24 and 48 hours post transient transfection via lipofectamine. Samples included parental cell line and cells transfected with vector containing *CCND3* native, optimised, disoptimised, fast, middling, or swap sequences, and an empty vector. TAT was used as a loading control.





4.4.1.2 Western Blot Analysis of Experimental Replicates of Protein Production from each *CCND3* Transcript Variant Sequence

Replicate samples of protein production from the different transcript variants of *CCND3* were analysed by western blot. For 37°C samples (Figure 4.7A), weak tubulin (TAT) signals were observed for each sample, especially for one of the Swap replicates. This sample also showed no evidence of any *CCND3* production. This could be from sample preparation, unsuccessful transfection, or genuine lack of expression. The latter is unlikely as other GvS samples demonstrate expression. At 24 hours, no modified sequences visually gave higher *CCND3* expression levels than Native replicates which was corroborated by densitometry analysis (Figure 4.7B). The Optimised transcript results were closest to Native and outperformed other transcript modifications by a statistically significant margin. Disoptimised *CCND3* transcript derived protein production appeared visually to be second highest, but once normalised in densitometry analysis to the tubulin (TAT) signal was shown to express comparable to Fast and Mid replicates.

By 48 hours (Figure 4.8A, Figure 4.8B), production of all modified transcript sequence derived CCND3 protein were statistically similar to Native replicates with the exception of the Middling transcript derived protein which showed increase production. It is noted that similarities of CCND3 results for those samples where there was high expression may be due to saturated signals.

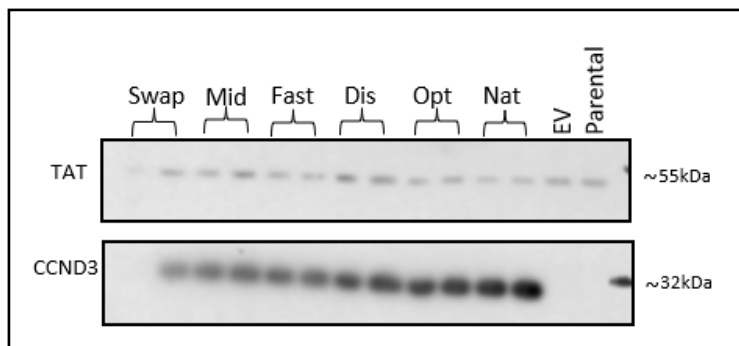


Figure 4.7A: Western blot analysis for CCND3 of Gs variant samples harvested from HEK 293 cells grown at 37°C, 24 hours post transient transfection via lipofectamine. Samples, in duplicate, included: single parental cell line and empty vector controls, Cells transfected with vector containing *CCND3* native, optimised, disoptimised, fast, middling, or swap sequences. TAT was used as a loading control.

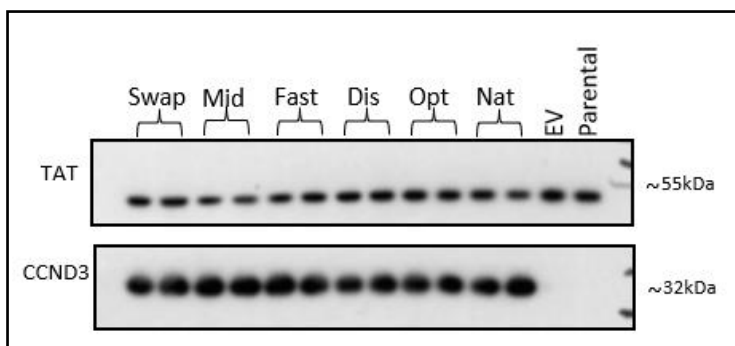


Figure 4.8A: Western blot analysis for CCND3 of Gs Variant Samples harvested from HEK 293 cells, grown at 37°C, 48 hours post transient transfection via lipofectamine. Samples, in duplicate, included: single parental cell line and empty vector controls, Cells transfected with vector containing *CCND3* native, optimised, disoptimised, fast, middling, or swap sequences. TAT was used as a loading control.

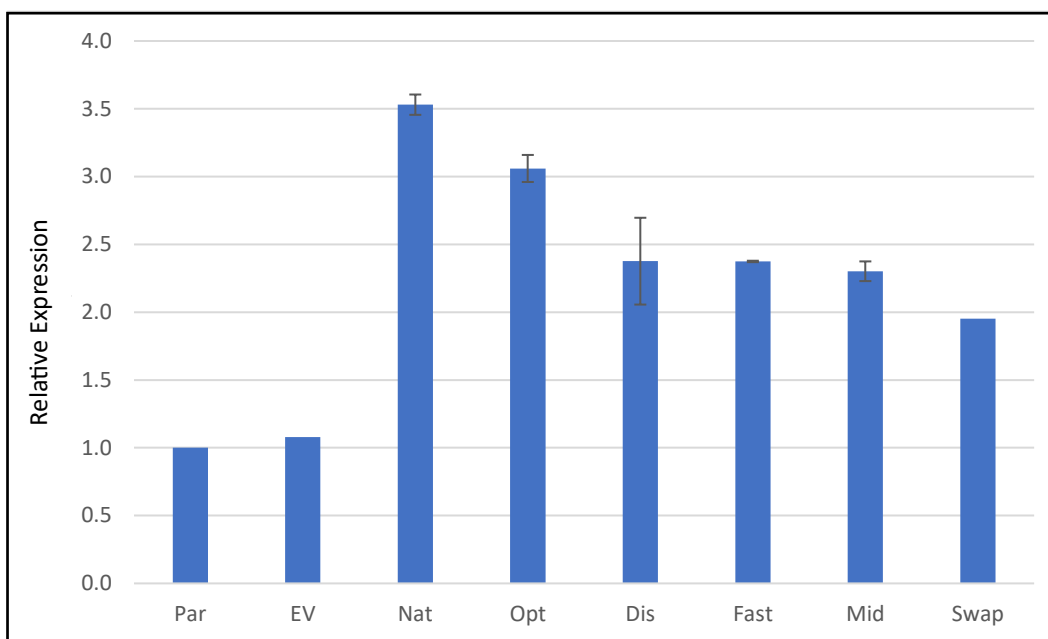


Figure 4.7B: Densitometry analysis of western blot CCND3 visualisation. Blot tested expression of the tagged, Gs variant of modified *CCND3*, harvested 24 hours post transient transfection from HEK293 cells. Expression of each sample has been normalized to TAT corresponding TAT loading control and made relative to untransfected parental cell line. Empty vector was included as additional control sample. N=1 for control samples, N=2 for non-control samples.

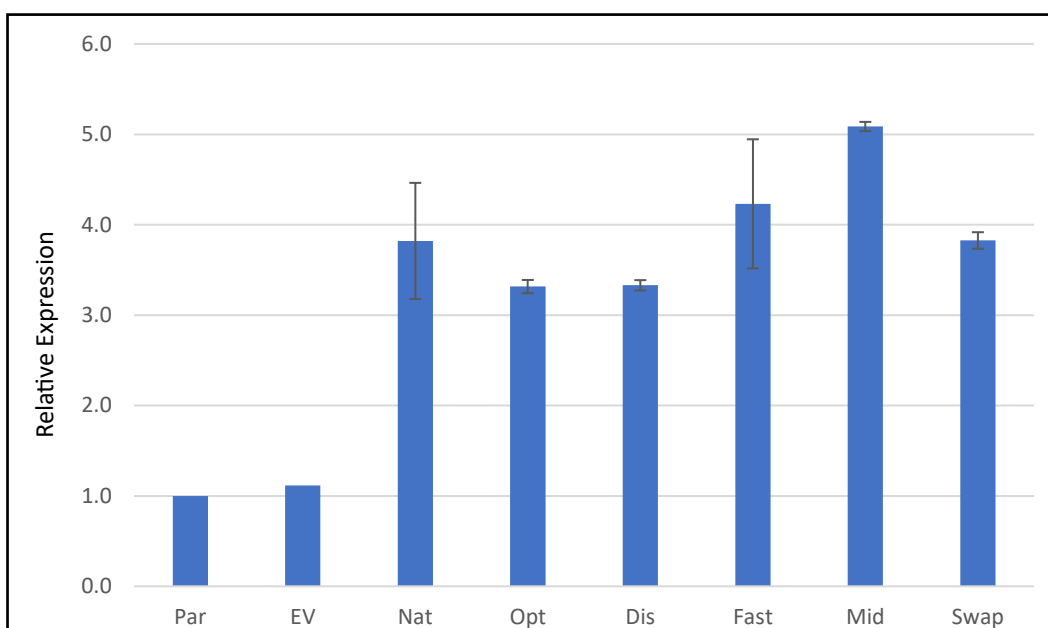


Figure 4.8B: Densitometry analysis of western blot CCND3 visualisation. Blot tested expression of the tagged, Gs variant of modified *CCND3*, harvested 48 hours post transient transfection from HEK293 cells. Expression of each sample has been normalized to TAT corresponding TAT loading control and made relative to untransfected parental cell line. Empty vector was included as additional control sample. N=1 for control samples, N=2 for non-control samples.

For analysis of samples cultured at 32°C, the Optimised transcript variant sequence samples were the only modification to yield a statistically significant increase in CCND3 protein production compared to the native transcript sequence, 24 hour post transfection (Figure 4.9A). increase in protein production for optimised sequences aligns with data reported in the previous section (Figure 4.6A) and was also present at the 48 hour time point for both data sets (Figure 4.6A, 4.10A). That result was not reflected in the densitometry analysis for previous study (Figure 4.6B), but is clearly visualised on western blot samples for CCND3 production if the tubulin (TAT) loading control variation is not taken into account.

At 24 hours, the Middling transcript variant sequences gave statistically lower protein production than all other *CCND3* expressing transcript samples except the disoptimised sequence which was predicted to be the slowest decoded sequence (Figure 4.9A). Disoptimised sequence variant derived samples showed high variance between replicates creating a high standard error of the mean score. At 48 hours post transfection (Figure 10A, 10B), Middling and Disoptimised transcript variant protein production was not significantly different to Native. Fast and Swap transcript variant sequences demonstrated the most drastic deviations from results in Figure 4.6A and showed no significant difference in protein production compared to native 24 hours post transfection. The Fast transcript variant appeared to result in the highest CCND3 protein production as assessed by densitometry but visually on the blot appeared comparable to Native if tubulin variation was discounted (Figure 4.6A, 4.6B). Tubulin values for western blot in Figure 4.6A may be unreliable as discussed in section 4.4.1.1. At 48 hours, fast transcript variants showed a small increase in expression compared to Native as visually seen in Figure 4.6A, whereas Swap protein production was greatly improved, also seen on western blot and not reflected on densitometry. It is noted that equal total protein loading in each lane by the Bradford assay was undertaken on all samples.

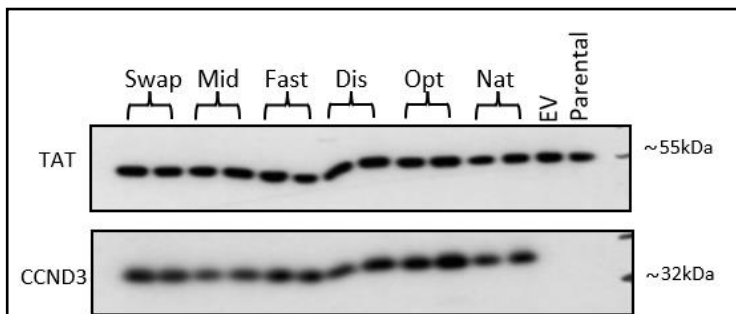


Figure 4.9A: Western blot analysis for CCND3 of Gs variant samples harvested from HEK 293 cells grown at 32°C, 24 hours post transient transfection via lipofectamine. Samples, in duplicate, included: single parental cell line and empty vector controls, Cells transfected with vector containing *CCND3* native, optimised, disoptimised, fast, middling, or swap sequences. TAT was used as a loading control.

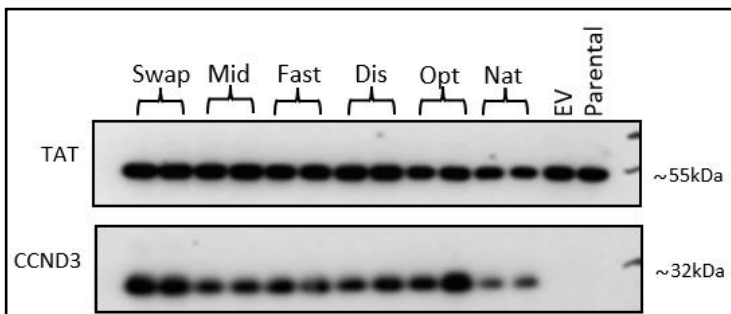
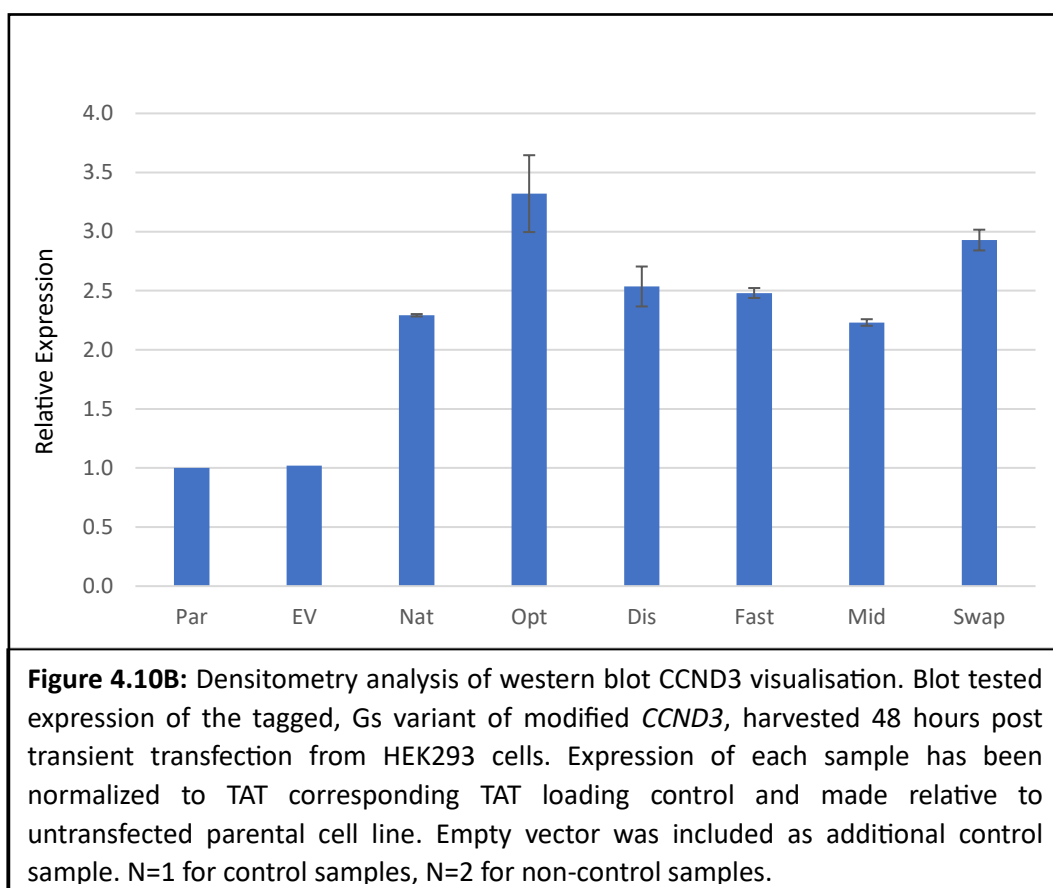
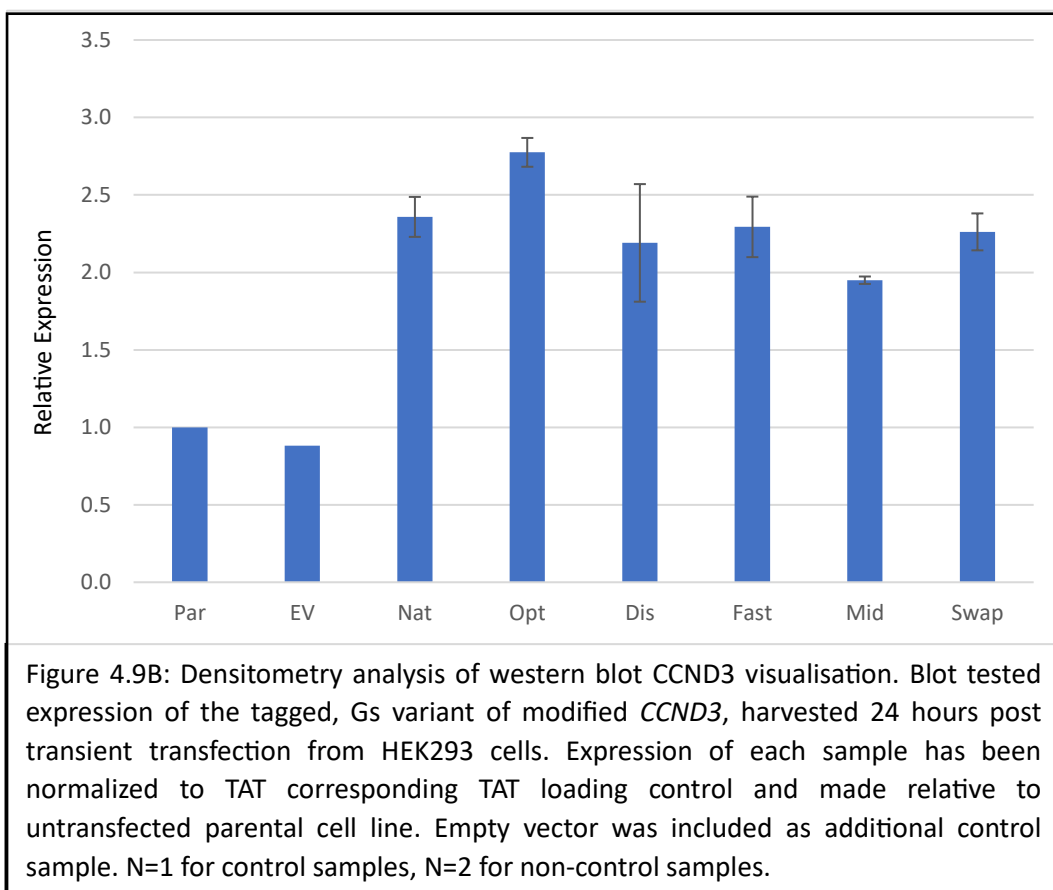


Figure 4.10A: Western blot analysis for CCND3 of Gs variant samples harvested from HEK 293 cells grown at 32°C, 48 hours post transient transfection via lipofectamine. Samples, in duplicate, included: single parental cell line and empty vector controls, Cells transfected with vector containing *CCND3* native, optimised, disoptimised, fast, middling, or swap sequences. TAT was used as a loading control.



4.4.2 Protein Production Analysis of CCND3 From Gv Transcript Variant Modifications

4.4.2.1 Comparison of CCND3 Protein Production at 24 and 48 h Post Transient Transfection from Different Transcript Variants

For cells grown at 37°C after transfection, analysis of cell lysates revealed the CCND3 protein production from Optimised and Disoptimised transcript variants showed similar expression profiles (Figure 4.11A, 4.11B). The CCND3 protein production was comparable to native transcript derived expression at 24 hours and greater by 48 hours from these transcript variants. This was not seen in disoptimised transcript variant sequences at 37°C. Fast transcript sequences gave the highest protein production of CCND3 at 24 hours and higher than the Native production at 48 hours. The Middling transcript variant also expressed better than seen in untagged G variant samples. Swap transcript variant sequences produced less CCND3 protein than native at both time points.

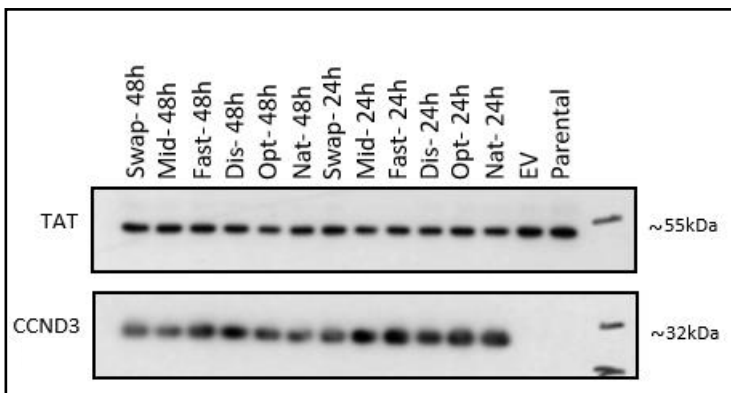


Figure 4.11A: Western blot analysis for V5 tagged CCND3 of Gv variant samples harvested from HEK 293 cells grown at 37°C, 24 and 48 hours post transient transfection via lipofectamine. Samples included parental cell line and Cells transfected with vector containing *CCND3* native, Optimised, Fast, Middling, or swap sequences, and an empty vector. TAT was used as a loading control.

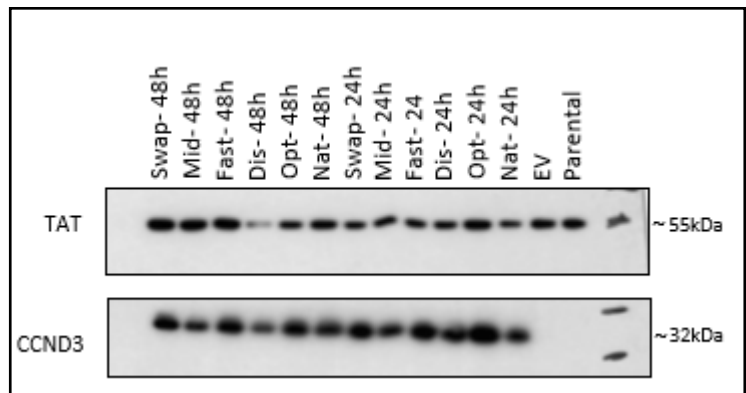
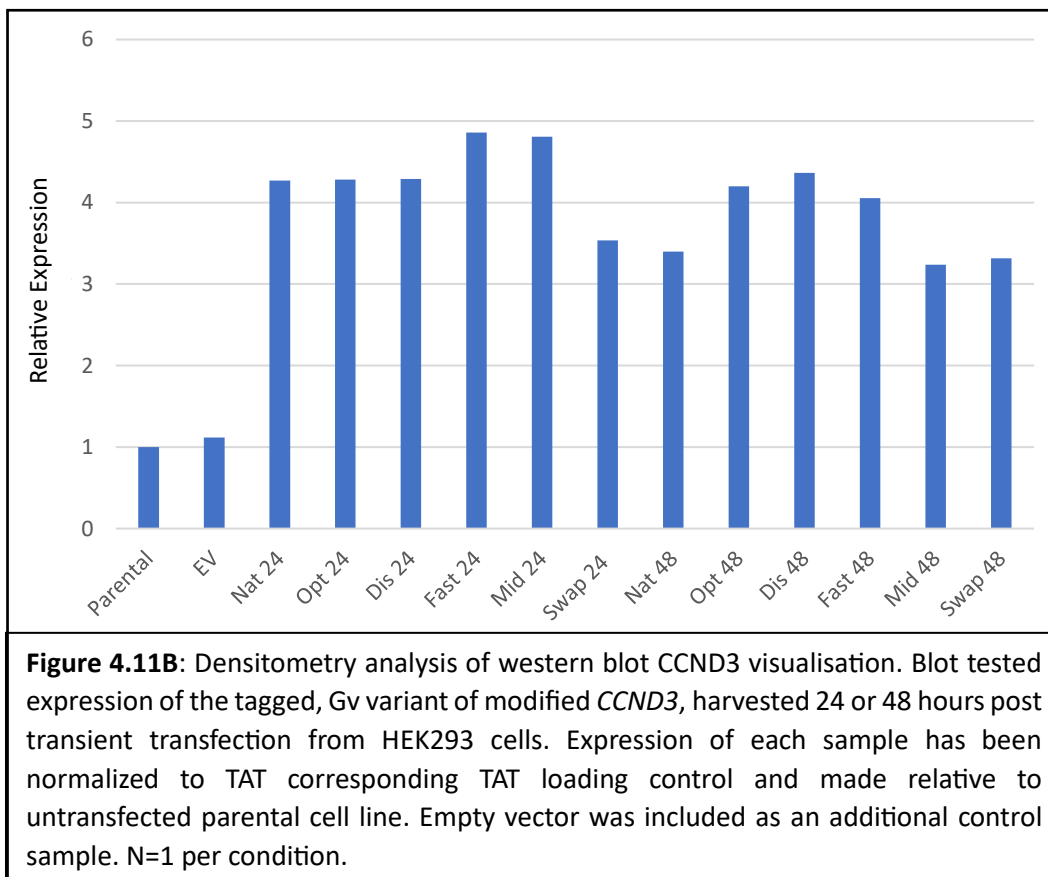
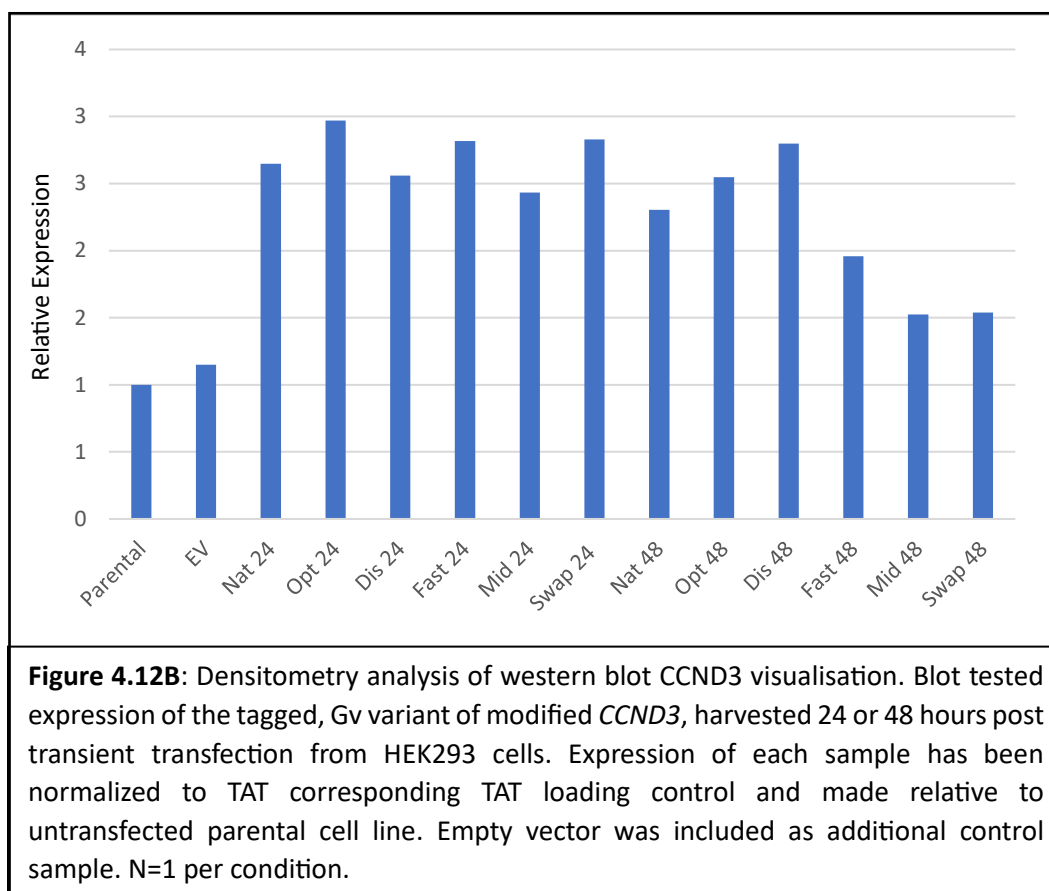


Figure 4.12A: Western blot analysis for V5 tagged CCND3 of Gv variant samples harvested from HEK 293 cells grown at 32°C, 24 and 48 hours post transient transfection via lipofectamine. Samples included parental cell line and cells transfected with vector containing *CCND3* native, optimised, disoptimised, fast, middling, or swap sequences, and an empty vector. TAT was used as a loading control.

For samples generated from cells grown at 32°C after transfection (Figure 4.12A, 4.12B), at the 24 hour time point the codon optimised transcript variant sequences lead to highest CCND3 protein production. Disoptimised and Middling transcript variant sequences gave reduced protein production compared to that from the Native sequence, whilst Fast and Swap transcript variant sequences gave a small increase in CCND3 protein production. At 48 hours post transfection, the protein production from the Native, optimised and swap transcript variants appeared comparable whereas expression from the Middling and Disoptimised transcript variants were low compared to other variants. The samples from cells transfected with the Disoptimised transcript variant had a poor TAT signal. Once expression was normalised to tubulin expression by densitometry analysis, the Disoptimised sample appeared to give the highest expression at this time point. High CCND3 signal for Swap sequence samples is mitigated by comparatively high tubulin (TAT) signal and once normalised, expression was on par with the Middling transcript variant which remained the lowest expressing construct at 48 hours for both 37°C and 32°C for Gs Variants.





4.4.2.2 Analysis of Experimental Replicates of each *CCND3* Sequence for Statistical Analysis.

For samples from 37°C cultured cells, the resolution of the bands on the western blot was poor for 24 hour replicate samples (Figure 4.13A), nevertheless Fast transcript variant replicates visually showed the highest *CCND3* protein production and densitometry analysis confirmed mean Fast production once normalised was greater than other samples although compared to Native replicates was not statistically significant (Figure 4.13B) (Figure 4.13B). Optimised transcript variant derived protein production was the second highest from the modified sequences but was not a lot different to protein production from the Middling, Swap and Disoptimised variants in that order. These last three modifications produced *CCND3* protein at a statistically reduced amount compare to Native.

At 48 hour time point (Figure 4.14A), the *CCND3* protein production from the Native replicates were varied but the tubulin (TAT) loadings were comparable and the tubulin band intensities across the blot were consistent. Both Native replicate *CCND3* band intensities visually appear higher than all modified sequence

bands, although densitometry showed no significant change between Native and Optimised samples (Figure 4.14B). All other modified sequence samples had significantly lower expression relative to Native. CCND3 protein production from Disoptimised, Fast, and Swap samples were comparable to one another, while the Middling variant gave the weakest expression.



Figure 4.13A: Western blot analysis for V5 tagged CCND3 of Gv variants. Samples harvested from HEK 293 cells, grown at 37°C, 24 hours post transient transfection via lipofectamine. Samples, in duplicate, included: single parental cell line and empty vector controls, Cells transfected with vector containing CCND3 native, optimised, disoptimised, fast, middling, or swap sequences. TAT was used as a loading control.

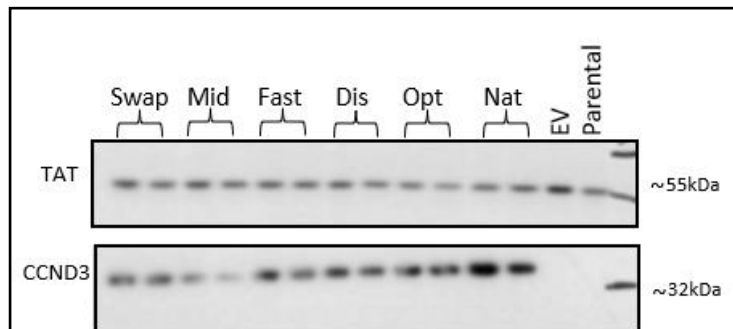


Figure 4.14A: Western blot analysis for V5 tagged CCND3 of Gv variant Samples harvested from HEK 293 cells, grown at 37°C, 48 hours post transient transfection via lipofectamine. Samples, in duplicate, included: single parental cell line and empty vector controls, Cells transfected with vector containing CCND3 native, optimised, disoptimised, fast, middling, or swap sequences. TAT was used as a loading control.

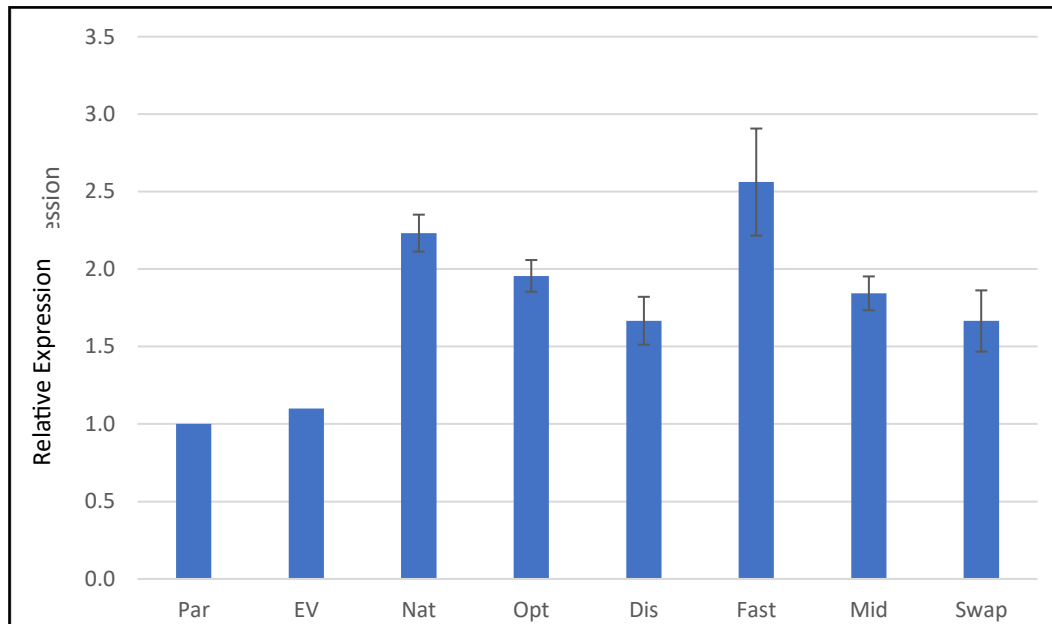


Figure 4.13B: Densitometry analysis of western blot CCND3 visualisation. Blot tested expression of the tagged, Gv variant of modified *CCND3*, harvested 24 hours post transient transfection from HEK293 cells. Expression of each sample has been normalized to TAT corresponding TAT loading control and made relative to untransfected parental cell line. Empty vector was included as additional control sample. N=1 for control samples, N=2 for non-control samples.

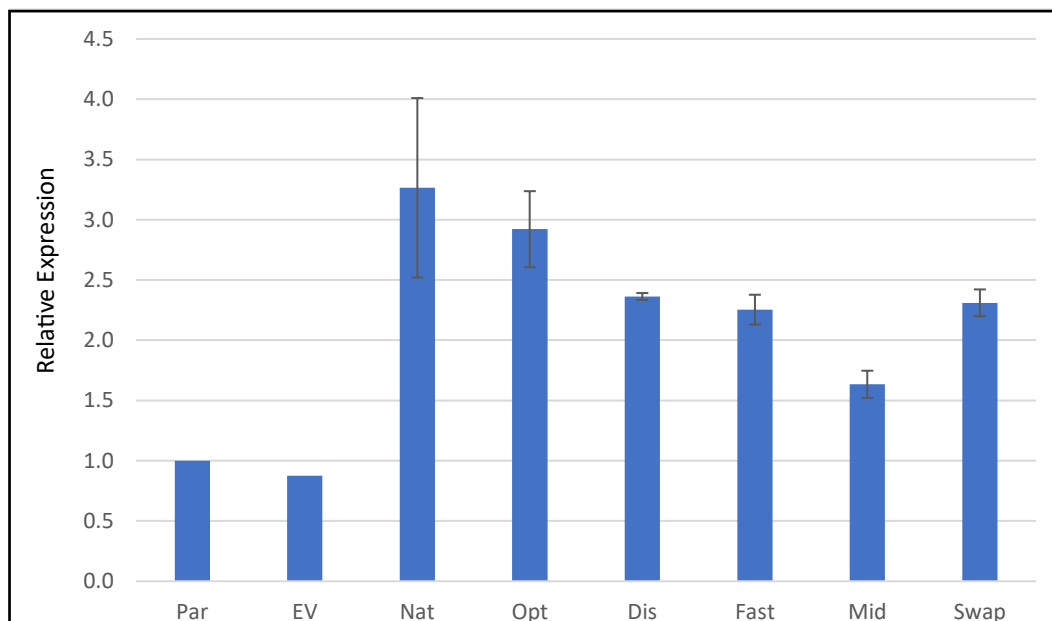


Figure 4.14B: Densitometry analysis of western blot CCND3 visualisation. Blot tested expression of the tagged, Gv variant of modified *CCND3*, harvested 48 hours post transient transfection from HEK293 cells. Expression of each sample has been normalized to TAT corresponding TAT loading control and made relative to untransfected parental cell line. Empty vector was included as additional control sample. N=1 for control samples, N=2 for non-control samples.

For samples taken from cells grown at 32°C, the protein production from the Optimised, Fast and Swap transcript variant sequences was increased in amounts of CCND3 compared to that from the Native sequence; the highest expression at 24 and 48 h being from the Optimised and Swap transcript variant sequences respectively (Figure 4.15A and Figure 4.16A). However, the tubulin (TAT) signal for Optimised 24 hour samples was also high, suggesting more protein was loaded in these samples. Once normalised, Optimised samples showed no significant change in CCND3 protein production compared to that from the Native transcript at either time point (Figure 4.15B, 4.16B). The Fast transcript derived protein production also showed no significant change compared to that from the Native sequence at 48 hours, but densitometry normalised expression at 24 h was significantly different compared to that from the Native. Based on densitometry, differences in expression between Swap and Native samples were not significant until 48 hours where CCND3 protein production from the Swap sequence was reduced compared to that from the Native (Figure 4.16B).

The protein production from the Disoptimised transcript variant at 24 hours was visually comparable to Native *CCND3* expression (Figure 4.15A). By 48 hours, the CCND3 protein signal appeared reduced compared to that from the Native sequence which was confirmed and corroborated by densitometry analysis. The Middling sequence appeared to visually result in reduced CCND3 protein compared to the Native at 24 hours but there was no significant difference upon densitometry analysis.

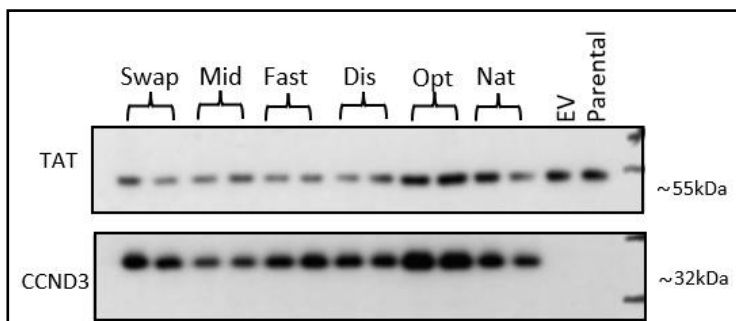


Figure 4.15A: Western blot analysis for CCND3 of Gv variant samples harvested from HEK 293 cells grown at 32°C, 24 hours post transient transfection via lipofectamine. Samples, in duplicate included: single parental cell line and empty vector controls, Cells transfected with vector containing *CCND3* native, optimised, disoptimised, fast, middling, or swap sequences. TAT was used as a loading control.

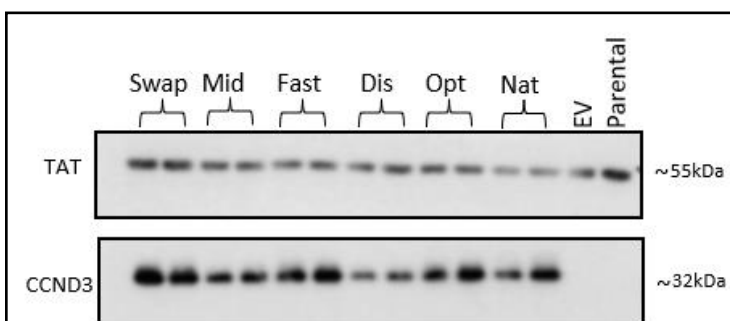
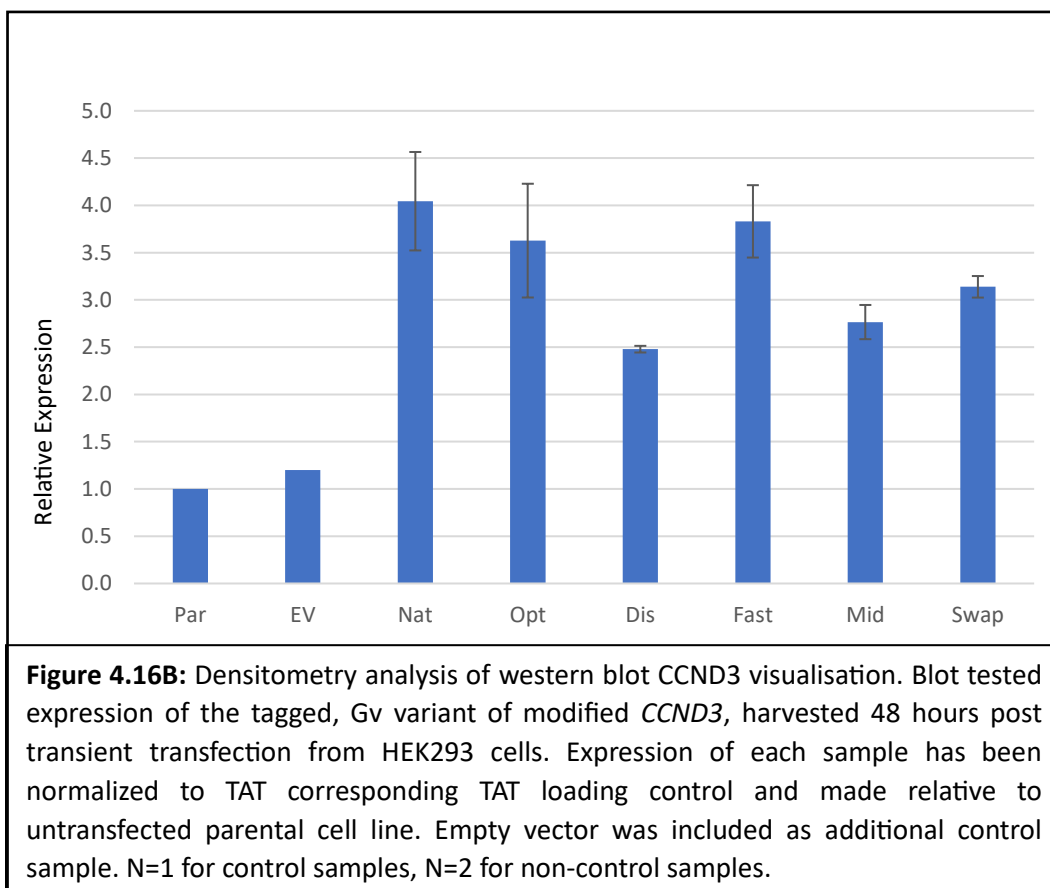
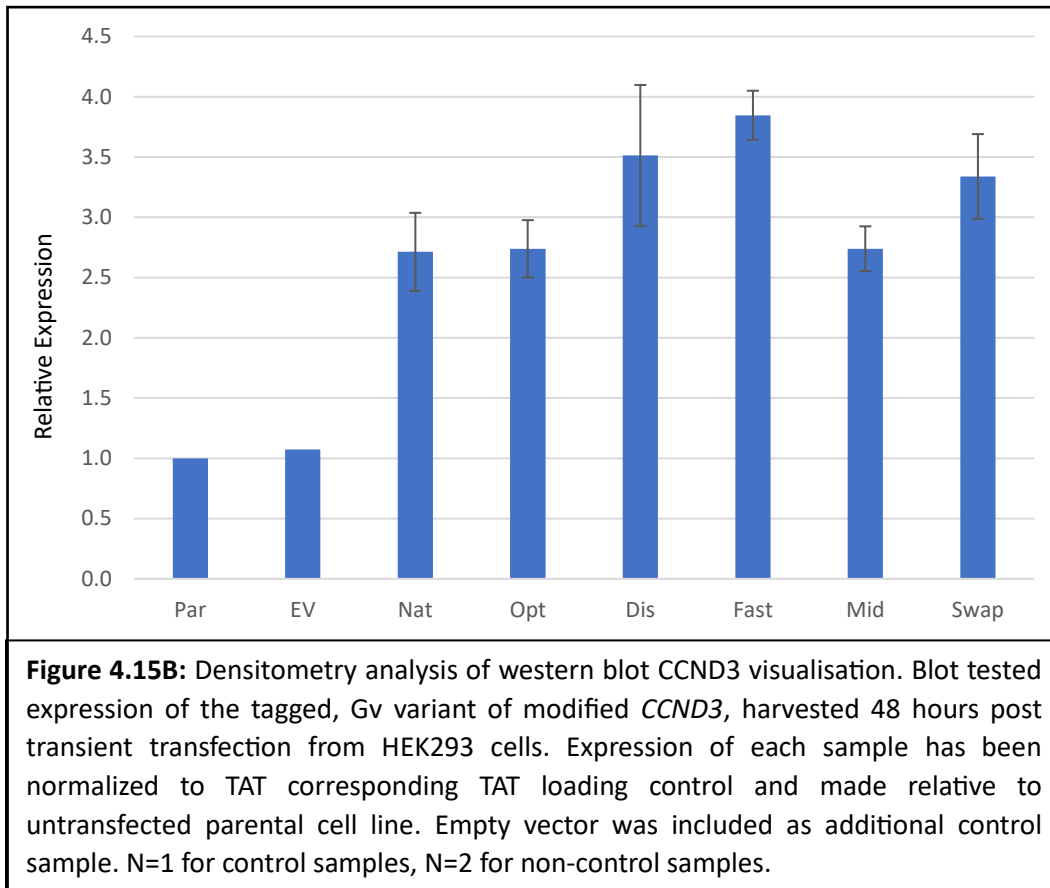


Figure 4.16A: Western blot analysis for CCND3 of samples Gv variant harvested from HEK 293 cells grown at 32°C, 48 hours post transient transfection via lipofectamine. Samples, in duplicate, included: single parental cell line and empty vector controls, Cells transfected with vector containing *CCND3* native, optimised, disoptimised, fast, middling, or swap sequences. TAT was used as a loading control.



4.4.3 Protein Production Analysis of CCND3 from Ts Transcript Variant Modifications

4.4.3.1 Comparison of Transient CCND3 Protein Production at 24 and 48 Hour Post Transfection from Different Transcript Variants

In the analysis of these samples, weak tubulin (TAT) bands were observed across the entirety of western blot (Figure 4.17A), with Empty Vector, Parental, Middling and SWAP transcript variants also at 48 h being particularly poor (Figure 4.17A). At the 24 hour time point, the Disoptimised and Middling transcript variant transfected samples produced less intense CCND3 protein signals than from the Native transcript, although the band intensity from the Disoptimised transcript did match that from the Native by the 48 hour time point. The Optimised, Fast, and Swap transcript variant transfected sample CCND3 protein production was increased compared to those from the Native at the 24 hour time point, with the Swap transcript variant *CCND3* expression having the highest relative change in expression compared to the Native as determined by densitometry analysis (Figure 4.17B). By 48 hours, the Optimised CCND3 protein production was comparable both visually and in by densitometry analysis to Native production. The codon Swap and Fast transcript variant derived CCND3 protein production was increased at this time compared to the native, with Fast transcript derived expression leading to the most intense western blot band and highest expression as determined by densitometry analysis. The relative production determined by densitometry for Swap transcript derived CCND3 protein at 48 hours may be exaggerated however, due to poor tubulin (TAT signal); visual CCND3 protein production observed on the western blot was comparable to that from the Native transcript.

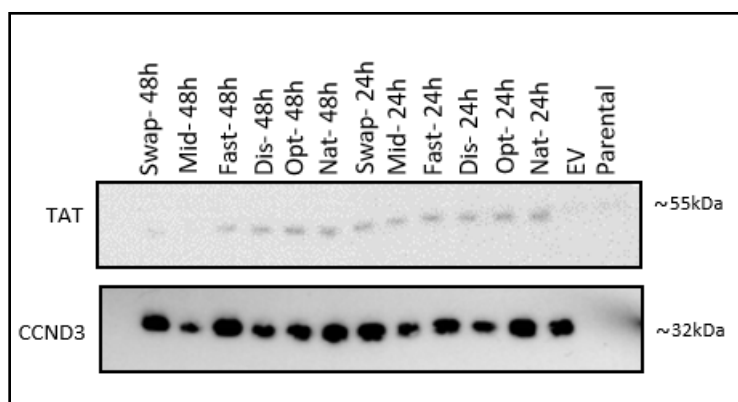


Figure 4.17A: Western blot analysis for CCND3 of Ts variant samples harvested from HEK 293 cells grown at 37°C, 24 and 48 hours post transient transfection via lipofectamine. Samples included parental cell line and cells transfected with vector containing *CCND3* native, optimised, disoptimised, fast, middling, or swap sequences, and an empty vector. TAT was used as a loading control.

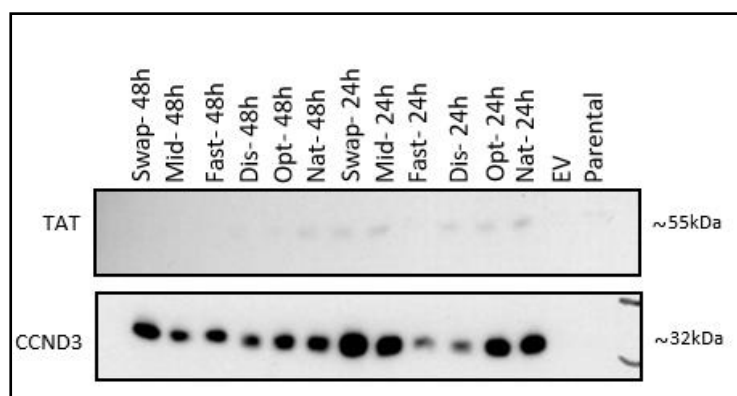
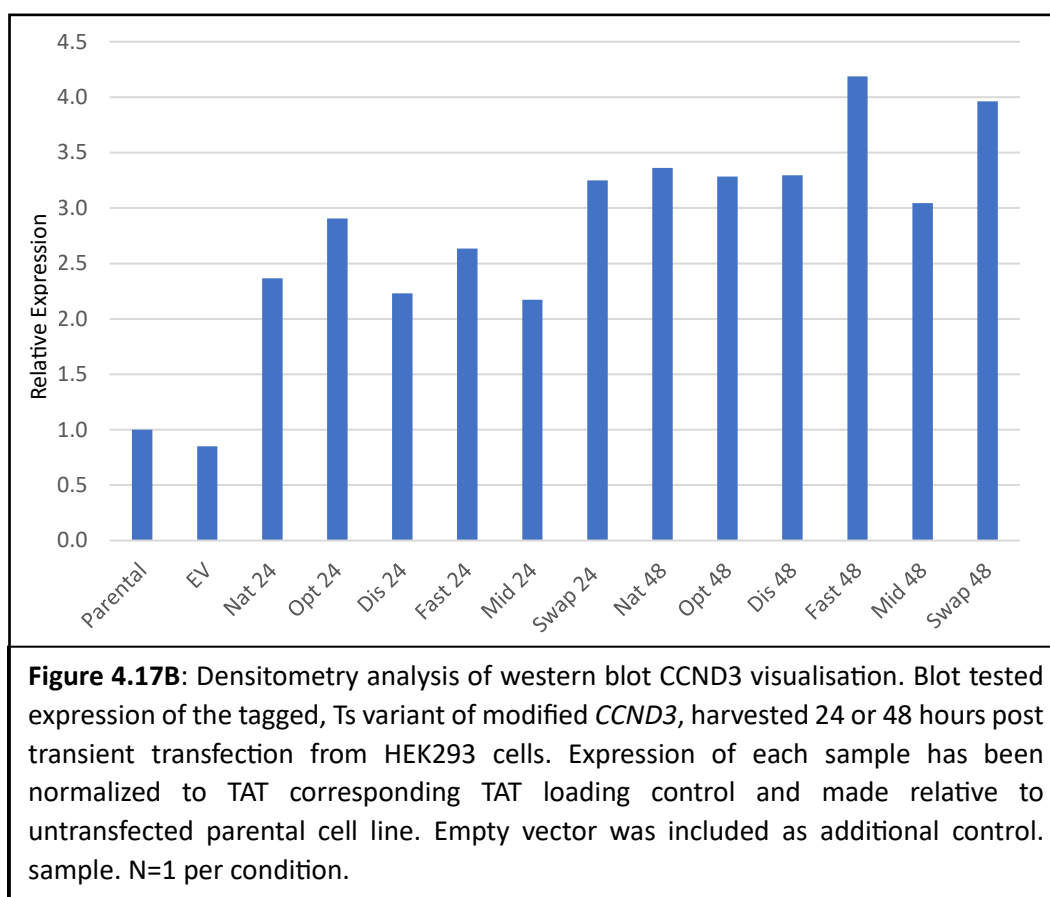


Figure 4.18A: Western blot analysis for CCND3 of Ts variant samples harvested from HEK 293 cells grown at 32°C, 24 and 48 hours post transient transfection via lipofectamine. Samples included parental cell line and cells transfected with vector containing *CCND3* native, optimised, disoptimised, fast, middling, or swap sequences, and an empty vector. TAT was used as a loading control.

It is noted that background signal across the tubulin blot of figure 4.18A will skew densitometry analysis (Figure 4.18B), with values progressively reducing towards the left-hand side as background reading increases. These samples had little to no tubulin (TAT) signal which may be caused by low protein loading (unlikely as Bradford assay was used to normalise total protein loading), depletion of anti-tubulin TAT antibody due to repeated use, or background shade masking signal. Additional densitometry analysis is included which has not been normalised to the tubulin loading control but assumes equal loading based upon the Bradford adjusted loading (Figure 4.18B). While data analysed and presented in this way comes with further caveats, it does mitigate background gradient issues and poor tubulin band visualisation.

Disoptimised and Fast transcript variant derived protein sample production was consistently lower than Native derived CCND3 protein production at both time points (24 and 48 h) and showed comparable expression to each other and Middling transcript variant samples at 48 hours. Middling transcript variant derived CCND3 protein production at 24 hours was higher than disoptimised and Fast transcript variant derived samples but less than Native transcript derived CCND3 protein. Optimised transcript variant CCND3 protein production was higher than that from Disoptimised and Fast transcript variant derived protein consistently and comparable to that from the Native transcript. The western blot analysis and subsequent visualisation of the intensity of the CCND3 bands suggests that the Swap sample was the highest expressed at 24 and 48 hours, however upon densitometry analysis the expression was less than native derived CCND3 at 48 h.



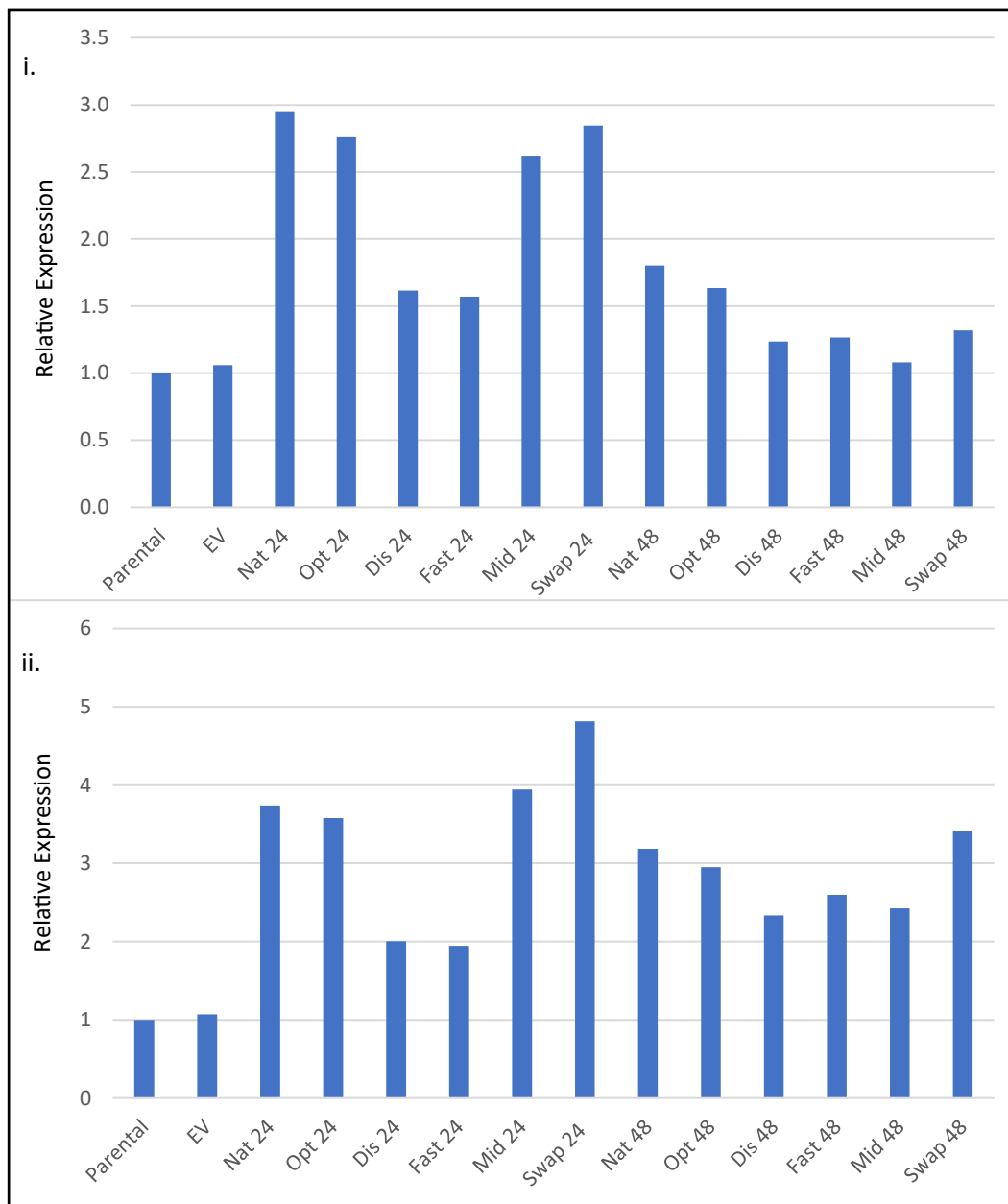


Figure 4.18B: Densitometry analysis of western blot CCND3 visualisation. Blot tested expression of the tagged, Ts variant of modified *CCND3*, harvested 24 or 48 hours post transient transfection from HEK293 cells. i. Expression of each sample has been normalized to corresponding TAT loading control and made relative to untransfected parental cell line. ii. Samples made relative to untransfected cell line without normalisation step. N= 1 per condition.

4.4.3.2 Western Blot Analysis of Experimental Replicates of Protein Production from each *CCND3* Sequence

Analysis of samples after 24 h at 37°C showed that the Fast transcript variant samples visually appeared to be the highest expressors with each other modified sequences expression being comparable to that from the native transcript sequence (Figure 4.19A). Once normalised to tubulin loading controls (TAT) by densitometry analysis, the Fast transcript variant derived protein production was still significantly higher than that from the Native transcript (Figure 4.19B). However, Optimised and Disoptimised transcript variants had the highest mean results, with statistical difference to native transcript derived *CCND3* expression. The Optimised and Fast transcript variant transfected samples band intensities were significantly different from each. As seen in Figure 4.17A, the Middling transcript variant samples showed reduced *CCND3* expression but unlike Figure 4.17A, the expression was comparable to Swap transcript variant sample *CCND3* expression.

The data in figure 4.20Bi. shows that protein production of replicates of transcript variant modified sequences at the 48 hour time point all demonstrate reduced *CCND3* protein production compared to that from the Native transcript. This general outcome is similar to the observations reported in Figure 4.19Bii, but contradicts data in Figure 4.19Bi for almost all samples. Figure 4.20Bii data shows far greater parallels between expression trends when comparing individual sequence expression and visual western blot signals (Figure 4.20A). In this case, visual Fast and Swap transcript variant derived protein production shows the highest *CCND3* expression out of the modified sequences.

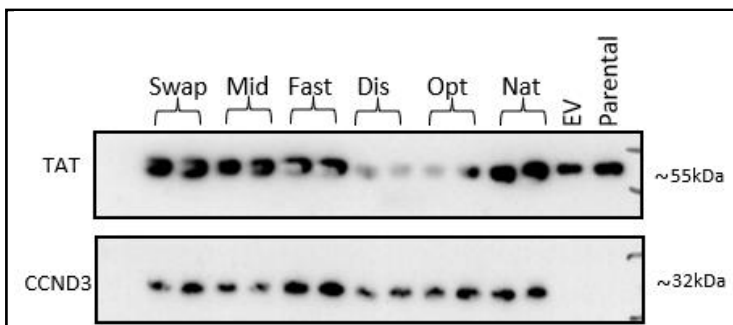


Figure 4.19A: Western blot analysis for *CCND3* for Ts variants. Samples harvested from HEK 293 cells, grown at 37°C, 24 hours post transient transfection via lipofectamine. Samples, in duplicate, included: single parental cell line and empty vector controls, Cells transfected with vector containing *CCND3* native, optimised, disoptimised, fast, middling, or swap sequences. TAT was used as a loading control.

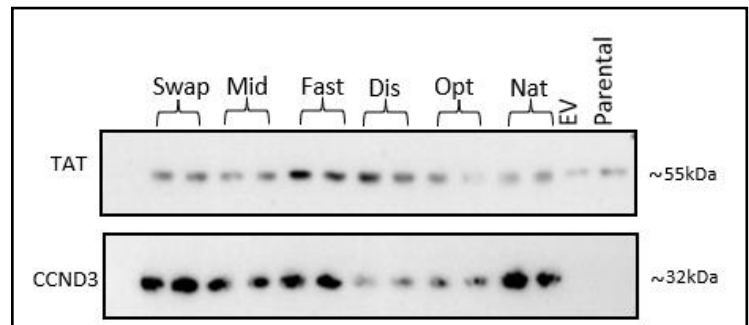
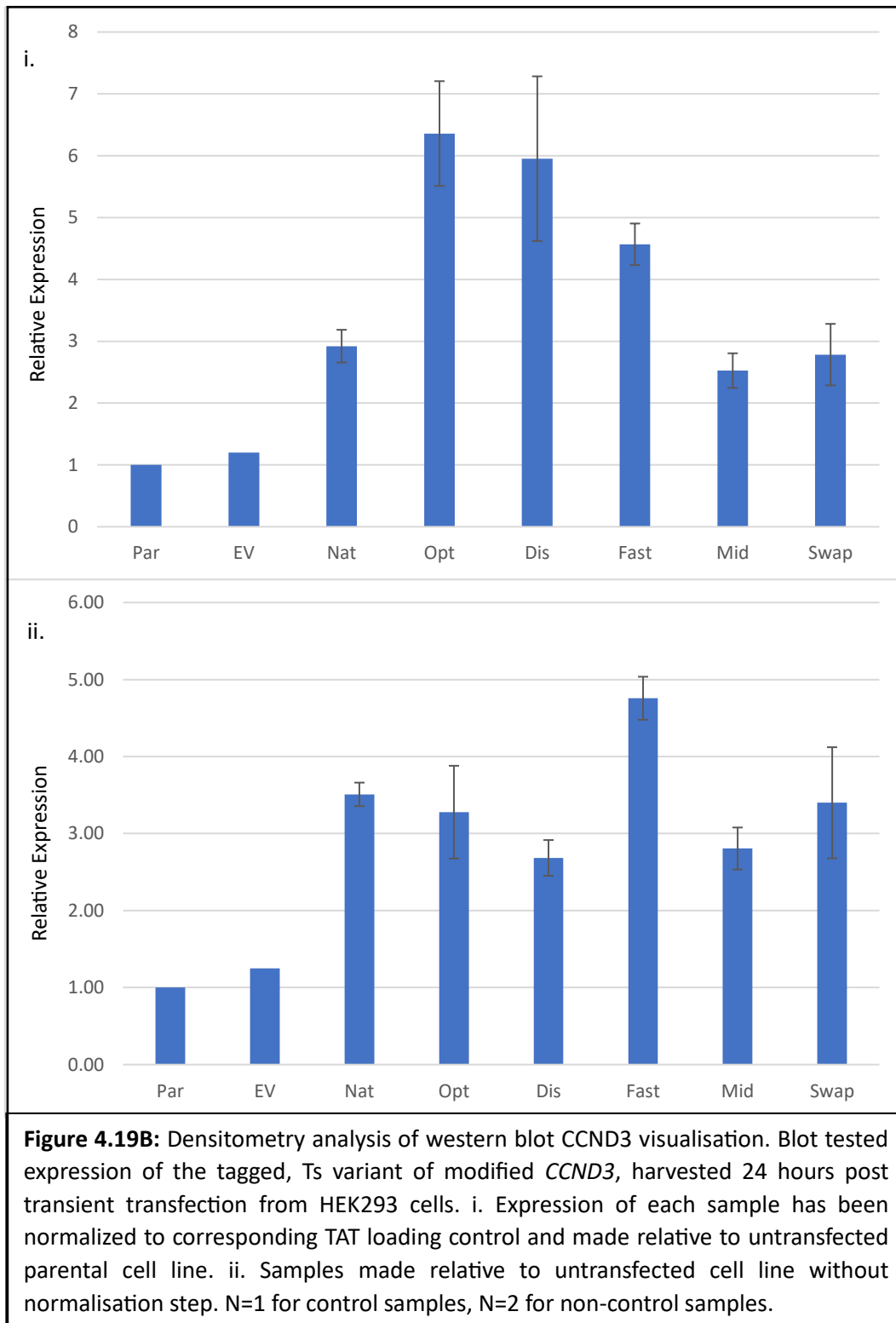
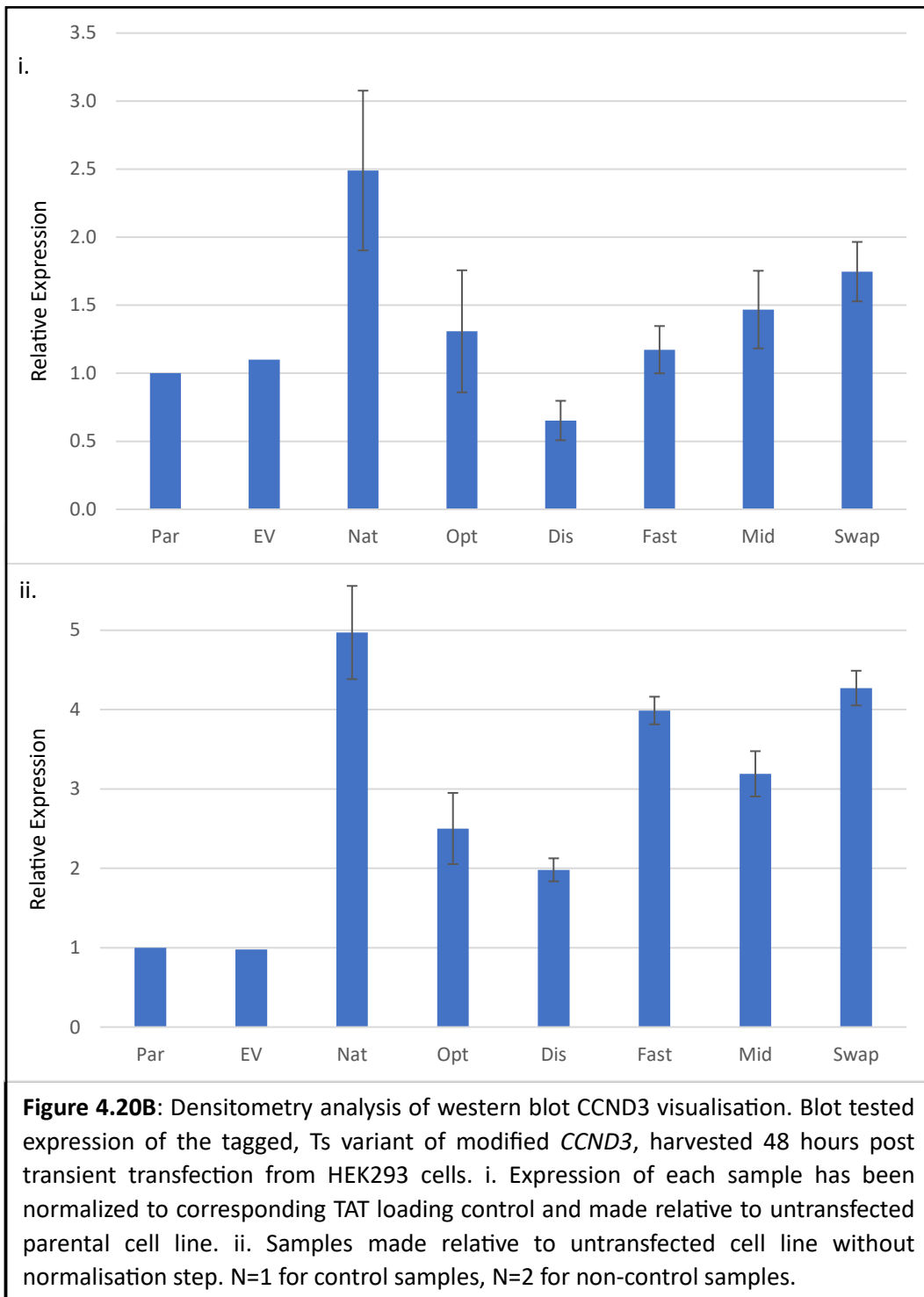


Figure 4.20A: Western blot analysis for *CCND3* of Ts variants. Samples harvested from HEK 293 cells, grown at 37°C, 48 hours post transient transfection via lipofectamine. Samples, in duplicate included: single parental cell line and empty vector controls, Cells transfected with vector containing *CCND3* native, optimised, disoptimised, fast, middling, or swap sequences. TAT was used as a loading control.





For samples collected after culture at 32°C, analysis of protein production by western blot showed that the swap transcript variant samples showed improved expression at 24 and 48 h compared to that from the Native transcript variant and achieved the highest expression of any modification at 24 hours (Figures 4.21A, 4.22A). This mirrors the results from the time comparison study (Figure 4.18). Statistically, the Native and Optimised transcript variant 24 hour samples were comparable, although Native samples showed high variation (Figure 4.21B). By 48 hours, the Optimised transcript samples were the highest CCND3 protein expressors (Figure 4.22B). Unlike the G variant experiments, the Middling transcript variant samples were comparable to the Fast transcript variant protein production at 24. As expected, the Disoptimised transcript variant gave less CCND3 protein production compared to that from the Native transcript variant at both time points.

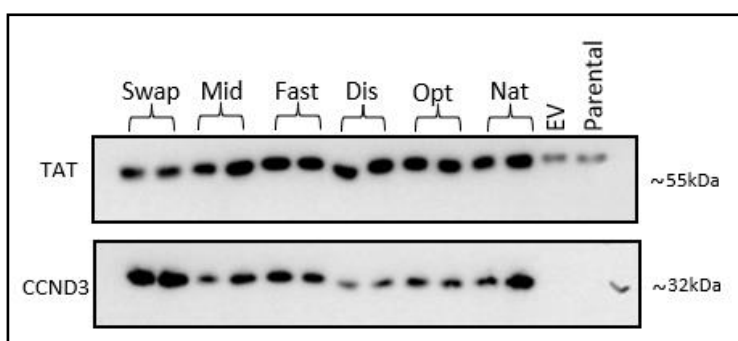


Figure 4.21A: Western blot analysis for CCND3 of Ts variants. Samples harvested from HEK 293 cells, grown at 32°C, 24 hours post transient transfection via lipofectamine. Samples, in duplicate, included: single parental cell line and empty vector controls, Cells transfected with vector containing *CCND3* native, optimised, disoptimised, fast, middling, or swap sequences. TAT was used as a loading control.

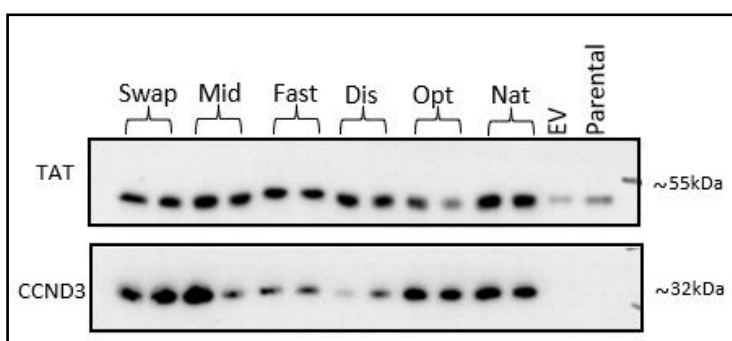
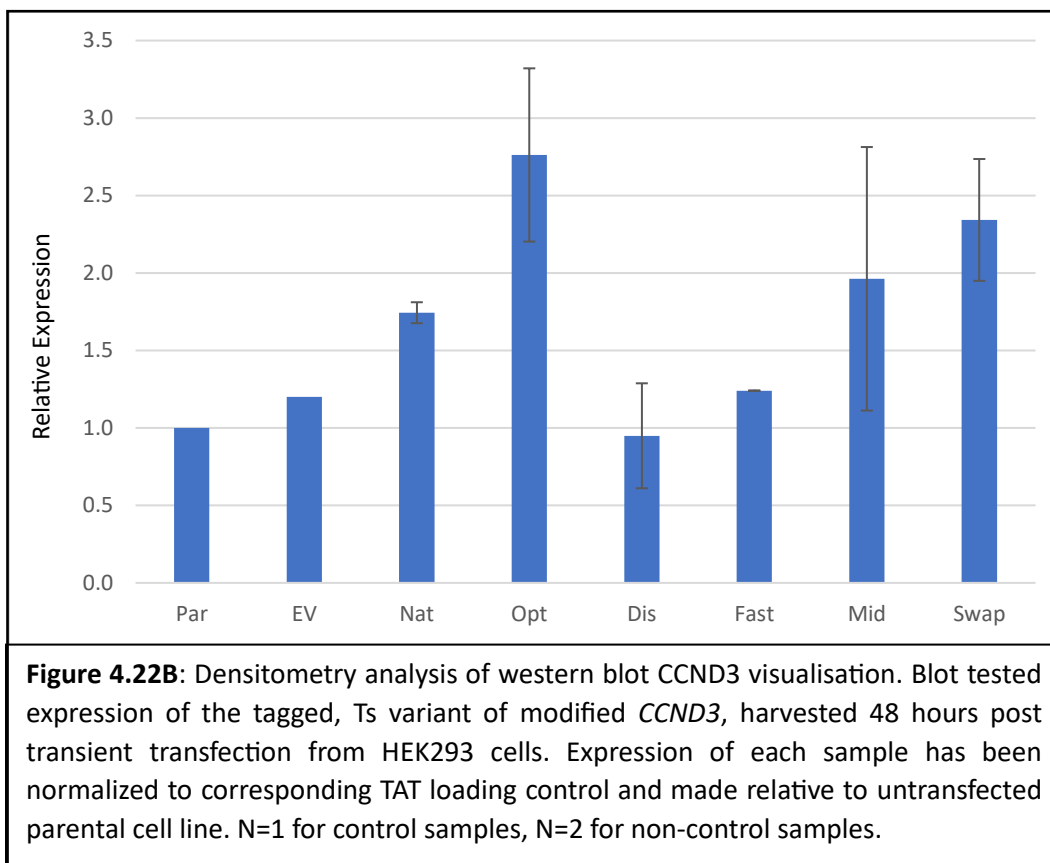
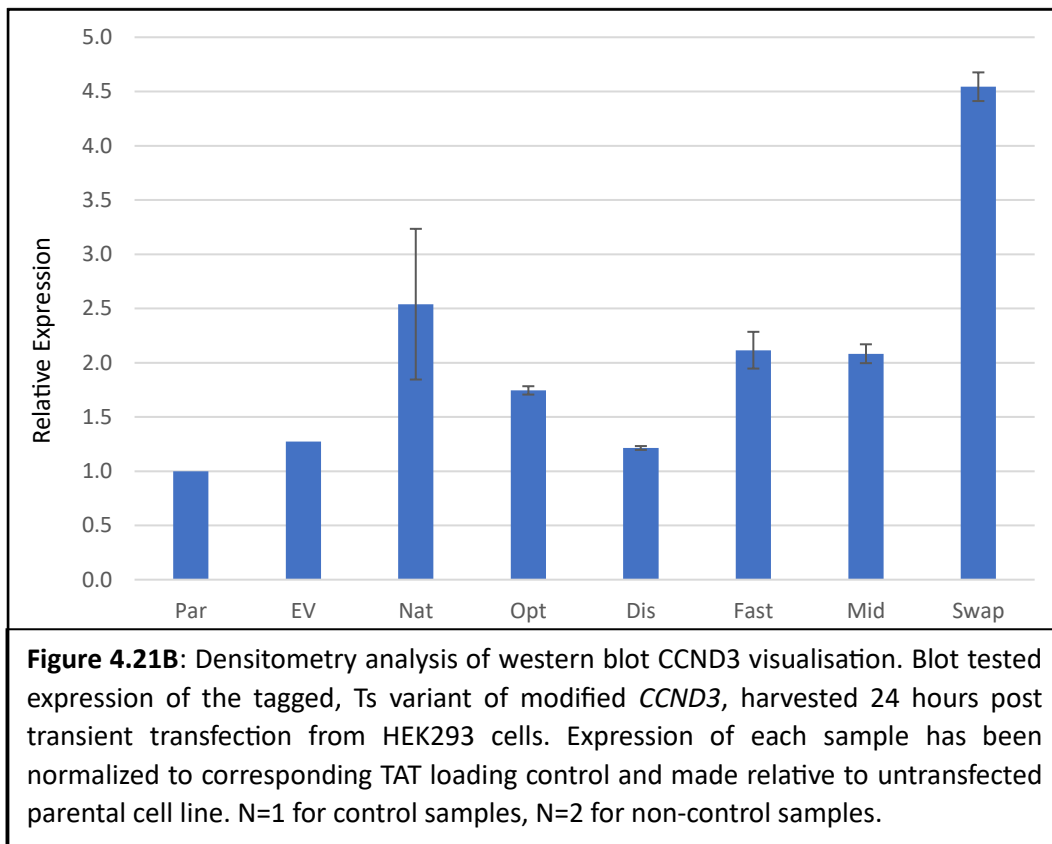


Figure 4.22A: Western blot analysis for CCND3 of Ts variants. Samples harvested from HEK 293 cells, grown at 32°C, 48 hours post transient transfection via lipofectamine. Samples, in duplicate, included: single parental cell line and empty vector controls, Cells transfected with vector containing *CCND3* native, optimised, disoptimised, fast, middling, or swap sequences. TAT was used as a loading control.



4.4.4 Analysis of CCND3 Protein Production from Tv Transcript Variant Modifications

4.4.4.1 Comparison of Transient Protein Production of CCND3 at 24 and 48 hour Post Transfection

For the Tv transcript variants, the highest *CCND3* expression as determined by densitometry at both time points was from the Fast transcript variant at 37°C (Figure 4.23A and Figure 4.23B). At 24 hours this was obvious visually on the western blot, and at 48 hour this transcript variant gave higher expression than that from the Disoptimised and Optimised transcript variant transfected samples. However, the apparent high expression at 48 hours from the Fast transcript variant is likely due to the low/lack of a tubulin band. The Middling transcript variant failed to give any observable protein production at 24 h and a weak protein band/signal at 48 h. The low or lack of expression from the Middling transcript variant was observed in all Tv variant western blots. Optimised and Disoptimised transcript variant *CCND3* protein production was comparable to that from the Native transcript at both time points visually, but appeared lower upon densitometry analysis at 24 hours. Swap transcript variant *CCND3* protein production was comparable to native at both time points based on the densitometry data analysis.

Whilst the Middling transcript variant derived *CCND3* protein production was absent at both time points in the 32°C western blot (Figure 4.24A), after 24 hours all other transcript modifications gave *CCND3* expression comparable to the Native transcript, both visually and after densitometry analysis (Figures 4.24A, 4.24B). After 48 hours, band smearing for Native, Optimised, and Disoptimised samples was the result of much higher expression and the presence impacted the densitometry analysis. Of the three, the Optimised transcript variant appeared to give the highest expression. Fast transcripts gave lower expression than the three described above, but more than that from the Swap transcript variant.

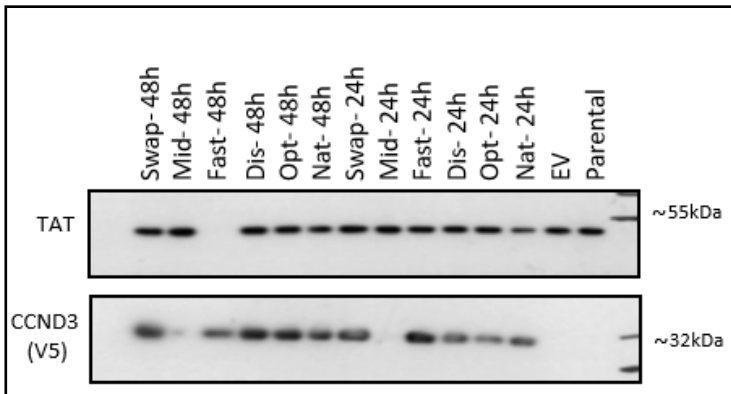


Figure 4.23A: Western blot analysis for V5 tagged CCND3 of Tv variant samples harvested from HEK 293 cells grown at 37°C, 24 and 48 hours post transient transfection via lipofectamine. Samples included parental cell line and Cells transfected with vector containing *CCND3* native, optimised, fast, middling, or swap sequences, and an empty vector. TAT was used as a loading control.

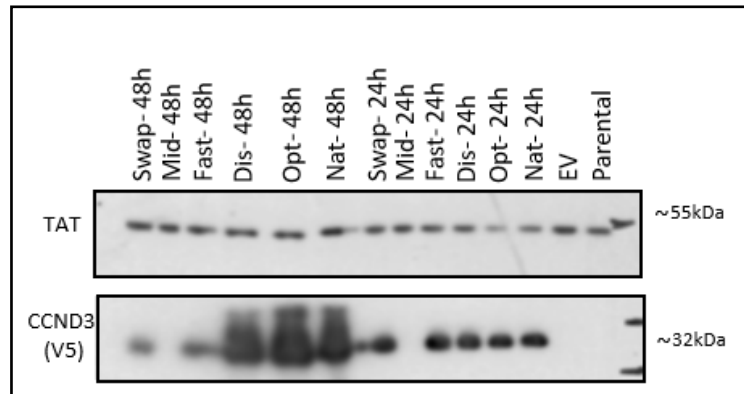


Figure 4.24A: Western blot analysis for V5 tagged CCND3 of Tv variant samples harvested from HEK 293 cells grown at 32°C, 24 and 48 hours post transient transfection via lipofectamine. Samples included parental cell line and Cells transfected with vector containing *CCND3* native, optimised, fast, middling, or swap sequences, and an empty vector. TAT was used as a loading control.

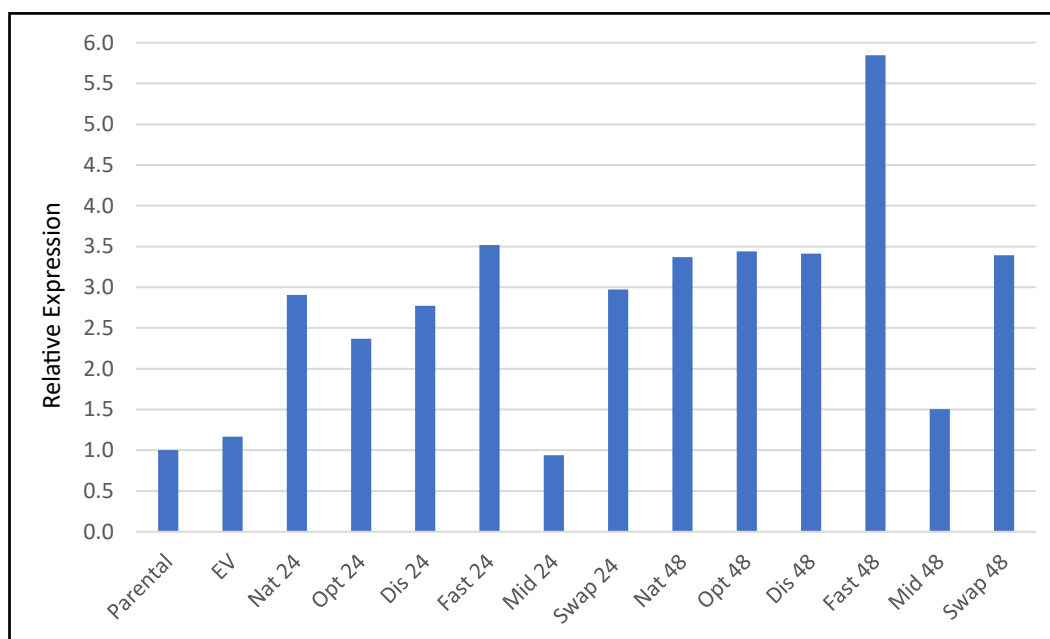


Figure 4.23B: Densitometry analysis of western blot CCND3 visualisation. Blot tested expression of the tagged, Tv variant of modified *CCND3*, harvested 24 or 48 hours post transient transfection from HEK293 cells. Expression of each sample has been normalized to corresponding TAT loading control and made relative to untransfected parental cell line. N= 1 per condition.

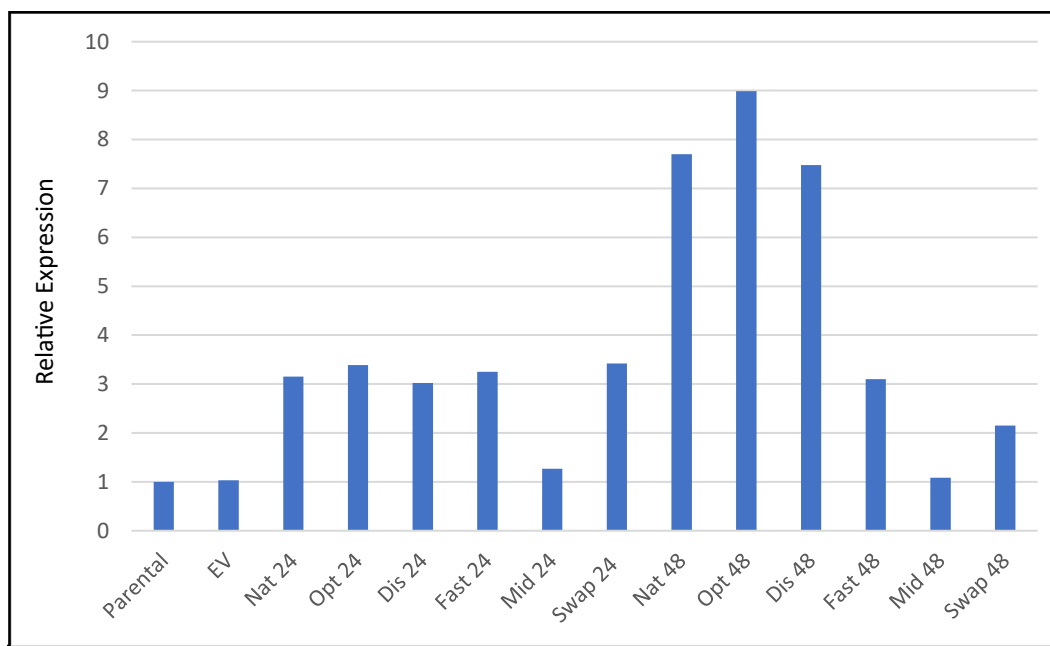
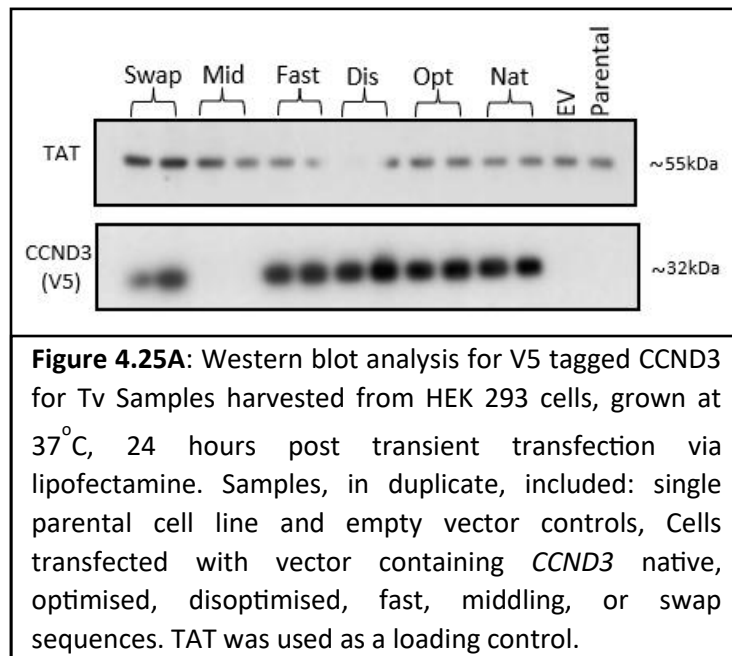


Figure 4.24B: Densitometry analysis of western blot CCND3 visualisation. Blot tested expression of the tagged, Tv variant of modified *CCND3*, harvested 24 or 48 hours post transient transfection from HEK293 cells. Expression of each sample has been normalized to corresponding TAT loading control and made relative to untransfected parental cell line. N=1 per condition.

4.4.4.2 Western Blot Analysis of Protein Production from Experimental Replicates of Each *CCND3* Transcript Variant Sequence

In Figure 4.25A, A band for *CCND3* protein production from the Middling transcript variant was again absent in both experimental replicates at 37°C, confirming that expression of this transcript was very low/undetectable in multiple experiments. Although the Disoptimised sequence yielded the highest *CCND3* protein expression relative to controls as shown by densitometry analysis (Figure 4.25B), this result partially reflects the low tubulin intensity band, and visually, Disoptimised *CCND3* expression was marginally increased compared to that from the Native transcript variant. This is reflected in Figure 4.25Bii when samples were not normalised to the tubulin (TAT) loading control and no statistical difference was seen between Native and Disoptimised samples. Optimised and Fast transcript variant mediated *CCND3* protein production was comparable to that from the Native transcript visually and upon densitometry analysis (Figure 4.25Bii) whilst the Swap transcript variant sequence was unanimously shown to result in reduced *CCND3* protein production.



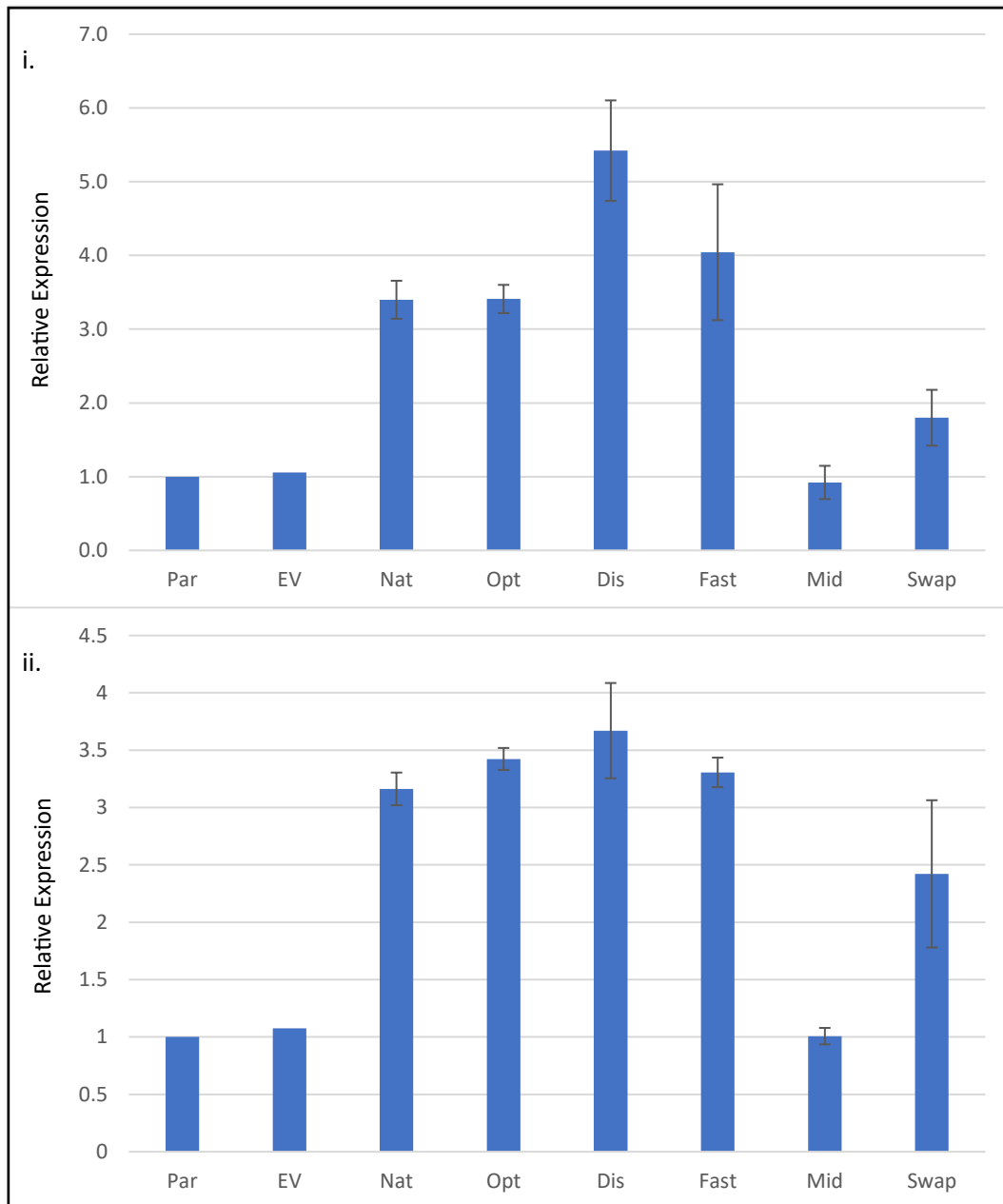


Figure 4.25B: Densitometry analysis of western blot CCND3 visualisation. Blot tested expression of the tagged, Tv variant of modified *CCND3*, harvested 24 hours post transient transfection from HEK293 cells. i. Expression of each sample has been normalized to corresponding TAT loading control and made relative to untransfected parental cell line. ii. Samples made relative to untransfected cell line without normalisation step. N=1 for control samples, N=2 for non-control samples.

Once again, at 32°C, like 37°C, CCND3 protein production from the Middling transcript variant was again absent in both experimental replicates (Figure 4.26A). For all these samples, tubulin expression was unsuccessfully visualised on western blot and as such *CCND3* expression values could not be normalised for densitometry analysis (Figure 4.26B). This likely reflects an issue with the tubulin antibody rather than the sample as CCND3 protein production was clearly present. As such, for this analysis CCND3 protein production was undertaken relative to Native samples for this study and an assumption made that all wells had equal protein loaded as normalised by the Bradford loading (Figure 4.26B and Figure 4.27B). Similar to the expression at 37°C after 24 hours, the Disoptimised transcript variant samples had the highest expression as determined by both densitometry analysis and visually on the western blot (Figure 4.26A). *CCND3* expression of the Optimised and Fast transcript variants was comparable to that from the Native, although the Optimised transcript sequence gave a small statistical increase in protein production. Swap transcript variant samples appeared to express the least CCND3 protein visually (apart from the Mid sample where no expression was observed), but this was not statistically significant compared to the expression from the Native transcript variant (Figure 4.26B).

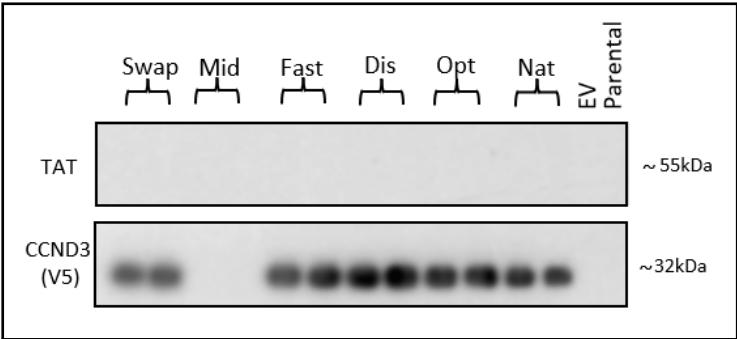


Figure 4.26A: Western blot analysis for V5 tagged CCND3 for Tv variants. Samples were harvested from HEK 293 cells, grown at 32°C, 24 hours post transient transfection via lipofectamine. Samples, in duplicate, included: single parental cell line and empty vector controls, Cells transfected with vector containing *CCND3* native, optimised, disoptimised, fast, middling, or swap sequences. TAT was used as a loading control.

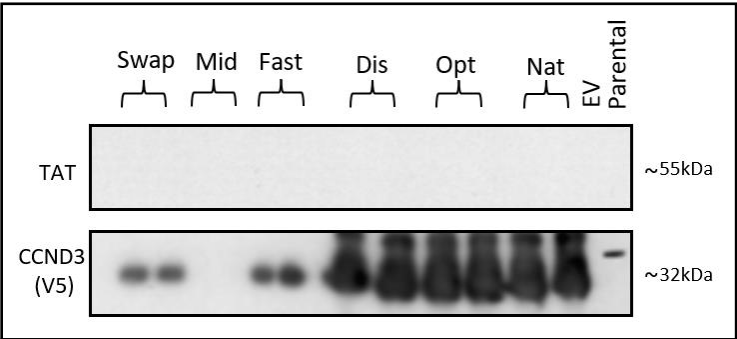
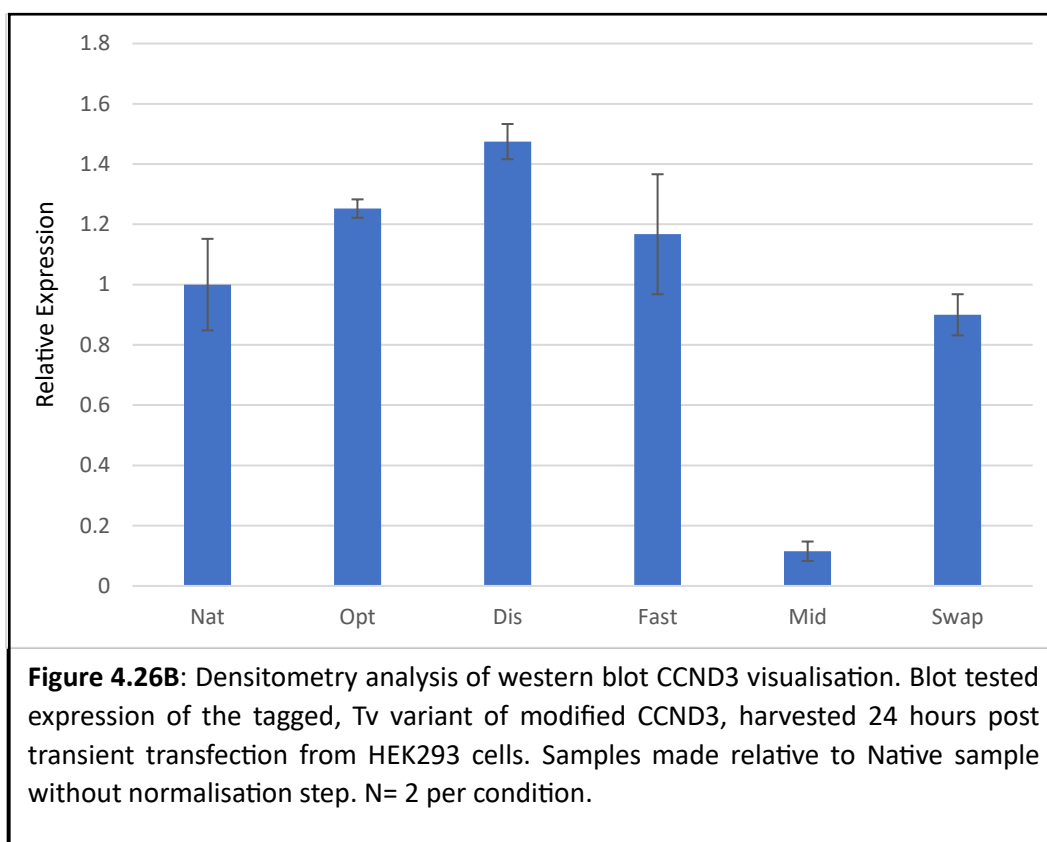


Figure 4.27: Western blot analysis for V5 tagged CCND3 of Tv variants. Samples were harvested from HEK 293 cells, grown at 32°C, 48 hours post transient transfection via lipofectamine. Samples, in duplicate, included: single parental cell line and empty vector controls, Cells transfected with vector containing *CCND3* native, optimised, disoptimised, fast, middling, or swap sequences. TAT was used as a loading control.



For samples taken at 32°C after 48 hours, the expression was again high for the Native, Optimised and Disoptimised transcript variant samples (Figure 4.27A) which made densitometry analysis difficult. Once again tubulin was not detected, likely due to an issue with the anti-tubulin antibody at this time. These issues make visual sample comparisons difficult and skew any densitometry analysis. However, it was clear that the Middling transcript variant samples failed to express detectable exogenous CCND3 and that the Fast and Swap transcript variants gave reduced CCND3 expression compared to the Native, Optimised and Disoptimised transcript variant sequences.

4.5 Determining any Impact of Stable *CCND3* Transcript Variant, and Subsequent Exogenous *CCND3* Protein Production, on the Growth Profiles of HEK293 Using the XCELLigence System

As CCND3 is involved with control of the cell cycle, it was important to determine if the over-expression of CCND3 in the model HEK293 cell system impacted the growth profile of cells. To investigate this, stably expressing pools of the different transcript variants were prepared. HEK293 cells were stably transfected

with Gv and Tv sets of the *CCND3* transcript variant modifications along with preparing cells transfected and selected with an empty vector (no *CCND3*) as a control. The Fsp1 restriction enzyme was used to linearise the plasmid, cutting in ampicillin gene, before transfection. Lipofectamine was used as a transfection reagent and cells were then incubated at 37°C for 3 weeks to recover with selection. Hygromycin was used as a selection pressure to remove unsuccessfully transfected cells from the population. Once stable pools were established, the xCELLigence system was used to monitor growth profiles over 6 days at 37°C where measurements were automatically taken every 30 minutes. Due to the slower growth, cultures incubated at 32°C were monitored over 10 days. The parental cell line and cells transfected with empty vector showed similar growth profiles at both temperatures, therefore any difference in growth profile observed in the *CCND3* transcript variant cell pools could be attributed to *CCND3* expression and not the transfection method or subsequent selection process applied.

Due to *CCND3*s involvement in the cell cycle, it was hypothesised that cells may grow faster when over-expressing *CCND3* due to a lack of cell cycle regulation concerning G1 to S phase transition. It was hypothesised that this could result in more rapid proliferation likely leading to a steep log (growth) phase, obtaining a higher stationary phase viable cell number, and a sharper decline or death phase compared to parental and empty vector samples. This however was not the case, and the opposite was observed. For both temperatures investigated, the growth profiles of the *CCND3* transcript variant expressing cell pools were similar in their nature to the controls and did not outgrow these, with the majority growing more slowly than the controls (Figure 4.28). There were two distinct growth profiles for *CCND3* transcript variant cell pools with a small difference for Tv recombinant *CCND3* expressors with the Middling, Fast and Swap transcript variant cell pools having a slower growth rate and achieving a lower cell density compared to the Native, Optimised and Disoptimised transcript variant expressing pools. These growth patterns did not directly correlate to expression of *CCND3* as higher *CCND3* expressing transcript variants did not necessarily grow to higher cell densities although the Native, Optimised and Disoptimised, that express the most *CCND3* did growth the most of the Tv transcript variants. Cells transfected with the Gv set of *CCND3* sequences achieved a maximum cell index (measure of cell density) of between 2.5 and 3, similar to the Tv set but without the clustering pattern of Native, Optimised and Disoptimised. Once again, the exogenous *CCND3* expressing pools did not reach the same cell index levels as parental or empty vector cultures. It may be the case that for stably transfected cell population, the burden of *CCND3* expression actually reduces cell growth or that the selection process inadvertently selects within each population for the lowest producing cohorts that have a growth advantage. This is discussed further in the discussion chapter of this thesis (Chapter 6).

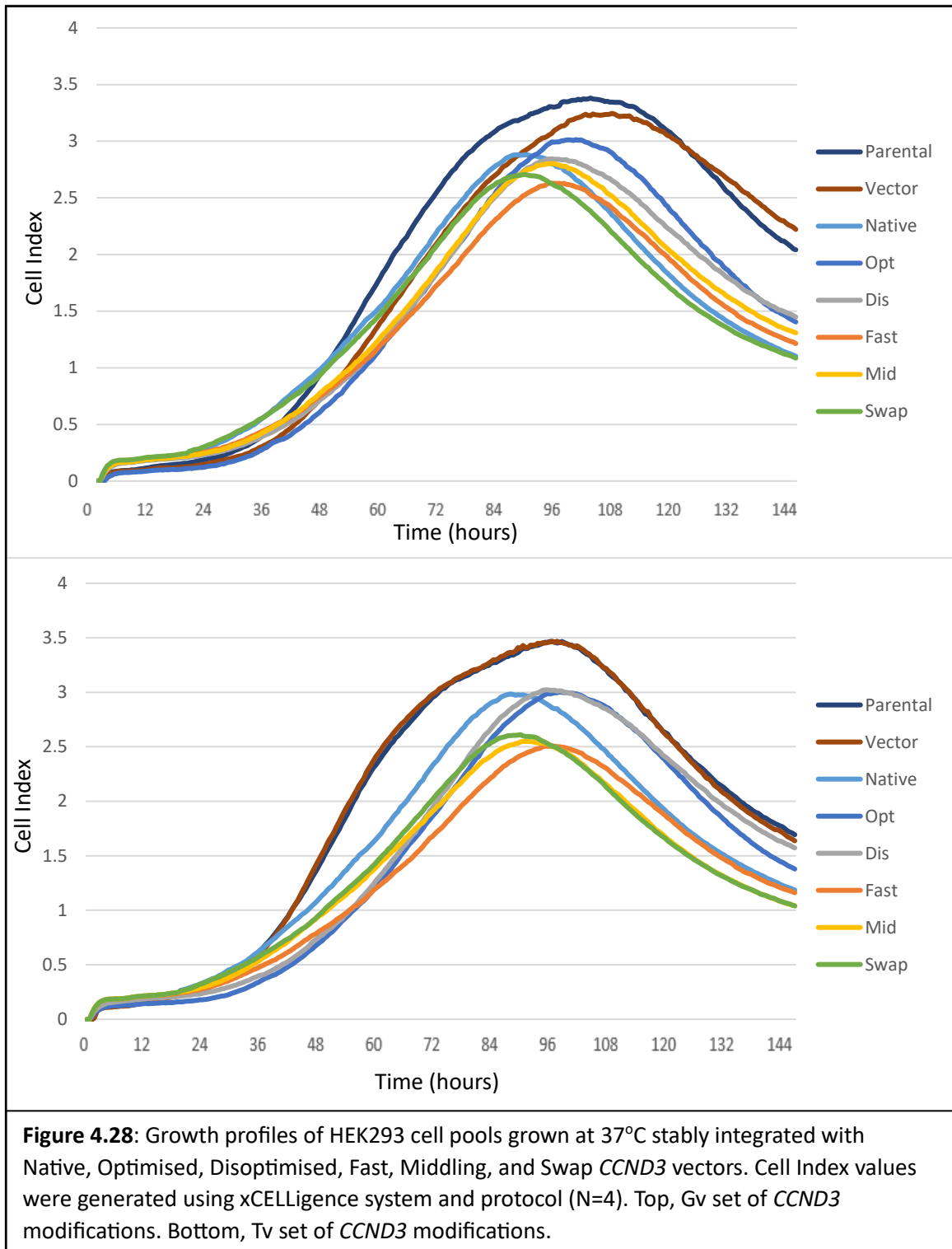
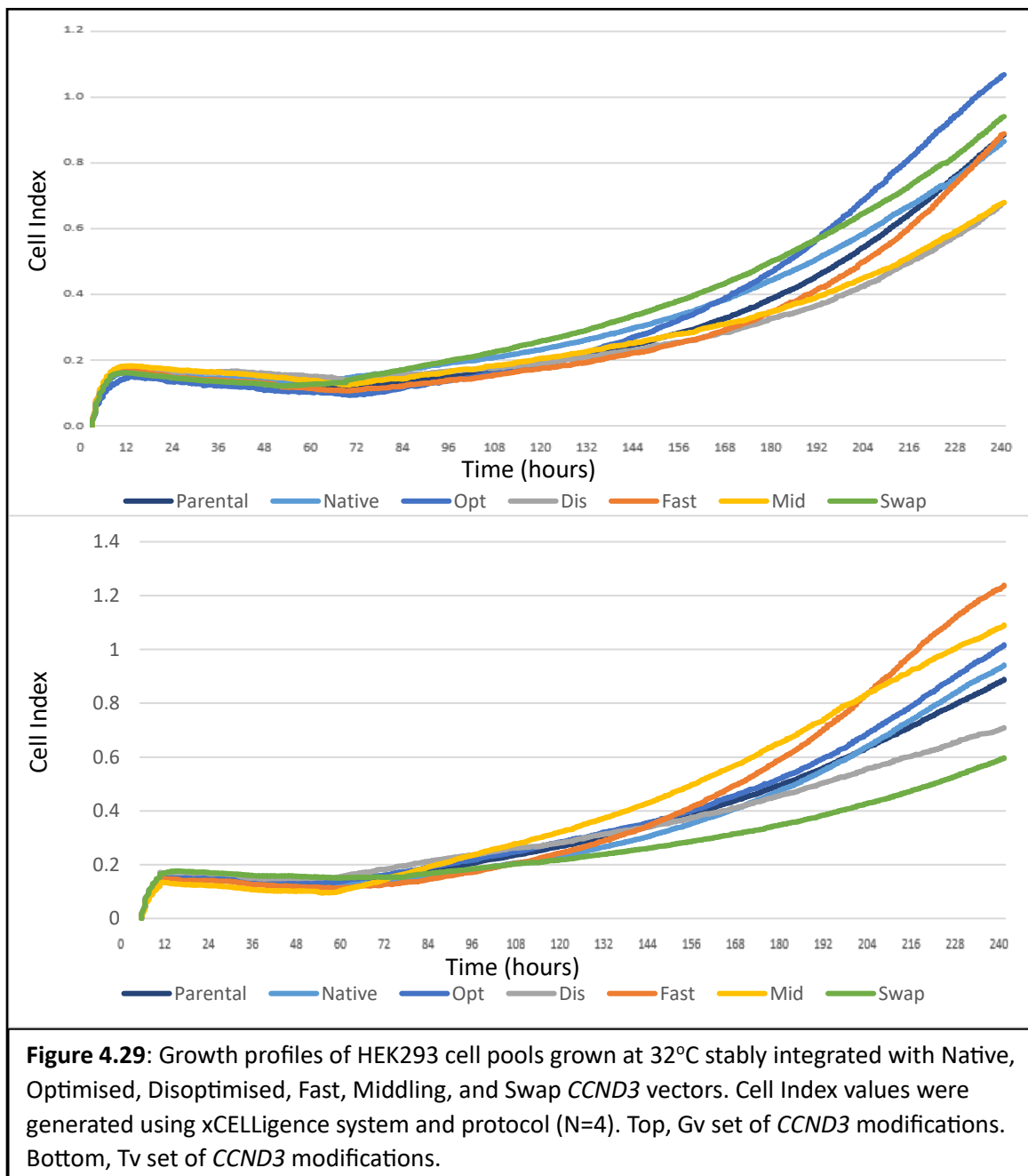


Figure 4.28: Growth profiles of HEK293 cell pools grown at 37°C stably integrated with Native, Optimised, Disoptimised, Fast, Middling, and Swap *CCND3* vectors. Cell Index values were generated using xCELLigence system and protocol (N=4). Top, Gv set of *CCND3* modifications. Bottom, Tv set of *CCND3* modifications.

The xCELLigence study also confirmed that cell growth was greatly reduced at 32°C compared to 37°C, with cells remaining in the lag phase for an extended period of time (Figure 4.29). This is typical for cultures grown at mildly hypothermic conditions and *CCND3* overexpression does not appear to alter this. Some growth was eventually achieved but after prolonged culture, far longer than the time points examined by western blot analysis of *CCND3* protein production. The gradient of this potential entry into log phase was lower than for cultures incubated at 37°C, further confirming that growth at 32°C is attenuated.



Chapter 5

Investigating Codon Transcript Variant Impact on GFP and Noggin Protein Production

5.1 Introduction

As discussed in depth in the Introduction chapter of this thesis, multiple factors and processes can impact the three basic steps of translation elongation and thus protein synthesis which include: codon decoding/tRNA selection, peptide bond formation, and translocation of mRNA-tRNA complex. Specific limiting processes or factors include slow peptidyl-transfer kinetics between specific amino acids such as Pro-Pro peptide formation, downstream secondary mRNA structures inhibiting ribosome progression, and availability of associated translocation machinery amongst others (Knight et al., 2015; Pavlov et al., 2009; Tholstrup et al., 2012). A particular potentially rate limiting factor investigated in this study is codon decoding times due to tRNA abundance and near-cognate tRNA abundance (Quax et al., 2015). As outlined elsewhere, ribosomes can experience pausing/stalling whilst waiting for the correct tRNA to enter the A site. Evidence suggests that initial codons succeeding the start codon have a greater impact on elongation rates in some mRNA transcripts (Chu et al., 2012). Codons in this region that correspond to abundant tRNAs enhance elongation rates whereas codons whose tRNA counterparts are limited hamper elongation, resulting in ribosomal stalling or slow elongation rates. This stalling not only impacts elongation rates, but can also feedback to hinder translation initiation as well and hence limit the number of ribosomes that are on an mRNA actively translating it at any one time (Chu, Kazana, et al., 2014).

Unlike *CCND3*, for both *Noggin* and *eGFP* studied in this chapter, the identified ROI and hence stretch of codons of interest is located at the 5' end of the coding sequence directly following the AUG start codon. These regions thus provide an additional insight into elongation control and how the decoding speed of this section impacts availability of the start codon and region around this; faster decoding in the 5' open reading frame is hypothesised to potentially increase initiation rates. The *Noggin* transcript was identified in other studies in the laboratory as having an elongation limiting initial 5' open reading frame sequence and the study of this would complement that of *CCND3* where the ROI was centrally in the ORF whilst *eGFP* is widely used as a reporter protein. The *eGFP* sequence has previously been optimised for mammalian expression therefore it was of interest to determine whether modifications to codon usage

focused on elongation control could further increase eGFP protein production compared to the initial ‘native’ sequence.

The codon usage in eGFP and *Noggin* transcript sequences were therefore modified at the 5’ end of the open reading frame to generate transcript sets encoding for the same amino acid sequence while using different codons as described in Chapter 3. Codon usage was determined by tRNA abundance and decoding times calculated for different codon variants using a previously published elongation model (Chu et al., 2012). Two sets of transcripts were generated for each model protein to assess whether protein secretion was impacted by modifying decoding speeds of the ER signal peptide that targets secretory proteins to the ER. To achieve this, the *Noggin* transcript was designed with and without its ER signal sequence and a secretory eGFP sequence was also designed as outlined in Chapter 3. For a peptide to enter the secretory pathway and ER co-translationally, the signal recognition particle (SRP) must bind the signal peptide as it emerges from the ribosome exit tunnel and attenuate mRNA translation until the SRP binds the SRP receptor on the ER, is released and elongation begins again as the peptide is ‘fed’ into the ER through the Sec61 translocon (Akopian et al., 2013; Linxweiler et al., 2017; Mason et al., 2000). It was hypothesised that changing the elongation speed of the signal peptide might interfere with SRP binding and whether attenuation of translation elongation would be impacted. To assess the impact of the different codon variants on protein synthesis, 6 well plates were seeded with HEK293 cells which were transfected with plasmids containing the appropriate recombinant DNA sequences or ‘empty’ for transfection control in duplicate. An un-transfected ‘parental’ cell line was also included as an additional control. Cell lysate was harvested for non-secretory samples and supernatants were used for secretory samples. Lysate samples were also harvested for secretory-GFP (sGFP) to assess total sGFP content, intracellularly and extracellularly. Tables 5.1 and 5.2 below are a reminder of the sequences designed, investigated and their abbreviations used in this chapter of work.

Table 5.1: eGFP transcript variant sequences designed and investigated in this chapter.

eGFP variants	Modified Sequence Abbreviations		
	Native	Optimised	Dis-optimised
Secretory-eGFP	<i>sGFPN</i>	<i>sGFPO</i>	<i>sGFPD</i>
eGFP	<i>GFPN</i>	<i>GFPO</i>	<i>GFPD</i>

Table 5.2: Noggin transcript variant sequences designed and investigated in this chapter.

Noggin variants	Modified Sequence Abbreviations			
	Native	Optimised	Dis- optimised	Swap
Modified Signal Sequence		<i>SNogO</i>	<i>SNogD</i>	<i>SNogS</i>
Removed Signal Sequence	<i>NogN</i>	<i>NogO</i>	<i>NogD</i>	<i>NogS</i>

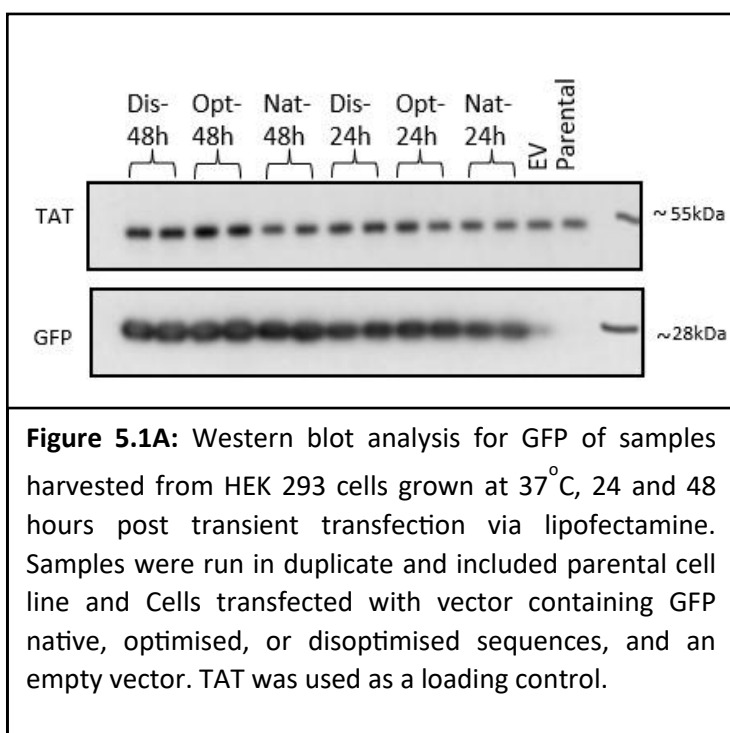
As with *CCND3* expression analysis in Chapter 4, two time points at two different incubation temperatures were studied: 24 and 48 hours and 32°C and 37°C respectively. The need for multiple time points was apparent as protein production trends between modified transcript sequences appeared to show time dependant variations. Cell growth was also monitored to determine any effect recombinant protein production had on HEK293 cell lines using cultures stably transfected with linearised *Noggin* or *eGFP* plasmids.

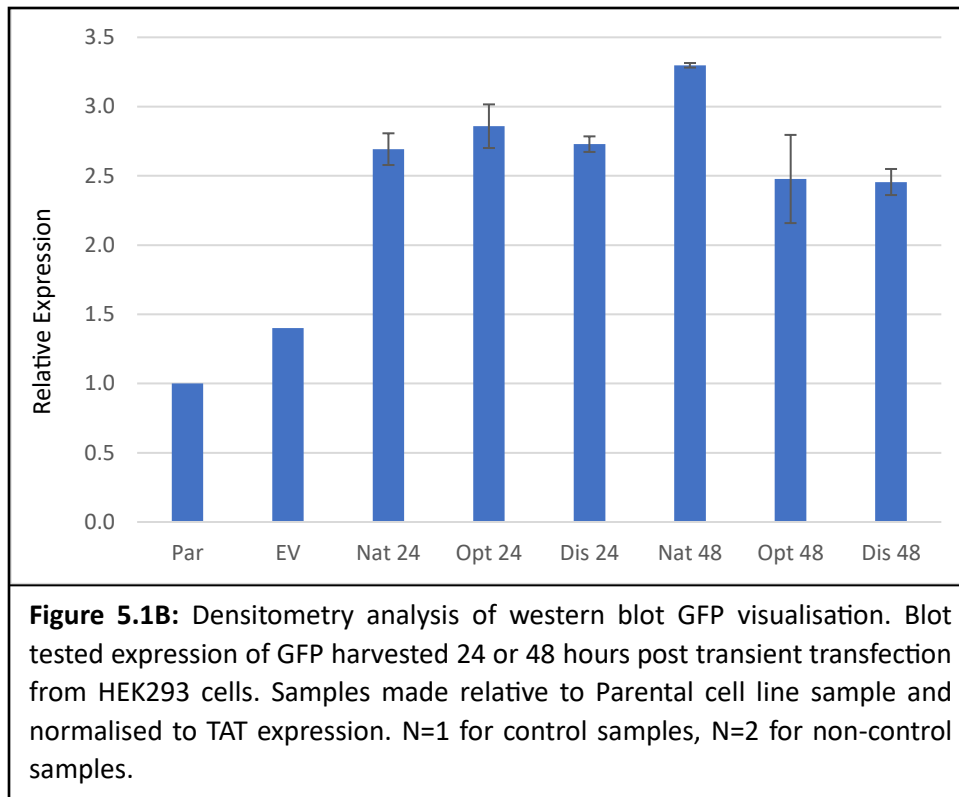
Protein production was analysed by western blotting and densitometry was carried out where appropriate. All western blots were performed following the parameters outlined in the Methods chapter (Chapter 2) and reduced samples were generated from cell lysates or supernatants, with protein levels adjusted for equal protein loading for better comparisons. Bradford assays were performed on cell lysates or supernatant for this purpose. Tubulin (TAT) was used as a loading control for cell lysate samples due to its ubiquitous and constitutive expression. Issues with the tubulin anti-TAT antibody efficacy became apparent across the duration of this study, impacting densitometry analysis. In some cases, relative densitometry was carried out without normalisation to tubulin to better represent expression results visually present on western blots. For all supernatant samples, no control protein was appropriate and a constant supernatant volume was loaded. Error bars were plotted using the standard error of the mean for western blots containing multiple repeats per experiment condition. No overlap of error bars between samples was determined to be an indicator that results were different in a statistically significant way. P-

values were also calculated and can be found in the appendix section. Densitometry analysis for Noggin native samples also lack error bars as $n=1$ for these samples on western blots

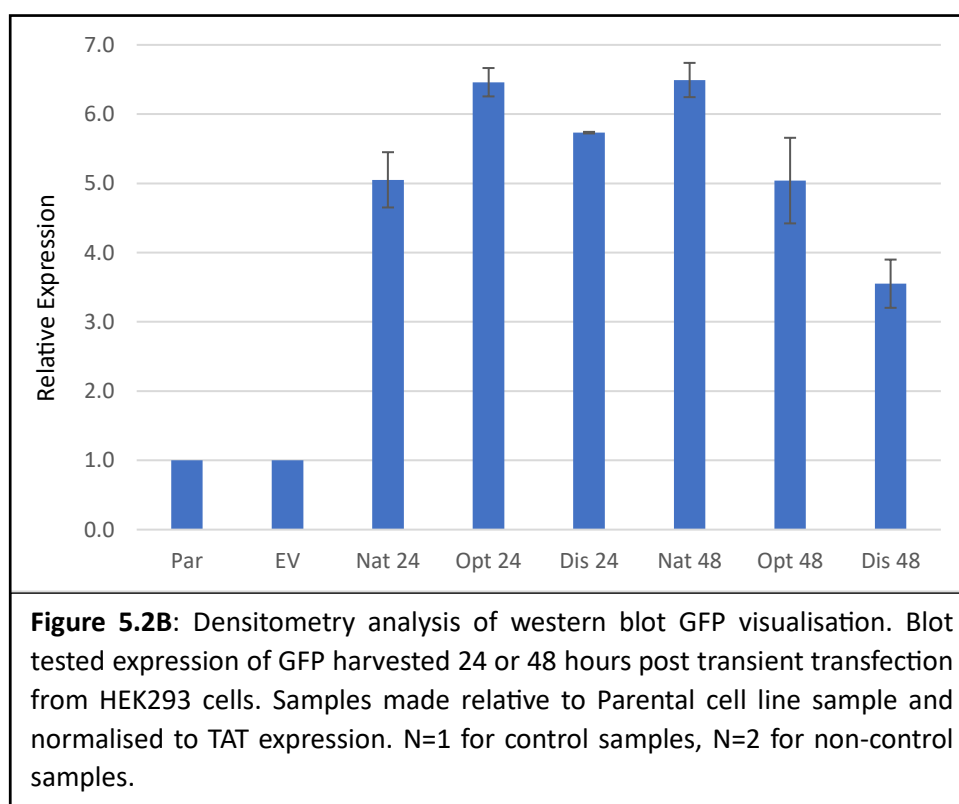
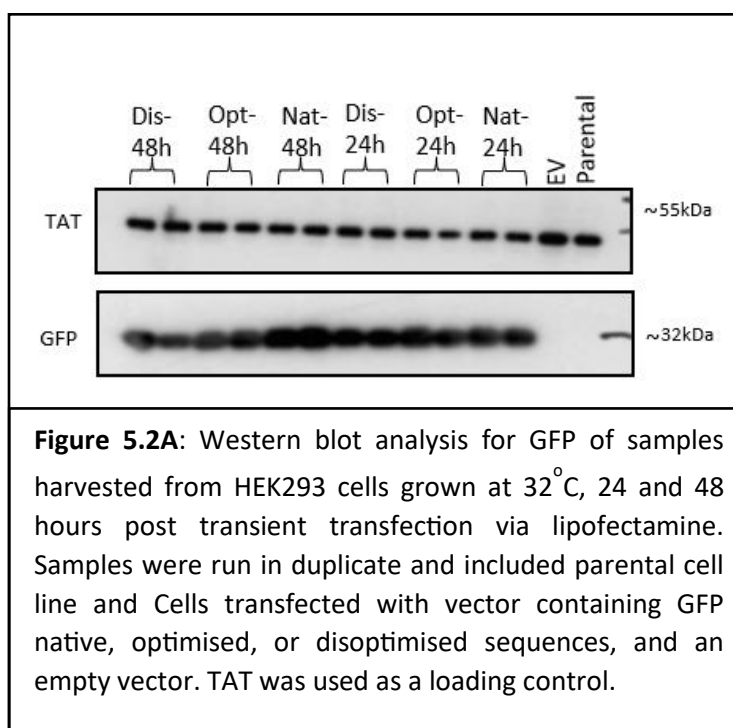
5.2 eGFP Protein Production from Different Transcript Variants in HEK293 Cells Incubated at 37°C or 32°C at 24 and 48 hour Post Transient Transfection

The different eGFP transcript variants were transfected into HEK293 cells and the resulting eGFP protein production monitored by western blot. The empty vector control lane of cells cultured at 37°C unexpectedly had a faint band present (Figure 5.1A). This sample should show no expression which was the case for all other blots and is thus most likely due to overspill of Native sample during SDS PAGE well loading rather than genuine Empty vector *GFP* expression. The parental samples showed no expression as expected, further confirming this as the likely answer. At 24 hours, the Optimised transcript variant samples visually showed stronger GFP signals than that from the Native, but once normalised to tubulin loading during densitometry analysis there was no statistical difference between these (Figure 5.1B). By the 48 hour time point, the optimised eGFP protein signal was less intense than that from the Native and shown to be statistically lower upon densitometry analysis. A similar trend was observed for the Disoptimised samples suggesting that manipulation of the codon usage of the ORF of eGFP at the 5' end did not have a major impact on the resulting protein synthesis.



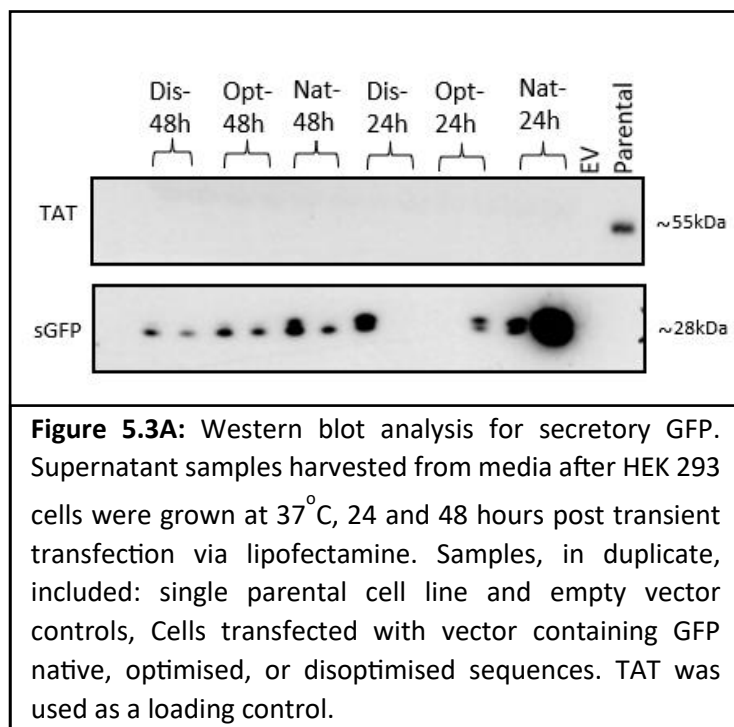


When samples that had been cultured at 32°C were investigated, no *eGFP* expression was observed for the Empty vector sample as expected, which supports the opinion that the faint signal in figure 5.1A was caused by well overspill whilst loading SDS PAGE. The tubulin anti-TAT expression across all samples was comparable except for Optimised 24 hour replicate 1 which was slightly reduced (Figure 5.2A). Optimised transcript variant mediated protein production of GFP at 24 hours was visually increased compared to that from the Native transcript which was corroborated to be statistically significant by densitometry analysis (Figure 5.2B). This was also the case for Disoptimised samples which exhibited higher expression than that from the Native transcript but lower than that from the Optimised transcript. By 48 hours after transfection, production of GFP was reduced compared to 24 h for both modified sequences compared to that from the Native visually on the western blot and this was confirmed by densitometry analysis. These trends are consistent with the trends observed in Figure 5.1A suggesting temperature change from 37 to 32°C does not impact the relative elongation speed between different modified sequences for *GFP* as observed previously for *CCND3* transcript variants where the ROI was in the middle of the ORF.



5.3 Secretory eGFP Protein Production from Different Transcript Variants in HEK293 Cells Incubated at 37°C or 32°C at 24 and 48 hour Post Transient Transfection

Expression of secretory eGFP (sGFP) from different transcript variants from cultures grown at 37°C was inconsistent after 24 hours, with high variance between replicates and several lanes showing no evidence of the presence of sGFP (one Disoptimised and Optimised sample, Figure 5.3A). Native expression from replicate one was also considerably higher than any of the other sample, including the second replicate (5-fold increase by densitometry analysis). The reason for this is not clear. Variation between the samples resulted in large SEM values as seen on the error bars. Figure 5.3Bii illustrates differences in apparent expression between samples with expression made relative to the Native 24 hour replicate 2. Secretory expression in general and the consistency between replicates was improved by 48 hours post transfection except for that from the Native transcript which remained with highly variability between the replicates. Optimised and Disoptimised transcript variants both gave less sGFP than that from the Native, with the optimised giving greater expression than from the Disoptimised transcript variant as expected. A lack of a tubulin band (TAT) in samples gave confidence that cells were not lysed during supernatant harvesting and therefore that the observed material was secreted eGFP.



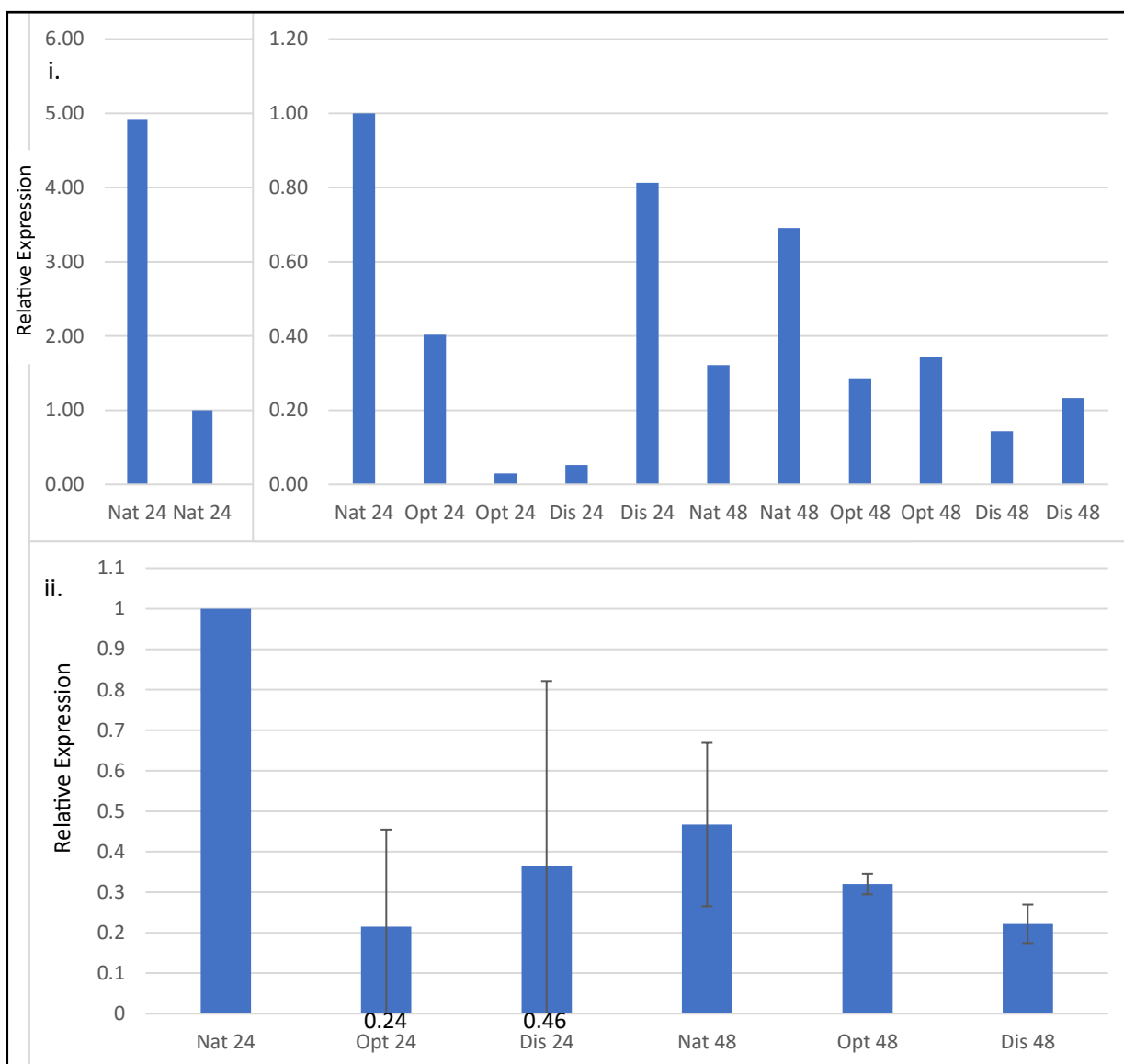
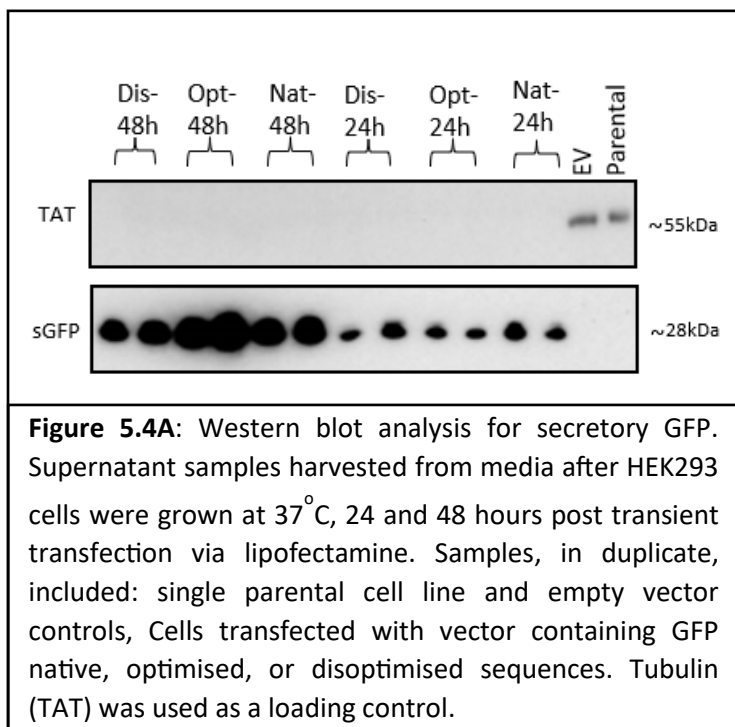
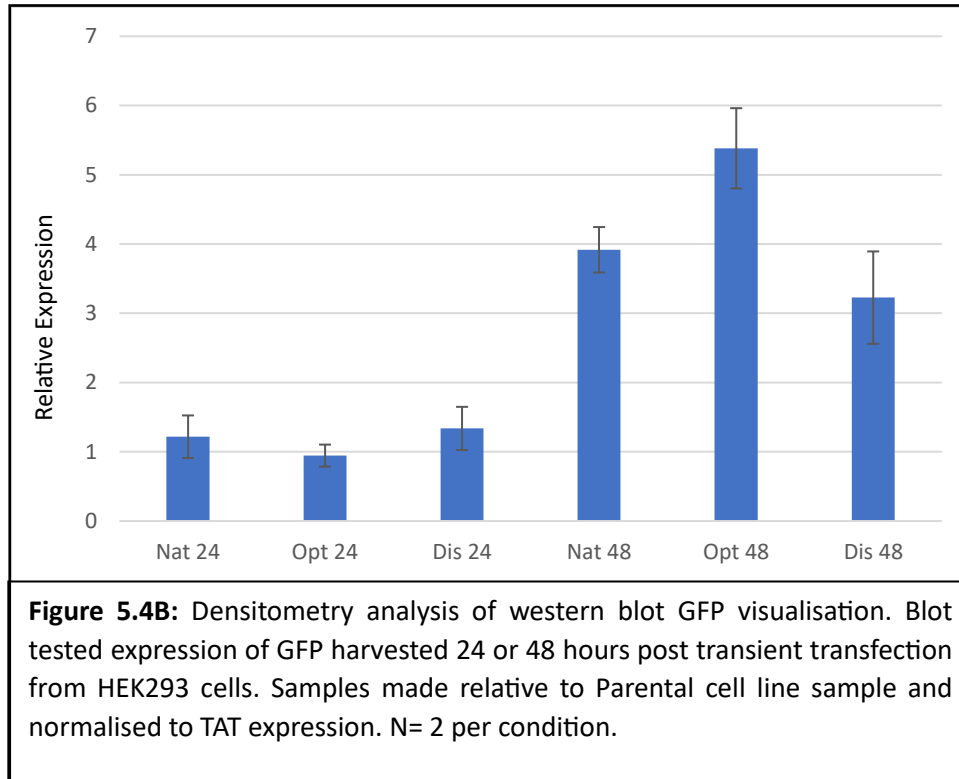


Figure 5.3B: Densitometry analysis of western blot GFP visualisation. Blot tested expression of GFP harvested 24 or 48 hours post transient transfection from HEK293 cells. i. Left, 24 hour Native replicate comparison relative to one another. Right, replicate expression comparisons relative to Native 24 hour replicate 2. N=1 for each condition. ii. Sample expression mean relative to Native 24 hour replicate 2, N= 1 for Nat 24, N= 2 for other conditions.

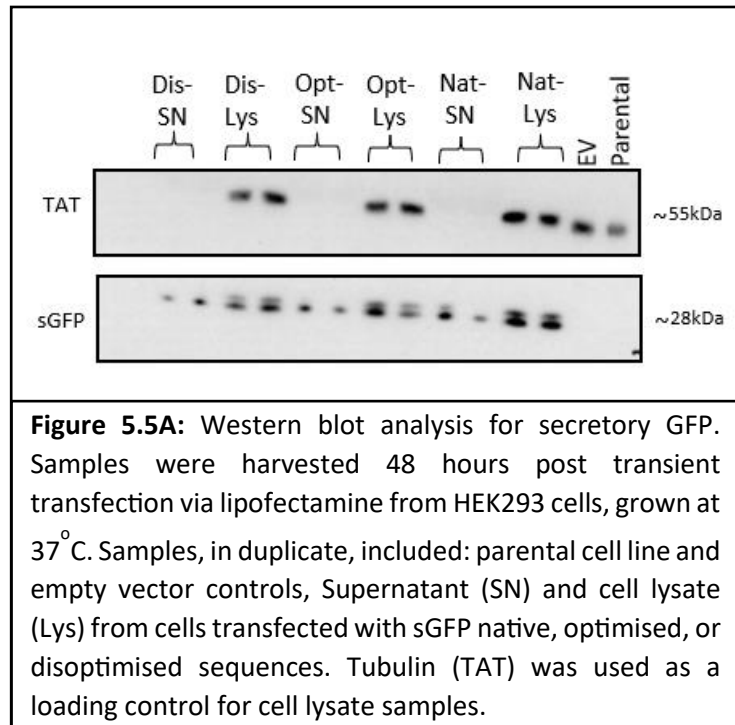
For those samples generated from cells cultured at 32°C after transfection, sGFP presence in supernatant samples was far more consistent at 24 hours and there was a stronger signal still, indicative of protein accumulation in the media, by 48 hours post-transfection (Figure 5.4A). After 24 hours, Optimised and Disoptimised transcript variant sequences cause no statistical change in protein production compared to that from the Native. However, by 48 hours both Optimised transcript variant replicates showed greatly increased expression compared to that from the Native and Disoptimised transcript variants. This visual trend was corroborated by densitometry analysis and shown to be statistically significant. *sGFPO* and intracellular *GFPO* samples displayed different trends in this regard which may reflect differences in the sensitivity of the optimisation when including or excluding the ER signal peptide and the additional steps involved in secretory expression compared to intracellular expression. Disoptimised transcript variant derived sGFP protein signals appeared weaker than Native at 48 hours (Figure 5.4A), but the difference was not statistically significant upon densitometry analysis (Figure 5.4B).





Cell lysate samples were also investigated of the secretory *eGFP* transfected cells to determine if intracellular material was present. For samples cultured at 37°C, for each *sGFP* transcript variant sequence, the *sGFP* band observed in the western blots in lysate samples was more intense than in supernatant samples (Figure 5.5A). However, it is difficult to interpret this in terms of concentration of *eGFP* as the lysate sample is concentrated from many cells in a small lysis volume whilst the supernatant samples are diluted by comparison in the culture media. Both Optimised and Disoptimised transcript variant modifications failed to result in increased *sGFP* expression in the lysate or supernatant samples relative to that from the Native transcript. The total *sGFP* expression by cells was assessed and confirmed that Native expression was greater than that from the Optimised transcript variant, which was greater than that from the Disoptimised transcript variant. Double banding was observed for *sGFP*, predominantly in lysate samples, but this was not observed in supernatant samples. This shift in apparent size is accounted for by the presence or absence of the ER signal sequence. The ER signal sequence is cleaved off after synthesis but before secretion of the protein, resulting in a protein of a smaller size once secreted compared to that containing the ER signal sequence. This confirms the protein in the supernatant was transiting through the

ER and was actively secreted from the cell as opposed to being released due to cell death or lysis during culture or harvest.



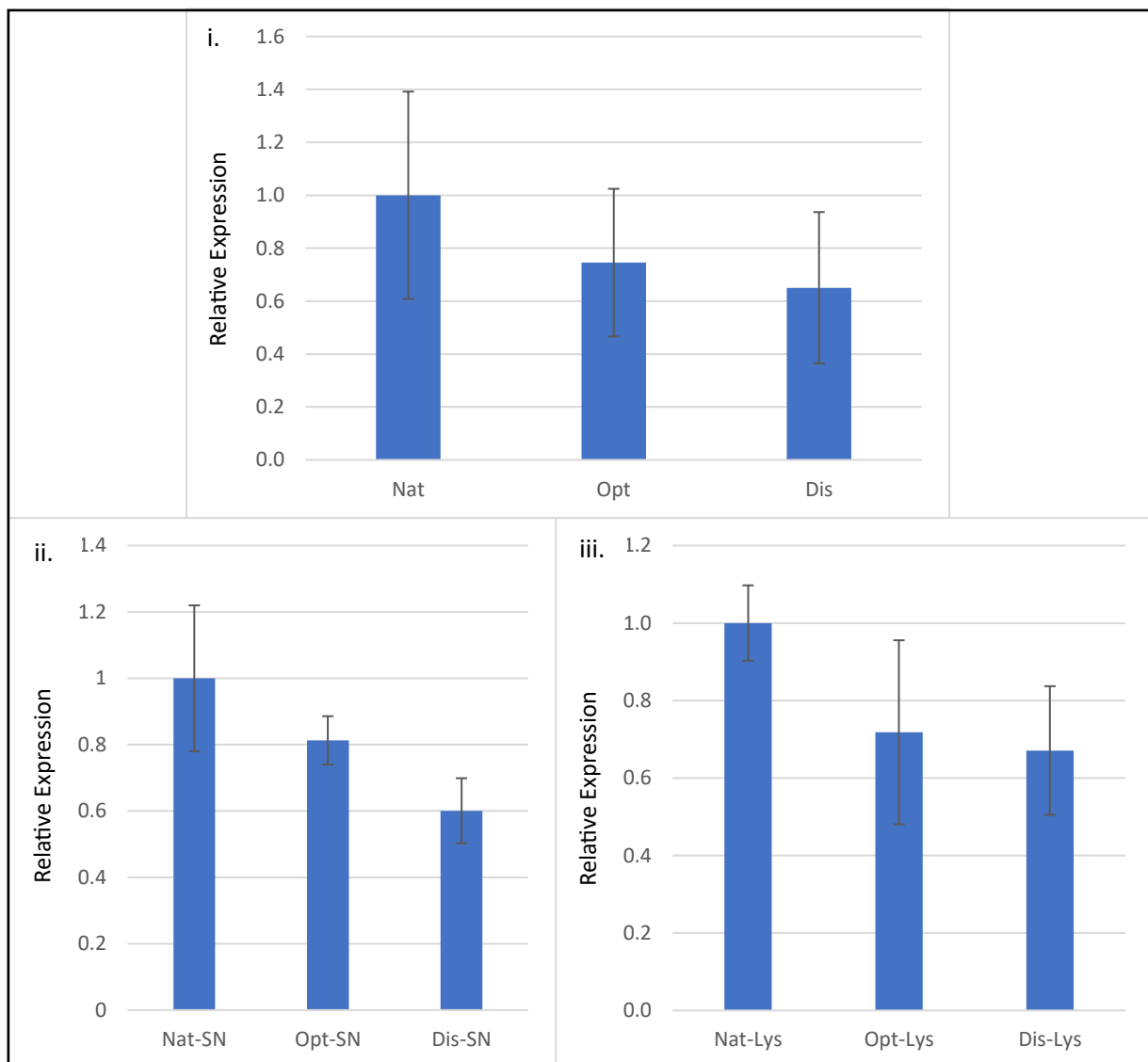
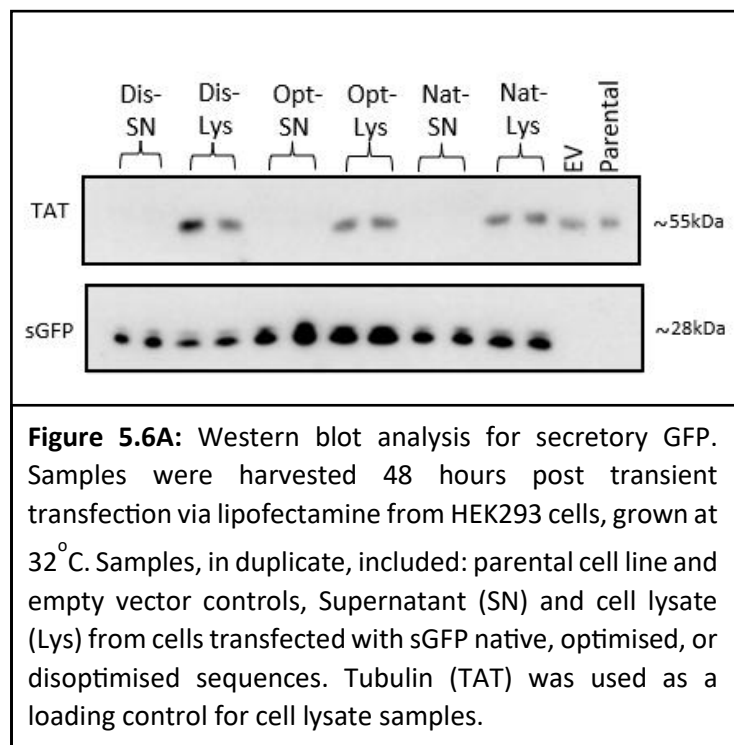
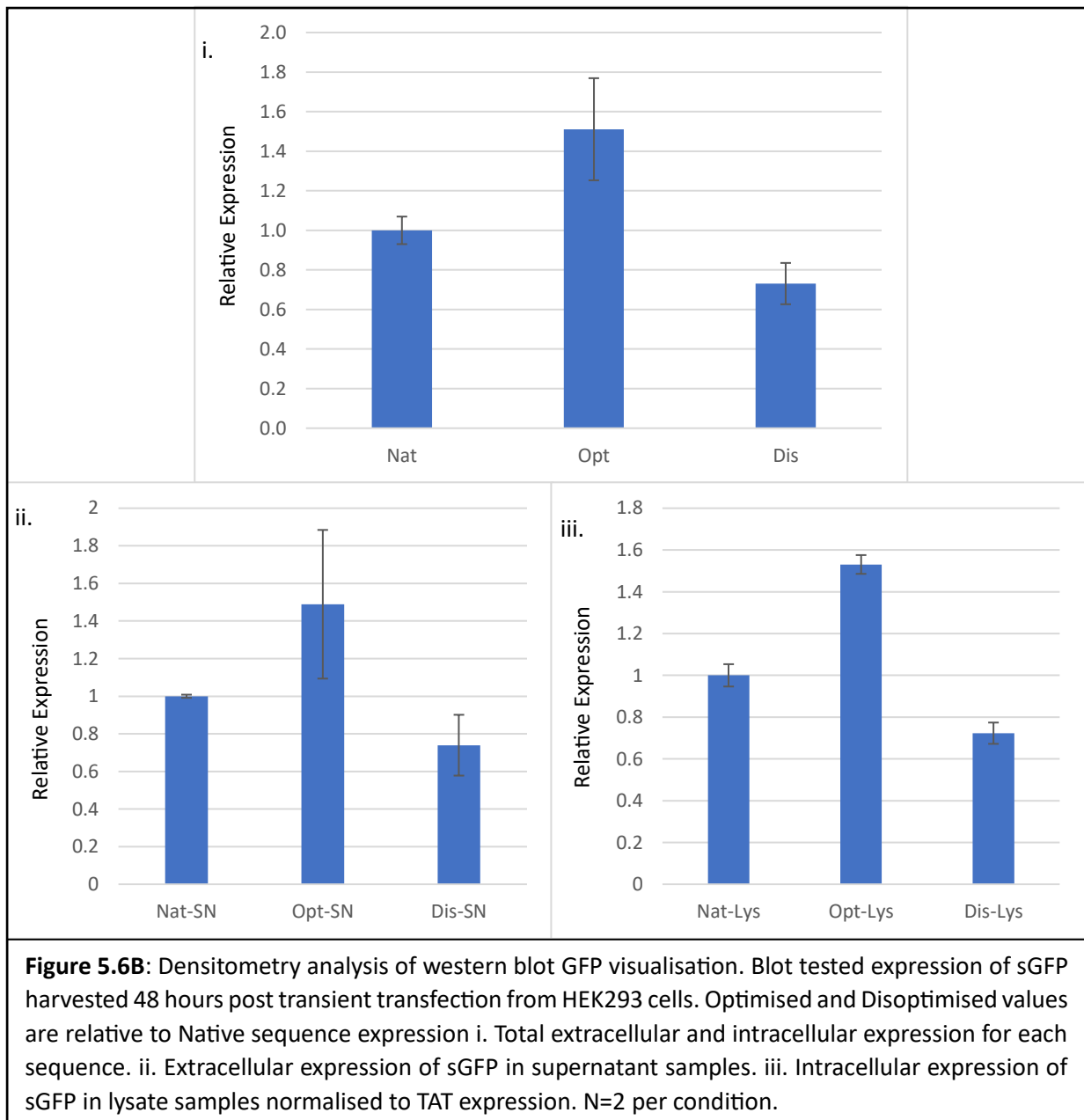


Figure 5.5B: Densitometry analysis of western blot GFP visualisation. Blot tested expression of GFP harvested 48 hours post transient transfection from HEK293 cells. Optimised and Disoptimised values are relative to Native sequence expression i. Total extracellular and intracellular expression for each sequence. ii. Extracellular expression of sGFP in supernatant samples. iii. Intracellular expression of sGFP in lysate samples normalised to TAT expression. N= 2 per condition.

When cells transfected and cultured at 32°C were analysed, once again the lack of a tubulin (TAT) signal in extracellular samples gives confidence that cells were not lysed during supernatant harvesting (Figure 4.6A). The samples cultured at 32°C resulted in a smaller disparity in sGFP levels between supernatant and lysate samples for each sGFP sequence than seen in the 37°C cultured cell samples (Figure 4.5A). This suggests the cells may be able to secrete the protein more efficiently at 32°C than at 37°C. Optimised transcript variant sGFP protein production had the strongest signal for both lysate and supernatant samples suggesting this sequence gave the highest protein production at 32°C. The enhanced expression over that from the Native transcript was confirmed to be statistically significant by densitometry analysis of lysate and supernatant samples as well as combined total sGFP expression. The Disoptimised transcript variant samples displayed poorer *GFP* expression than other sequences and were statistically lower than that from the Native transcript for lysate, supernatant and combined total intracellular and secretory expression. These observations closely align with the results from the 48 hour samples in figure 5.4A. Double bands close in weight were again observed, however these appeared to be present in all samples unlike at 37°C. The presence in both lysate and supernatant samples of the heavier band, that presumably contains the ER signal sequence, suggests that the processing and cleavage of the ER signal sequence at the reduced temperature may be compromised compared to this process at 37°C. Alternatively, the band could be the result of unexpected post-translational modification(s).





5.4 Assessment of the Impact of *GFP* Transcript Variant Expression on HEK293 Growth using the xCELLigence Instrument

To determine if the expression of the *GFP* transcript variants had any differential impact on the growth of the HEK293 model system, HEK293 cells were stably transfected with linearised *eGFP* and *sGFP* transcript variant constructs along with an empty vector as a control (Figure 5.7). The *FspI* restriction enzyme was used to linearise the plasmid, cutting in a single site in the plasmid approximately midway through the ampicillin gene. Lipofectamine was used as a transfection reagent and cells were then incubated at 37°C for 3 weeks to recover. Hygromycin was used as a selection pressure (and added 24 h post-transfection) to remove unsuccessfully transfected cells from the population. The xCELLigence culture system was then used to monitor growth profiles of the resulting cell pools where readings were automatically taken every 30 minutes. *sGFP* transcript variant transfected cultures exhibited the greatest variation in growth profiles when compared to any other group of cell pools. All *sGFP* transcript variant pools obtained reduced maximum cell indexes compared to the parental HEK293 cell line, with the *sGFP*-N transcript variant pools performing poorest, reaching slightly over half the cell index of parental cultures (Figure 5.7). *sGFP*-N cell pools also experienced an extended lag time compared to other samples. For the intracellular expressing pools, comparable maximum cell indexes were obtained for all cell pools and the parental cell line, but they varied in the time taken to achieve this (Figure 5.7).

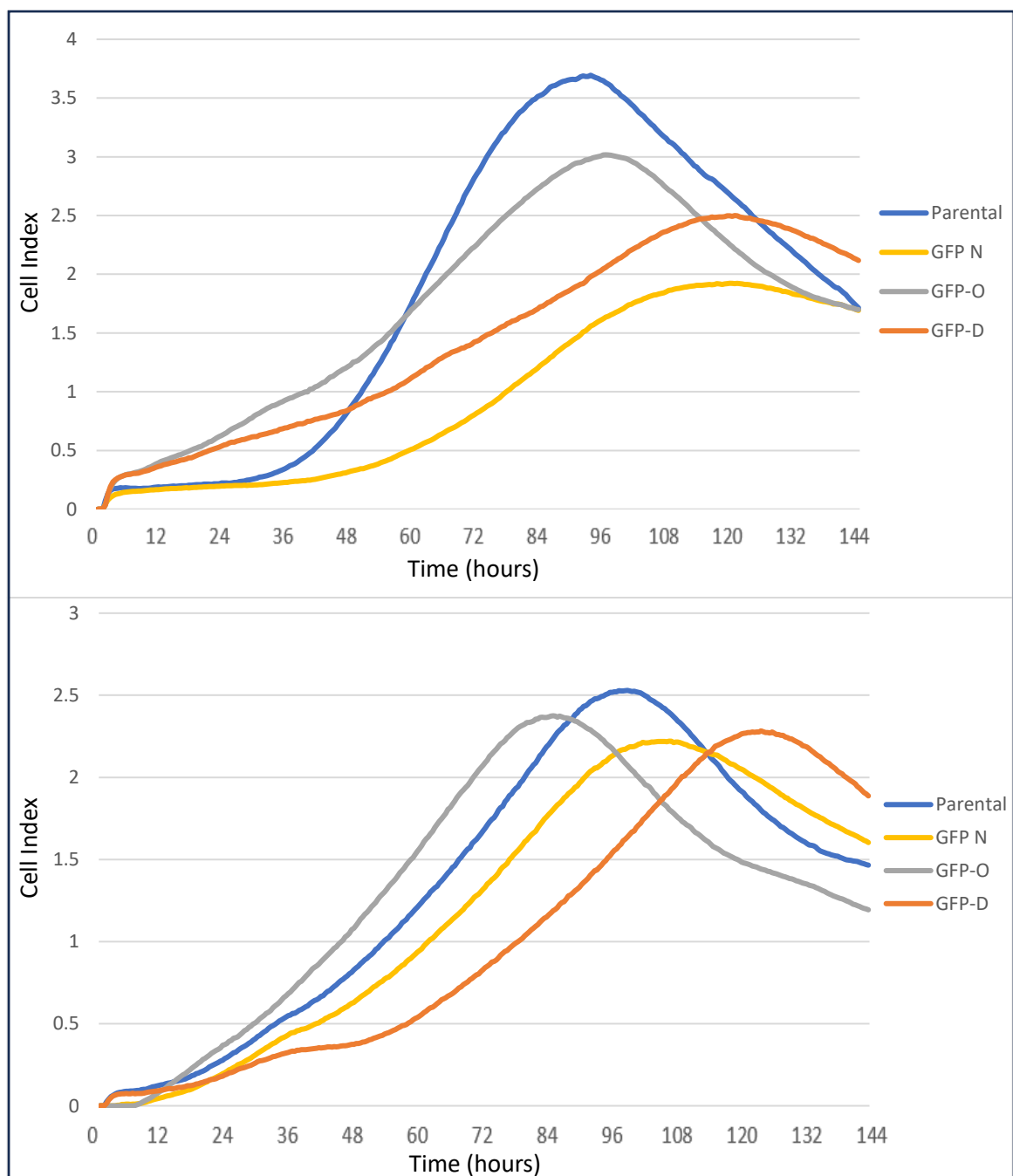
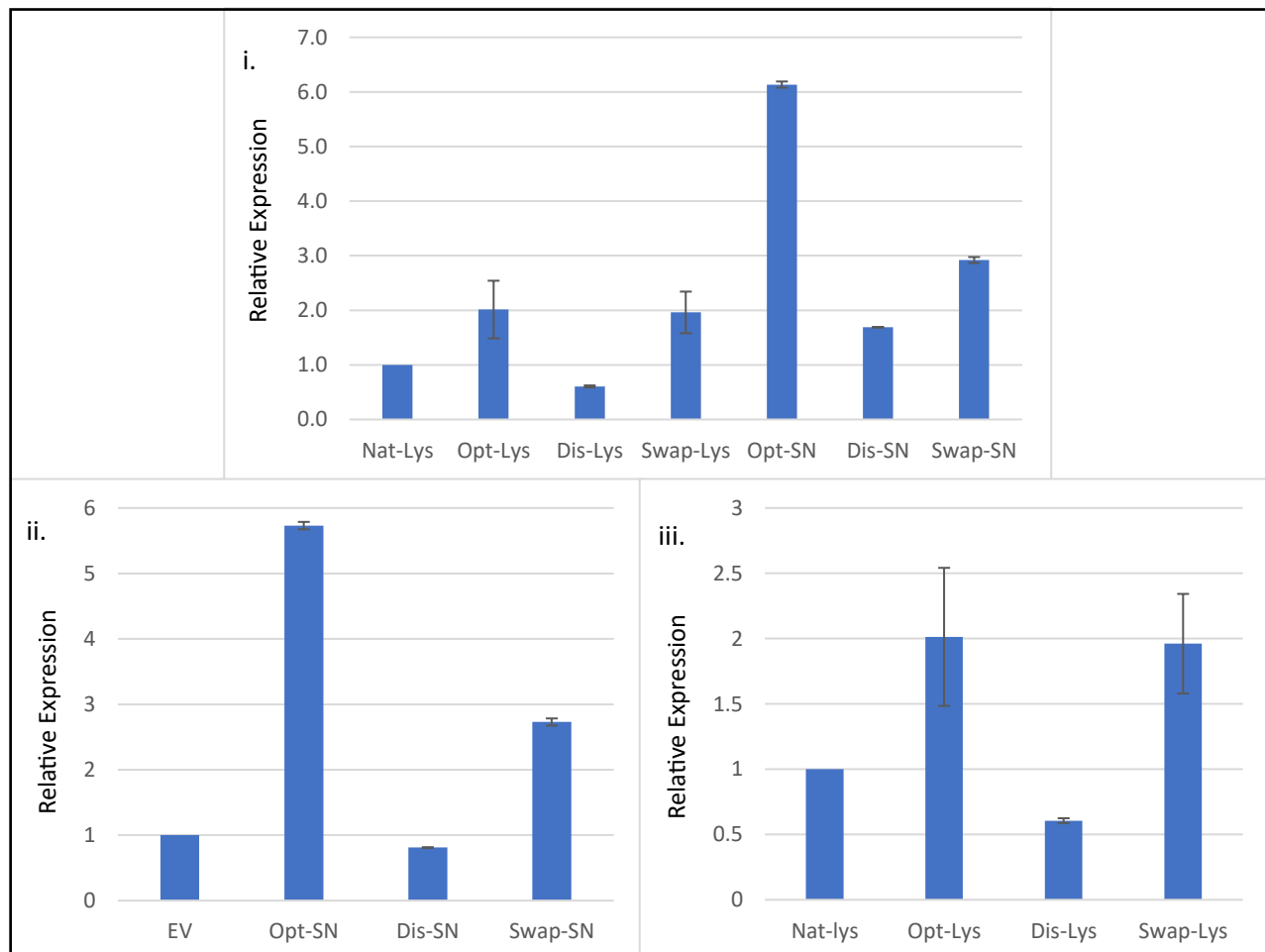
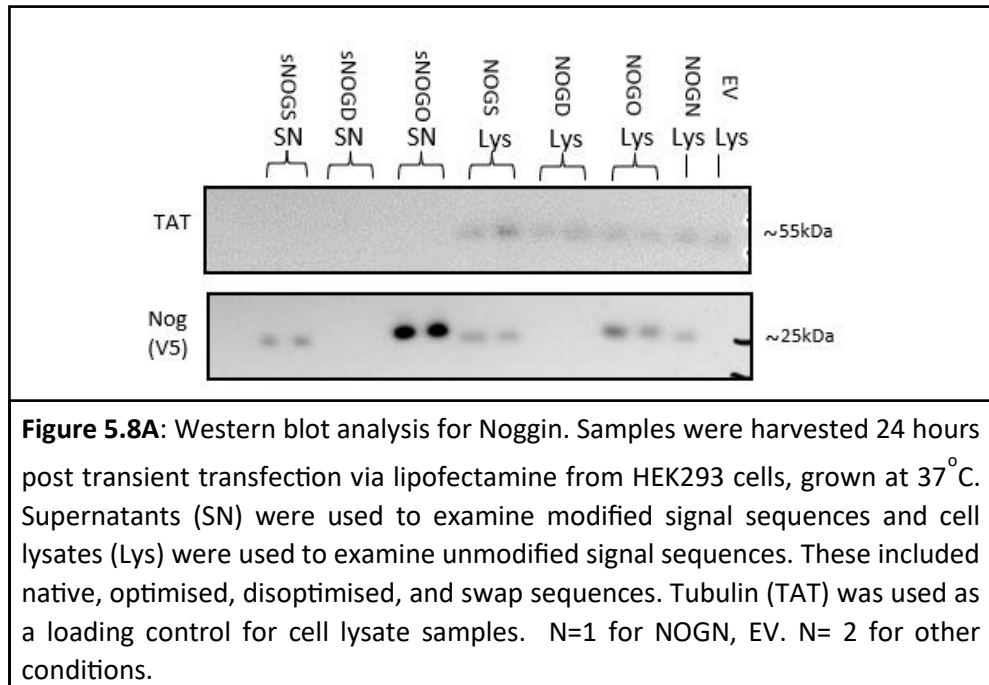


Figure 5.7: Growth profiles of HEK293 cell pools grown at 37°C stably integrated with Native, Optimised, and Disoptimised eGFP vectors. Cell Index values were generated using xCELLigence system and protocol (N=4). Top, Secretory eGFP modifications. Bottom, eGFP modifications.

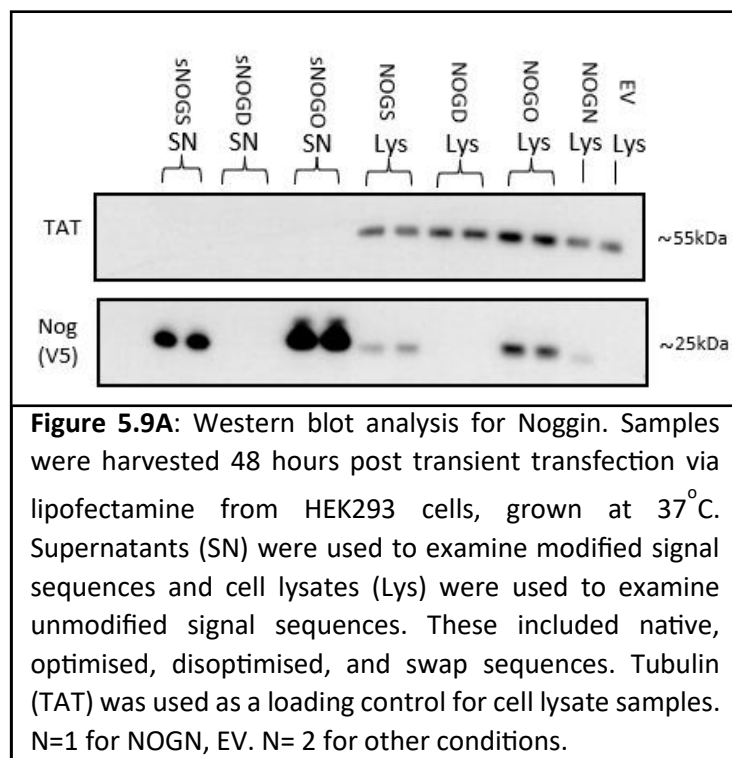
5.5 *Noggin* Expression from 5' Open Reading Frame Codon Transcript Variants Encoding for the Same Protein

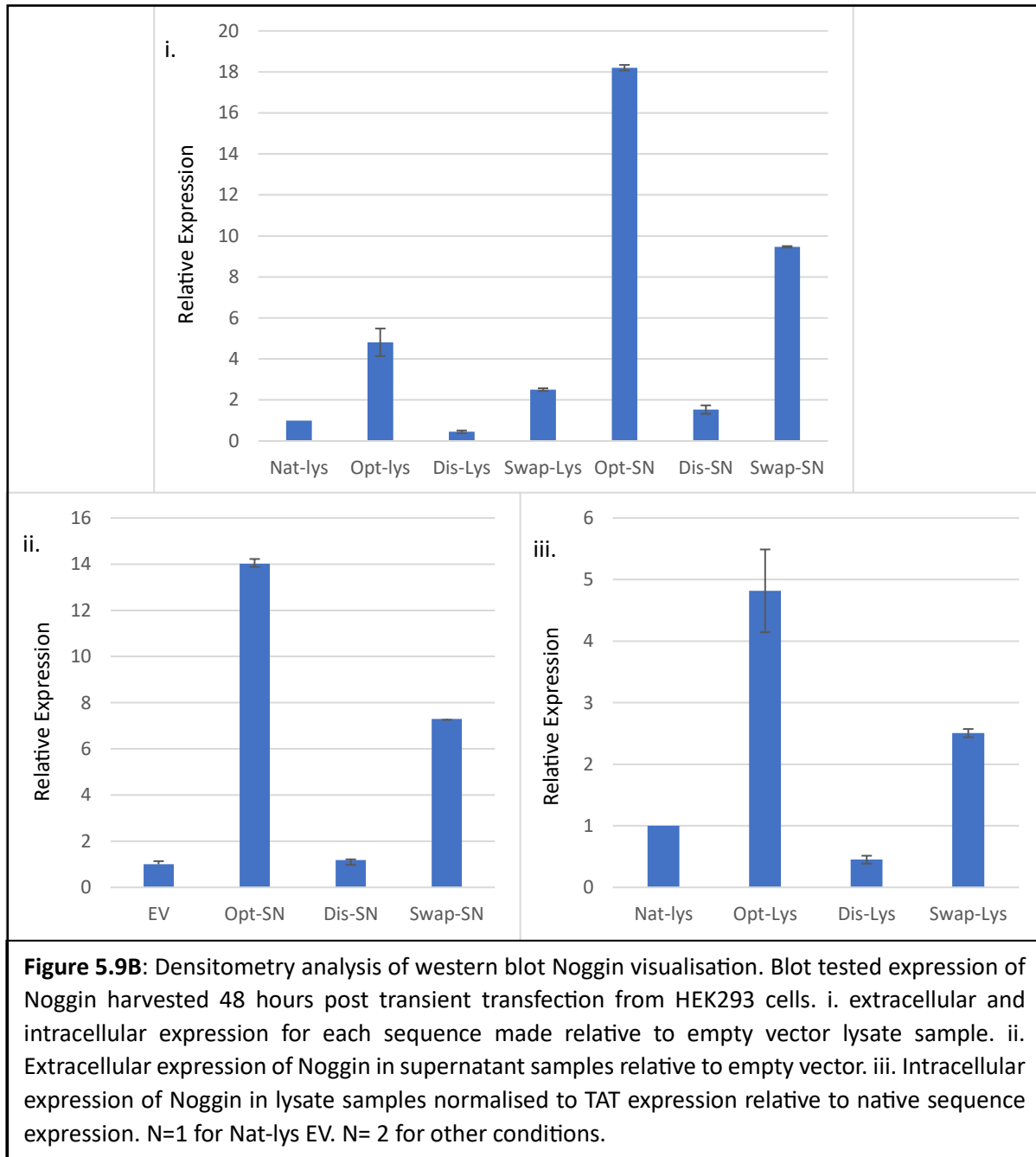
5.5.1 Comparison of *Noggin* Protein Production from Different Transcript Variants in HEK293 Cells Cultured at 37°C or 32°C, 24 and 48 hour Post Transient Transfection

As for the GFP samples described earlier in this chapter, western blot evaluation was used to assess *Noggin* protein production (NOG) of samples harvested from HEK293 cultures transfected with the different *Noggin* transcript variants and then incubated at 37°C for 24 hours post transfection. An empty vector transfection sample showed no NOG-V5 signal (or any other signal) and therefore any V5 signal, at the size expected for the NOG protein, relates solely to recombinant *NOG* expression. Both of the disoptimised *Noggin* transcript variant sequences, that with the modified signal sequence and removed signal sequence showed no evidence of exogenous *NOG* expression (Figure 5.8A) suggesting that disoptimisation through codon manipulation of the 5' end of the ORF was detrimental to translation elongation and protein synthesis to the extent that this prevented any observable protein synthesis. This was seen throughout all NOG western blots. The correct sequence was confirmed and to be in frame and thus the observation can be attributed to the codon usage impacting protein production. Reduced mRNA stability due to changes in the transcript or rapid degradation of the protein due to faulty folding are possible explanations examined further in the discussion chapter of the thesis. In lysate samples, the Native transcript sequence resulted in observable expression, but both the Optimised and Swap transcript variant sequences gave increased protein production visually and this was confirmed by densitometry analysis (Figure 5.8B). There was no statistically significant difference between the Optimised and Swap transcript variant samples in the intracellular lysate material. However, when investigating codon modified signal sequence supernatant samples, the Optimised transcript variant clearly displayed the highest secretory expression and achieved a statistically significant increase in protein production compared to Swap samples (Figure 5.8A).

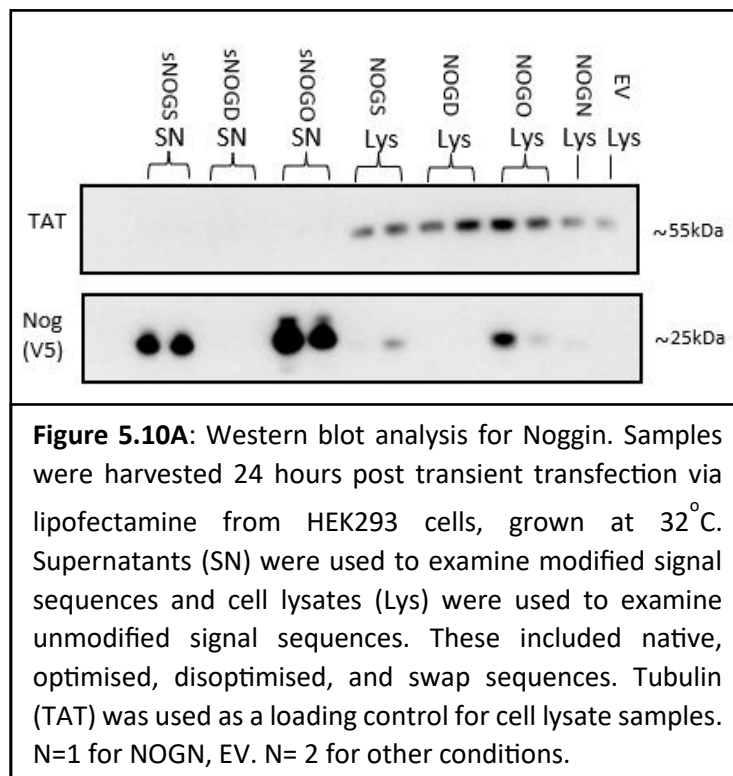


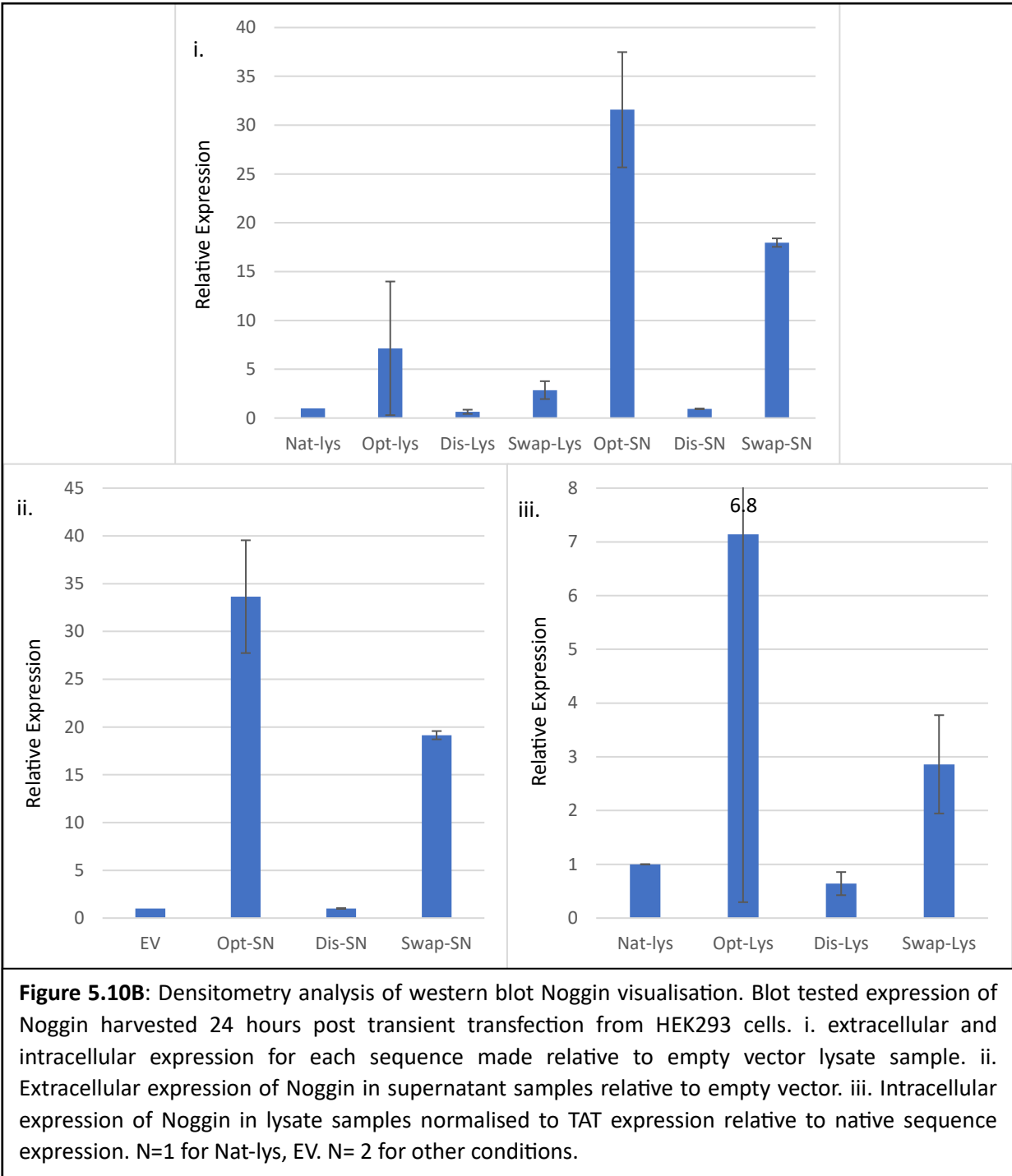
When samples cultured at 37°C for 48 h were evaluated for NOG protein production there was once again no band present for those samples transfected with Disoptimised transcript variants (Figure 5.9A). All other samples showed the presence of more intense NOG bands compared to figure 5.8A under the same conditions (equal exposure time and protein loaded). This suggests maintained expression over the 48 h period from the modified sequences which was not observed previously from some *CCND3* or GFP transcript variants (examples include Figure 4.18A and Figure 5.2A). As for the 24 h sample, the Optimised transcript variant continued to yield the most intense band for both the transcript containing and lacking the modified signal sequence, and the increase in expression was statistically significant compared to other transcript variant samples as determined by densitometry analysis. The Swap transcript variant also produced a higher protein production as determined by the band intensity compared to the Native transcript variant in lysate samples. Densitometry analysis (Figure 5.8B and 5.9B) provide further evidence for consistent expression over time as expression for optimised and Swap supernatant samples maintained an expression ratio of 2:1 (6:3 at 24 hours and 18:9 at 48 hours). The Optimised transcript variant supernatant samples had a faint additional band present above the mature NOG band size which may be due to protein where the ER signal sequence was not processed/cleaved.



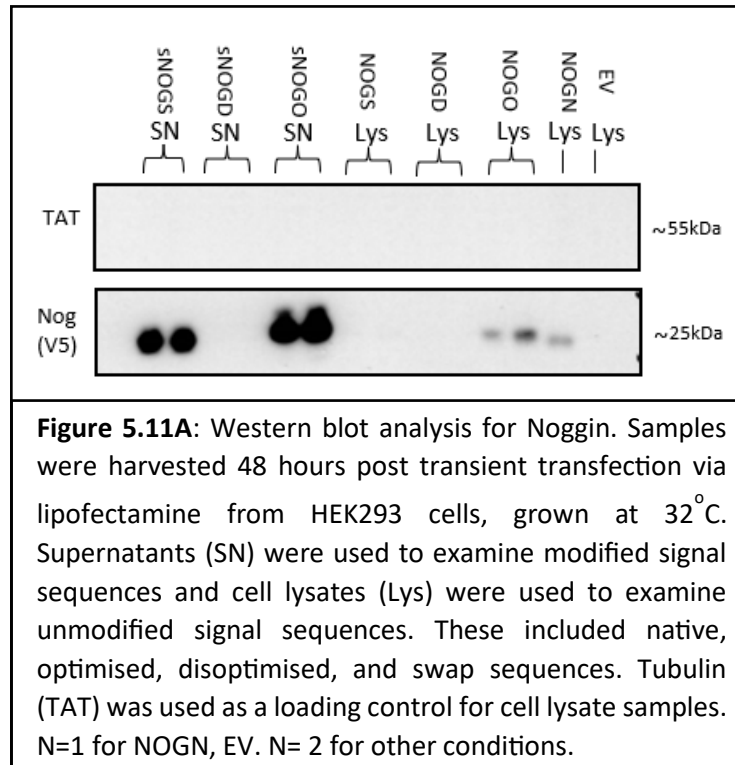


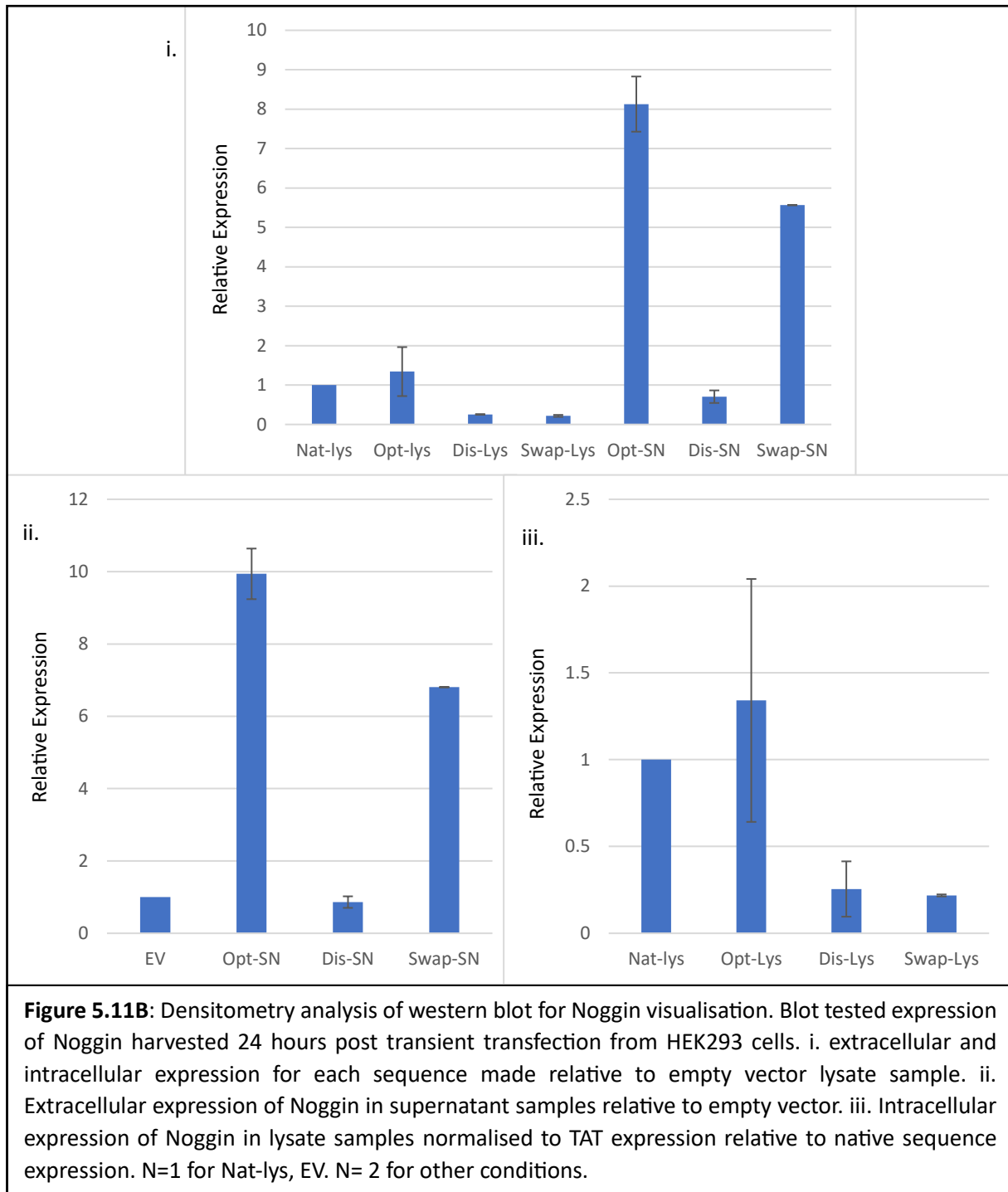
When western blot was used to evaluate *NOG* expression of samples harvested from HEK293 cultures incubated at 32°C 24 hours post transfection the disoptimised sequence samples gave no signal (Figure 5.10A) as expected in line with the results at 37°C. The Native transcript derived *NOG* protein band intensity was weak but visible and the disparity between the protein production from the Native transcript and the Optimised transcript variant in lysate samples was drastic for one replicate but much more modest for the other. There was much greater variation between *NOG-O* expression signals than the tubulin (TAT) signals would suggest for replicate samples. The swap (*NOG-S*) lysate replicates also showed some variation in expression at 32°C but generally appeared to result in less expression than the optimised (*NOG-O*) replicates. These fluctuations were much less apparent in s*NOG* supernatant samples with Swap transcript variant derived protein production being much higher than that from the Native transcript whilst the expression was even higher in the Optimised transcript variant samples (Figure 5.10A). Densitometry analysis (Figure 5.10B) confirmed the same expression trends from the different transcript variants was observed at 32°C as observed in samples from cultures grown at 37°C in that the Optimised transcript variants (both modified signal sequence and no signal sequence samples) gave higher statistically significant *NOG* protein production than other Sequences.





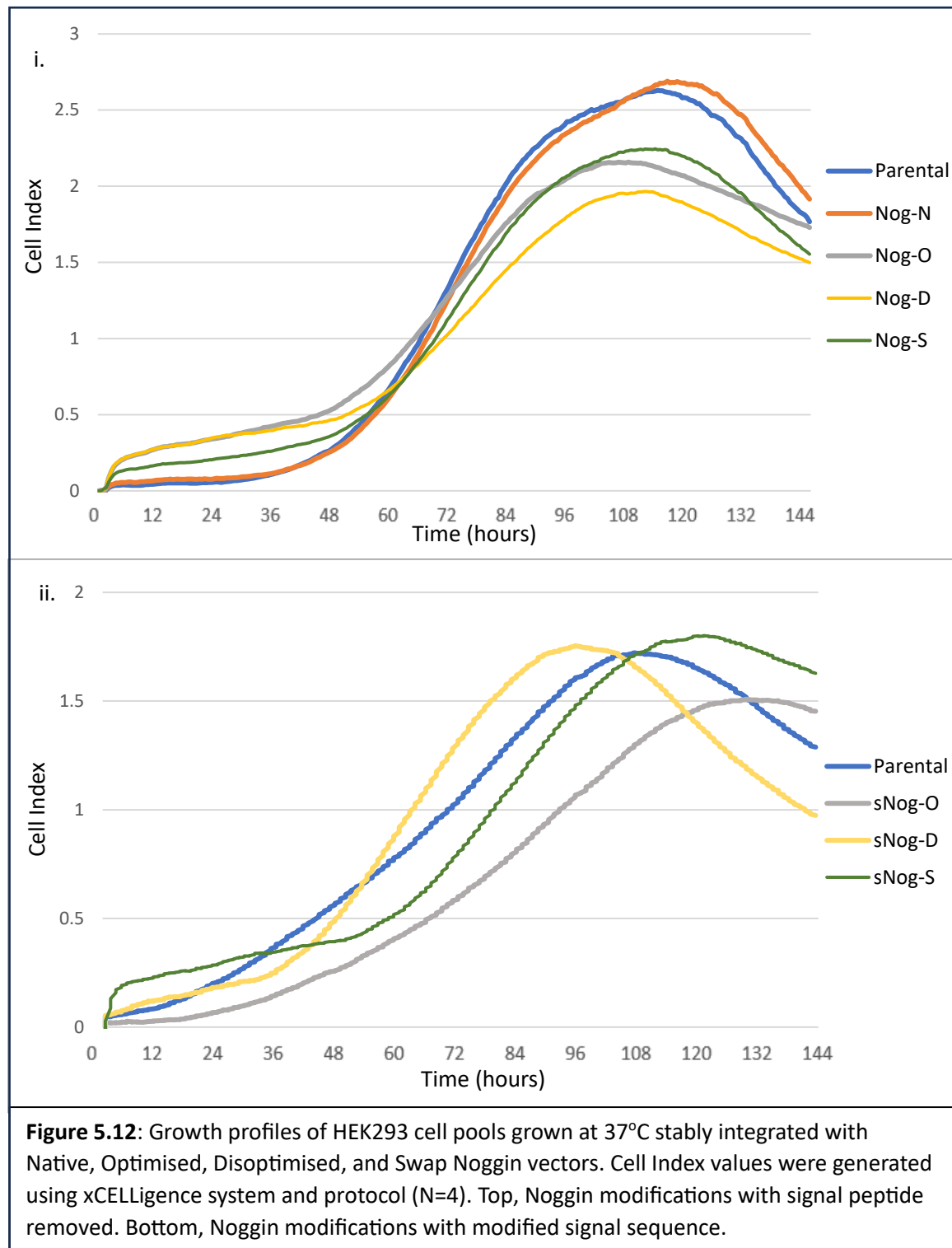
Western blot evaluation of NOG protein production of samples harvested from HEK293 cultures incubated at 32°C for 48 h hours post transfection (Figure 5.11A) showed a similar trend to that of the 24 h samples. It is noted that in this blot tubulin band signals were absent in lysate samples. Rather than genuine absence of protein being the cause, this result is most likely due to antibody depletion as this antibody had been continually reused and no new antibody was available. For the intracellular samples there was no evidence of *NOG-S* expression. Again, when secretory NOG was investigated the codon optimise variant *SNOG-O* showed the highest production in supernatant samples and the elevated amount present was statistically significant in amount compared to the Swap transcript variant derived protein amount (Figure 5.11).





5.5.2 Assessment of the Impact of *Noggin* Transcript Variant Expression on HEK293 Growth using the xCELLigence Instrument

As outlined for the GFP experiment (see section 5.4) HEK293 cells were stably transfected with the *Noggin* transcript constructs along with an empty vector as a control after FspI restriction enzyme digestion to linearise the plasmid. After the cell pool generation process, the xCELLigence system was used to monitor the growth profiles with readings taken automatically every 30 minutes. Exogenous intracellular *Noggin* production (no signal peptide) did have some impact on the subsequent growth profiles (Figure 5.12) although none of these were dramatic. At 37°C the native pool grew very similar to the parental cell line and obtained a similar maximum cell index value whilst the modified transcript variants containing cell pools all obtained a lower cell index. When the secretory *NOG* cell pools were investigated (containing an ER signal sequence), the cell pool expressing the optimised transcript variant grew the slowest and obtained the lowest maximum cell index suggesting that the exogenous expression and protein secretion of the optimised transcript did have some impact on the growth profile of this pool. This may reflect the fact that this construct and transcript variant was the most highly expressed of all the variants and this potentially imposes the greatest load on the HEK293 cell. This is discussed and explored more in the discussion chapter of this thesis (Chapter 6).



Chapter 6

Discussion

The work described in this thesis set out to investigate the impact of manipulation of codon usage to generate mRNA transcript variants that encoded for the same amino acid sequence. Unlike other work investigating codon usage and the impact on protein synthesis, the work described here focussed upon manipulation of target sequences within transcript whereby in the native transcript the target area was identified as containing a stretch of codons that were decoded slowly and thus the polypeptide elongation speed across such stretches of codons is predicted to be slow (Chu et al., 2012). The target transcripts included one where the predicted slow decoded section was towards the middle of the open reading frame (*CCND3*) and one where the predicted slowly decoded sequence was at the start of the sequence (*NOG*). Interestingly, for the *NOG* sequence this was further complicated by the fact the protein this encodes for is targeted into the ER and thus has an ER signal sequence at the 5' end of the ORF. As described in Chapter 5, translation of transcripts that contain an ER signal sequence is attenuated once the signal sequence polypeptide emerges from the ribosome upon binding of the signal recognition particle (SRP) and is only resumed once the polypeptide is docked in the Sec61 translocon in the ER and the SRP is released so that ribosome movement and elongation can resume (Linxweiler et al., 2017; Mason et al., 2000; von Heijne, 1985). The 3rd model transcript was the reporter gene *eGFP*. The codon manipulation was shown to be able to impact protein synthesis in a sequence and position determined manner. These findings are discussed in more detail below.

6.1. Comparing codon modifications and transcript variants of all transcripts tested and the impact on protein synthesis

When transcript codon modification was undertaken on the model transcripts investigated in this study, the impact on subsequent protein production was monitored and confirmed that a large impact on protein production could be achieved by altering specific regions of mRNA transcripts rather than entire sequence. The modifications were undertaken based on an elongation model and tRNA abundance previously published (Chu et al., 2012) to generate transcript for the same amino acid sequence but with vastly different predicted elongation speeds across the specific region of interest. The impact on protein

synthesis, as determined using western blotting, was clearest when modifying codon usage immediately proceeding the AUG start codon of previously unoptimized, 'native' (endogenous) mRNA transcripts such as observed when manipulating the *Noggin* transcript sequence. This data agrees with the hypothesis that the relationship between mRNA translation initiation control and elongation control are more closely linked than previously thought, whereby the initial elongation speed or rate can impact the rate of initiation. In support of this hypothesis, Tuller et al. identified that the first 30-50 codons of many transcripts are translated with low efficiency (enriched with rare codons) as an evolutionarily conserved mechanism to reduce the number of ribosomes able to simultaneously translate transcripts (Tuller et al., 2010). This mechanism of control helps mitigate ribosome collisions downstream during elongation, increasing overall translation rates and expression (Tuller et al., 2010). Further, whilst Sejour et al. further confirmed the presence of a slowly translated region towards the start (5' end of the ORF) of many transcripts, they also found contrary evidence that genes with rapidly translated 5' ends led to increased protein production compared to slow 5' end elongation speed translated transcripts (Sejour et al., 2022). The results presented in this chapter, specifically on the optimised *Noggin* sequences further support these conclusions where these resulted in greatly increased protein production compared to both native and disoptimised codon variant transcript sequences by only modifying 5' end codons.

It therefore seems, at least in the case of the *Noggin* 5' modified transcript variant, that changing slowly decoded codons as predicted by their tRNA abundance at the 5' ORF of a sequence for the fastest decoded codons based on the elongation model used in this study results in enhanced protein synthesis. If this results in enhanced initiation too, and more ribosomes translating a transcript, future work could investigate if this results in a greater number of ribosomes being loaded onto the target transcript. This could be monitored using polysome profiling or RiboSeq analysis in order to determine whether or not there are more ribosomes on such modified transcripts. If this was the case, a greater number of ribosomes on a transcript may also protect the transcript from degradation, impacting the half-life of the transcript and extending the 'productive half-life' of the transcript whereby it may be available for translation and protein synthesis longer than for transcripts with fewer ribosomes associated with it.

When the *eGFP* transcript was codon modified this did show some changes in the subsequent protein production as determined by western blot, however the impact on expression is likely lessened when

compared to that observed when modifying the *Noggin* transcript due to the *GFP* sequence already being highly optimised for mammalian cell line expression (Guohong, 1996). The *CCND3* study, whereby the transcript sequence was modified away from the 5' ORF and the region of interest was located towards the middle of the ORF, highlights the complexities involved in elongational control. Further variations were also seen in the response between the G variant and T *CCND3* variant sequences which consist of a single nucleotide change. Nevertheless, the work on the *CCND3* transcript also showed that whilst alterations to mRNA codons immediately following the AUG start codon show the most drastic changes in subsequent protein production, alterations to regions identified further downstream as slowly elongated can also impact subsequent protein production rates. However, the relationship between codon usage, predicted elongation speed and protein synthesis for the region of interest in the *CCND3* transcript was not as clear as that for the 5' codon modified *Noggin* transcript.

Regardless of the transcript investigated, optimising or manipulation of codon usage in ROIs based on the elongation model rationale in all three model transcripts resulted in enhanced protein production in multiple instances. This confirms that the elongation model (Chu et al., 2012) can successfully identify regions predicted as slowly elongated during translation and where alterations in codon usage to codons that are decoded faster can impact subsequent protein production rates. Further, greater improvements on the observed protein production levels from transcript variants compared to the native or endogenous transcript were observed when investigated under mildly hypothermic culture conditions (32°C). Fast sequences, in addition to the 'optimised' sequences also in most cases resulted in increased protein production, although often not to the same degree as that observed from the optimised sequences.

The main exception to enhanced protein production from the optimised transcript sequences was for the non-secreted *eGFP* transcript which has already been highly optimised since its conception and wide use in mammalian cell line studies (Guohong, 1996). As outlined above, protein production from the *GFP* transcript may therefore not be as sensitive to codon alterations made in this study as other transcripts. Interestingly, the secretory form of GFP did support increased protein production for cells transfected with optimised transcripts, most apparent in cultures incubated at 32°C. This trend was again observed in *Noggin* transcript variant studies. sGFP protein production was not consistently observed until 48 h post-transfection among replicates incubated at 37°C (Figure 5.3A) which was not the case for cultures

incubated at 32°C (Figure 5.4A). The mechanisms behind transcript codon alterations impacting production of proteins destined for the ER and secretion compared to those which are not, is unclear but the potential impact can be predicted or hypothesised based upon the mechanism by which such polypeptides are directed to the ER.

With regard specifically to those transcripts that encode for ER and secretory pathway destined polypeptides, after translation initiation is instigated by ribosome subunits assembling around the AUG start codon in the cytosol, elongation commences beginning with the 5' signal sequence. The subsequent translated N-terminus signal peptide is bound by the signal recognition particle (SRP) once it emerges from the ribosome exit tunnel, temporarily attenuating translation whilst the mRNA-ribosome-polypeptide complex is transported to the endoplasmic reticulum (ER). SRP receptors on the surface of the ER bind SRP allowing for ribosome-translocon tethering, and once SRP dissociates from the signal peptide, elongation can resume (Tajima et al., 1986; von Heijne, 1985). The synthesised polypeptide passes through the translocon channel into the ER lumen as translation and elongation occurs (co-translational translocation) for folding and further processing for delivery to other organelles or secretion out of the cell (Barlowe & Miller, 2013; Walter & Blobel, 1981). Therefore, increasing the rate of signal peptide synthesis or elongation may result in more mRNA-ribosome-polypeptide complexes reaching the ER and thusly increasing protein production. Also, swift translocation progression of bound ER ribosomes along mRNA sequence may allow for non-translating ribosomes still associated with the ER to begin initiation afresh (Potter & Nicchitta, 2000). Due to their bound nature as oppose to cytosolic ribosomes, the mRNA transcript would require enough of its sequence length to be decoded by the first instance of elongation before enough of the 3' end of the transcript is available for another ribosome to begin initiation. Further, if a transcript is anchored to the ER via a translating ribosome, for another ribosome to be translating the same transcript and its polypeptide to be anchored into a translocon, sufficient transcript length must be available to 'bridge' between the two anchored ribosomes as the transcript is 'fed' through the ribosomes. Finally, increasing the elongation speed of an ER signal peptide may not always be appropriate. It might be hypothesised that if the ER signal sequence emerges too fast from the ribosome, an SRP may not be able to recognise and bind the sequence and attenuate translation rapidly enough and hence translation might continue and the resulting polypeptide remain cytoplasmic in nature. This is clearly not desirable for a protein that should be ER targeted and therefore it is likely that for a secretory protein, the elongation speed of the ER signal peptide might need to be balanced between optimised and as fast as possible and

that which ensures all ER signal peptides emerging from ribosomes are SRP bound. Further work would be required to determine if this is the case. Regardless, the modified transcript sequences investigated here for secretory polypeptides were able to alter secretory protein production. Disoptimised transcript sequences for *sGFP* lead to decreased secretory protein production compared to that from the native sequence and these mechanisms described above could also explain this outcome.

The disoptimised transcript variant sequences displayed poor protein production compared to native in the majority of studies. For the *CCND3* transcript variant study this was generally the case although there were some exceptions as discussed below. As expected, low expression was observed from HEK293 cells transfected with disoptimised *Noggin* and *GFP* transcript sequences. Global reduction in translation caused by mild hypothermic conditions (cold shock response) is thought to be less impactful on codons with slower decoding speeds (Bastide et al., 2017). As such, it was hypothesised that differences between expression of native and disoptimised sequences would be more favourable for disoptimised sequences in cultures incubated at 32°C. There were instances for *CCND3* transcripts where this can be observed; GVK *CCND3* disoptimised transcripts did show more comparable protein production compared to native transcript sequences when transfected cells were cultured at 32°C rather than 37°C incubation temperatures (Figures 4.13, 4.15). However, other *CCND3* transcript variant studies did not display this characteristic. This was also not the case for *Noggin* or *GFP* disoptimised transcript variants. When the disoptimised *GFP* transcript sequence was transfected into HEK293 cells and incubated at 32°C there was comparable protein production between the Disoptimised and Native transcript samples at 24 h, but this was not maintained at 48 h (Figures 5.2B, 5.4B). Temperature had little effect on the trends observed for *sGFP*, and Disoptimised *Noggin* sequences lacked detectable protein production signals whatsoever, showing that disoptimisation of the 5' ORF of the *Noggin* transcript was extremely detrimental to protein synthesis.

The Swap codon transcript variants also improved protein production compared to that from the native endogenous transcript in multiple cases for both the *CCND3* and *Noggin* transcripts. It was hypothesised that this modification could show that specific placement of slow or fast codons throughout transcript are essential for correct ribosomal translocation and assist elongation control. Thus, following on from this, it was thought that by altering these codons only, this might destabilise ribosomal translocation leading to

a change in protein production from these transcripts. Swapping codons with fast decoding speeds in the native transcript for slow codons in the variant transcript could result in more ribosome collisions over the region of interest, 'bottlenecking' ribosomes and effectively stalling translation. On-the-other-hand, swapping codons with slow decoding speeds in the native transcript for fast codons in the variant transcript could result in less stalling around the region of interest, however this may also result in collisions downstream if the region was important in regulating translocation rates. Results described in this thesis suggest that these regions are important for limiting/controlling protein production and elongation speeds, and modifying the codons can lead to higher protein production despite slower codons being used in areas previously occupied by fast codons. The location of fast/slow codons therefore appears not as simple as merely their presence within these regions of interest in the transcript, in fact, swapping specific codons resulted in greater protein production in multiple instances.

A further variation was to generate so termed 'Middling' transcript variant sequences, which for *CCND3* consistently displayed poor protein production, often resulting in the weakest western blot signal from any variant which were further confirmed by densitometry analysis. HEK293 cell cultures transfected with Middling *CCND3* sequences showed little temperature dependent response. This adds to confidence that fast/slow codons are more heavily impacted by temperature than average codons (Bastide et al., 2017). If based purely on tRNA abundance, results might be expected to show consistently greater protein production from Middling transcript variant sequences compared to Disoptimised transcript variant sequences, at least for samples from cultures grown at 37°C. This not being the case suggests other factors also impact ribosomal translocation management and protein synthesis rates.

As described previously, some transcript sequences failed to result in detectable signs of model protein production consistently, despite extensive checks being made to confirm the correct transcript sequence in ORF and thus ruling this out as a potential reason for lack of expression. The consistent lack of NOG-D and sNOG-D protein production from the respective transcripts, despite successful sequence and open reading frame checks, suggests 3 possible causes relating to transfection, mRNA stability, and translation/protein functionality. The former could be caused by unsuccessful transfection failing to introduce template linearised plasmid into cells and is unlikely. No additional *Fspn1* restriction sites were

introduced by incorporating modified sequences and the same protocol was followed as with successful transfections using other transcript bearing plasmids.

mRNA stability issues could be culpable; mRNA stability is heavily influenced by nucleotide sequence which were directly altered (Presnyak et al., 2015; Wu et al., 2019). *NOG-D*, *sNOG-D* and *TvM* modified sequences all resulted in no expression after transfection and have fewer secondary structures in mRNA structure predictions. Predicted mRNA structures for both *Noggin* disoptimised sequences in particular are highly comparable to one another and vastly different to those predicted for the native sequence. Thus, a lack of stabilising structures may have rendered these transcripts highly unstable reducing their persistence within cells and thereby preventing their translation and production of protein at a detectable amount at a similar intensity to other modified sequences.

Secondary mRNA structures, if not impacting mRNA transcript half-life stability, could still impact translation efficiency. Secondary structures can stall ribosomal progression along a transcript, in some instances indefinitely (high amount hydrogen bonding) if not resolved by cell responses such as programmed ribosomal frameshifting, as seen in Gag-Pol fusion by HIV (Jacks et al., 1988; Schuller & Green, 2018; Wilson et al., 1988). This can render mRNA unreadable, leading to premature termination of translation and ribosome rescue. This would lead to nonfunctional truncated protein production, uncomplimentary to the V5 antibody used in western blotting, and destined for degradation within the cell via the proteasome (Dikic, 2017). As mentioned above, the predicted sequences for these transcripts display a reduced number of secondary structures, but specific regions could still impact the translocation step of elongation in this manner. However, other modified sequences displayed greater variation in mRNA structure compared to native transcripts and contained far more, albeit differently located, secondary structures. It is surprising that these transcripts such as *sNOG-S* did show increased expression if secondary structures did negatively impact translocation of ribosomes this greatly.

The final explanation is the abundance of the tRNA for the particular codons used. It is possible that these are at such low concentrations, particularly of charged tRNA, that the elongation step is slowed to such an extent it is almost halted. If the correct tRNA cannot be sampled in the ribosome site, or incorrect tRNAs

continually sample and need to be rejected, this could more-or-less attenuate translation and thus limit protein synthesis from these transcripts.

6.2. The Potential Implications and Applications of Region-Specific Codon Optimisation

As stated in section 6.1, region of interest transcript variant optimised sequences for all model transcripts resulted improved protein production and compared to native at 32°C this was enhanced for optimised sequences. The effect of mildly hypothermic culturing conditions for bioreactor cultures of mammalian cells producing secreted biotherapeutic proteins are well documented. Mild hypothermic culturing can provide multiple benefits including maintaining high culture viability over prolonged incubation times leading to increases in product yields and quality for some products by reducing intramolecular aggregation (Ahn et al., 2008; Goey et al., 2017; Kaufmann et al., 1999; Tharmalingam et al., 2008). The codon optimisation approach investigated in this thesis is highly relevant to this industry as a result. Lower culture temperatures lead to reduced global protein synthesis, but not necessarily for transcripts optimised in this manner. Although industry does use codon optimisation strategies for transcripts for recombinant biotherapeutic proteins, this tends to be optimisation of complete transcripts based upon codon usage in the genome rather than targeting specific sites predicted to decode slowly and using tRNA abundancies to codon optimise. Using this region-specific approach and tRNA abundancies linked to the elongation model used in these studies could result in greater increases in recombinant protein production. This either reduces production costs or increases usable product, both desirable.

Alongside potential applications in biotherapeutic protein production from cultured mammalian cells, elongation control design of transcript may have important applications in health areas too. For example, Noggin is highly relevant to neurodegeneration, being shown to rescue stem cell loss typical in Alzheimer's disease (Díaz-Moreno et al., 2018). Thus, another application for modifying codons in this manner is gene therapy and RNA vaccine use. In gene therapy, functional genes are introduced *in vivo* aiming to treat a wide array of conditions including cancers, cystic fibrosis, aids and heart disease to name a few (Walsh & Walsh, 2022). Such an approach could be used to look to express appropriate amounts of Noggin in patients with early neurodegenerative disease. Gene therapy delivery is often achieved via use of a viral vector such as adeno-associated virus (AAV) with an AAV genome containing the desired gene for expression. The predominant method for regulating gene and hence protein production from such vectors

is through using different promoters upstream of the gene of interest. This is a rather blunt instrument and usually involves using either tissue specific promoters or promoters of different strength to try and tune the transcriptional activity, and hence amount of transcript available for translation and thus protein synthesis. The codon optimisation method deployed in this study could provide an additional axis to modulate protein production, either up or down, depending upon the desired amount of protein to be produced, resulting in more effective therapy. Further, the design could be tissue or cell specific assuming that the tRNA abundancies of the target cell or tissue are known. As de/increases in protein production were achieved through codon manipulation in cultures grown at 37°C, it is likely that this ability to tune expression would also occur *in vivo*. Tuning expression for specific tissue types via promoter selection has already been proffered (Nieuwenhuis et al., 2021; Zheng & Baum, 2008), which could also be achieved via codon usage if this was done alongside a tRNA abundance study as mentioned above and below.

The same approach could also be applied to the growing field of RNA vaccines. There are multiple benefits to mRNA mediated vaccination over conventional viral methods including the rapid uptake and expression of this method of vaccination, as well as high yields in *in vitro* reactions of mRNAs at bench scale easing the burden on production and scalability (Pardi et al., 2018). Potency could be potentially be increased further however, by implementing optimisation for elongation in a tissue or cell specific manner, as outlined in this study.

6.3. Further Supplementary Studies to Potentially Improve the Ability to Design Transcript Variant Sequences to Tune Protein Production to Desired Amounts

The work in this thesis has demonstrated how region-specific codon manipulation of a transcript can be used to tune subsequent protein production up or down. To further aid the understanding of this work, RNA sequencing and RiboSeq approaches would provide valuable insight into the stability of modified transcripts and whether transcripts that resulted in no detectable expression of their respective protein, such as *NOG-D*, were caused because no transcript was present or mRNA instability issues. The half-life of transcripts could also be determined of the different transcripts using methods previously described (Presnyak et al., 2015). As described above, the differences in protein production achieved from transfected modified variant transcript sequences could be caused by codon usage and elongation control or varying stability of individual mRNA transcripts. This further study would aid in confirming changes in

protein production are predominantly caused by altering codon usage and hence elongation speed of the growing polypeptide and not due to differences in mRNA stability.

Altering the abundance of specific tRNAs within cells and reassessing the effect sequence modifications have on subsequent protein production could also be investigated to see if this results in highly tailored protein production. As previously discussed, codons with slower decoding times typically correspond to a rarer tRNA within the tRNA pool of a cell. This increases the stochastic process of a cognate tRNA entering the A site during elongation. If the population for a less abundant tRNA were to be increased, this may lead to improved protein production from disoptimised sequences. Not only would this supplement knowledge gained during the studies undertaken in this thesis, but if used in conjunction with mildly hypothermic conditions in an industrial setting, production of recombinant proteins may also be improved whilst limiting other cellular functions. Aminoacyl-tRNA levels would need to also be confirmed to ensure overexpression of less abundant tRNAs did not result in only increasing the amount of uncharged tRNA (lacking amino acid) which would have the opposite effect. Further, manipulation of tRNA levels may change the expression of endogenous transcripts in the cell which may have unforeseen consequences. Some such studies were initiated during this PhD focussing on the model transcript, RTN3 which has a transcript ROI predicted to be slowly decoded immediately following the AUG start codon. The codons in this ROI sequence were disoptimised to correspond to rarer tRNAs, and thus tRNA sequences were identified and designed encoding the cognate tRNAs for these rarer codons for overexpression (Figure 6.1, Table 6.1). This would provide evidence that overexpression of less abundant tRNAs does increase subsequent protein production from codon elongation disoptimised sequences. These studies are currently ongoing at the University of Kent.

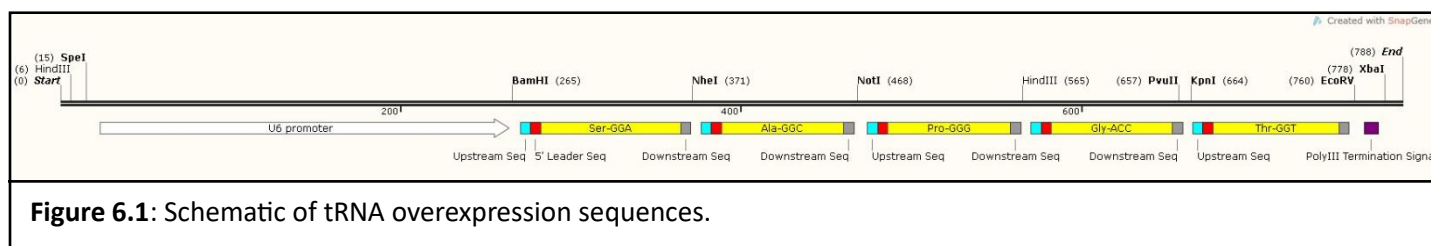


Table 6.1: First 40 codons, their respective amino acids, and predicted decoding times (s) for the disoptimised RTN3 sequence.

Amino Acid	Slowest Codon Sequence	Decoding Time
M	ATG	0.034593776
A	GCC	1.242491409
E	GAG	0.0957373
P	CCC	0.886127431
S	TCC	1.273908192
A	GCC	1.242491409
A	GCC	1.242491409
T	ACC	0.737232591
Q	CAA	0.04093071
S	TCC	1.273908192
H	CAT	0.834141556
S	TCC	1.273908192
I	ATC	0.35187311
S	TCC	1.273908192
S	TCC	1.273908192
S	TCC	1.273908192
S	TCC	1.273908192
F	TTT	0.862058028
G	GGT	0.854143037
A	GCC	1.242491409
E	GAG	0.0957373
P	CCC	0.886127431
S	TCC	1.273908192
A	GCC	1.242491409
P	CCC	0.886127431
G	GGT	0.854143037
G	GGT	0.854143037
G	GGT	0.854143037
G	GGT	0.854143037
S	TCC	1.273908192
P	CCC	0.886127431
G	GGT	0.854143037
A	GCC	1.242491409
C	TGT	0.843332863
P	CCC	0.886127431
A	GCC	1.242491409
L	CTC	1.089311091
G	GGG	0.235110809
T	ACC	0.737232591
K	AAA	0.036161963

Investigating other model transcripts relevant to the industrial setting would also improve confidence in the design of sequences based on predicted decoding times for enhanced protein synthesis. Furthermore, modifying the entire mRNA sequence as well as highlighted ROI areas relevant to elongation control could also help determine whether this method of optimisation is most effective by identifying specific areas of elongation control, or is transferable to entire transcript length.

It is noted here that tubulin protein visualisation was poor for multiple western blots reported within this thesis. This arose due to repeated use, and eventual depletion of, the anti-tubulin antibody. Issues caused by scarcity of reagents brought on by the Covid-19 pandemic and subsequent time constraints prevented further antibody from being generated for use. All samples used for western blots contained equal total protein based on Bradford assays. This is shown to be reliable for samples where antibody is not diminished and typically consistent tubulin signals between samples were observed. This gives some

confidence that the observed results where a tubulin protein band visualisation was poor are still valid in terms of the patterns of protein production observed for the protein of interest. However, in the future this should be repeated to further confirm the results. Switching antibody for an altogether different loading control would also overcome this issue, such as β -actin. Moreover, the inclusion of either Coomassie staining or Ponceau staining would have provided additional confidence in equal protein loading between samples and greatly mitigated the impact caused by poor anti-tubulin antibody reads. These should also be carried out alongside additional western blots in future experiments.

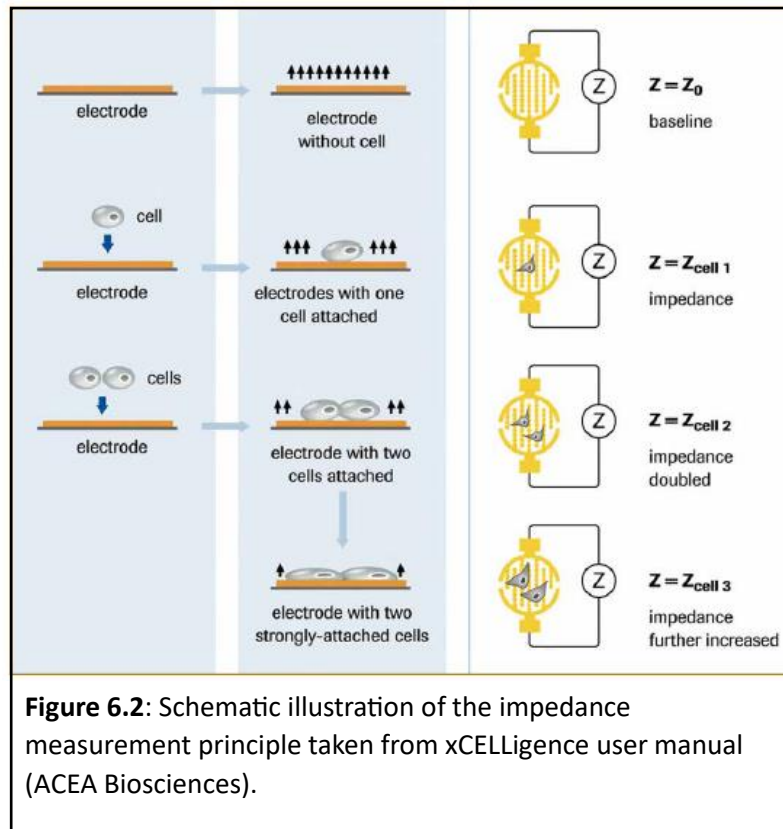
6.4. Limitations of this Study

There are a number of limitations to the work produced, that do not detract from the work, but in retrospect could have been, or should in the future, be addressed. The inclusion of a Native *Noggin* transcript with an ER signal sequence would have been useful to include for better comparison of secreted samples. This does not invalidate the secretory *Noggin* results which shows that modified signal sequence transcripts displayed expression profiles similar to those observed for non-secreted (lysate) samples and that the codon manipulation of this sequence impacted the observed protein production. It would be hypothesised that Optimised and Swap transcript variant derived secreted samples would have shown improved expression compared to native *Noggin* sequence expression.

It is surprising that increased cell proliferation did not result from increases in *CCND3* expression. It is likely that this was caused by the burden of *CCND3* expression actually reducing cell growth or that the selection process inadvertently selected within each population for the lowest producing cohorts that have a growth advantage. However, other explanations could be possible. It is possible that exogenously produced *CCND3* was non-functional, perhaps due to misfolding caused by codon alterations impacting translation rates or the presence of the V5 tag at the C-terminal. However, if this was the case, HEK293 cells transfected with native *CCND3* sequence would be expected to show increased proliferation as no modifications were made between it and the endogenous *CCND3* sequence. Other control mechanisms may also be active in the cell preventing rapid G1 to S phase transition whereby overexpression of *CCND3* alone was inadequate to enhance this transition in HEK293 cells.

Furthermore, with regard to the growth kinetics, the xCELLigence system used to monitor growth measures electrode impedance of sensors across the base of each well and adhered cells act as insulators

at the electrode/solution interface leading to increased impedance (Figure 6.2) (ACEA Biosciences, 2013). Cell index is derived from changes in impedance caused in this manner and does not distinguish between number of cells and area covered. If the *CCND3* overexpressing cells transition from G1 to S phase without marked increases in size typical of G1 phase growth, the engineered *CCND3* over-expressing cells may be achieving higher cell concentrations but covering a similar or smaller surface area.



6.5. Final Conclusions

This project aimed to assess the impact of codon usage and manipulation in specific regions of target transcripts highlight by an elongation model formulated by Dr Tobias von der Haar as being elongation controlled, that is to say in these regions elongation is predicted to proceed slowly and hence impact protein synthesis rates. These investigations were undertaken on a number of model transcripts and several sequence variants were designed to evaluate this. The main conclusions of this project are;

1. Alterations made to a target mRNA transcript in a specific region, whilst conserving amino acid sequence, based on an elongation model can be used to tune or impact protein production rates, with codon optimised sequences (those that correspond to most abundant tRNAs) yielding consistently increased protein production whilst disoptimised sequences reduced protein production. The elongation model is therefore an effective tool in transcript sequence design although can be further developed through cell and tissue specific models as discussed elsewhere in the discussion.
2. Other transcript modification options led to a variety of impacts on protein production, highlighting the complicated nature of elongation control. Middling sequences resulted in poorest protein production from a *CCND3* model transcript, despite disoptimised sequences containing slower decoded codons. This suggests the presence of fast and slowly decoded codons can be beneficial and might assist in smooth ribosome translocation along the mRNA, leading to better protein production from these transcripts.
3. The biggest impact on protein production was for codon alterations made immediately after the AUG start codon. This supports the theory that these areas are often integral to elongation control. This could work to either increase or decrease AUG codon availability for initiation, suggesting that control of initiation and elongation are linked.
4. Some alterations prevented expression all together suggesting that changes to these areas can completely prevent translation through multiple potential mechanisms including impacting mRNA stability and creating additional secondary structures unconducive to ribosome translocation.
5. The rationale behind generating these model transcripts could be applied into the field of recombinant protein drug production coupled with mildly hypothermic bioreactor conditions to increase drug yield or reduce production cost.
6. The findings are also relevant for RNA-based vaccines to allow fine tuning of protein production from transcripts as well as for potential gene therapy applications. Insights into effects modification have on mRNA stability would be key to this application.

References

- ACEA Biosciences. (2013, January). *RTCA DP Instrument Operator's Manual*. http://www.cytometrie-imagerie-saint-antoine.org/media/4046/RTCA_DP_System_Operator_Manual_v3.pdf
- Adomavicius, T., Guaita, M., Zhou, Y., Jennings, M. D., Latif, Z., Roseman, A. M., & Pavitt, G. D. (2019). The structural basis of translational control by eIF2 phosphorylation. *Nature Communications*, 10(1), 2136. <https://doi.org/10.1038/s41467-019-10167-3>
- Agilent. (n.d.). *Agilent Primer Design*. <https://www.agilent.com/store/quickchange-primer-design-help.jsp>. Retrieved September 29, 2023, from <https://www.agilent.com/store/quickchange-primer-design-help.jsp>
- Ahn, W. S., Jeon, J.-J., Jeong, Y.-R., Lee, S. J., & Yoon, S. K. (2008). Effect of culture temperature on erythropoietin production and glycosylation in a perfusion culture of recombinant CHO cells. *Biotechnology and Bioengineering*, 101(6), 1234–1244. <https://doi.org/10.1002/bit.22006>
- Akopian, D., Shen, K., Zhang, X., & Shan, S. (2013). Signal Recognition Particle: An Essential Protein-Targeting Machine. *Annual Review of Biochemistry*, 82(1), 693–721. <https://doi.org/10.1146/annurev-biochem-072711-164732>
- Al-fageeh, M. B. (2008). *Regulation of Cold-Inducible RNA Binding Protein (CIRP) Expression*.
- Al-Fageeh, M. B., & Smales, C. M. (2006). Control and regulation of the cellular responses to cold shock: the responses in yeast and mammalian systems. *Biochemical Journal*, 397(2), 247–259. <https://doi.org/10.1042/BJ20060166>
- Al-Fageeh, M. B., & Smales, C. M. (2009). Cold-inducible RNA binding protein (CIRP) expression is modulated by alternative mRNAs. *RNA (New York, N.Y.)*, 15(6), 1164–1176. <https://doi.org/10.1261/rna.1179109>
- Amrani, N., Ganesan, R., Kervestin, S., Mangus, D. A., Ghosh, S., & Jacobson, A. (2004). A faux 3'-UTR promotes aberrant termination and triggers nonsense- mediated mRNA decay. *Nature*, 432(7013), 112–118. <https://doi.org/10.1038/nature03060>
- Barlowe, C. K., & Miller, E. A. (2013). Secretory Protein Biogenesis and Traffic in the Early Secretory Pathway. *Genetics*, 193(2), 383–410. <https://doi.org/10.1534/genetics.112.142810>

- Barnes, L. M., Bentley, C. M., & Dickson, A. J. (2003). Stability of protein production from recombinant mammalian cells. *Biotechnology and Bioengineering*, 81(6), 631–639.
<https://doi.org/10.1002/bit.10517>
- Barthelme, D., Dinkelaker, S., Albers, S.-V., Londei, P., Ermler, U., & Tampé, R. (2011). Ribosome recycling depends on a mechanistic link between the FeS cluster domain and a conformational switch of the twin-ATPase ABCE1. *Proceedings of the National Academy of Sciences*, 108(8), 3228–3233. <https://doi.org/10.1073/pnas.1015953108>
- Bastide, A., Peretti, D., Knight, J. R. P., Grosso, S., Spriggs, R. V, Pichon, X., Sbarrato, T., Roobol, A., Roobol, J., Vito, D., Bushell, M., von der Haar, T., Smales, C. M., Mallucci, G. R., & Willis, A. E. (2017). RTN3 Is a Novel Cold-Induced Protein and Mediates Neuroprotective Effects of RBM3. *Current Biology*, 27(5), 638–650. <https://doi.org/10.1016/j.cub.2017.01.047>
- Behrmann, E., Loerke, J., Budkevich, T. V., Yamamoto, K., Schmidt, A., Penczek, P. A., Vos, M. R., Bürger, J., Mielke, T., Scheerer, P., & Spahn, C. M. T. (2015). Structural Snapshots of Actively Translating Human Ribosomes. *Cell*, 161(4), 845–857. <https://doi.org/10.1016/j.cell.2015.03.052>
- Ben-Shem, A., Garreau de Loubresse, N., Melnikov, S., Jenner, L., Yusupova, G., & Yusupov, M. (2011). The Structure of the Eukaryotic Ribosome at 3.0 Å Resolution. *Science*, 334(6062), 1524–1529.
<https://doi.org/10.1126/science.1212642>
- Benne, R., & Hershey, J. W. (1978). The mechanism of action of protein synthesis initiation factors from rabbit reticulocytes. *The Journal of Biological Chemistry*, 253(9), 3078–3087.
- Beringer, M., Bruell, C., Xiong, L., Pfister, P., Bieling, P., Katunin, V. I., Mankin, A. S., Böttger, E. C., & Rodnina, M. V. (2005). Essential Mechanisms in the Catalysis of Peptide Bond Formation on the Ribosome. *Journal of Biological Chemistry*, 280(43), 36065–36072.
<https://doi.org/10.1074/jbc.M507961200>
- Beringer, M., & Rodnina, M. V. (2007). The Ribosomal Peptidyl Transferase. *Molecular Cell*, 26(3), 311–321. <https://doi.org/10.1016/j.molcel.2007.03.015>
- Beznosková, P., Cuchalová, L., Wagner, S., Shoemaker, C. J., Gunišová, S., von der Haar, T., & Valášek, L. S. (2013). Translation Initiation Factors eIF3 and HCR1 Control Translation Termination and Stop Codon Read-Through in Yeast Cells. *PLoS Genetics*, 9(11), e1003962.
<https://doi.org/10.1371/journal.pgen.1003962>

- Beznosková, P., Gunišová, S., & Valášek, L. S. (2016). Rules of UGA-N decoding by near-cognate tRNAs and analysis of readthrough on short uORFs in yeast. *RNA*, 22(3), 456–466. <https://doi.org/10.1261/rna.054452.115>
- Beznosková, P., Wagner, S., Jansen, M. E., von der Haar, T., & Valášek, L. S. (2015). Translation initiation factor eIF3 promotes programmed stop codon readthrough. *Nucleic Acids Research*, 43(10), 5099–5111. <https://doi.org/10.1093/nar/gkv421>
- Blanchard, S. C., Gonzalez, R. L., Kim, H. D., Chu, S., & Puglisi, J. D. (2004). tRNA selection and kinetic proofreading in translation. *Nature Structural & Molecular Biology*, 11(10), 1008–1014. <https://doi.org/10.1038/nsmb831>
- Bonetti, B., Fu, L., Moon, J., & Bedwell, D. M. (1995). The Efficiency of Translation Termination is Determined by a Synergistic Interplay Between Upstream and Downstream Sequences in *Saccharomyces cerevisiae*. *Journal of Molecular Biology*, 251(3), 334–345. <https://doi.org/10.1006/jmbi.1995.0438>
- Brogna, S., & Wen, J. (2009). Nonsense-mediated mRNA decay (NMD) mechanisms. *Nature Structural & Molecular Biology*, 16(2), 107–113. <https://doi.org/10.1038/nsmb.1550>
- Brown, A., Shao, S., Murray, J., Hegde, R. S., & Ramakrishnan, V. (2015). Structural basis for stop codon recognition in eukaryotes. *Nature*, 524(7566), 493–496. <https://doi.org/10.1038/nature14896>
- Brown, G. C. (1991). Total cell protein concentration as an evolutionary constraint on the metabolic control distribution in cells. *Journal of Theoretical Biology*, 153(2), 195–203. [https://doi.org/10.1016/S0022-5193\(05\)80422-9](https://doi.org/10.1016/S0022-5193(05)80422-9)
- Brunet, L. J., McMahon, J. A., McMahon, A. P., & Harland, R. M. (1998). Noggin, Cartilage Morphogenesis, and Joint Formation in the Mammalian Skeleton. *Science*, 280(5368), 1455–1457. <https://doi.org/10.1126/science.280.5368.1455>
- Budge, J. D., Knight, T. J., Povey, J., Roobol, J., Brown, I. R., Singh, G., Dean, A., Turner, S., Jaques, C. M., Young, R. J., Racher, A. J., & Smales, C. M. (2020a). Data for engineering lipid metabolism of Chinese hamster ovary (CHO) cells for enhanced recombinant protein production. *Data in Brief*, 29, 105217. <https://doi.org/10.1016/j.dib.2020.105217>
- Budge, J. D., Knight, T. J., Povey, J., Roobol, J., Brown, I. R., Singh, G., Dean, A., Turner, S., Jaques, C. M.,

- Young, R. J., Racher, A. J., & Smales, C. M. (2020b). Engineering of Chinese hamster ovary cell lipid metabolism results in an expanded ER and enhanced recombinant biotherapeutic protein production. *Metabolic Engineering*, 57, 203–216. <https://doi.org/10.1016/j.ymben.2019.11.007>
- Carlberg, U., Nilsson, A., & Nygård, O. (1990). Functional properties of phosphorylated elongation factor 2. *European Journal of Biochemistry*, 191(3), 639–645. <https://doi.org/10.1111/j.1432-1033.1990.tb19169.x>
- Carvalho, M. D. G. D. C., Carvalho, J. F., & Merrick, W. C. (1984). Biological characterization of various forms of elongation factor 1 from rabbit reticulocytes. *Archives of Biochemistry and Biophysics*, 234(2), 603–611. [https://doi.org/10.1016/0003-9861\(84\)90310-2](https://doi.org/10.1016/0003-9861(84)90310-2)
- Cheng, H., Jiang, W., Phillips, F. M., Haydon, R. C., Peng, Y., Zhou, L., Luu, H. H., An, N., Breyer, B., Vanichakarn, P., Szatkowski, J. P., Park, J. Y., & He, T. C. (2003). Osteogenic activity of the fourteen types of human bone morphogenetic proteins (BMPs). *Journal of Bone and Joint Surgery*, 85(8), 1544–1552. <https://doi.org/10.2106/00004623-200308000-00017>
- Choi, J., & Puglisi, J. D. (2017). Three tRNAs on the ribosome slow translation elongation. *Proceedings of the National Academy of Sciences*, 114(52), 13691–13696. <https://doi.org/10.1073/pnas.1719592115>
- Chu, D., Kazana, E., Bellanger, N., Singh, T., Tuite, M. F., & von der Haar, T. (2014). Translation elongation can control translation initiation on eukaryotic mRNAs. *The EMBO Journal*, 33(1), 21–34. <https://doi.org/10.1002/embj.201385651>
- Chu, D., Thompson, J., & von der Haar, T. (2014). Charting the dynamics of translation. *BioSystems*, 119(1), 1–9. <https://doi.org/10.1016/j.biosystems.2014.02.005>
- Chu, D., Zabet, N., & von der Haar, T. (2012). A novel and versatile computational tool to model translation. *Bioinformatics (Oxford, England)*, 28(2), 292–293. <https://doi.org/10.1093/bioinformatics/btr650>
- Cormack, B. P., Valdivia, R. H., & Falkow, S. (1996). FACS-optimized mutants of the green fluorescent protein (GFP). *Gene*, 173(1), 33–38. [https://doi.org/10.1016/0378-1119\(95\)00685-0](https://doi.org/10.1016/0378-1119(95)00685-0)
- Costamagna, D., Mommaerts, H., Sampaolesi, M., & Tylzanowski, P. (2016). Noggin inactivation affects the number and differentiation potential of muscle progenitor cells in vivo. *Scientific Reports*, 6(1),

31949. <https://doi.org/10.1038/srep31949>
- Dang, F., Nie, L., & Wei, W. (2021). Ubiquitin signaling in cell cycle control and tumorigenesis. *Cell Death & Differentiation*, 28(2), 427–438. <https://doi.org/10.1038/s41418-020-00648-0>
- Danna, K., & Nathans, D. (1971). Specific cleavage of simian virus 40 DNA by restriction endonuclease of *Hemophilus influenzae*. *Proceedings of the National Academy of Sciences*, 68(12), 2913–2917. <https://doi.org/10.1073/pnas.68.12.2913>
- DeCero, S. A., Winslow, C. H., & Coburn, J. (2020). Method to Overcome Inefficiencies in Site-Directed Mutagenesis of A/T-Rich DNA. *Journal of Biomolecular Techniques : JBT*, jbt.20-3103-003. <https://doi.org/10.7171/jbt.20-3103-003>
- des Georges, A., Hashem, Y., Unbehauen, A., Grassucci, R. A., Taylor, D., Hellen, C. U. T., Pestova, T. V., & Frank, J. (2014). Structure of the mammalian ribosomal pre-termination complex associated with eRF1•eRF3•GDPNP. *Nucleic Acids Research*, 42(5), 3409–3418. <https://doi.org/10.1093/nar/gkt1279>
- Dever, T. E., Dinman, J. D., & Green, R. (2018). Translation Elongation and Recoding in Eukaryotes. *Cold Spring Harbor Perspectives in Biology*, 10(8), a032649. <https://doi.org/10.1101/cshperspect.a032649>
- Dever, T. E., & Green, R. (2012). The Elongation, Termination, and Recycling Phases of Translation in Eukaryotes. *Cold Spring Harbor Perspectives in Biology*, 4(7), a013706–a013706. <https://doi.org/10.1101/cshperspect.a013706>
- Díaz-Moreno, M., Armenteros, T., Gradari, S., Hortigüela, R., García-Corzo, L., Fontán-Lozano, Á., Trejo, J. L., & Mira, H. (2018). Noggin rescues age-related stem cell loss in the brain of senescent mice with neurodegenerative pathology. *Proceedings of the National Academy of Sciences*, 115(45), 11625–11630. <https://doi.org/10.1073/pnas.1813205115>
- Dikic, I. (2017). Proteasomal and Autophagic Degradation Systems. *Annual Review of Biochemistry*, 86(1), 193–224. <https://doi.org/10.1146/annurev-biochem-061516-044908>
- Dinman, J. D. (2012). Mechanisms and implications of programmed translational frameshifting. *Wiley Interdisciplinary Reviews: RNA*, 3(5), 661–673. <https://doi.org/10.1002/wrna.1126>
- Doma, M. K., & Parker, R. (2006). Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation

- elongation. *Nature*, 440(7083), 561–564. <https://doi.org/10.1038/nature04530>
- Dong, J., Lai, R., Nielsen, K., Fekete, C. A., Qiu, H., & Hinnebusch, A. G. (2004). The Essential ATP-binding Cassette Protein RLI1 Functions in Translation by Promoting Preinitiation Complex Assembly. *Journal of Biological Chemistry*, 279(40), 42157–42168. <https://doi.org/10.1074/jbc.M404502200>
- Elf, J., Nilsson, D., Tenson, T., & Ehrenberg, M. (2003). Selective charging of tRNA isoacceptors explains patterns of codon usage. *Science*, 300(5626), 1718–1722. <https://doi.org/10.1126/science.1083811>
- Elkon, R., Zlotorynski, E., Zeller, K. I., & Agami, R. (2010). Major role for mRNA stability in shaping the kinetics of gene induction. *BMC Genomics*, 11, 259. <https://doi.org/10.1186/1471-2164-11-259>
- Engeland, K. (2022). Cell cycle regulation: p53-p21-RB signaling. *Cell Death & Differentiation*, 29(5), 946–960. <https://doi.org/10.1038/s41418-022-00988-z>
- Ensembl. (n.d.). *rs386700585 SUBSTITUTION - CCND3 Transcript*.
https://www.ensembl.org/Homo_sapiens/Variation/Explore?Db=core;G=ENSG00000112576;R=6:41934956-42048968;T=ENST00000372991;V=rs386700585;Vdb=variation;Vf=237523992.
 Retrieved September 29, 2023, from
https://www.ensembl.org/Homo_sapiens/Variation/Explore?db=core;g=ENSG00000112576;r=6:41934956-42048968;t=ENST00000372991;v=rs386700585;vdb=variation;vf=237523992
- Fischer, N., Neumann, P., Konevega, A. L., Bock, L. V., Ficner, R., Rodnina, M. V., & Stark, H. (2015). Structure of the E. coli ribosome–EF-Tu complex at 3 Å resolution by Cs-corrected cryo-EM. *Nature*, 520(7548), 567–570. <https://doi.org/10.1038/nature14275>
- Frolova, L., Le Goff, X., Zhouravleva, G., Davydova, E., Philippe, M., & Kisselev, L. (1996). Eukaryotic polypeptide chain release factor eRF3 is an eRF1- and ribosome-dependent guanosine triphosphatase. *RNA (New York, N.Y.)*, 2(4), 334–341.
- Frolova, L., Seit-Nebi, A., & Kisselev, L. (2002). Highly conserved NIKS tetrapeptide is functionally essential in eukaryotic translation termination factor eRF1. *RNA*, 8(2), 129–136. <https://doi.org/10.1017/S1355838202013262>
- Frolova, L. Y., Tsivkovskii, R. Y., Sivolobova, G. F., Oparina, N. Y., Serpinsky, O. I., Blinov, V. M., Tatkov, S. I., & Kisselev, L. L. (1999). Mutations in the highly conserved GGQ motif of class I polypeptide release factors abolish ability of human eRF1 to trigger peptidyl-tRNA hydrolysis. *RNA*, 5(8), 1014–

1020. <https://doi.org/10.1017/S135583829999043X>

Gamble, C. E., Brule, C. E., Dean, K. M., Fields, S., & Grayhack, E. J. (2016). Adjacent Codons Act in Concert to Modulate Translation Efficiency in Yeast. *Cell*, 166(3), 679–690. <https://doi.org/10.1016/j.cell.2016.05.070>

Gao, Y.-G., Selmer, M., Dunham, C. M., Weixlbaumer, A., Kelley, A. C., & Ramakrishnan, V. (2009). The Structure of the Ribosome with Elongation Factor G Trapped in the Posttranslocational State. *Science*, 326(5953), 694–699. <https://doi.org/10.1126/science.1179709>

Gebauer, F., Preiss, T., & Hentze, M. W. (2012). From Cis-Regulatory Elements to Complex RNPs and Back. *Cold Spring Harbor Perspectives in Biology*, 4(7), a012245–a012245. <https://doi.org/10.1101/cshperspect.a012245>

Geggier, P., Dave, R., Feldman, M. B., Terry, D. S., Altman, R. B., Munro, J. B., & Blanchard, S. C. (2010). Conformational Sampling of Aminoacyl-tRNA during Selection on the Bacterial Ribosome. *Journal of Molecular Biology*, 399(4), 576–595. <https://doi.org/10.1016/j.jmb.2010.04.038>

Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R.-Y., Algire, M. A., Benders, G. A., Montague, M. G., Ma, L., Moodie, M. M., Merryman, C., Vashee, S., Krishnakumar, R., Assad-Garcia, N., Andrews-Pfannkoch, C., Denisova, E. A., Young, L., Qi, Z.-Q., Segall-Shapiro, T. H., ... Venter, J. C. (2010). Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Science*, 329(5987), 52–56. <https://doi.org/10.1126/science.1190719>

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, 6(5), 343–345. <https://doi.org/10.1038/nmeth.1318>

Gnirke, A., Geigenmüller, U., Rheinberger, H. J., & Nierhaus, L. H. (1989). The allosteric three-site model for the ribosomal elongation cycle. Analysis with a heteropolymeric mRNA. *The Journal of Biological Chemistry*, 264(13), 7291–7301.

Godet, A.-C., David, F., Hantelys, F., Tatin, F., Lacazette, E., Garmy-Susini, B., & Prats, A.-C. (2019). IRES Trans-Acting Factors, Key Actors of the Stress Response. *International Journal of Molecular Sciences*, 20(4), 924. <https://doi.org/10.3390/ijms20040924>

Goey, C. H., Tsang, J. M. H., Bell, D., & Kontoravdi, C. (2017). Cascading effect in bioprocessing-The

- impact of mild hypothermia on CHO cell behavior and host cell protein composition. *Biotechnology and Bioengineering*, 114(12), 2771–2781. <https://doi.org/10.1002/bit.26437>
- Graille, M., Chaillet, M., & van Tilbeurgh, H. (2008). Structure of Yeast Dom34. *Journal of Biological Chemistry*, 283(11), 7145–7154. <https://doi.org/10.1074/jbc.M708224200>
- Gratacós, F. M., & Brewer, G. (2010). The role of AUF1 in regulated mRNA decay. *WIREs RNA*, 1(3), 457–473. <https://doi.org/10.1002/wrna.26>
- Gregio, A. P. B., Cano, V. P. S., Avaca, J. S., Valentini, S. R., & Zanelli, C. F. (2009). eIF5A has a function in the elongation step of translation in yeast. *Biochemical and Biophysical Research Communications*, 380(4), 785–790. <https://doi.org/10.1016/j.bbrc.2009.01.148>
- Guan, S., & Rosenecker, J. (2017). Nanotechnologies in delivery of mRNA therapeutics using nonviral vector-based delivery systems. *Gene Therapy*, 24(3), 133–143. <https://doi.org/10.1038/gt.2017.5>
- Guohong, Z. (1996). An Enhanced Green Fluorescent Protein Allows Sensitive Detection of Gene Transfer in Mammalian Cells. *BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS*, 227, 707–711.
- Gurvich, O. L., Baranov, P. V., Gesteland, R. F., & Atkins, J. F. (2005). Expression Levels Influence Ribosomal Frameshifting at the Tandem Rare Arginine Codons AGG_AGG and AGA_AGA in *Escherichia coli*. *Journal of Bacteriology*, 187(12), 4023–4032. <https://doi.org/10.1128/JB.187.12.4023-4032.2005>
- Gutierrez, E., Shin, B.-S., Woolstenhulme, C. J., Kim, J.-R., Saini, P., Buskirk, A. R., & Dever, T. E. (2013). eIF5A Promotes Translation of Polyproline Motifs. *Molecular Cell*, 51(1), 35–45. <https://doi.org/10.1016/j.molcel.2013.04.021>
- Guydosh, N. R., & Green, R. (2014). Dom34 Rescues Ribosomes in 3' Untranslated Regions. *Cell*, 156(5), 950–962. <https://doi.org/10.1016/j.cell.2014.02.006>
- Guydosh, N. R., & Green, R. (2017). Translation of poly(A) tails leads to precise mRNA cleavage. *RNA*, 23(5), 749–761. <https://doi.org/10.1261/rna.060418.116>
- Hanson, G., & Collier, J. (2017). Codon optimality, bias and usage in translation and mRNA decay. *Nature Reviews Molecular Cell Biology*, 19(1), 20–30. <https://doi.org/10.1038/nrm.2017.91>
- Harigaya, Y., & Parker, R. (2016). Analysis of the association between codon optimality and mRNA

- stability in *Schizosaccharomyces pombe*. *BMC Genomics*, 17(1), 895.
<https://doi.org/10.1186/s12864-016-3237-6>
- Herzog, V. A., Reichholf, B., Neumann, T., Rescheneder, P., Bhat, P., Burkard, T. R., Wlotzka, W., von Haeseler, A., Zuber, J., & Ameres, S. L. (2017). Thiol-linked alkylation of RNA to assess expression dynamics. *Nature Methods*, 14(12), 1198–1204. <https://doi.org/10.1038/nmeth.4435>
- Heuer, A., Gerovac, M., Schmidt, C., Trowitzsch, S., Preis, A., Kötter, P., Berninghausen, O., Becker, T., Beckmann, R., & Tampé, R. (2017). Structure of the 40S–ABCE1 post-splitting complex in ribosome recycling and translation initiation. *Nature Structural & Molecular Biology*, 24(5), 453–460.
<https://doi.org/10.1038/nsmb.3396>
- Hinnebusch, A. G. (2011). Molecular Mechanism of Scanning and Start Codon Selection in Eukaryotes. *Microbiology and Molecular Biology Reviews : MMBR*, 75(3), 434.
<https://doi.org/10.1128/MMBR.00008-11>
- Hinnebusch, A. G. (2014). The Scanning Mechanism of Eukaryotic Translation Initiation. *Annual Review of Biochemistry*, 83(1), 779–812. <https://doi.org/10.1146/annurev-biochem-060713-035802>
- Hinnebusch, A. G., & Lorsch, J. R. (2012). The Mechanism of Eukaryotic Translation Initiation: New Insights and Challenges. *Cold Spring Harbor Perspectives in Biology*, 4(10).
<https://doi.org/10.1101/CSHPERSPECT.A011544>
- Hochegger, H., Takeda, S., & Hunt, T. (2008). Cyclin-dependent kinases and cell-cycle transitions: does one fit all? *Nature Reviews Molecular Cell Biology*, 9(11), 910–916.
<https://doi.org/10.1038/nrm2510>
- Hu, X., Bryington, M., Fisher, A. B., Liang, X., Zhang, X., Cui, D., Datta, I., & Zuckerman, K. S. (2002). Ubiquitin/Proteasome-dependent Degradation of D-type Cyclins Is Linked to Tumor Necrosis Factor-induced Cell Cycle Arrest. *Journal of Biological Chemistry*, 277(19), 16528–16537.
<https://doi.org/10.1074/jbc.M109929200>
- Ivanov, A., Mikhailova, T., Eliseev, B., Yeramala, L., Sokolova, E., Susorov, D., Shuvalov, A., Schaffitzel, C., & Alkalaeva, E. (2016). PABP enhances release factor recruitment and stop codon recognition during translation termination. *Nucleic Acids Research*, 44(16), 7766–7776.
<https://doi.org/10.1093/nar/gkw635>

- Jacks, T., Power, M. D., Masiarz, F. R., Luciw, P. A., Barr, P. J., & Varmus, H. E. (1988). Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature*, 331(6153), 280–283.
<https://doi.org/10.1038/331280a0>
- Jackson, R. J. (2013). The Current Status of Vertebrate Cellular mRNA IRESs. *Cold Spring Harbor Perspectives in Biology*, 5(2), a011569–a011569. <https://doi.org/10.1101/cshperspect.a011569>
- Jackson, R. J., Hellen, C. U. T., & Pestova, T. V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Reviews Molecular Cell Biology* 2010 11:2, 11(2), 113–127.
<https://doi.org/10.1038/nrm2838>
- Januszyk, K., & Lima, C. D. (2014). The eukaryotic RNA exosome. *Current Opinion in Structural Biology*, 24, 132–140. <https://doi.org/10.1016/j.sbi.2014.01.011>
- Jao, D. L.-E., & Chen, K. Y. (2006). Tandem affinity purification revealed the hypusine-dependent binding of eukaryotic initiation factor 5A to the translating 80S ribosomal complex. *Journal of Cellular Biochemistry*, 97(3), 583–598. <https://doi.org/10.1002/jcb.20658>
- Jefferis, R. (2009). Glycosylation as a strategy to improve antibody-based therapeutics. *Nature Reviews Drug Discovery*, 8(3), 226–234. <https://doi.org/10.1038/nrd2804>
- Joazeiro, C. A. P. (2017). Ribosomal Stalling During Translation: Providing Substrates for Ribosome-Associated Protein Quality Control. *Annual Review of Cell and Developmental Biology*, 33(1), 343–368. <https://doi.org/10.1146/annurev-cellbio-111315-125249>
- Johnson, D. G. (1998). Role of E2F in cell cycle control and cancer. *Frontiers in Bioscience*, 3(4), A291.
<https://doi.org/10.2741/A291>
- Jung, H., Gkogkas, C. G., Sonenberg, N., & Holt, C. E. (2014). Remote control of gene function by local translation. *Cell*, 157(1), 26–40. <https://doi.org/10.1016/j.cell.2014.03.005>
- Karikó, K., Muramatsu, H., Welsh, F. A., Ludwig, J., Kato, H., Akira, S., & Weissman, D. (2008). Incorporation of Pseudouridine Into mRNA Yields Superior Nonimmunogenic Vector With Increased Translational Capacity and Biological Stability. *Molecular Therapy*, 16(11), 1833–1840.
<https://doi.org/10.1038/mt.2008.200>
- Katahira, J. (2015). Nuclear Export of Messenger RNA. *Genes*, 6(2), 163–184.
<https://doi.org/10.3390/genes6020163>

- Kauffman, K. J., Webber, M. J., & Anderson, D. G. (2016). Materials for non-viral intracellular delivery of messenger RNA therapeutics. *Journal of Controlled Release*, 240, 227–234.
<https://doi.org/10.1016/j.jconrel.2015.12.032>
- Kaufmann, H., Mazur, X., Fussenegger, M., & Bailey, J. E. (1999). Influence of low temperature on productivity, proteome and protein phosphorylation of CHO cells. *Biotechnology and Bioengineering*, 63(5), 573–582. [https://doi.org/10.1002/\(SICI\)1097-0290\(19990605\)63:5<573::AID-BIT7>3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1097-0290(19990605)63:5<573::AID-BIT7>3.0.CO;2-Y)
- Kaushal, A. (2023). Innate immune regulations and various siRNA modalities. *Drug Delivery and Translational Research*. <https://doi.org/10.1007/s13346-023-01361-4>
- Khoshnevis, S., Gross, T., Rotte, C., Baierlein, C., Ficner, R., & Krebber, H. (2010). The iron–sulphur protein RNase L inhibitor functions in translation termination. *EMBO Reports*, 11(3), 214–219.
<https://doi.org/10.1038/embor.2009.272>
- Kilchert, C., Wittmann, S., & Vasiljeva, L. (2016). The regulation and functions of the nuclear RNA exosome complex. *Nature Reviews Molecular Cell Biology*, 17(4), 227–239.
<https://doi.org/10.1038/nrm.2015.15>
- Knight, J. R. P., Bastide, A., Roobol, A., Roobol, J., Jackson, T. J., Utami, W., Barrett, D. A., Smales, C. M., & Willis, A. E. (2015). Eukaryotic elongation factor 2 kinase regulates the cold stress response by slowing translation elongation. *Biochemical Journal*, 465(2), 227–238.
<https://doi.org/10.1042/BJ20141014>
- Knight, J. R. P., Garland, G., Pöyry, T., Mead, E., Vlahov, N., Sfakianos, A., Grosso, S., De-Lima-Hedayioglu, F., Mallucci, G. R., von der Haar, T., Smales, C. M., Sansom, O. J., & Willis, A. E. (2020). Control of translation elongation in health and disease. *Disease Models & Mechanisms*, 13(3).
<https://doi.org/10.1242/dmm.043208>
- Kolupaeva, V. G., Unbehauen, A., Lomakin, I. B., Hellen, C. U. T., & Pestova, T. V. (2005). Binding of eukaryotic initiation factor 3 to ribosomal 40S subunits and its role in ribosomal dissociation and anti-association. *RNA*, 11(4), 470–486. <https://doi.org/10.1261/rna.7215305>
- Kozak, M. (1986). Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proceedings of the National Academy of Sciences*, 83(9), 2850–2854.
<https://doi.org/10.1073/pnas.83.9.2850>

- Łabno, A., Tomecki, R., & Dziembowski, A. (2016). Cytoplasmic RNA decay pathways - Enzymes and mechanisms. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1863(12), 3125–3147. <https://doi.org/10.1016/j.bbamcr.2016.09.023>
- Lai, M.-C., Lee, Y.-H. W., & Tarn, W.-Y. (2008). The DEAD-Box RNA Helicase DDX3 Associates with Export Messenger Ribonucleoproteins as well as Tip-associated Protein and Participates in Translational Control. *Molecular Biology of the Cell*, 19(9), 3847–3858. <https://doi.org/10.1091/mbc.e07-12-1264>
- Legate, K. R. (2000). Nucleotide dependent binding of the GTPase domain of the signal recognition particle receptor beta-subunit to the alpha-subunit. *Journal of Biological Chemistry*. <https://doi.org/10.1074/jbc.M003215200>
- Linxweiler, M., Schick, B., & Zimmermann, R. (2017). Let's talk about Secs: Sec61, Sec62 and Sec63 in signal transduction, oncology and personalized medicine. *Signal Transduction and Targeted Therapy*, 2(1), 17002. <https://doi.org/10.1038/sigtrans.2017.2>
- Lucent, D., Snow, C. D., Aitken, C. E., & Pande, V. S. (2010). Non-Bulk-Like Solvent Behavior in the Ribosome Exit Tunnel. *PLoS Computational Biology*, 6(10), e1000963. <https://doi.org/10.1371/journal.pcbi.1000963>
- Luo, Y., Na, Z., & Slavoff, S. A. (2018). P-Bodies: Composition, Properties, and Functions. *Biochemistry*, 57(17), 2424–2431. <https://doi.org/10.1021/acs.biochem.7b01162>
- Luria, S. E., & Human, M. L. (1952). A NONHEREDITARY, HOST-INDUCED VARIATION OF BACTERIAL VIRUSES. *Journal of Bacteriology*, 64(4), 557–569. <https://doi.org/10.1128/jb.64.4.557-569.1952>
- Mangus, D. A., Evans, M. C., & Jacobson, A. (2003). Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression. *Genome Biology*, 4(7), 223. <https://doi.org/10.1186/gb-2003-4-7-223>
- Maracci, C., Peske, F., Dannies, E., Pohl, C., & Rodnina, M. V. (2014). Ribosome-induced tuning of GTP hydrolysis by a translational GTPase. *Proceedings of the National Academy of Sciences*, 111(40), 14418–14423. <https://doi.org/10.1073/pnas.1412676111>
- Marcelino, J., Sciortino, C. M., Romero, M. F., Ulatowski, L. M., Ballock, R. T., Economides, A. N., Eimon, P. M., Harland, R. M., & Warman, M. L. (2001). Human disease-causing *NOG* missense mutations:

- Effects on noggin secretion, dimer formation, and bone morphogenetic protein binding. *Proceedings of the National Academy of Sciences*, 98(20), 11353–11358.
<https://doi.org/10.1073/pnas.201367598>
- Mason, N., Ciufo, L. F., & Brown, J. D. (2000). Elongation arrest is a physiologically important function of signal recognition particle. *The EMBO Journal*, 19(15), 4164–4174.
<https://doi.org/10.1093/emboj/19.15.4164>
- Mauro, V. P. (2018). Codon Optimization in the Production of Recombinant Biotherapeutics: Potential Risks and Considerations. *BioDrugs*, 32(1), 69–81. <https://doi.org/10.1007/s40259-018-0261-x>
- McClary, B., Zinshteyn, B., Meyer, M., Jouanneau, M., Pellegrino, S., Yusupova, G., Schuller, A., Reyes, J. C. P., Lu, J., Guo, Z., Ayinde, S., Luo, C., Dang, Y., Romo, D., Yusupov, M., Green, R., & Liu, J. O. (2017). Inhibition of Eukaryotic Translation by the Antitumor Natural Product Agelastatin A. *Cell Chemical Biology*, 24(5), 605-613.e5. <https://doi.org/10.1016/j.chembiol.2017.04.006>
- McHugh, K. P., Xu, J., Aron, K. L., Borys, M. C., & Li, Z. J. (2020). Effective temperature shift strategy development and scale confirmation for simultaneous optimization of protein productivity and quality in Chinese hamster ovary cells. *Biotechnology Progress*, 36(3).
<https://doi.org/10.1002/btpr.2959>
- Mead, E. J. E. J., Masterton, R. J. R. J., Von Der Haar, T., Tuite, M. F. M. F., & Smales, C. M. M. (2014). Control and regulation of mRNA translation. *Biochemical Society Transactions*, 42(1), 151–154.
<https://doi.org/10.1042/BST20130259>
- Meydan, S., & Guydosh, N. R. (2021). A cellular handbook for collided ribosomes: surveillance pathways and collision types. *Current Genetics*, 67(1), 19–26. <https://doi.org/10.1007/s00294-020-01111-w>
- Mierzejewska, K., Siwek, W., Czapinska, H., Kaus-Drobek, M., Radlinska, M., Skowronek, K., Bujnicki, J. M., Dadlez, M., & Bochtler, M. (2014). Structural basis of the methylation specificity of R.DpnI. *Nucleic Acids Research*, 42(13), 8745–8754. <https://doi.org/10.1093/nar/gku546>
- Miller, W. A., Wang, Z., & Treder, K. (2007). The amazing diversity of cap-independent translation elements in the 3'-untranslated regions of plant viral RNAs. *Biochemical Society Transactions*, 35(6), 1629–1633. <https://doi.org/10.1042/BST0351629>
- Millevoi, S., & Vagner, S. (2010). Molecular mechanisms of eukaryotic pre-mRNA 3' end processing

- regulation. *Nucleic Acids Research*, 38(9), 2757–2774. <https://doi.org/10.1093/nar/gkp1176>
- Moazed, D., & Noller, H. F. (1989). Intermediate states in the movement of transfer RNA in the ribosome. *Nature*, 342(6246), 142–148. <https://doi.org/10.1038/342142a0>
- Muhlrad, D., Decker, C. J., & Parker, R. (1994). Deadenylation of the unstable mRNA encoded by the yeast MFA2 gene leads to decapping followed by 5'–&3' digestion of the transcript. *Genes & Development*, 8(7), 855–866. <https://doi.org/10.1101/gad.8.7.855>
- Mukherjee, C., Bakthavachalu, B., & Schoenberg, D. R. (2014). The Cytoplasmic Capping Complex Assembles on Adapter Protein Nck1 Bound to the Proline-Rich C-Terminus of Mammalian Capping Enzyme. *PLoS Biology*, 12(8), e1001933. <https://doi.org/10.1371/journal.pbio.1001933>
- Namy, O., Moran, S. J., Stuart, D. I., Gilbert, R. J. C., & Brierley, I. (2006). A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature*, 441(7090), 244–247. <https://doi.org/10.1038/nature04735>
- Narula, A., Ellis, J., Taliaferro, J. M., & Rissland, O. S. (2019). Coding regions affect mRNA stability in human cells. *Rna*, rna.073239.119. <https://doi.org/10.1261/rna.073239.119>
- Nieuwenhuis, B., Haenzi, B., Hilton, S., Carnicer-Lombarte, A., Hobo, B., Verhaagen, J., & Fawcett, J. W. (2021). Optimization of adeno-associated viral vector-mediated transduction of the corticospinal tract: comparison of four promoters. *Gene Therapy*, 28(1–2), 56–74. <https://doi.org/10.1038/s41434-020-0169-1>
- Ou, K.-C., Wang, C.-Y., Liu, K.-T., Chen, Y.-L., Chen, Y.-C., Lai, M.-D., & Yen, M.-C. (2014). Optimization protein productivity of human interleukin-2 through codon usage, gene copy number and intracellular tRNA concentration in CHO cells. *Biochemical and Biophysical Research Communications*, 454(2), 347–352. <https://doi.org/10.1016/j.bbrc.2014.10.097>
- Pan, Q., Shai, O., Lee, L. J., Frey, B. J., & Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics*, 40(12), 1413–1415. <https://doi.org/10.1038/ng.259>
- Pardi, N., Hogan, M. J., Porter, F. W., & Weissman, D. (2018). mRNA vaccines — a new era in vaccinology. *Nature Reviews Drug Discovery*, 17(4), 261–279. <https://doi.org/10.1038/nrd.2017.243>

- Park, M. H., Cooper, H. L., & Folk, J. E. (1981). Identification of hypusine, an unusual amino acid, in a protein from human lymphocytes and of spermidine as its biosynthetic precursor. *Proceedings of the National Academy of Sciences of the United States of America*, 78(5), 2869–2873. <https://doi.org/10.1073/pnas.78.5.2869>
- Park, M. H., Nishimura, K., Zanelli, C. F., & Valentini, S. R. (2010). Functional significance of eIF5A and its hypusine modification in eukaryotes. *Amino Acids*, 38(2), 491–500. <https://doi.org/10.1007/s00726-009-0408-7>
- Park, S., Park, J. M., Kim, S., Kim, J.-A., Shepherd, J. D., Smith-Hicks, C. L., Chowdhury, S., Kaufmann, W., Kuhl, D., Ryazanov, A. G., Haganir, R. L., Linden, D. J., & Worley, P. F. (2008). Elongation Factor 2 and Fragile X Mental Retardation Protein Control the Dynamic Translation of Arc/Arg3.1 Essential for mGluR-LTD. *Neuron*, 59(1), 70–83. <https://doi.org/10.1016/j.neuron.2008.05.023>
- Passmore, L. A., & Collier, J. (2022). Roles of mRNA poly(A) tails in regulation of eukaryotic gene expression. *Nature Reviews Molecular Cell Biology*, 23(2), 93–106. <https://doi.org/10.1038/s41580-021-00417-y>
- Pavlov, M. Y., Watts, R. E., Tan, Z., Cornish, V. W., Ehrenberg, M., & Forster, A. C. (2009). Slow peptide bond formation by proline and other *N*-alkylamino acids in translation. *Proceedings of the National Academy of Sciences*, 106(1), 50–54. <https://doi.org/10.1073/pnas.0809211106>
- Pelechano, V., & Alepuz, P. (2017). eIF5A facilitates translation termination globally and promotes the elongation of many non polyproline-specific tripeptide sequences. *Nucleic Acids Research*, 45(12), 7326–7338. <https://doi.org/10.1093/nar/gkx479>
- Peltz, S. W., Brown, A. H., & Jacobson, A. (1993). mRNA destabilization triggered by premature translational termination depends on at least three cis-acting sequence elements and one trans-acting factor. *Genes & Development*, 7(9), 1737–1754. <https://doi.org/10.1101/gad.7.9.1737>
- Peretti, D., Bastide, A., Radford, H., Verity, N., Molloy, C., Martin, M. G., Moreno, J. A., Steinert, J. R., Smith, T., Dinsdale, D., Willis, A. E., & Mallucci, G. R. (2015). RBM3 mediates structural plasticity and protective effects of cooling in neurodegeneration. *Nature*, 518(7538), 236–239. <https://doi.org/10.1038/nature14142>
- Pestova, T. V., Borukhov, S. I., & Hellen, C. U. T. (1998). Eukaryotic ribosomes require initiation factors 1 and 1A to locate initiation codons. *Nature*, 394(6696), 854–859. <https://doi.org/10.1038/29703>

- Pisarev, A. V., Hellen, C. U. T., & Pestova, T. V. (2007). Recycling of Eukaryotic Posttermination Ribosomal Complexes. *Cell*, 131(2), 286–299. <https://doi.org/10.1016/j.cell.2007.08.041>
- Pisarev, A. V., Skabkin, M. A., Pisareva, V. P., Skabkina, O. V., Rakotondrafara, A. M., Hentze, M. W., Hellen, C. U. T., & Pestova, T. V. (2010). The Role of ABCE1 in Eukaryotic Posttermination Ribosomal Recycling. *Molecular Cell*, 37(2), 196–210. <https://doi.org/10.1016/j.molcel.2009.12.034>
- Pittman, Y. R., Valente, L., Jeppesen, M. G., Andersen, G. R., Patel, S., & Kinzy, T. G. (2006). Mg²⁺ and a Key Lysine Modulate Exchange Activity of Eukaryotic Translation Elongation Factor 1B α . *Journal of Biological Chemistry*, 281(28), 19457–19468. <https://doi.org/10.1074/jbc.M601076200>
- Plotkin, J. B., & Kudla, G. (2011). Synonymous but not the same. *National Review of Genetics*, 12(1), 32–42. <https://doi.org/10.1038/nrg2899.Synonymous>
- Potter, M. D., & Nicchitta, C. V. (2000). Regulation of Ribosome Detachment from the Mammalian Endoplasmic Reticulum Membrane. *Journal of Biological Chemistry*, 275(43), 33828–33835. <https://doi.org/10.1074/jbc.M005294200>
- Prasher, D. C., Eckenrode, V. K., Ward, W. W., Prendergast, F. G., & Cormier, M. J. (1992). Primary structure of the Aequorea victoria green-fluorescent protein. *Gene*, 111(2), 229–233. [https://doi.org/10.1016/0378-1119\(92\)90691-H](https://doi.org/10.1016/0378-1119(92)90691-H)
- Pratt, A. J., & MacRae, I. J. (2009). The RNA-induced Silencing Complex: A Versatile Gene-silencing Machine. *Journal of Biological Chemistry*, 284(27), 17897–17901. <https://doi.org/10.1074/jbc.R900012200>
- Preis, A., Heuer, A., Barrio-Garcia, C., Hauser, A., Eyler, D. E., Berninghausen, O., Green, R., Becker, T., & Beckmann, R. (2014). Cryoelectron Microscopic Structures of Eukaryotic Translation Termination Complexes Containing eRF1-eRF3 or eRF1-ABCE1. *Cell Reports*, 8(1), 59–65. <https://doi.org/10.1016/j.celrep.2014.04.058>
- Presnyak, V., Alhusaini, N., Chen, Y.-H. H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K. E. E., Graveley, B. R. R., & Collier, J. (2015). Codon Optimality Is a Major Determinant of mRNA Stability. *Cell*, 160(6), 1111–1124. <https://doi.org/10.1016/j.cell.2015.02.029>
- Qiang, X., Yang, W.-L., Wu, R., Zhou, M., Jacob, A., Dong, W., Kuncewitch, M., Ji, Y., Yang, H., Wang, H., Fujita, J., Nicastro, J., Coppa, G. F., Tracey, K. J., & Wang, P. (2013). Cold-inducible RNA-binding

- protein (CIRP) triggers inflammatory responses in hemorrhagic shock and sepsis. *Nature Medicine*, October, 1–9. <https://doi.org/10.1038/nm.3368>
- Quax, T. E. F., Claassens, N. J., Söll, D., & van der Oost, J. (2015). Codon Bias as a Means to Fine-Tune Gene Expression. *Molecular Cell*, 59(2), 149–161. <https://doi.org/10.1016/j.molcel.2015.05.035>
- Raghavan, A. (2002). Genome-wide analysis of mRNA decay in resting and activated primary human T lymphocytes. *Nucleic Acids Research*, 30(24), 5529–5538. <https://doi.org/10.1093/nar/gkf682>
- Ramanathan, A., Robb, G. B., & Chan, S.-H. (2016). mRNA capping: biological functions and applications. *Nucleic Acids Research*, 44(16), 7511–7526. <https://doi.org/10.1093/nar/gkw551>
- Ratje, A. H., Loerke, J., Mikolajka, A., Brünner, M., Hildebrand, P. W., Starosta, A. L., Dönhöfer, A., Connell, S. R., Fucini, P., Mielke, T., Whitford, P. C., Onuchic, J. N., Yu, Y., Sanbonmatsu, K. Y., Hartmann, R. K., Penczek, P. A., Wilson, D. N., & Spahn, C. M. T. (2010). Head swivel on the ribosome facilitates translocation by means of intra-subunit tRNA hybrid sites. *Nature*, 468(7324), 713–716. <https://doi.org/10.1038/nature09547>
- Renana, S., & Tamir, T. (2014). Modelling the efficiency of codon-tRNA interactions based on codon usage bias. *DNA Research*, 21(5), 511–525. <https://doi.org/10.1093/dnares/dsu017>
- Roberts, R. J. (2005). How restriction enzymes became the workhorses of molecular biology. *Proceedings of the National Academy of Sciences*, 102(17), 5905–5908. <https://doi.org/10.1073/pnas.0500923102>
- Rodnina, M. V., & Wintermeyer, W. (2009). Recent mechanistic insights into eukaryotic ribosomes. *Current Opinion in Cell Biology*, 21(3), 435–443. <https://doi.org/10.1016/j.ceb.2009.01.023>
- Rong, L., Livingstone, M., Sukarieh, R., Petroulakis, E., Gingras, A.-C., Crosby, K., Smith, B., Polakiewicz, R. D., Pelletier, J., Ferraiuolo, M. A., & Sonenberg, N. (2008). Control of eIF4E cellular localization by eIF4E-binding proteins, 4E-BPs. *RNA*, 14(7), 1318–1327. <https://doi.org/10.1261/rna.950608>
- Ross, J. (1995). mRNA stability in mammalian cells. *Microbiological Reviews*, 59(3), 423–450. <http://www.ncbi.nlm.nih.gov/pubmed/9204770>
- Rouwendaal*, G. J. A., Mendes, O., Wolbert, E. J. H., & Douwe de Boer, A. (1997). Enhanced expression in tobacco of the gene encoding green fluorescent protein by modification of its codon usage. *Plant Molecular Biology*, 33(6), 989–999. <https://doi.org/10.1023/A:1005740823703>

- Saini, P., Eyler, D. E., Green, R., & Dever, T. E. (2009). Hypusine-containing protein eIF5A promotes translation elongation. *Nature*, 459(7243), 118–121. <https://doi.org/10.1038/nature08034>
- Sainsbury, S., Bernecky, C., & Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. *Nature Reviews Molecular Cell Biology*, 16(3), 129–143. <https://doi.org/10.1038/nrm3952>
- Salas-Marco, J., & Bedwell, D. M. (2004). GTP Hydrolysis by eRF3 Facilitates Stop Codon Decoding during Eukaryotic Translation Termination. *Molecular and Cellular Biology*, 24(17), 7769–7778. <https://doi.org/10.1128/MCB.24.17.7769-7778.2004>
- Sasikumar, A. N., Perez, W. B., & Kinzy, T. G. (2012). The many roles of the eukaryotic elongation factor 1 complex. *WIREs RNA*, 3(4), 543–555. <https://doi.org/10.1002/wrna.1118>
- Sawai, C. M., Freund, J., Oh, P., Ndiaye-Lobry, D., Bretz, J. C., Strikoudis, A., Genesca, L., Trimarchi, T., Kelliher, M. A., Clark, M., Soulier, J., Chen-Kiang, S., & Aifantis, I. (2012). Therapeutic Targeting of the Cyclin D3:CDK4/6 Complex in T Cell Leukemia. *Cancer Cell*, 22(4), 452–465. <https://doi.org/10.1016/j.ccr.2012.09.016>
- Schmidt, C., Becker, T., Heuer, A., Braunger, K., Shanmuganathan, V., Pech, M., Berninghausen, O., Wilson, D. N., & Beckmann, R. (2016). Structure of the hypusinylated eukaryotic translation factor eIF-5A bound to the ribosome. *Nucleic Acids Research*, 44(4), 1944–1951. <https://doi.org/10.1093/nar/gkv1517>
- Schneider-Poetsch, T., Usui, T., Kaida, D., & Yoshida, M. (2010). Garbled messages and corrupted translations. *Nature Chemical Biology*, 6(3), 189–198. <https://doi.org/10.1038/nchembio.326>
- Schreier, M. H., & Staehelin, T. (1973). Initiation of eukaryotic protein synthesis: [Met-tRNA^f.40S Ribosome] initiation complex catalysed by purified initiation factors in the absence of mRNA. *Nature New Biology*, 242(115), 35–38. <https://doi.org/10.1038/newbio242035a0>
- Schuller, A. P., & Green, R. (2018). Roadblocks and resolutions in eukaryotic translation. *Nature Reviews Molecular Cell Biology*, 19(8), 1–16. <https://doi.org/10.1038/s41580-018-0011-4>
- Schuller, A. P., Wu, C. C.-C., Dever, T. E., Buskirk, A. R., & Green, R. (2017). eIF5A Functions Globally in Translation Elongation and Termination. *Molecular Cell*, 66(2), 194–205.e5. <https://doi.org/10.1016/j.molcel.2017.03.003>

- Sejour, R., Leatherwood, J., & Futcher, B. (2022). *Slow Translation of the 5' ends of genes may be a non-adaptive consequence of their rapid evolutionary turn-over and slow loss of rare codons.*
- Sharp, P. M., & Li, W. H. (1987). The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research*, 15(3), 1281–1295. <https://doi.org/10.1093/nar/15.3.1281>
- Shatkin, A. J. (1976). Capping of eucaryotic mRNAs. In *Cell* (Vol. 9, Issue 4 PART 2, pp. 645–653). [https://doi.org/10.1016/0092-8674\(76\)90128-8](https://doi.org/10.1016/0092-8674(76)90128-8)
- Sherr, C. J. (1995). D-type cyclins. *Trends in Biochemical Sciences*, 20(5), 187–190. [https://doi.org/10.1016/S0968-0004\(00\)89005-2](https://doi.org/10.1016/S0968-0004(00)89005-2)
- Sherr, C. J., & Roberts, J. M. (2004). Living with or without cyclins and cyclin-dependent kinases. *Genes & Development*, 18(22), 2699–2711. <https://doi.org/10.1101/gad.1256504>
- Shimomura, O., Johnson, F. H., & Saiga, Y. (1962). Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan, Aequorea. *Journal of Cellular and Comparative Physiology*, 59(3), 223–239. <https://doi.org/10.1002/jcp.1030590302>
- Shoji, S., Walker, S. E., & Fredrick, K. (2009). Ribosomal Translocation: One Step Closer to the Molecular Mechanism. *ACS Chemical Biology*, 4(2), 93–107. <https://doi.org/10.1021/cb8002946>
- Singh, C. R., Lee, B., Udagawa, T., Mohammad-Qureshi, S. S., Yamamoto, Y., Pavitt, G. D., & Asano, K. (2006). An eIF5/eIF2 complex antagonizes guanine nucleotide exchange by eIF2B during translation initiation. *The EMBO Journal*, 25(19), 4537–4546. <https://doi.org/10.1038/sj.emboj.7601339>
- Sivan, G., Aviner, R., & Elroy-Stein, O. (2011). Mitotic Modulation of Translation Elongation Factor 1 Leads to Hindered tRNA Delivery to Ribosomes. *Journal of Biological Chemistry*, 286(32), 27927–27935. <https://doi.org/10.1074/jbc.M111.255810>
- Skabkin, M. A., Skabkina, O. V., Dhote, V., Komar, A. A., Hellen, C. U. T., & Pestova, T. V. (2010). Activities of Ligatin and MCT-1/DENR in eukaryotic translation initiation and ribosomal recycling. *Genes & Development*, 24(16), 1787–1801. <https://doi.org/10.1101/gad.1957510>
- Smith, H. O., & Welcox, K. W. (1970). A Restriction enzyme from Hemophilus influenzae. *Journal of Molecular Biology*, 51(2), 379–391. [https://doi.org/10.1016/0022-2836\(70\)90149-X](https://doi.org/10.1016/0022-2836(70)90149-X)
- Song, H., Mugnier, P., Das, A. K., Webb, H. M., Evans, D. R., Tuite, M. F., Hemmings, B. A., & Barford, D.

- (2000). The Crystal Structure of Human Eukaryotic Release Factor eRF1—Mechanism of Stop Codon Recognition and Peptidyl-tRNA Hydrolysis. *Cell*, 100(3), 311–321. [https://doi.org/10.1016/S0092-8674\(00\)80667-4](https://doi.org/10.1016/S0092-8674(00)80667-4)
- Spahn, C. M., Gomez-Lorenzo, M. G., Grassucci, R. A., Jørgensen, R., Andersen, G. R., Beckmann, R., Penczek, P. A., Ballesta, J. P., & Frank, J. (2004). Domain movements of elongation factor eEF2 and the eukaryotic 80S ribosome facilitate tRNA translocation. *The EMBO Journal*, 23(5), 1008–1019. <https://doi.org/10.1038/sj.emboj.7600102>
- Spriggs, K. A., Bushell, M., & Willis, A. E. (2010). Translational Regulation of Gene Expression during Conditions of Cell Stress. *Molecular Cell*, 40(2), 228–237. <https://doi.org/10.1016/j.molcel.2010.09.028>
- Tajima, S., Lauffer, L., Rath, V. L., & Walter, P. (1986). The signal recognition particle receptor is a complex that contains two distinct polypeptide chains. *The Journal of Cell Biology*, 103(4), 1167–1178. <https://doi.org/10.1083/jcb.103.4.1167>
- Taylor, D. J., Nilsson, J., Merrill, A. R., Andersen, G. R., Nissen, P., & Frank, J. (2007). Structures of modified eEF2-80S ribosome complexes reveal the role of GTP hydrolysis in translocation. *The EMBO Journal*, 26(9), 2421–2431. <https://doi.org/10.1038/sj.emboj.7601677>
- Temperley, R., Richter, R., Dennerlein, S., Lightowlers, R. N., & Chrzanowska-Lightowlers, Z. M. (2010). Hungry Codons Promote Frameshifting in Human Mitochondrial Ribosomes. *Science*, 327(5963), 301–301. <https://doi.org/10.1126/science.1180674>
- Tharmalingam, T., Sunley, K., & Butler, M. (2008). High yields of monomeric recombinant β -interferon from macroporous microcarrier cultures under hypothermic conditions. *Biotechnology Progress*, 24(4), 832–838. <https://doi.org/10.1002/btpr.8>
- Thess, A., Grund, S., Mui, B. L., Hope, M. J., Baumhof, P., Fotin-Mleczek, M., & Schlake, T. (2015). Sequence-engineered mRNA Without Chemical Nucleoside Modifications Enables an Effective Protein Therapy in Large Animals. *Molecular Therapy*, 23(9), 1456–1464. <https://doi.org/10.1038/mt.2015.103>
- Tholstrup, J., Oddershede, L. B., & Sørensen, M. A. (2012). mRNA pseudoknot structures can act as ribosomal roadblocks. *Nucleic Acids Research*, 40(1), 303–313. <https://doi.org/10.1093/nar/gkr686>

- Trabuco, L. G., Schreiner, E., Eargle, J., Cornish, P., Ha, T., Luthey-Schulten, Z., & Schulten, K. (2010). *The role of L1 stalk:tRNA interaction in the ribosome elongation cycle*.
<https://doi.org/10.1016/j.jmb.2010.07.056>
- Tsuboi, T., Kuroha, K., Kudo, K., Makino, S., Inoue, E., Kashima, I., & Inada, T. (2012). Dom34:Hbs1 Plays a General Role in Quality-Control Systems by Dissociation of a Stalled Ribosome at the 3' End of Aberrant mRNA. *Molecular Cell*, 46(4), 518–529. <https://doi.org/10.1016/j.molcel.2012.03.013>
- Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., Pan, T., Dahan, O., Furman, I., & Pilpel, Y. (2010). An Evolutionarily Conserved Mechanism for Controlling the Efficiency of Protein Translation. *Cell*, 141(2), 344–354. <https://doi.org/10.1016/j.cell.2010.03.031>
- Unbehaun, A., Borukhov, S. I., Hellen, C. U. T., & Pestova, T. V. (2004). Release of initiation factors from 48S complexes during ribosomal subunit joining and the link between establishment of codon-anticodon base-pairing and hydrolysis of eIF2-bound GTP. *Genes & Development*, 18(24), 3078–3093. <https://doi.org/10.1101/gad.1255704>
- van den Elzen, A. M. G., Schuller, A., Green, R., & Séraphin, B. (2014). Dom34-Hbs1 mediated dissociation of inactive 80S ribosomes promotes restart of translation after stress. *The EMBO Journal*, n/a-n/a. <https://doi.org/10.1002/emboj.201386123>
- Vélez-Cruz, R., & Johnson, D. (2017). The Retinoblastoma (RB) Tumor Suppressor: Pushing Back against Genome Instability on Multiple Fronts. *International Journal of Molecular Sciences*, 18(8), 1776. <https://doi.org/10.3390/ijms18081776>
- Vicens, Q., Kieft, J. S., & Rissland, O. S. (2018). Revisiting the Closed-Loop Model and the Nature of mRNA 5'–3' Communication. *Molecular Cell*, 72(5), 805–812. <https://doi.org/10.1016/j.molcel.2018.10.047>
- von Heijne, G. (1985). Signal sequences. *Journal of Molecular Biology*, 184(1), 99–105. [https://doi.org/10.1016/0022-2836\(85\)90046-4](https://doi.org/10.1016/0022-2836(85)90046-4)
- Wahl, M. C., & Moller, W. (2002). Structure and Function of the Acidic Ribosomal Stalk Proteins. *Current Protein and Peptide Science*, 3(1), 93–106. <https://doi.org/10.2174/1389203023380756>
- Walden, W. E., & Thach, R. E. (1986). Translational control of gene expression in a normal fibroblast. Characterization of a subclass of mRNAs with unusual kinetic properties. *Biochemistry*, 25(8),

- 2033–2041. <https://doi.org/10.1021/bi00356a030>
- Walsh, G., & Walsh, E. (2022). Biopharmaceutical benchmarks 2022. *Nature Biotechnology*, 40(12), 1722–1760. <https://doi.org/10.1038/s41587-022-01582-x>
- Walter, P., & Blobel, G. (1981). Translocation of proteins across the endoplasmic reticulum III. Signal recognition protein (SRP) causes signal sequence-dependent and site-specific arrest of chain elongation that is released by microsomal membranes. *The Journal of Cell Biology*, 91(2), 557–561. <https://doi.org/10.1083/jcb.91.2.557>
- Wang, Z., Jiao, X., Carr-Schmid, A., & Kiledjian, M. (2002). The hDcp2 protein is a mammalian mRNA decapping enzyme. *Proceedings of the National Academy of Sciences*, 99(20), 12663–12668. <https://doi.org/10.1073/pnas.192445599>
- Weisser, M., & Ban, N. (2019). Extensions, Extra Factors, and Extreme Complexity: Ribosomal Structures Provide Insights into Eukaryotic Translation. *Cold Spring Harbor Perspectives in Biology*, 11(9), a032367. <https://doi.org/10.1101/cshperspect.a032367>
- Wells, S. E., Hillner, P. E., Vale, R. D., & Sachs, A. B. (1998). Circularization of mRNA by eukaryotic translation initiation factors. *Mol Cell*, 2(1), 135–140. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9702200
- Wilson, W., Braddock, M., Adams, S. E., Rathjen, P. D., Kingsman, S. M., & Kingsman, A. J. (1988). HIV expression strategies: Ribosomal frameshifting is directed by a short sequence in both mammalian and yeast systems. *Cell*, 55(6), 1159–1169. [https://doi.org/10.1016/0092-8674\(88\)90260-7](https://doi.org/10.1016/0092-8674(88)90260-7)
- Wiśniewski, J. R., Hein, M. Y., Cox, J., & Mann, M. (2014). A “Proteomic Ruler” for Protein Copy Number and Concentration Estimation without Spike-in Standards. *Molecular & Cellular Proteomics*, 13(12), 3497–3506. <https://doi.org/10.1074/mcp.M113.037309>
- Wohlgemuth, I., Beringer, M., & Rodnina, M. V. (2006). Rapid peptide bond formation on isolated 50S ribosomal subunits. *EMBO Reports*, 7(7), 699–703. <https://doi.org/10.1038/sj.embor.7400732>
- Wohlgemuth, I., Brenner, S., Beringer, M., & Rodnina, M. V. (2008). Modulation of the Rate of Peptidyl Transfer on the Ribosome by the Nature of Substrates. *Journal of Biological Chemistry*, 283(47), 32229–32235. <https://doi.org/10.1074/jbc.M805316200>

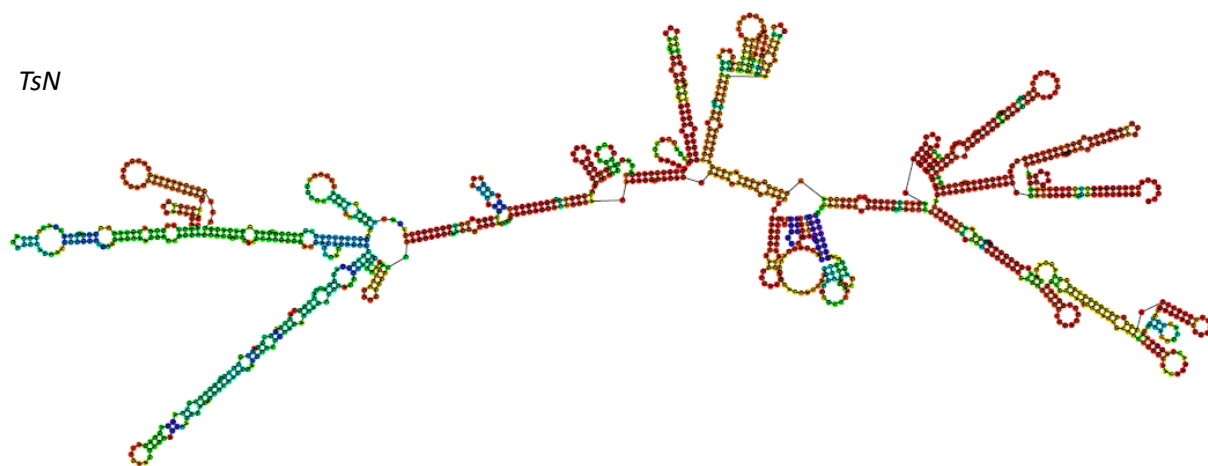
- Woods, A., Sherwin, T., Sasse, R., Macrae, T. H., Baines, A. J., & Gull, K. (1989). Definition of individual components within the cytoskeleton of *Trypanosoma brucei* by a library of monoclonal antibodies. *Journal of Cell Science*, 93(3), 491–500. <https://doi.org/10.1242/jcs.93.3.491>
- Wu, Q., & Bazzini, A. A. (2018). Systems to study codon effect on post-transcriptional regulation of gene expression. *Methods*, 137, 82–89. <https://doi.org/10.1016/j.ymeth.2017.11.006>
- Wu, Q., Medina, S. G., Kushawah, G., DeVore, M. L., Castellano, L. A., Hand, J. M., Wright, M., & Bazzini, A. A. (2019). Translation affects mRNA stability in a codon-dependent manner in human cells. *ELife*, 8, 1–22. <https://doi.org/10.7554/eLife.45396>
- Wurm, J. P., & Sprangers, R. (2019). Dcp2: an mRNA decapping enzyme that adopts many different shapes and forms. *Current Opinion in Structural Biology*, 59, 115–123. <https://doi.org/10.1016/j.sbi.2019.07.009>
- Xu, J., Tang, P., Yongky, A., Drew, B., Borys, M. C., Liu, S., & Li, Z. J. (2019). Systematic development of temperature shift strategies for Chinese hamster ovary cells based on short duration cultures and kinetic modeling. *MAbs*, 11(1), 191–204. <https://doi.org/10.1080/19420862.2018.1525262>
- Young, D. J., Gydosh, N. R., Zhang, F., Hinnebusch, A. G., & Green, R. (2015). Rli1/ABCE1 Recycles Terminating Ribosomes and Controls Translation Reinitiation in 3'UTRs In Vivo. *Cell*, 162(4), 872–884. <https://doi.org/10.1016/j.cell.2015.07.041>
- Zakin, L., & De Robertis, E. M. (2010). Extracellular regulation of BMP signaling. *Current Biology*, 20(3), R89–R92. <https://doi.org/10.1016/j.cub.2009.11.021>
- Zheng, C., & Baum, B. J. (2008). Evaluation of Promoters for Use in Tissue-Specific Gene Delivery. In *Gene Therapy Protocols* (pp. 205–219). Humana Press. https://doi.org/10.1007/978-1-60327-248-3_13
- Zhong, P., & Huang, H. (2017). Recent progress in the research of cold-inducible RNA-binding protein. *Future Science OA*, 3(4), FSO246. <https://doi.org/10.4155/fsoa-2017-0077>
- Zimmerman, L. B., De Jesús-Escobar, J. M., & Harland, R. M. (1996). The Spemann Organizer Signal noggin Binds and Inactivates Bone Morphogenetic Protein 4. *Cell*, 86(4), 599–606. [https://doi.org/10.1016/S0092-8674\(00\)80133-6](https://doi.org/10.1016/S0092-8674(00)80133-6)

Appendices

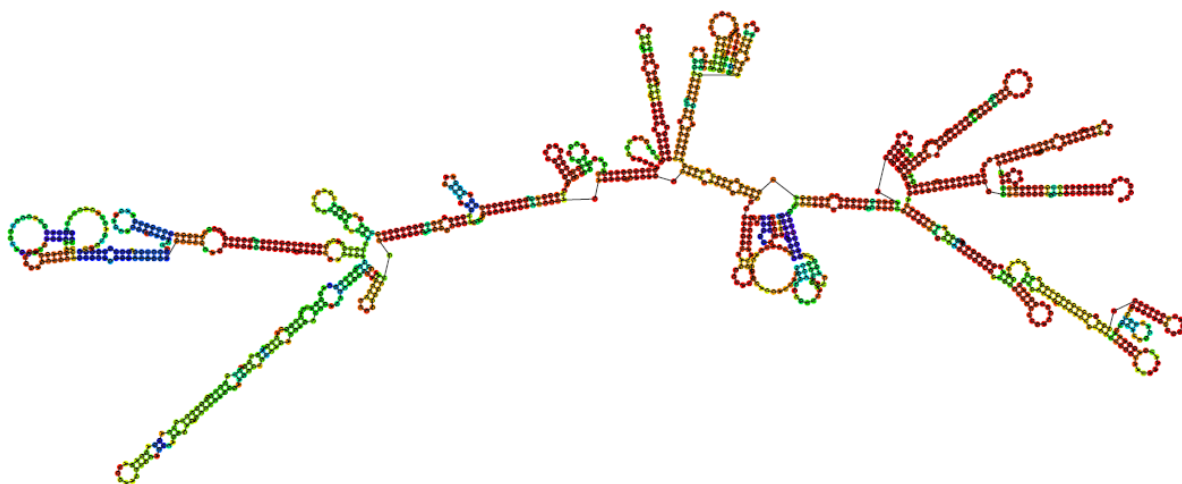
7.1 *CCND3* mRNA Secondary Structures

Ts Variants

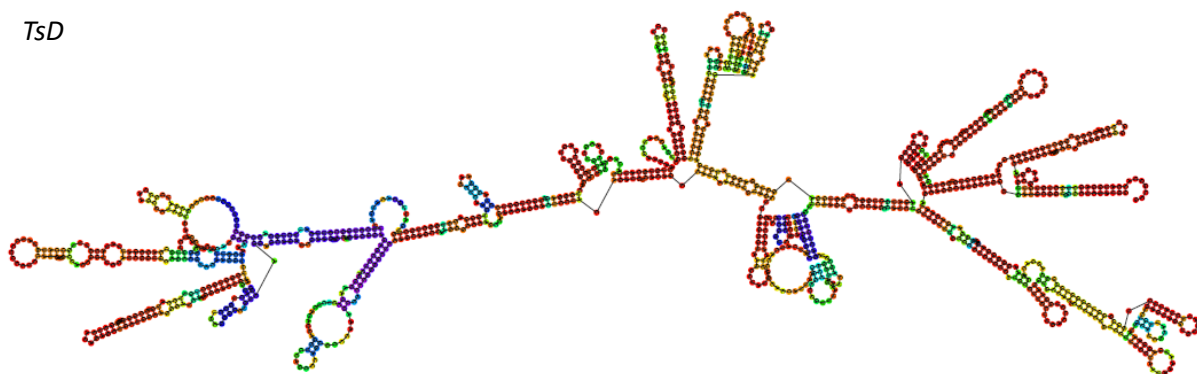
TsN



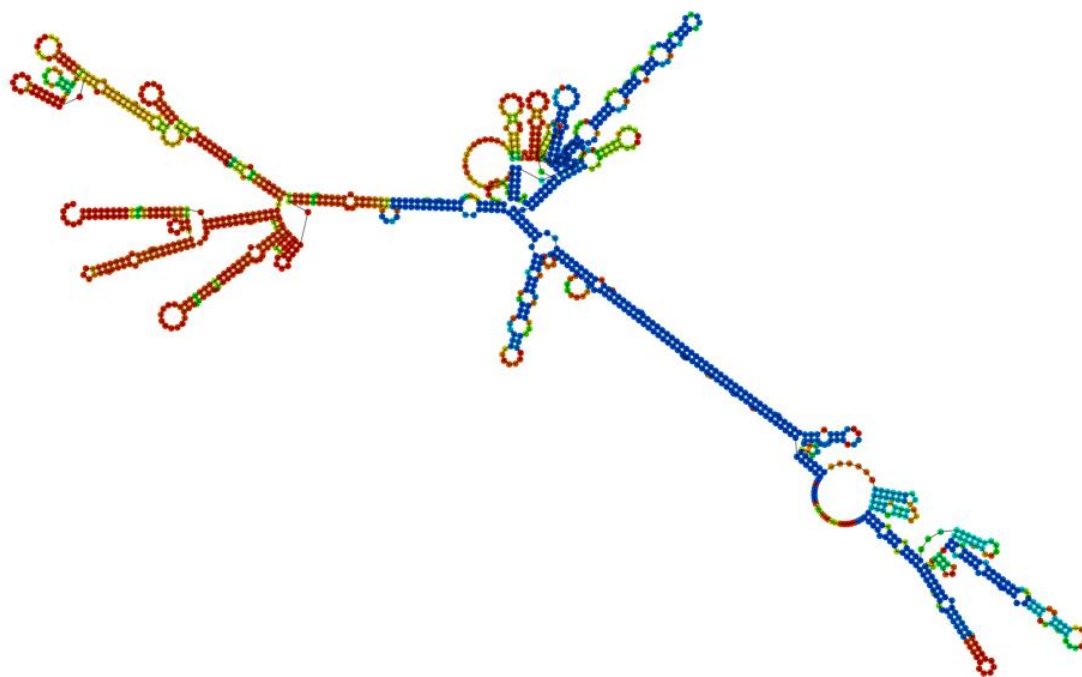
TsO



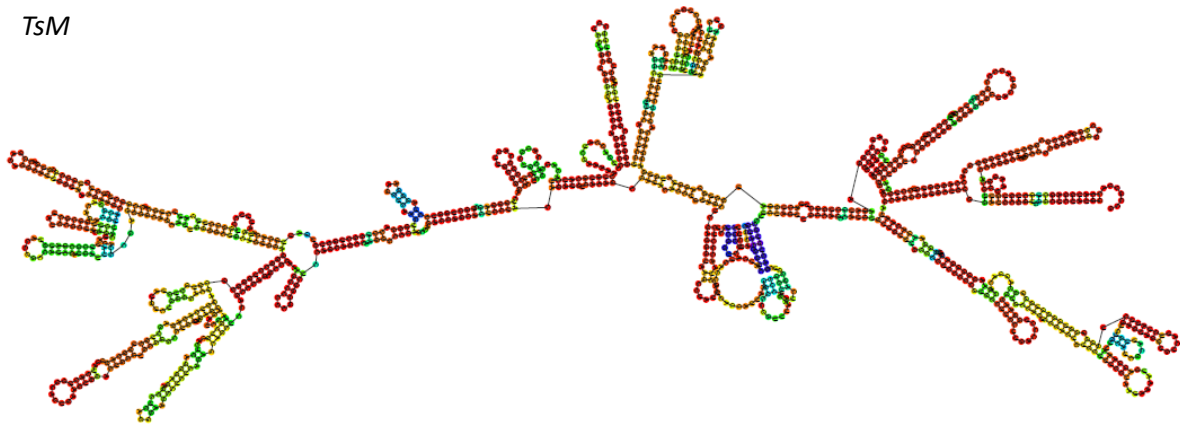
TsD



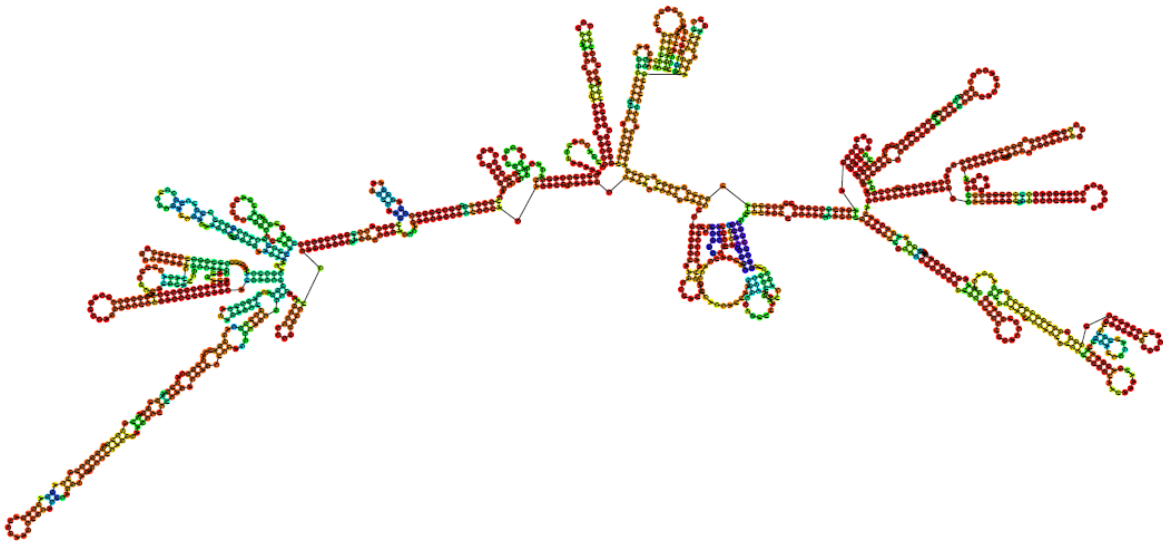
TsF



TsM

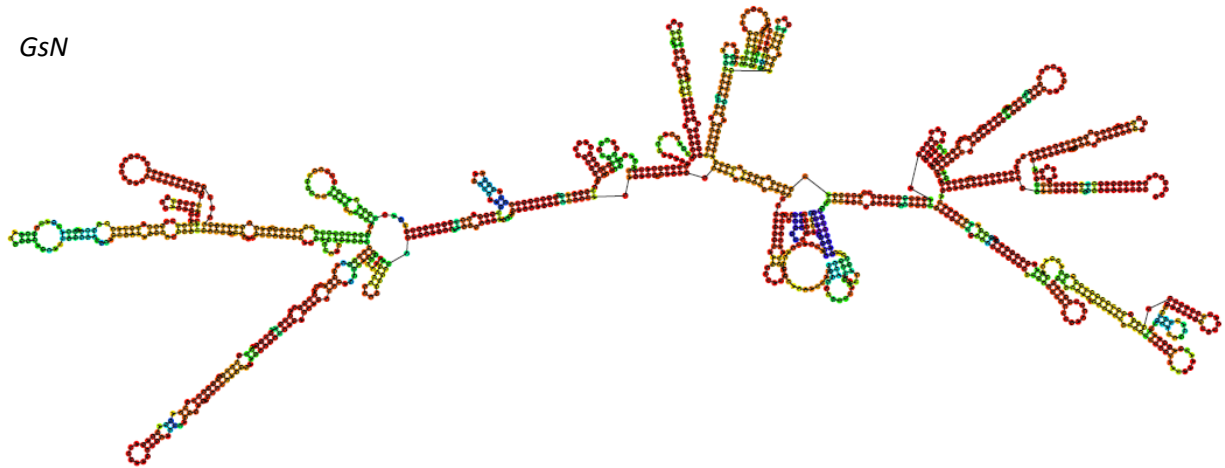


TsS

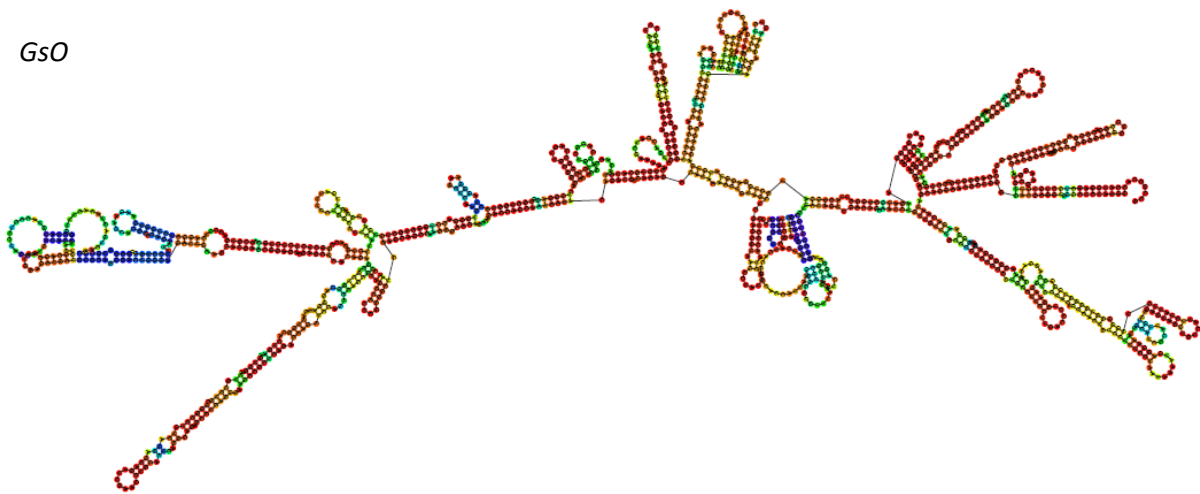


Gs Variants

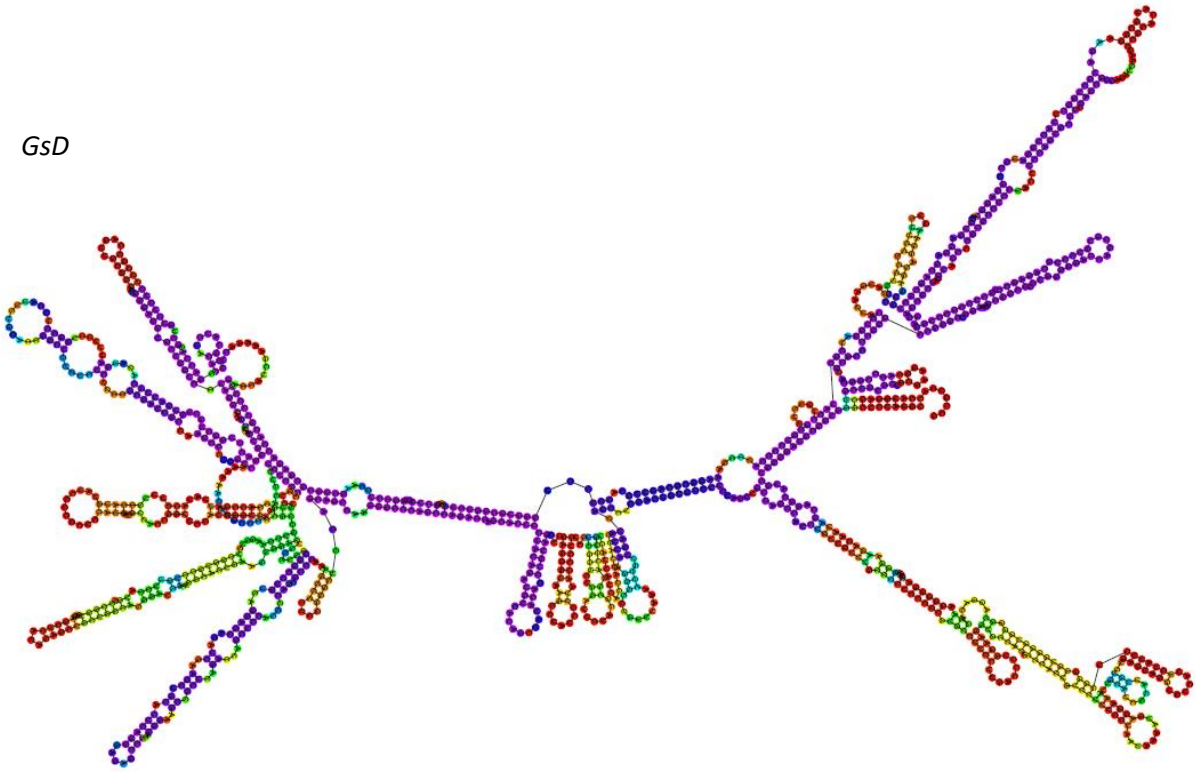
GsN



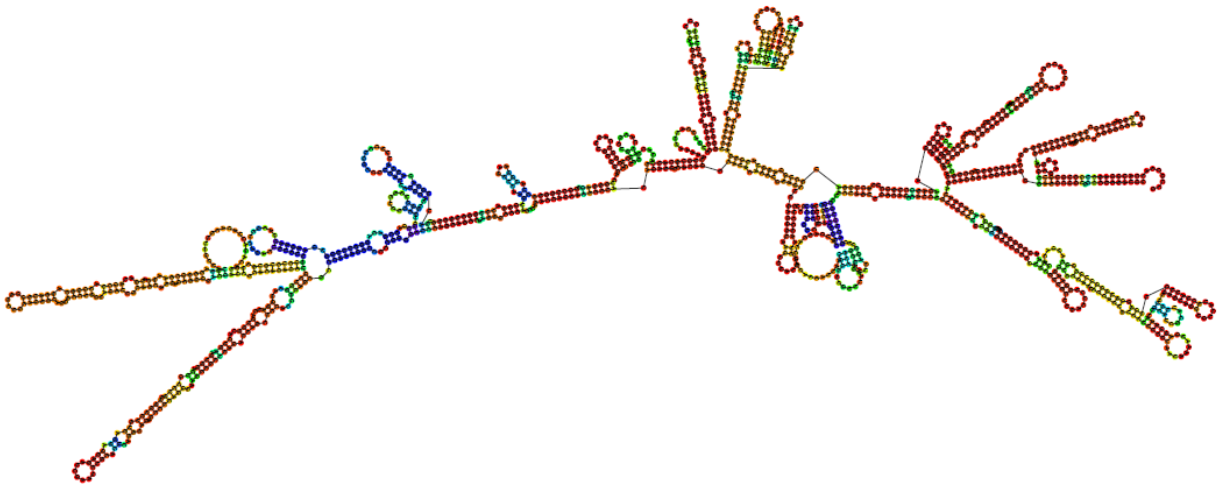
GsO



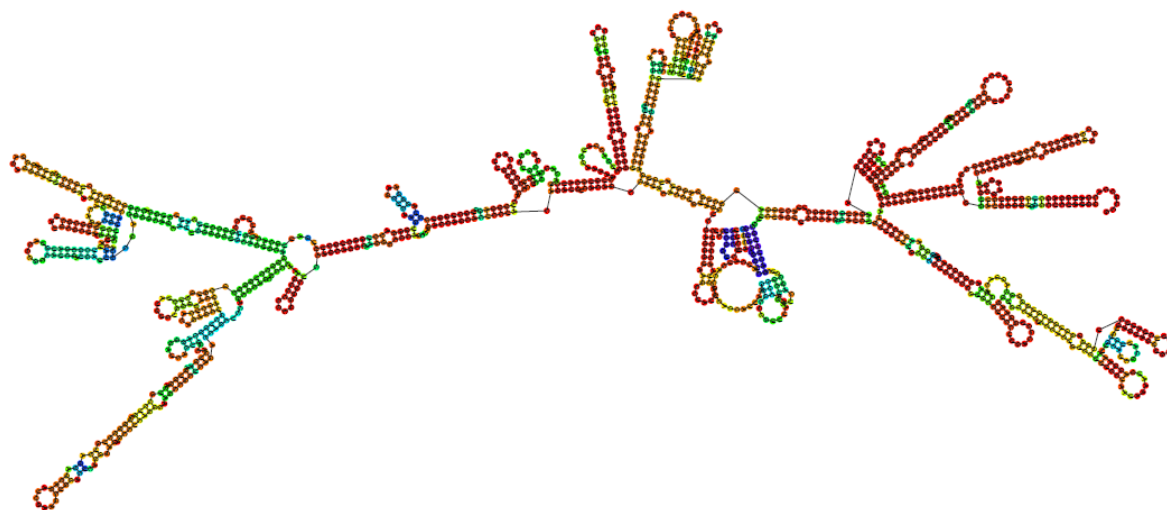
GsD



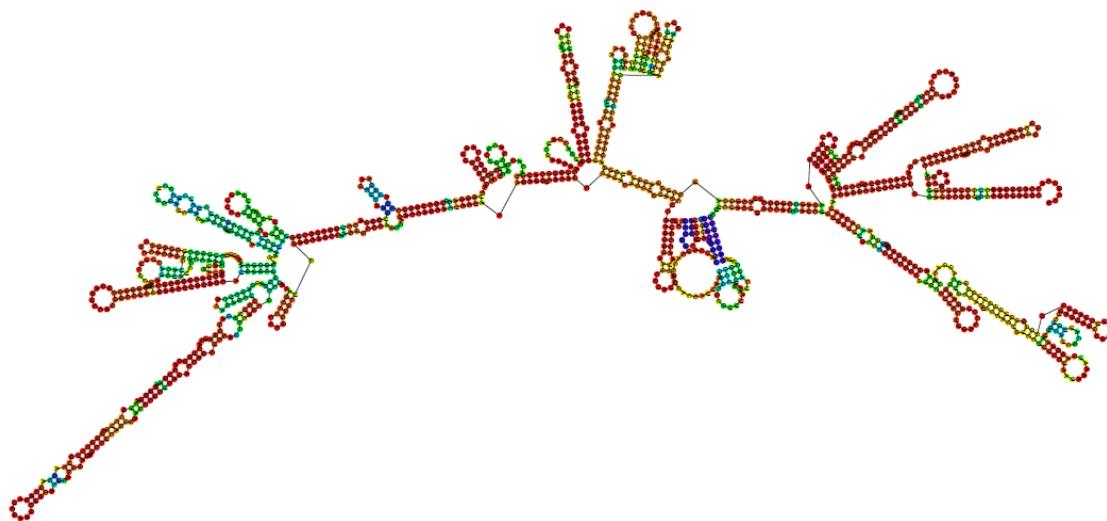
GsF



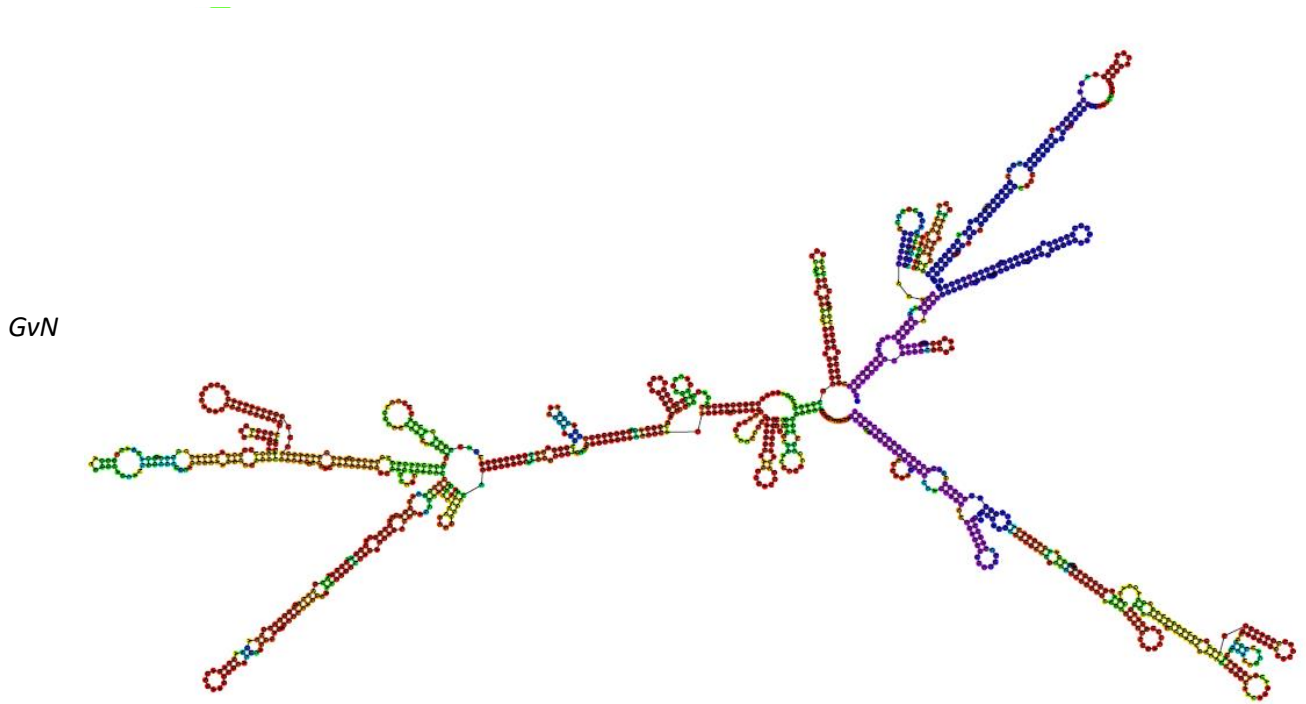
GsM



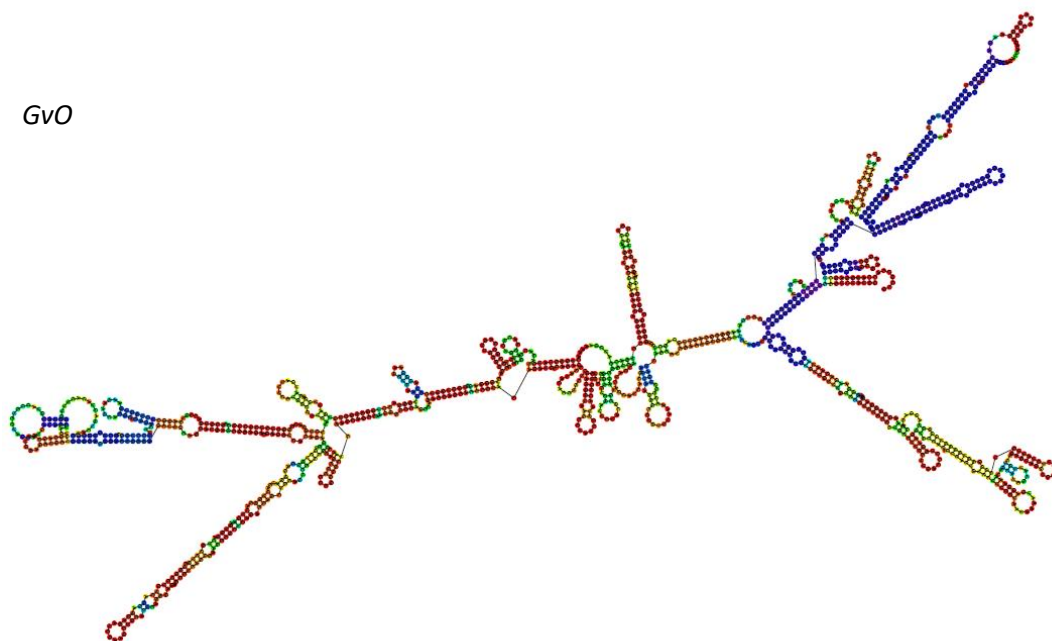
GsS



Gv Variants

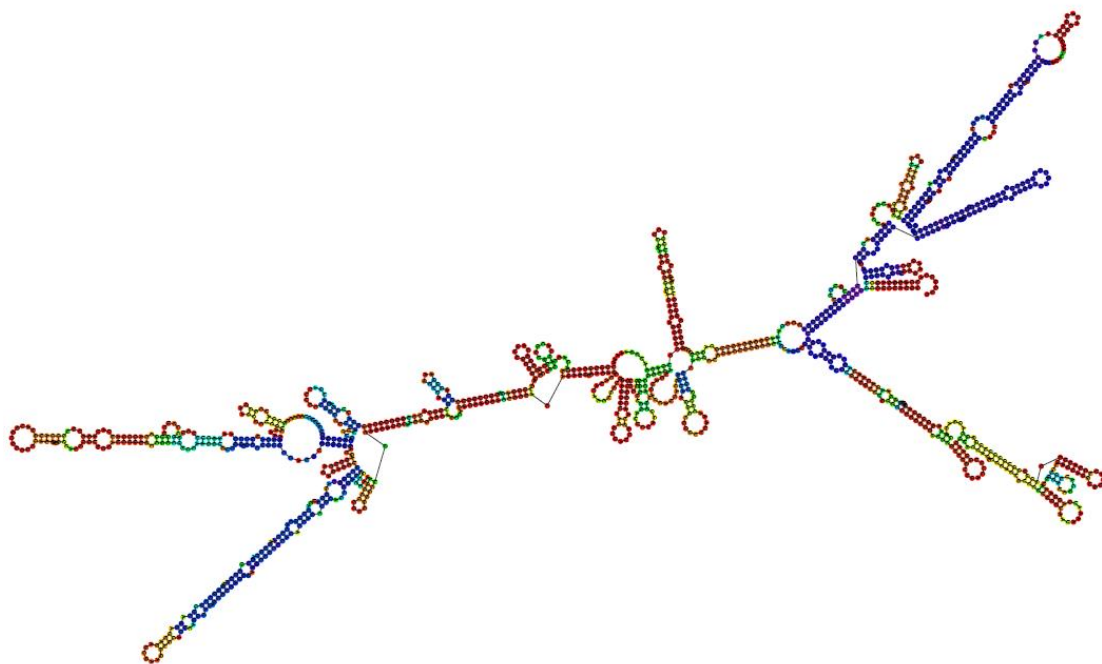


GvN

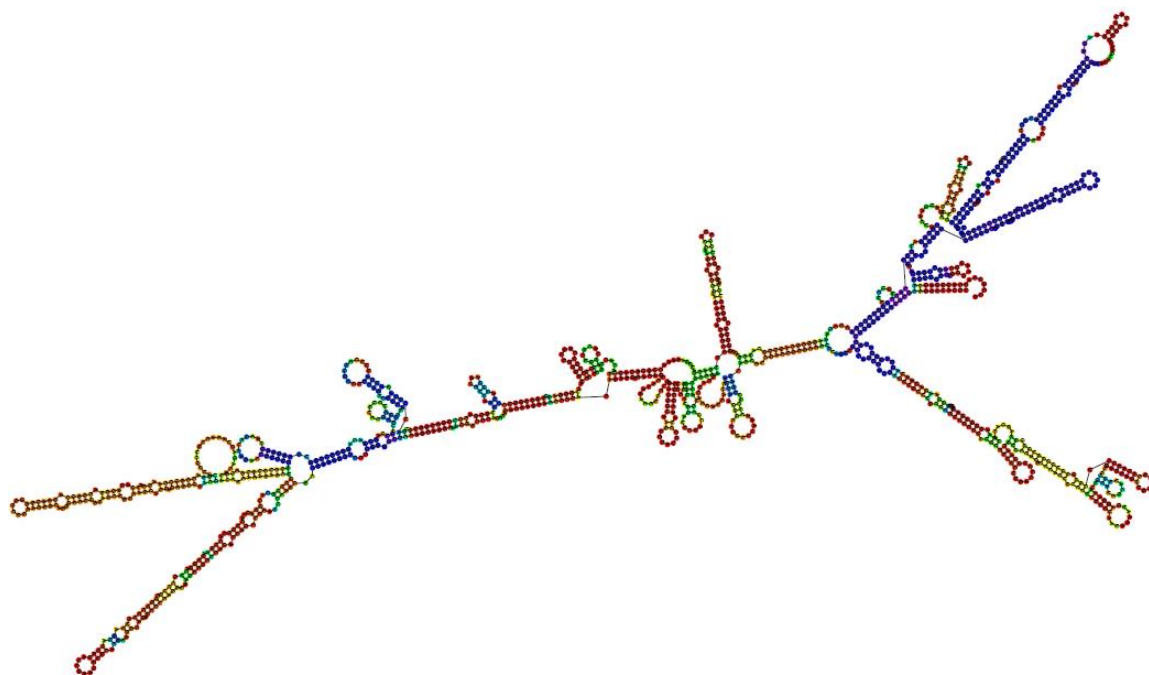


GvO

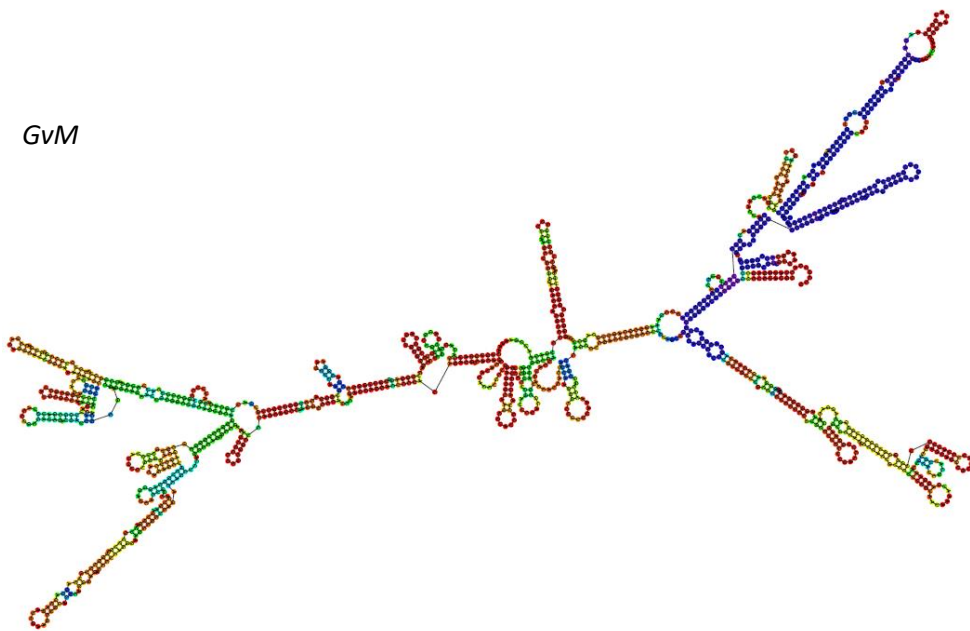
GvD



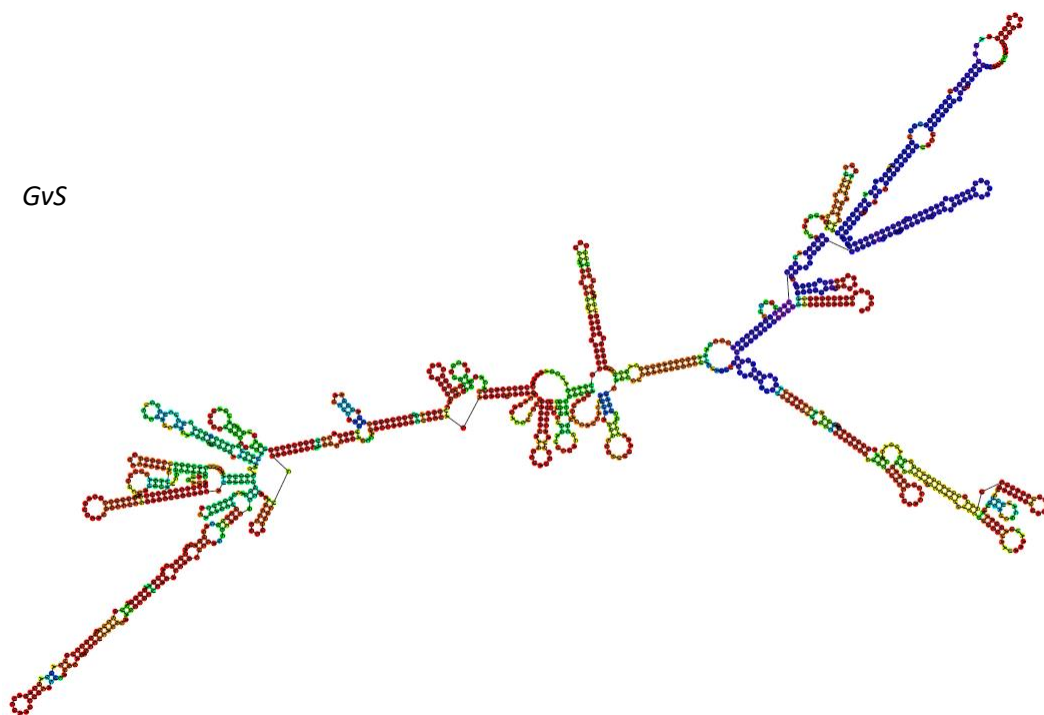
GvF



GvM



GvS



7.2 Gene Sequences

7.2.1 *CCND3* Sequences

Tv Variants

TvN

ATGGAGCTGCTGTGTTGCGAAGGCACCCGGCACGCGCCCCGGGCGGGCCGGACCCGCGGCTGCTGGGGGACCA
GCGTGTCTGCAGAGCCTGCTCCGCCTGGAGGAGCGCTACGTACCCCGCGCCTCCTACTTCCAGTGCGTGCAGCGG
GAGATCAAGCCGCACATGCGGAAGATGCTGGCTTACTGGATGCTGGAGGTATGTGAGGAGCAGCGCTGTGAGGAG
GAAGTCTTCCCCCTGGCCATGAACTACCTGGATCGCTACCTGTCTTTCGTCCCCACCCGAAAGGCGCAGTTGCAGCT
CCTGGGTGCGGTCTGCATGCTGCTGGCCTCCAAGCTGCGCGAGACCACGCCCCTGACCATCGAAAACTGTGCATCT
ACACCGACCACGCTGTCTCTCCCCGCCAGTTGCGGGACTGGGAGGTGCTGGTCCTAGGGAAGCTCAAGTGGGACCT
GGCTGCTGTGATTGCACATGATTTCTGGCCTTCATTCTGCACCGGCTCTCTTGCCCCGTGACCGACAGGCCCTTGGT
CAAAAAGCATGCCAGACCTTTTTGGCCCTCTGTGCTACAGATTATACCTTTGCCATGTACCCGCCATCCATGATGCC
ACGGGCAGCATTGGGGCTGCAGTGCAAGGCCTGGGTGCCTGCTCCATGTCCGGGGATGAGCTCACAGAGCTGCTG
GCAGGGATCACTGGCACTGAAGTGGACTGCCTGCGGGCCTGTCAGGAGCAGATCGAAGCTGCACTCAGGGAGAGC
CTCAGGGAAGCCTCTCAGACCAGCTCCAGCCCAGCGCCCAAAGCCCCCGGGGCTCCAGCAGCCAAGGGCCCAGC
CAGACCAGCACTCCTACAGATGTCACAGCCATACACCTGCTCGAGGGTAAGCCTATCCCTAACCTCTCCTCGGTCTC
GATTCTACGCGTACCGGTCATCATCACCATCACCAT

TvO

ATGGAGCTGCTGTGTTGCGAAGGCACCCGGCACGCGCCCCGGGCGGGCCGGACCCGCGGCTGCTGGGGGACCA
GCGTGTCTGCAGAGCCTGCTCCGCCTGGAGGAGCGCTACGTACCCCGCGCCTCCTACTTCCAGTGCGTGCAGCGG
GAGATCAAGCCGCACATGCGGAAGATGCTGGCTTACTGGATGCTGGAGGTATGTGAGGAGCAGCGCTGTGAGGAG
GAAGTCTTCCCCCTGGCCATGAACTACCTGGATCGCTACCTGTCTTTCGTCCCCACCCGAAAGGCGCAGTTGCAGCT
CCTGGGTGCGGTCTGCATGCTGCTGGCCTCCAAGCTGCGCGAGACCACGCCCCTGACCATCGAAAACTGTGCATCT
ACACCGACCACGCTGTCTCTCCCCGCCAGTTGCGGGACTGGGAGGTGCTGGTCCTAGGGAAGCTCAAGTGGGACCT
GGCTGCTGTGATTGCACATGATTTCTGGCCTTCATTCTGCACCGGCTCTCTTGCCCTCGTGACCGTCAGGCTCTGGT
GAAGAAGCACGCTCAGACTTTCCTGGCTCTGTGCGCTACTGACTACACTTTCGCTATGTACCCTCCTAGCATGATAGCT
ACTGGCAGCATTGGAGCTGCTGTGCAGGGACTGGGAGCTTGCAGCATGAGCGGAGACGAGCTCACAGAGCTGCTG
GCAGGGATCACTGGCACTGAAGTGGACTGCCTGCGGGCCTGTCAGGAGCAGATCGAAGCTGCACTCAGGGAGAGC
CTCAGGGAAGCCTCTCAGACCAGCTCCAGCCCAGCGCCCAAAGCCCCCGGGGCTCCAGCAGCCAAGGGCCCAGC
CAGACCAGCACTCCTACAGATGTCACAGCCATACACCTGCTCGAGGGTAAGCCTATCCCTAACCTCTCCTCGGTCTC
GATTCTACGCGTACCGGTCATCATCACCATCACCAT

TvD

ATGGAGCTGCTGTGTTGCGAAGGCACCCGGCACGCGCCCCGGGCGGGCCGGACCCGCGGCTGCTGGGGGACCA
GCGTGTCTGCAGAGCCTGCTCCGCCTGGAGGAGCGCTACGTACCCCGCGCCTCCTACTTCCAGTGCGTGCAGCGG
GAGATCAAGCCGCACATGCGGAAGATGCTGGCTTACTGGATGCTGGAGGTATGTGAGGAGCAGCGCTGTGAGGAG
GAAGTCTTCCCCCTGGCCATGAACTACCTGGATCGCTACCTGTCTTTCGCTCCCCACCCGAAAGGCGCAGTTGCAGCT
CCTGGGTGCGGTCTGCATGCTGCTGGCCTCCAAGCTGCGCGAGACCACGCCCCTGACCATCGAAAACTGTGCATCT
ACACCGACCACGCTGTCTCTCCCCGCCAGTTGCGGGACTGGGAGGTGCTGGTCCTAGGGAAGCTCAAGTGGGACCT
GGCTGCTGTGATTGCACATGATTTCTGGCCTTCATTCTGCACCGGCTCTCTCTCCCCCGCGATCGCCAAGCCCTCGTC
AAAAACATGCCCAAACCTTTCTCGCCCTCTGTGCCACCGATTATACCTTTGCCATGTATCCCCCTCCATGATCGCCAC
CGGTTCCATCGGTGCCGCCGTCCAAGGTCTCGGTGCCTGTTCCATGTCCGGTGATGAGCTCACAGAGCTGCTGGCAG
GGATCACTGGCACTGAAGTGGACTGCCTGCGGGCCTGTCAGGAGCAGATCGAAGCTGCACTCAGGGAGAGCCTCA
GGGAAGCCTCTCAGACCAGCTCCAGCCCAGCGCCCAAAGCCCCCGGGGCTCCAGCAGCCAAGGGCCCCAGCCAGA
CCAGCACTCTACAGATGTACAGCCATACACCTGCTCGAGGGTAAGCCTATCCCTAACCTCTCCTCGGTCTCGATT
TACGCGTACCGGTCATCATCACCATCACCAT

TvF

ATGGAGCTGCTGTGTTGCGAAGGCACCCGGCACGCGCCCCGGGCGGGCCGGACCCGCGGCTGCTGGGGGACCA
GCGTGTCTGCAGAGCCTGCTCCGCCTGGAGGAGCGCTACGTACCCCGCGCCTCCTACTTCCAGTGCGTGCAGCGG
GAGATCAAGCCGCACATGCGGAAGATGCTGGCTTACTGGATGCTGGAGGTATGTGAGGAGCAGCGCTGTGAGGAG
GAAGTCTTCCCCCTGGCCATGAACTACCTGGATCGCTACCTGTCTTTCGCTCCCCACCCGAAAGGCGCAGTTGCAGCT
CCTGGGTGCGGTCTGCATGCTGCTGGCCTCCAAGCTGCGCGAGACCACGCCCCTGACCATCGAAAACTGTGCATCT
ACACCGACCACGCTGTCTCTCCCCGCCAGTTGCGGGACTGGGAGGTGCTGGTCCTAGGGAAGCTCAAGTGGGACCT
GGCTGCTGTGATTGCACATGATTTCTGGCCTTCATTCTGCACCGGCTCTCTCTGCCTCGTGACCGACAGGCTTTGGT
GAAGAAGCACGCTCAGACTTTCTTGCTCTGTGCGCTACAGACTACACTTTTCGCTATGTACCCGCCAAGCATGATAGC
TACGGGCAGCATTGGGGCTGCAGTGAGGGCCTGGGAGCTTGAGCATGAGCGGGGACGAGCTCACAGAGCTGC
TGGCAGGGATCACTGGCACTGAAGTGGACTGCCTGCGGGCCTGTCAGGAGCAGATCGAAGCTGCACTCAGGGAGA
GCCTCAGGGAAGCCTCTCAGACCAGCTCCAGCCCAGCGCCCAAAGCCCCCGGGGCTCCAGCAGCCAAGGGCCCCA
GCCAGACCAGCACTCTACAGATGTACAGCCATACACCTGCTCGAGGGTAAGCCTATCCCTAACCTCTCCTCGGTC
TCGATTCTACGCGTACCGGTCATCATCACCATCACCAT

TvM

ATGGAGCTGCTGTGTTGCGAAGGCACCCGGCACGCGCCCCGGGCGGGCCGGACCCGCGGCTGCTGGGGGACCA
GCGTGTCTGCAGAGCCTGCTCCGCCTGGAGGAGCGCTACGTACCCCGCGCCTCCTACTTCCAGTGCGTGCAGCGG
GAGATCAAGCCGCACATGCGGAAGATGCTGGCTTACTGGATGCTGGAGGTATGTGAGGAGCAGCGCTGTGAGGAG
GAAGTCTTCCCCCTGGCCATGAACTACCTGGATCGCTACCTGTCTTTCGCTCCCCACCCGAAAGGCGCAGTTGCAGCT
CCTGGGTGCGGTCTGCATGCTGCTGGCCTCCAAGCTGCGCGAGACCACGCCCCTGACCATCGAAAACTGTGCATCT
ACACCGACCACGCTGTCTCTCCCCGCCAGTTGCGGGACTGGGAGGTGCTGGTCCTAGGGAAGCTCAAGTGGGACCT
GGCTGCTGTGATTGCACATGATTTCTGGCCTTCATTCTGCACCGGCTCTCTCTGCCGCGTGACCGACAGGCATTGGT
AAAGAAGCACGCACAGACGTTCTTGCGCTATGCGCTACAGACTACACATTTCGCGATGTACCCGCCATCGATGATTGC
GACGGGCAGCATTGGGGCTGCAGTGAGGGCCTGGGCGCGTGCTCTATGTCTGGGGACGAGCTCACAGAGCTGCT

GGCAGGGATCACTGGCACTGAAGTGGA CTGCCTGCGGGCCTGTCAGGAGCAGATCGAAGCTGCACTCAGGGAGA
GCCTCAGGGAAGCCGCTCAGACCAGCTCCAGCCCAGCGCCCAAAGCCCCCGGGGCTCCAGCAGCCAAGGGCCCA
GCCAGACCAGCACTCCTACAGATGTCACAGCCATACACCTGCTCGAGGGTAAGCCTATCCCTAACCTCTCCTCGGTC
TCGATTCTACGCGTACCGGTCATCATCACCATCACCAT

TvS

ATGGAGCTGCTGTGTTGCGAAGGCACCCGGCACGCGCCCCGGGCGGGCCGGACCCGCGGCTGCTGGGGGACCA
GCGTGTCCTGCAGAGCCTGCTCCGCCTGGAGGAGCGCTACGTACCCCGCGCCTCCTACTTCCAGTGCGTGCAGCGG
GAGATCAAGCCGCACATGCGGAAGATGCTGGCTTACTGGATGCTGGAGGTATGTGAGGAGCAGCGCTGTGAGGAG
GAAGTCTTCCCCCTGGCCATGAACTACCTGGATCGCTACCTGTCTTTCGTCCCCACCCGAAAGGCGCAGTTGCAGCT
CCTGGGTGCGGTCTGCATGCTGCTGGCCTCCAAGCTGCGCGAGACCACGCCCCTGACCATCGAAAACTGTGCATCT
ACACCGACCACGCTGTCTCTCCCCGCCAGTTGCGGGACTGGGAGGTGCTGGTCCTAGGGAAGCTCAAGTGGGACCT
GGCTGCTGTGATTGCACATGATTTCCTGGCCTTCATTCTGCACCGGCTCTCTCTCCCTCGCGATCGACAAGCTTTGGTG
AAGAAACACGCTCAAACCTTTCTTGCTCTGTGCGCCACAGACTACACTTTTCGTATGTATCCGCCAAGCATGATAGCT
ACGGGCTCCATCGGGGCGCAGTCCAGGGCCTCGGAGCTTGTAGCATGAGCGGGGACGAGCTCACAGAGCTGCTG
GCAGGGATCACTGGCACTGAAGTGGA CTGCCTGCGGGCCTGTCAGGAGCAGATCGAAGCTGCACTCAGGGAGAGC
CTCAGGGAAGCCTCTCAGACCAGCTCCAGCCCAGCGCCCAAAGCCCCCGGGGCTCCAGCAGCCAAGGGCCCCAGC
CAGACCAGCACTCCTACAGATGTCACAGCCATACACCTGCTCGAGGGTAAGCCTATCCCTAACCTCTCCTCGGTCTC
GATTCTACGCGTACCGGTCATCATCACCATCACCAT

Ts Variants

TsN

ATGGAGCTGCTGTGTTGCGAAGGCACCCGGCACGCGCCCCGGGCGGGCCGGACCCGCGGCTGCTGGGGGACCA
GCGTGTCCTGCAGAGCCTGCTCCGCCTGGAGGAGCGCTACGTACCCCGCGCCTCCTACTTCCAGTGCGTGCAGCGG
GAGATCAAGCCGCACATGCGGAAGATGCTGGCTTACTGGATGCTGGAGGTATGTGAGGAGCAGCGCTGTGAGGAG
GAAGTCTTCCCCCTGGCCATGAACTACCTGGATCGCTACCTGTCTTTCGTCCCCACCCGAAAGGCGCAGTTGCAGCT
CCTGGGTGCGGTCTGCATGCTGCTGGCCTCCAAGCTGCGCGAGACCACGCCCCTGACCATCGAAAACTGTGCATCT
ACACCGACCACGCTGTCTCTCCCCGCCAGTTGCGGGACTGGGAGGTGCTGGTCCTAGGGAAGCTCAAGTGGGACCT
GGCTGCTGTGATTGCACATGATTTCCTGGCCTTCATTCTGCACCGGCTCTCTTGCCCCGTGACCGACAGGCCTTGGT
CAAAAAGCATGCCAGACCTTTTTGGCCCTCTGTGCTACAGATTATACCTTTGCCATGTACCCGCCATCCATGATCGCC
ACGGGCAGCATTGGGGCTGCAGTGCAAGGCCTGGGTGCCTGCTCCATGTCCGGGGATGAGCTCACAGAGCTGCTG
GCAGGGATCACTGGCACTGAAGTGGA CTGCCTGCGGGCCTGTCAGGAGCAGATCGAAGCTGCACTCAGGGAGAGC
CTCAGGGAAGCCTCTCAGACCAGCTCCAGCCCAGCGCCCAAAGCCCCCGGGGCTCCAGCAGCCAAGGGCCCCAGC
CAGACCAGCACTCCTACAGATGTCACAGCCATACACCTGTAG

TsO

ATGGAGCTGCTGTGTTGCGAAGGCACCCGGCACGCGCCCCGGGCCGGGCCGGACCCGCGGCTGCTGGGGGACCA
GCGTGTCCTGCAGAGCCTGCTCCGCCTGGAGGAGCGCTACGTACCCCGCGCCTCCTACTTCCAGTGCGTGCAGCGG
GAGATCAAGCCGCACATGCGGAAGATGCTGGCTTACTGGATGCTGGAGGTATGTGAGGAGCAGCGCTGTGAGGAG
GAAGTCTTCCCCCTGGCCATGAACTACCTGGATCGCTACCTGTCTTTCGTCCCCACCCGAAAGGCGCAGTTGCAGCT
CCTGGGTGCGGTCTGCATGCTGCTGGCCTCCAAGCTGCGCGAGACCACGCCCCTGACCATCGAAAACTGTGCATCT
ACACCGACCACGCTGTCTCTCCCCGCCAGTTGCGGGACTGGGAGGTGCTGGTCCTAGGGAAGCTCAAGTGGGACCT
GGCTGCTGTGATTGCACATGATTTCTGGCCTTATTCTGCACCGGCTCTCTCTGCCTCGTGACCGTCAGGCTCTGGT
GAAGAAGCACGCTCAGACTTTCCTGGCTCTGTGCGCTACTGACTACACTTTCGCTATGTACCCTCCTAGCATGATAGCT
ACTGGCAGCATTGGAGCTGCTGTGCAGGGACTGGGAGCTTGAGCATGAGCGGAGACGAGCTCACAGAGCTGCTG
GCAGGGATCACTGGCACTGAAGTGGACTGCCTGCGGGCCTGTCAGGAGCAGATCGAAGCTGCACTCAGGGAGAGC
CTCAGGGAAGCCTCTCAGACCAGCTCCAGCCCAGCGCCCAAAGCCCCCGGGGCTCCAGCAGCCAAGGGCCCAGC
CAGACCAGCACTCCTACAGATGTCACAGCCATACACCTGTAG

TsD

ATGGAGCTGCTGTGTTGCGAAGGCACCCGGCACGCGCCCCGGGCCGGGCCGGACCCGCGGCTGCTGGGGGACCA
GCGTGTCCTGCAGAGCCTGCTCCGCCTGGAGGAGCGCTACGTACCCCGCGCCTCCTACTTCCAGTGCGTGCAGCGG
GAGATCAAGCCGCACATGCGGAAGATGCTGGCTTACTGGATGCTGGAGGTATGTGAGGAGCAGCGCTGTGAGGAG
GAAGTCTTCCCCCTGGCCATGAACTACCTGGATCGCTACCTGTCTTTCGTCCCCACCCGAAAGGCGCAGTTGCAGCT
CCTGGGTGCGGTCTGCATGCTGCTGGCCTCCAAGCTGCGCGAGACCACGCCCCTGACCATCGAAAACTGTGCATCT
ACACCGACCACGCTGTCTCTCCCCGCCAGTTGCGGGACTGGGAGGTGCTGGTCCTAGGGAAGCTCAAGTGGGACCT
GGCTGCTGTGATTGCACATGATTTCTGGCCTTATTCTGCACCGGCTCTCTCTCCCCCGCGATCGCCAAGCCCTCGTC
AAAAACATGCCCAAACCTTTCTCGCCCTCTGTGCCACCGATTATACCTTTGCCATGTATCCCCCTCCATGATCGCCAC
CGGTTCCATCGGTGCCGCCGTCCAAGGTCTCGGTGCCTGTTCCATGTCCGGTGATGAGCTCACAGAGCTGCTGGCAG
GGATCACTGGCACTGAAGTGGACTGCCTGCGGGCCTGTCAGGAGCAGATCGAAGCTGCACTCAGGGAGAGCCTCA
GGGAAGCCTCTCAGACCAGCTCCAGCCCAGCGCCCAAAGCCCCCGGGGCTCCAGCAGCCAAGGGCCCAGCCAGA
CCAGCACTCCTACAGATGTCACAGCCATACACCTGTAG

TsF

ATGGAGCTGCTGTGTTGCGAAGGCACCCGGCACGCGCCCCGGGCCGGGCCGGACCCGCGGCTGCTGGGGGACCA
GCGTGTCCTGCAGAGCCTGCTCCGCCTGGAGGAGCGCTACGTACCCCGCGCCTCCTACTTCCAGTGCGTGCAGCGG
GAGATCAAGCCGCACATGCGGAAGATGCTGGCTTACTGGATGCTGGAGGTATGTGAGGAGCAGCGCTGTGAGGAG
GAAGTCTTCCCCCTGGCCATGAACTACCTGGATCGCTACCTGTCTTTCGTCCCCACCCGAAAGGCGCAGTTGCAGCT
CCTGGGTGCGGTCTGCATGCTGCTGGCCTCCAAGCTGCGCGAGACCACGCCCCTGACCATCGAAAACTGTGCATCT
ACACCGACCACGCTGTCTCTCCCCGCCAGTTGCGGGACTGGGAGGTGCTGGTCCTAGGGAAGCTCAAGTGGGACCT
GGCTGCTGTGATTGCACATGATTTCTGGCCTTATTCTGCACCGGCTCTCTCTGCCTCGTGACCGACAGGCTTTGGT
GAAGAAGCACGCTCAGACTTTCCTGGCTCTGTGCGCTACAGACTACACTTTCGCTATGTACCCGCCAAGCATGATAGC
TACGGGCAGCATTGGGGCTGCAGTGCAGGGCCTGGGAGCTTGAGCATGAGCGGGGACGAGCTCACAGAGCTGC
TGGCAGGGATCACTGGCACTGAAGTGGACTGCCTGCGGGCCTGTCAGGAGCAGATCGAAGCTGCACTCAGGGAGA

GCCTCAGGGAAGCCTCTCAGACCAGCTCCAGCCCAGCGCCCAAAGCCCCCGGGGCTCCAGCAGCCAAGGGCCCA
GCCAGACCAGCACTCCTACAGATGTCACAGCCATACACCTGTAG

TsM

ATGGAGCTGCTGTGTTGCGAAGGCACCCGGCACGCGCCCCGGGCGGGCCGGACCCGCGGCTGCTGGGGGACCA
GCGTGTCCTGCAGAGCCTGCTCCGCCTGGAGGAGCGCTACGTACCCCGCGCCTCCTACTTCCAGTGCGTGCAGCGG
GAGATCAAGCCGCACATGCGGAAGATGCTGGCTTACTGGATGCTGGAGGTATGTGAGGAGCAGCGCTGTGAGGAG
GAAGTCTTCCCCCTGGCCATGAACTACCTGGATCGCTACCTGTCTTTCGCTCCCCACCCGAAAGGCGCAGTTGCAGCT
CCTGGGTGCGGTCTGCATGCTGCTGGCCTCCAAGCTGCGCGAGACCACGCCCCTGACCATCGAAAACTGTGCATCT
ACACCGACCACGCTGTCTCTCCCCGCCAGTTGCGGGACTGGGAGGTGCTGGTCCTAGGGAAGCTCAAGTGGGACCT
GGCTGCTGTGATTGCACATGATTTCTGGCCTTCACTTCTGCACCGGCTCTCTTGCCGCGTGACCGACAGGCATTGGT
AAAGAAGCACGCACAGACGTTCTTGGCGCTATGCGCTACAGACTACACATTCGCGATGTACCCGCCATCGATGATTGC
GACGGGCAGCATTGGGGCTGCAGTGAGGGCCTGGGCGCGTGCTCTATGTCTGGGGACGAGCTCACAGAGCTGCT
GGCAGGGATCACTGGCACTGAAGTGGACTGCCTGCGGGCCTGTCAGGAGCAGATCGAAGCTGCACTCAGGGAGA
GCCTCAGGGAAGCCTCTCAGACCAGCTCCAGCCCAGCGCCCAAAGCCCCCGGGGCTCCAGCAGCCAAGGGCCCA
GCCAGACCAGCACTCCTACAGATGTCACAGCCATACACCTGTAG

TsS

ATGGAGCTGCTGTGTTGCGAAGGCACCCGGCACGCGCCCCGGGCGGGCCGGACCCGCGGCTGCTGGGGGACCA
GCGTGTCCTGCAGAGCCTGCTCCGCCTGGAGGAGCGCTACGTACCCCGCGCCTCCTACTTCCAGTGCGTGCAGCGG
GAGATCAAGCCGCACATGCGGAAGATGCTGGCTTACTGGATGCTGGAGGTATGTGAGGAGCAGCGCTGTGAGGAG
GAAGTCTTCCCCCTGGCCATGAACTACCTGGATCGCTACCTGTCTTTCGCTCCCCACCCGAAAGGCGCAGTTGCAGCT
CCTGGGTGCGGTCTGCATGCTGCTGGCCTCCAAGCTGCGCGAGACCACGCCCCTGACCATCGAAAACTGTGCATCT
ACACCGACCACGCTGTCTCTCCCCGCCAGTTGCGGGACTGGGAGGTGCTGGTCCTAGGGAAGCTCAAGTGGGACCT
GGCTGCTGTGATTGCACATGATTTCTGGCCTTCACTTCTGCACCGGCTCTCTCTCCCTCGCGATCGACAAGCTTTGGTG
AAGAAACACGCTCAAACCTTTCTTGGCTCTGTGCGCCACAGACTACACTTTTCGCTATGTATCCGCCAAGCATGATAGCT
ACGGGCTCCATCGGGGCCGAGTCCAGGGCCTCGGAGCTTGTAGCATGAGCGGGGACGAGCTCACAGAGCTGCTG
GCAGGGATCACTGGCACTGAAGTGGACTGCCTGCGGGCCTGTCAGGAGCAGATCGAAGCTGCACTCAGGGAGAGC
CTCAGGGAAGCCTCTCAGACCAGCTCCAGCCCAGCGCCCAAAGCCCCCGGGGCTCCAGCAGCCAAGGGCCCAGC
CAGACCAGCACTCCTACAGATGTCACAGCCATACACCTGTAG

Gv Variants

GvN

ATGGAGCTGCTGTGTTGCGAAGGCACCCGGCACGCGCCCCGGGCGGGCCGGACCCGCGGCTGCTGGGGGACCA
GCGTGTCTGCAGAGCCTGCTCCGCCTGGAGGAGCGCTACGTACCCCGCGCCTCCTACTTCCAGTGCGTGCAGCGG
GAGATCAAGCCGCACATGCGGAAGATGCTGGCTTACTGGATGCTGGAGGTATGTGAGGAGCAGCGCTGTGAGGAG
GAAGTCTTCCCCCTGGCCATGAACTACCTGGATCGCTACCTGTCTTTCGTCCCCACCCGAAAGGCGCAGTTGCAGCT
CCTGGGTGCGGTCTGCATGCTGCTGGCCTCCAAGCTGCGCGAGACCACGCCCCTGACCATCGAAAACTGTGCATCT
ACACCGACCACGCTGTCTCTCCCCGCCAGTTGCGGGACTGGGAGGTGCTGGTCCTAGGGAAGCTCAAGTGGGACCT
GGCTGCTGTGATTGCACATGATTTCTGGCCTTCATTCTGCACCGGCTCTCTGCCCCGTGACCGACAGGCCTTGGT
CAAAAAGCATGCCAGACCTTTTTGGCCCTCTGTGCTACAGATTATACCTTTGCCATGTACCCGCCATCCATGATCGCC
ACGGGACAGCATTGGGGCTGCAGTGCAAGGCCTGGGTGCCTGCTCCATGTCCGGGGATGAGCTCACAGAGCTGCTG
GCAGGGATCACTGGCACTGAAGTGGACTGCCTGCGGGCCTGTCAGGAGCAGATCGAAGCTGCACTCAGGGAGAGC
CTCAGGGAAGCCGCTCAGACCAGCTCCAGCCCAGCGCCCAAAGCCCCCGGGCTCCAGCAGCCAAGGGCCCCAGC
CAGACCAGCACTCTACAGATGTACAGCCATACACCTGCTCGAGGGTAAGCCTATCCCTAACCTCTCCTCGGTCTC
GATTCTACGCGTACCGGTCATCATCACCATCACCAT

GvO

ATGGAGCTGCTGTGTTGCGAAGGCACCCGGCACGCGCCCCGGGCGGGCCGGACCCGCGGCTGCTGGGGGACCA
GCGTGTCTGCAGAGCCTGCTCCGCCTGGAGGAGCGCTACGTACCCCGCGCCTCCTACTTCCAGTGCGTGCAGCGG
GAGATCAAGCCGCACATGCGGAAGATGCTGGCTTACTGGATGCTGGAGGTATGTGAGGAGCAGCGCTGTGAGGAG
GAAGTCTTCCCCCTGGCCATGAACTACCTGGATCGCTACCTGTCTTTCGTCCCCACCCGAAAGGCGCAGTTGCAGCT
CCTGGGTGCGGTCTGCATGCTGCTGGCCTCCAAGCTGCGCGAGACCACGCCCCTGACCATCGAAAACTGTGCATCT
ACACCGACCACGCTGTCTCTCCCCGCCAGTTGCGGGACTGGGAGGTGCTGGTCCTAGGGAAGCTCAAGTGGGACCT
GGCTGCTGTGATTGCACATGATTTCTGGCCTTCATTCTGCACCGGCTCTCTGCTCTGACCGTCAGGCTCTGGT
GAAGAAGCACGCTCAGACTTTCCTGGCTCTGTGCGCTACTGACTACACTTTCGCTATGTACCCTCCTAGCATGATAGCT
ACTGGCAGCATTGGAGCTGCTGTGCAGGGACTGGGAGCTTGCAAGCATGAGCGGAGACGAGCTCACAGAGCTGCTG
GCAGGGATCACTGGCACTGAAGTGGACTGCCTGCGGGCCTGTCAGGAGCAGATCGAAGCTGCACTCAGGGAGAGC
CTCAGGGAAGCCGCTCAGACCAGCTCCAGCCCAGCGCCCAAAGCCCCCGGGCTCCAGCAGCCAAGGGCCCCAGC
CAGACCAGCACTCTACAGATGTACAGCCATACACCTGCTCGAGGGTAAGCCTATCCCTAACCTCTCCTCGGTCTC
GATTCTACGCGTACCGGTCATCATCACCATCACCAT

GvD

ATGGAGCTGCTGTGTTGCGAAGGCACCCGGCACGCGCCCCGGGCGGGCCGGACCCGCGGCTGCTGGGGGACCA
GCGTGTCTGCAGAGCCTGCTCCGCCTGGAGGAGCGCTACGTACCCCGCGCCTCCTACTTCCAGTGCGTGCAGCGG
GAGATCAAGCCGCACATGCGGAAGATGCTGGCTTACTGGATGCTGGAGGTATGTGAGGAGCAGCGCTGTGAGGAG
GAAGTCTTCCCCCTGGCCATGAACTACCTGGATCGCTACCTGTCTTTCGTCCCCACCCGAAAGGCGCAGTTGCAGCT
CCTGGGTGCGGTCTGCATGCTGCTGGCCTCCAAGCTGCGCGAGACCACGCCCCTGACCATCGAAAACTGTGCATCT
ACACCGACCACGCTGTCTCTCCCCGCCAGTTGCGGGACTGGGAGGTGCTGGTCCTAGGGAAGCTCAAGTGGGACCT
GGCTGCTGTGATTGCACATGATTTCTGGCCTTCATTCTGCACCGGCTCTCTCTCCCCCGCGATCGCCAAGCCCTCGTC
AAAAACATGCCAAACCTTTCTCGCCCTCTGTGCCACCGATTATACCTTTGCCATGTATCCCCCTCCATGATCGCCAC

CGGTTCCATCGGTGCCGCCGTCCAAGGTCTCGGTGCCTGTTCCATGTCCGGTGATGAGCTCACAGAGCTGCTGGCAG
GGATCACTGGCACTGAAGTGGACTGCCTGCGGGCCTGTCAGGAGCAGATCGAAGCTGCACTCAGGGAGAGCCTCA
GGGAAGCCGCTCAGACCAGCTCCAGCCCAGCGCCCAAAGCCCCCGGGGCTCCAGCAGCCAAGGGCCCCAGCCAG
ACCAGCACTCCTACAGATGTCACAGCCATACACCTGCTCGAGGGTAAGCCTATCCCTAACCCCTCTCCTCGGTCTCGATT
CTACGCGTACCGGTCATCATCACCATCACCAT

GvF

ATGGAGCTGCTGTGTTGCGAAGGCACCCGGCACGCGCCCCGGGCGGGCCGGACCCGCGGCTGCTGGGGGACCA
GCGTGTCCTGCAGAGCCTGCTCCGCCTGGAGGAGCGCTACGTACCCCGCGCCTCCTACTTCCAGTGCGTGCAGCGG
GAGATCAAGCCGCACATGCGGAAGATGCTGGCTTACTGGATGCTGGAGGTATGTGAGGAGCAGCGCTGTGAGGAG
GAAGTCTTCCCCCTGGCCATGAACTACCTGGATCGCTACCTGTCTTTCGTCCCCACCCGAAAGGCGCAGTTGCAGCT
CCTGGGTGCGGTCTGCATGCTGCTGGCCTCCAAGCTGCGCGAGACCACGCCCCTGACCATCGAAAACTGTGCATCT
ACACCGACCACGCTGTCTCTCCCGCCAGTTGCGGGACTGGGAGGTGCTGGTCCTAGGGAAGCTCAAGTGGGACCT
GGCTGCTGTGATTGCACATGATTTCTGGCCTTATTCTGCACCGGCTCTCTCTGCCTCGTGACCGACAGGCTTTGGT
GAAGAAGCACGCTCAGACTTTCTTGGCTCTGTGCGCTACAGACTACACTTTCGTATGTACCCGCCAAGCATGATAGC
TACGGGCAGCATTGGGGCTGCAGTGCAGGGCCTGGGAGCTTGAGCATGAGCGGGGACGAGCTCACAGAGCTGC
TGGCAGGGATCACTGGCACTGAAGTGGACTGCCTGCGGGCCTGTCAGGAGCAGATCGAAGCTGCACTCAGGGAGA
GCCTCAGGGAAGCCGCTCAGACCAGCTCCAGCCCAGCGCCCAAAGCCCCCGGGGCTCCAGCAGCCAAGGGCCCA
GCCAGACCAGCACTCCTACAGATGTCACAGCCATACACCTGCTCGAGGGTAAGCCTATCCCTAACCCCTCTCCTCGGTC
TCGATTCTACGCGTACCGGTCATCATCACCATCACCAT

GvM

ATGGAGCTGCTGTGTTGCGAAGGCACCCGGCACGCGCCCCGGGCGGGCCGGACCCGCGGCTGCTGGGGGACCA
GCGTGTCCTGCAGAGCCTGCTCCGCCTGGAGGAGCGCTACGTACCCCGCGCCTCCTACTTCCAGTGCGTGCAGCGG
GAGATCAAGCCGCACATGCGGAAGATGCTGGCTTACTGGATGCTGGAGGTATGTGAGGAGCAGCGCTGTGAGGAG
GAAGTCTTCCCCCTGGCCATGAACTACCTGGATCGCTACCTGTCTTTCGTCCCCACCCGAAAGGCGCAGTTGCAGCT
CCTGGGTGCGGTCTGCATGCTGCTGGCCTCCAAGCTGCGCGAGACCACGCCCCTGACCATCGAAAACTGTGCATCT
ACACCGACCACGCTGTCTCTCCCGCCAGTTGCGGGACTGGGAGGTGCTGGTCCTAGGGAAGCTCAAGTGGGACCT
GGCTGCTGTGATTGCACATGATTTCTGGCCTTATTCTGCACCGGCTCTCTCTGCCGCGTGACCGACAGGCATTGGT
AAAGAAGCACGCACAGACGTTCTTGGCGCTATGCGCTACAGACTACACATTCGCGATGTACCCGCCATCGATGATTGC
GACGGGCAGCATTGGGGCTGCAGTGCAGGGCCTGGGCGCGTGCTCTATGTCTGGGGACGAGCTCACAGAGCTGCT
GGCAGGGATCACTGGCACTGAAGTGGACTGCCTGCGGGCCTGTCAGGAGCAGATCGAAGCTGCACTCAGGGAGA
GCCTCAGGGAAGCCGCTCAGACCAGCTCCAGCCCAGCGCCCAAAGCCCCCGGGGCTCCAGCAGCCAAGGGCCCA
GCCAGACCAGCACTCCTACAGATGTCACAGCCATACACCTGCTCGAGGGTAAGCCTATCCCTAACCCCTCTCCTCGGTC
TCGATTCTACGCGTACCGGTCATCATCACCATCACCAT

GvS

ATGGAGCTGCTGTGTTGCGAAGGCACCCGGCACGCGCCCCGGGCCGGGCCGGACCCGCGGCTGCTGGGGGACCA
GCGTGTCTGCAGAGCCTGCTCCGCCTGGAGGAGCGCTACGTACCCCGCGCCTCCTACTTCCAGTGCGTGCAGCGG
GAGATCAAGCCGCACATGCGGAAGATGCTGGCTTACTGGATGCTGGAGGTATGTGAGGAGCAGCGCTGTGAGGAG
GAAGTCTTCCCCCTGGCCATGAACTACCTGGATCGCTACCTGTCTTTCGTCCCCACCCGAAAGGCGCAGTTGCAGCT
CCTGGGTGCGGTCTGCATGCTGCTGGCCTCCAAGCTGCGCGAGACCACGCCCCTGACCATCGAAAACTGTGCATCT
ACACCGACCACGCTGTCTCTCCCCGCCAGTTGCGGGACTGGGAGGTGCTGGTCCTAGGGAAGCTCAAGTGGGACCT
GGCTGCTGTGATTGCACATGATTTCTGGCCTTATTCTGCACCGGCTCTCTCTCCCTCGCGATCGACAAGCTTTGGTG
AAGAAACACGCTCAAACCTTTCTTGCTCTGTGCGCCACAGACTACACTTTTCGTATGTATCCGCCAAGCATGATAGCT
ACGGGCTCCATCGGGGCCGAGTCCAGGGCCTCGGAGCTTGTAGCATGAGCGGGGACGAGCTCACAGAGCTGCTG
GCAGGGATCACTGGCACTGAAGTGGACTGCCTGCGGGCCTGTCAGGAGCAGATCGAAGCTGCACTCAGGGAGAGC
CTCAGGGAAGCCGCTCAGACCAGCTCCAGCCCAGCGCCCAAAGCCCCCGGGGCTCCAGCAGCCAAGGGCCCCAGC
CAGACCAGCACTCTACAGATGTACAGCCATACACCTGCTCGAGGGTAAGCCTATCCCTAACCTCTCCTCGGTCTC
GATTCTACGCGTACCGGTCATCATCACCATCACCAT

Gs Variants

GsN

ATGGAGCTGCTGTGTTGCGAAGGCACCCGGCACGCGCCCCGGGCCGGGCCGGACCCGCGGCTGCTGGGGGACCA
GCGTGTCTGCAGAGCCTGCTCCGCCTGGAGGAGCGCTACGTACCCCGCGCCTCCTACTTCCAGTGCGTGCAGCGG
GAGATCAAGCCGCACATGCGGAAGATGCTGGCTTACTGGATGCTGGAGGTATGTGAGGAGCAGCGCTGTGAGGAG
GAAGTCTTCCCCCTGGCCATGAACTACCTGGATCGCTACCTGTCTTTCGTCCCCACCCGAAAGGCGCAGTTGCAGCT
CCTGGGTGCGGTCTGCATGCTGCTGGCCTCCAAGCTGCGCGAGACCACGCCCCTGACCATCGAAAACTGTGCATCT
ACACCGACCACGCTGTCTCTCCCCGCCAGTTGCGGGACTGGGAGGTGCTGGTCCTAGGGAAGCTCAAGTGGGACCT
GGCTGCTGTGATTGCACATGATTTCTGGCCTTATTCTGCACCGGCTCTCTTGCCCCGTGACCGACAGGCCTTGCT
CAAAAAGCATGCCCAGACCTTTTTGGCCCTCTGTGCTACAGATTATACCTTTGCCATGTACCCGCCATCCATGATCGCC
ACGGGACAGCATTGGGGCTGCAAGGCTGGGTGCTGCTCCATGTCCGGGGATGAGCTCACAGAGCTGCTG
GCAGGGATCACTGGCACTGAAGTGGACTGCCTGCGGGCCTGTCAGGAGCAGATCGAAGCTGCACTCAGGGAGAGC
CTCAGGGAAGCCTCTCAGACCAGCTCCAGCCCAGCGCCCAAAGCCCCCGGGGCTCCAGCAGCCAAGGGCCCCAGC
CAGACCAGCACTCTACAGATGTACAGCCATACACCTGTAG

GsO

ATGGAGCTGCTGTGTTGCGAAGGCACCCGGCACGCGCCCCGGGCCGGGCCGGACCCGCGGCTGCTGGGGGACCA
GCGTGTCTGCAGAGCCTGCTCCGCCTGGAGGAGCGCTACGTACCCCGCGCCTCCTACTTCCAGTGCGTGCAGCGG
GAGATCAAGCCGCACATGCGGAAGATGCTGGCTTACTGGATGCTGGAGGTATGTGAGGAGCAGCGCTGTGAGGAG
GAAGTCTTCCCCCTGGCCATGAACTACCTGGATCGCTACCTGTCTTTCGTCCCCACCCGAAAGGCGCAGTTGCAGCT
CCTGGGTGCGGTCTGCATGCTGCTGGCCTCCAAGCTGCGCGAGACCACGCCCCTGACCATCGAAAACTGTGCATCT
ACACCGACCACGCTGTCTCTCCCCGCCAGTTGCGGGACTGGGAGGTGCTGGTCCTAGGGAAGCTCAAGTGGGACCT
GGCTGCTGTGATTGCACATGATTTCTGGCCTTATTCTGCACCGGCTCTCTTGCCCTCGTGACCGTCAGGCTCTGGT
GAAGAAGCACGCTCAGACTTTCTGGCTCTGTGCGCTACTGACTACACTTTTCGTATGTACCTCCTAGCATGATAGCT
ACTGGCAGCATTGGAGCTGCTGTGCAGGGACTGGGAGCTTGCAAGCATGAGCGGAGACGAGCTCACAGAGCTGCTG

GCAGGGATCACTGGCACTGAAGTGGACTGCCTGCGGGCCTGTCAGGAGCAGATCGAAGCTGCACTCAGGGAGAGC
CTCAGGGAAGCCTCTCAGACCAGCTCCAGCCCAGCGCCCAAAGCCCCCGGGGCTCCAGCAGCCAAGGGCCCAGC
CAGACCAGCACTCCTACAGATGTCACAGCCATACACCTGTAG

GsD

ATGGAGCTGCTGTGTTGCGAAGGCACCCGGCACGCGCCCCGGGCGGGCCGGACCCGCGGCTGCTGGGGGACCA
GCGTGTCTGCAGAGCCTGCTCCGCCTGGAGGAGCGCTACGTACCCCGCGCCTCCTACTTCCAGTGCGTGCAGCGG
GAGATCAAGCCGCACATGCGGAAGATGCTGGCTTACTGGATGCTGGAGGTATGTGAGGAGCAGCGCTGTGAGGAG
GAAGTCTTCCCCCTGGCCATGAAGTACCTGGATCGCTACCTGTCTTTCGTCCCCACCCGAAAGGCGCAGTTGCAGCT
CCTGGGTGCGGTCTGCATGCTGCTGGCCTCCAAGCTGCGCGAGACCACGCCCCTGACCATCGAAAACTGTGCATCT
ACACCGACCACGCTGTCTCTCCCCGCCAGTTGCGGGACTGGGAGGTGCTGGTCCTAGGGAAGCTCAAGTGGGACCT
GGCTGCTGTGATTGCACATGATTTCTGGCCTTCATTCTGCACCGGCTCTCTCTCCCCCGCGATCGCCAAGCCCTCGTC
AAAAACATGCCAAACCTTTCTCGCCCTCTGTGCCACCGATTATACCTTTGCCATGTATCCCCCTCCATGATCGCCAC
CGGTTCCATCGGTGCCGCGTCCAAGGTCTCGGTGCCTGTTCCATGTCCGGTGATGAGCTCACAGAGCTGCTGGCAG
GGATCACTGGCACTGAAGTGGACTGCCTGCGGGCCTGTCAGGAGCAGATCGAAGCTGCACTCAGGGAGAGCCTCA
GGGAAGCCGCTCAGACCAGCTCCAGCCCAGCGCCCAAAGCCCCCGGGGCTCCAGCAGCCAAGGGCCCAGCCAG
ACCAGCACTCCTACAGATGTCACAGCCATACACCTGTAG

GsF

ATGGAGCTGCTGTGTTGCGAAGGCACCCGGCACGCGCCCCGGGCGGGCCGGACCCGCGGCTGCTGGGGGACCA
GCGTGTCTGCAGAGCCTGCTCCGCCTGGAGGAGCGCTACGTACCCCGCGCCTCCTACTTCCAGTGCGTGCAGCGG
GAGATCAAGCCGCACATGCGGAAGATGCTGGCTTACTGGATGCTGGAGGTATGTGAGGAGCAGCGCTGTGAGGAG
GAAGTCTTCCCCCTGGCCATGAAGTACCTGGATCGCTACCTGTCTTTCGTCCCCACCCGAAAGGCGCAGTTGCAGCT
CCTGGGTGCGGTCTGCATGCTGCTGGCCTCCAAGCTGCGCGAGACCACGCCCCTGACCATCGAAAACTGTGCATCT
ACACCGACCACGCTGTCTCTCCCCGCCAGTTGCGGGACTGGGAGGTGCTGGTCCTAGGGAAGCTCAAGTGGGACCT
GGCTGCTGTGATTGCACATGATTTCTGGCCTTCATTCTGCACCGGCTCTCTCTGCCTCGTGACCGACAGGCTTTGGT
GAAGAAGCACGCTCAGACTTTCTTGCTCTGTGCGCTACAGACTACACTTTTGCTATGTACCCGCCAAGCATGATAGC
TACGGGCAGCATTGGGGCTGCAGTGAGGGCCTGGGAGCTTGAGCATGAGCGGGGACGAGCTCACAGAGCTGC
TGGCAGGGATCACTGGCACTGAAGTGGACTGCCTGCGGGCCTGTCAGGAGCAGATCGAAGCTGCACTCAGGGAGA
GCCTCAGGGAAGCCGCTCAGACCAGCTCCAGCCCAGCGCCCAAAGCCCCCGGGGCTCCAGCAGCCAAGGGCCCA
GCCAGACCAGCACTCCTACAGATGTCACAGCCATACACCTGTAG

GsM

ATGGAGCTGCTGTGTTGCGAAGGCACCCGGCACGCGCCCCGGGCGGGCCGGACCCGCGGCTGCTGGGGGACCA
GCGTGTCTGCAGAGCCTGCTCCGCCTGGAGGAGCGCTACGTACCCCGCGCCTCCTACTTCCAGTGCGTGCAGCGG
GAGATCAAGCCGCACATGCGGAAGATGCTGGCTTACTGGATGCTGGAGGTATGTGAGGAGCAGCGCTGTGAGGAG
GAAGTCTTCCCCCTGGCCATGAACTACCTGGATCGCTACCTGTCTTTCGTCCCCACCCGAAAGGCGCAGTTGCAGCT
CCTGGGTGCGGTCTGCATGCTGCTGGCCTCCAAGCTGCGCGAGACCACGCCCCTGACCATCGAAAACTGTGCATCT
ACACCGACCACGCTGTCTCTCCCCGCCAGTTGCGGGACTGGGAGGTGCTGGTCCTAGGGAAGCTCAAGTGGGACCT
GGCTGCTGTGATTGCACATGATTTCTGGCCTTCATTCTGCACCGGCTCTCTTGCCGCGTGACCGACAGGCATTGGT
AAAGAAGCACGCACAGACGTTCTTGGCGCTATGCGCTACAGACTACACATTCGCGATGTACCCGCCATCGATGATTGC
GACGGGACGATTGGGGCTGCAGTGCAGGGCCTGGGCGCGTGCTCTATGTCTGGGGACGAGCTCACAGAGCTGCT
GGCAGGGATCACTGGCACTGAAGTGGACTGCCTGCGGGCCTGTCAGGAGCAGATCGAAGCTGCACTCAGGGAGA
GCCTCAGGGAAGCCGCTCAGACCAGCTCCAGCCAGCGCCCAAAGCCCCCGGGGCTCCAGCAGCCAAGGGCCCA
GCCAGACCAGCACTCTACAGATGTCACAGCCATACACCTGTAG

GsS

ATGGAGCTGCTGTGTTGCGAAGGCACCCGGCACGCGCCCCGGGCGGGCCGGACCCGCGGCTGCTGGGGGACCA
GCGTGTCTGCAGAGCCTGCTCCGCCTGGAGGAGCGCTACGTACCCCGCGCCTCCTACTTCCAGTGCGTGCAGCGG
GAGATCAAGCCGCACATGCGGAAGATGCTGGCTTACTGGATGCTGGAGGTATGTGAGGAGCAGCGCTGTGAGGAG
GAAGTCTTCCCCCTGGCCATGAACTACCTGGATCGCTACCTGTCTTTCGTCCCCACCCGAAAGGCGCAGTTGCAGCT
CCTGGGTGCGGTCTGCATGCTGCTGGCCTCCAAGCTGCGCGAGACCACGCCCCTGACCATCGAAAACTGTGCATCT
ACACCGACCACGCTGTCTCTCCCCGCCAGTTGCGGGACTGGGAGGTGCTGGTCCTAGGGAAGCTCAAGTGGGACCT
GGCTGCTGTGATTGCACATGATTTCTGGCCTTCATTCTGCACCGGCTCTCTCTCCCTCGCGATCGACAAGCTTTGGTG
AAGAAACACGCTCAAACCTTTCTTGGCTCTGTGCGCCACAGACTACACTTTCGCTATGTATCCGCCAAGCATGATAGCT
ACGGGCTCCATCGGGGCCGAGTCCAGGGCCTCGGAGCTTGTAGCATGAGCGGGGACGAGCTCACAGAGCTGCTG
GCAGGGATCACTGGCACTGAAGTGGACTGCCTGCGGGCCTGTCAGGAGCAGATCGAAGCTGCACTCAGGGAGAGC
CTCAGGGAAGCCGCTCAGACCAGCTCCAGCCAGCGCCCAAAGCCCCCGGGGCTCCAGCAGCCAAGGGGCCAGC
CAGACCAGCACTCTACAGATGTCACAGCCATACACCTGTAG

7.2.2 eGFP Sequences

eGFP

GFP-N

ATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGC
CACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCA
CCGGCAAGCTGCCGTGCCCTGGCCACCCCTCGTGACCACCTGACCTACGGCGTGCAAGTCTTACGCCGTACCCC
GACCACATGAAGCAGCAGACTTCTTCAAGTCCGCCATGCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAA
GGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCAGGGGCGACACCCTGGTGAACCGCATCGAGCTGAA
GGGCATCGACTTCAAGGAGGACGGCAATATCCTGGGGCACAAGCTGGAGTACAACAGCCACAACGTCTAT
ATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAAGTTCAGATCCGCCACAACATCGAGGACGGCAGCGTG

CAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTGA
GCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGC
CGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAA

eGFP-O

ATGGTGAGCAAGGGAGAAGAACTGTTCACTGGAGTGGTGCCTATACTGGTGGAAGTGGACGGAGACGTGAACGGA
CACAAGTTCAGCGTGAGCGGAGAAGGAGAAGGAGACGCTACTTACGGAAAGCTGACTCTGAAGTTCATATGACTA
CCGGCAAGCTGCCCCGTGCCCTGGCCACCCCTCGTGACCACCCTGACCTACGGCGTGCAAGTGCTTCAGCCGCTACCCC
GACCACATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAA
GGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAA
GGGCATCGACTTCAAGGAGGACGGCAATATCCTGGGGCACAAGCTGGAGTACAATAACAGCCACAACGTCTAT
ATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTG
CAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTGA
GCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGC
CGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAA

eGFP-D

ATGGTCTCAAAGGTGAGGAGCTCTTTACCGGTGTCGTCCCCATCCTCGTCGAGCTCGATGGTGATGTCAATGGTCAT
AAATTTTCCGTCTCCGGTGAGGGTGAGGGTGATGCCACCTATGGTAAACTCACCTCAAATTTATCTGTACCACCGGC
AAGCTGCCCCGTGCCCTGGCCACCCCTCGTGACCACCCTGACCTACGGCGTGCAAGTGCTTCAGCCGCTACCCCGACCA
CATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACG
ACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCA
TCGACTTCAAGGAGGACGGCAATATCCTGGGGCACAAGCTGGAGTACAATAACAGCCACAACGTCTATATCATG
GCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGACGCTC
GCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCACCC
AGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGAT
CACTCTCGGCATGGACGAGCTGTACAAGTAA

sGFP-N

ATGAAGCTCTCCCTGGTGCCGCGATGCTGCTGCTCAGCGCGGCGGGCTGTGAGCAAGGGCGAGGAGCTG
TTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCGGCGAG
GGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGAAGCTGCCCGTGCCCTGGC
CCACCCTCGTGACCACCCTGACCTACGGCGTGCAAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCAGCACTT
TCAAGTCCGCCATGCCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCC
GCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGAC
GGCAATATCCTGGGGCACAAGCTGGAGTACAATAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAA
CGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGACGCTCGCCGACCACTACCAGCAG
AACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAG

ACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGA
GCTGTACAAGTAA

sGFP-O

ATGAAGCTCTCCCTGGTGGCCGCGATGCTGCTGCTGCTCAGCGCGGCGCGGGCTGTGAGCAAGGGAGAAGAACTG
TTCAGTGGAGTGGTGCCTATACTGGTGGAACTGGACGGAGACGTGAACGGACACAAGTTCAGCGTGAGCGGAGAA
GGAGAAGGAGACGCTACTTACGGAAAGCTGACTCTGAAGTTCATATGCACTACCGGCAAGCTGCCCCGTGCCCTGGC
CCACCCTCGTGACCACCCTGACCTACGGCGTGAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTC
TTCAAGTCCGCCATGCCCCAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCC
GCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGAC
GGCAATATCTGGGGCACAAGCTGGAGTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAA
CGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGACGCTCGCCGACCACTACCAGCAG
AACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAG
ACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGA
GCTGTACAAGTAA

sGFP-D

ATGAAGCTCTCCCTGGTGGCCGCGATGCTGCTGCTGCTCAGCGCGGCGCGGGCTGTCTCAAAGGTGAGGAGCTCT
TTACCGGTGTCTGTCCTCCATCTCGTCGAGCTCGATGGTGATGTCAATGGTCATAAATTTCCGTCTCCGGTGAGGGTG
AGGGTGATGCCACCTATGGTAACTCACCTCAAATTTATCTGTACCACCGCAAGCTGCCCCGTGCCCTGGCCCCACC
TCGTGACCACCCTGACCTACGGCGTGAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAG
TCCGCCATGCCCCAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCG
AGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAATA
TCCTGGGGCACAAGCTGGAGTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCAT
CAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGACGCTCGCCGACCACTACCAGCAGAACACC
CCCATCGGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCA
ACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTA
CAAGTAA

7.2.3 Noggin sequences

NOG

NOG-N

ATGGAGCGCTGCCCCAGCCTAGGGGTACCCCTCTACGCCCTGGTGGTGGTCCTGGGGCTGCGGGCGACACCGGCC
GGCGGCCAGCACTATCTCCACATCCGCCCGGCACCCAGCGACAACCTGCCCTGGTGGACCTCATCGAACACCCAGA
CCCTATCTTTGACCCCAAGGAAAAGGATCTGAACGAGACGCTGCTGCGCTCGCTGCTCGGGGGCCACTACGACCCA
GGCTTCATGGCCACCAGCCACCTGAAGATAGACCTGGTGGTGGTGGCGGAGCCGCAAGTGGAGCCGAAGATTTG
GCCGAACCTTGATCAATTGCTCAGACAACGCCCCAGCGGAGCGATGCCTTCCGAAATTAAGGGATTGGAATTTTCAGA
AGGGCTGGCTCAAGGGAAGAAGCAACGGCTGTCAAAGAACTTAGAAGAAAACCTCAAATGTGGCTCTGGAGTCA

AACCTTTTGTCTGTACTTTATGCTTGGAATGATCTCGGGTCCAGGTTCTGGCCCCGGTATGTCAAAGTCGGTTCCTGC
TTTTCCAAAAGATCTTGTAGTGTCCCTGAAGGGATGGTATGTAAACCTAGCAAAAGCGTCCATCTGACCGTTTTGAGA
TGGAGATGCCAAAGACGGGGCGGACAACGATGTGGGTGGATTCCAATTCAATATCCAATTATATCAGAATGTAAATGT
TCCTGTCTCGAGGGTAAGCCTATCCCTAACCTCTCCTCGGTCTCGATTCTACGCGTACCGGTCATCATCACCATACCA
TTAA

NOG-O

ATGGAGCGCTGCCCCAGCCTAGGGGTACCCCTCTACGCCCTGGTGGTGGTCCTGGGGCTGCGGGCGACACCGGCC
GGCGGCCAGCACTACCTGCACATAGCTCCTGCTCCTAGCGACAACCTGCCTCTGGTGGACCTGATAGAACACCCCTGA
CCCTATATTCGACCCTAAGGAAAAGGACCTGAACGAACTCTGCTGCGTAGCCTGCTGGGAGGACACTACGACCCTG
GATTCATGGCCACCAGCCACCTGAAGATAGACCTGGTGGTGGTGGCGGAGCCGCAGGTGGAGCCGAAGATTTGG
CCGAACCTTGATCAATTGCTCAGACAACGCCCCAGCGGAGCGATGCCTTCCGAAATTAAGGGATTGGAATTTTCAGAA
GGGCTGGCTCAAGGGAAGAAGCAACGGCTGTCAAAGAACTTAGAAGAAAACCTCAAATGTGGCTCTGGAGTCAA
ACCTTTTGTCTGTACTTTATGCTTGGAATGATCTCGGGTCCAGGTTCTGGCCCCGGTATGTCAAAGTCGGTTCCTGCT
TTTCCAAAAGATCTTGTAGTGTCCCTGAAGGGATGGTATGTAAACCTAGCAAAAGCGTCCATCTGACCGTTTTGAGAT
GGAGATGCCAAAGACGGGGCGGACAACGATGTGGGTGGATTCCAATTCAATATCCAATTATATCAGAATGTAAATGTT
CCTGTCTCGAGGGTAAGCCTATCCCTAACCTCTCCTCGGTCTCGATTCTACGCGTACCGGTCATCATCACCATCACCAT
TAA

NOG-D

ATGGAGCGCTGCCCCAGCCTAGGGGTACCCCTCTACGCCCTGGTGGTGGTCCTGGGGCTGCGGGCGACACCGGCC
GGCGGCCAGCATTATCTCCATATCCGCCCCGCCCCCTCCGATAATCTCCCCCTCGTGGATCTCATCGAGCATCCCGATCC
CATCTTTGATCCCAAAGAGAAAGATCTCAATGAGTCCCTCCTCCGCTCCCTCCTCGGTGGTCATTATGATCCCGGTTTT
ATGGCCACCAGCCCACCTGAAGATAGACCTGGTGGTGGTGGCGGAGCCGCAGGTGGAGCCGAAGATTTGGCCGAA
CTTGATCAATTGCTCAGACAACGCCCCAGCGGAGCGATGCCTTCCGAAATTAAGGGATTGGAATTTTCAGAAGGGCT
GGCTCAAGGGAAGAAGCAACGGCTGTCAAAGAACTTAGAAGAAAACCTCAAATGTGGCTCTGGAGTCAAACCTT
TTGTCCTGTACTTTATGCTTGGAATGATCTCGGGTCCAGGTTCTGGCCCCGGTATGTCAAAGTCGGTTCCTGCTTTTCC
AAAAGATCTTGTAGTGTCCCTGAAGGGATGGTATGTAAACCTAGCAAAAGCGTCCATCTGACCGTTTTGAGATGGAG
ATGCCAAAGACGGGGCGGACAACGATGTGGGTGGATTCCAATTCAATATCCAATTATATCAGAATGTAAATGTTCTGT
TCTCGAGGGTAAGCCTATCCCTAACCTCTCCTCGGTCTCGATTCTACGCGTACCGGTCATCATCACCATCACCATTAA

NOG-S

ATGGAGCGCTGCCCCAGCCTAGGGGTACCCCTCTACGCCCTGGTGGTGGTCCTGGGGCTGCGGGCGACACCGGCC
GGCGGCCAGCATTACCTGCATATACGTCCGGCACCTTCCGATAATCTCCCTCTCGTCGATCTGATAGAGCATCCAGATC
CCATATTCGATCCTAAAGAGAAAAGACCTCAATGAAACGCTCCTCCGTTCTGCTCCTGGGGGGCCATTATGATCCAGGCT
TTATGGCCACCAGCCCACCTGAAGATAGACCTGGTGGTGGTGGCGGAGCCGCAGGTGGAGCCGAAGATTTGGCCG
AATTGATCAATTGCTCAGACAACGCCCCAGCGGAGCGATGCCTTCCGAAATTAAGGGATTGGAATTTTCAGAAGG
GCTGGCTCAAGGGAAGAAGCAACGGCTGTCAAAGAACTTAGAAGAAAACCTCAAATGTGGCTCTGGAGTCAAAC
CTTTTGTCTGTACTTTATGCTTGGAATGATCTCGGGTCCAGGTTCTGGCCCCGGTATGTCAAAGTCGGTTCCTGCTTT

TCCAAAAGATCTTGTAGTGTCCCTGAAGGGATGGTATGTAAACCTAGCAAAAGCGTCCATCTGACCGTTTTGAGATGG
AGATGCCAAAGACGGGGCGGACAACGATGTGGGTGGATTCCAATTCAATATCCAATTATATCAGAATGTAAATGTTCC
TGTCTCGAGGGTAAGCCTATCCCTAACCTCTCCTCGGTCTCGATTCTACGCGTACCGGTCATCATCACCATCACCATT
A

sNOG-O

ATGGAACGTTGCCCTAGCCTGGGAGTGA CTCTGTACGCTCTGGTGGTGGTGGTGGGACTGCGTGCTACTCCTGCTGG
AGGACAGCACTACCTGCACATAGCTCCTGCTCCTAGCGACAACCTGCCTCTGGTGGACCTGATAGAACACCCTGACC
CTATATTCGACCCTAAGGAAAAGGACCTGAACGAAACTCTGCTGCGTAGCCTGCTGGGAGGACACTACGACCCTGGA
TTCATGGCCACCAGCCACCTGAAGATAGACCTGGTGGTGGTGGCGGAGCCGAGGTGGAGCCGAAGATTTGGCC
GAACTTGATCAATTGCTCAGACAACGCCCCAGCGGAGCGATGCCTTCCGAAATTAAGGGATTGGAATTTTCAGAAG
GGCTGGCTCAAGGGAAGAAGCAACGGCTGTCAAAGAACTTAGAAGAAAACCTCAAATGTGGCTCTGGAGTCAAA
CCTTTTGTCTGTACTTTATGCTTGAATGATCTCGGGTCCAGGTTCTGGCCCCGGTATGTCAAAGTCGGTTCCTGCTT
TTCCAAAAGATCTTGTAGTGTCCCTGAAGGGATGGTATGTAAACCTAGCAAAAGCGTCCATCTGACCGTTTTGAGATG
GAGATGCCAAAGACGGGGCGGACAACGATGTGGGTGGATTCCAATTCAATATCCAATTATATCAGAATGTAAATGTTT
CTGTCTCGAGGGTAAGCCTATCCCTAACCTCTCCTCGGTCTCGATTCTACGCGTACCGGTCATCATCACCATCACCATT
AA

sNOG-D

ATGGAGCGCTGTCCCTCCCTCGGTGTCACCCTCTATGCCCTCGTCGTCGTCCTCGGTCTCCGCGCCACCCCCGCCGGT
GGTCAACATTATCTCCATATCCGCCCCGCCCTCCGATAATCTCCCCCTCGTGATCTCATCGAGCATCCCGATCCCAT
CTTTGATCCCAAAGAGAAAGATCTCAATGAGTCCCTCCTCCGCTCCCTCCTCGGTGGTCATTATGATCCCGTTTTATG
GCCACCAGCCACCTGAAGATAGACCTGGTGGTGGTGGCGGAGCCGAGGTGGAGCCGAAGATTTGGCCGAACCTT
GATCAATTGCTCAGACAACGCCCCAGCGGAGCGATGCCTTCCGAAATTAAGGGATTGGAATTTTCAGAAGGGCTGG
CTCAAGGGAAGAAGCAACGGCTGTCAAAGAACTTAGAAGAAAACCTCAAATGTGGCTCTGGAGTCAAACCTTTTG
TCCTGTACTTTATGCTTGAATGATCTCGGGTCCAGGTTCTGGCCCCGGTATGTCAAAGTCGGTTCCTGCTTTTCCAAA
AGATCTTGTAGTGTCCCTGAAGGGATGGTATGTAAACCTAGCAAAAGCGTCCATCTGACCGTTTTGAGATGGAGATG
CCAAAGACGGGGCGGACAACGATGTGGGTGGATTCCAATTCAATATCCAATTATATCAGAATGTAAATGTTCTGTCT
CGAGGGTAAGCCTATCCCTAACCTCTCCTCGGTCTCGATTCTACGCGTACCGGTCATCATCACCATCACCATTAA

sNOG-S

ATGGAACGTTGTCCTTCCCTAGGGGTGA CTCTGTATGCTCTCGTCGTCGTCGCTCGGGCTCCGGGCGACACCGGCTGG
CGGCCAACATTACCTGCATATACGTCCGGCACCTTCCGATAATCTCCCTCTCGTCGATCTGATAGAGCATCCAGATCCCA
TATTCGATCCTAAAGAGAAAGACCTCAATGAAACGCTCCTCCGTTGCTCTCTGGGGGGCCATTATGATCCAGGCTTTA
TGGCCACCAGCCACCTGAAGATAGACCTGGTGGTGGTGGCGGAGCCGAGGTGGAGCCGAAGATTTGGCCGAAC
TTGATCAATTGCTCAGACAACGCCCCAGCGGAGCGATGCCTTCCGAAATTAAGGGATTGGAATTTTCAGAAGGGCTG
GCTCAAGGGAAGAAGCAACGGCTGTCAAAGAACTTAGAAGAAAACCTCAAATGTGGCTCTGGAGTCAAACCTTTT
GTCCTGTACTTTATGCTTGAATGATCTCGGGTCCAGGTTCTGGCCCCGGTATGTCAAAGTCGGTTCCTGCTTTTCCAA
AAGATCTTGTAGTGTCCCTGAAGGGATGGTATGTAAACCTAGCAAAAGCGTCCATCTGACCGTTTTGAGATGGAGAT

GCCAAAGACGGGGCGGACAACGATGTGGGTGGATTCCAATTCAATATCCAATTATATCAGAATGTAAATGTTCTGTC
TCGAGGGTAAGCCTATCCCTAACCCCTCCTCGGTCTCGATTCTACGCGTACCGGTCATCATCACCATCACCATTAA

7.4 P-values for Chapter 4 and Chapter 5 Densitometry Data.

These P-values were calculated for Densitometry charts containing samples with a minimum of N=2. Type 2, 2 tailed T-tests were performed comparing two data groups at a time.

Table 7.1: P-Values comparing samples found in figure 4.7B.

	Native	Opt.	Dis.	Fast	Middling
Native	1	0.03357	0.038966	0.002105	0.0036172
Opt.	0.03357	1	0.104693	0.010426	0.0131042
Dis.	0.038966	0.104693	1	0.96199	0.76085692
Fast	0.002105	0.010426	0.96199	1	0.31427801
Middling	0.003617	0.013104	0.760857	0.314278	1

Table 7.2: P-Values comparing samples found in figure 4.8B.

	Native	Opt.	Dis.	Fast	Middling	Swap
Native	1	0.364363	0.372442	0.603359	0.1129851	0.960371
Opt.	0.364363	1	0.86632	0.202835	0.0012678	0.025719
Dis.	0.372442	0.86632	1	0.206127	0.0009518	0.023053
Fast	0.603359	0.202835	0.206127	1	0.2453105	0.477771
Middling	0.112985	0.001268	0.000952	0.24531	1	0.003436
Swap	0.960371	0.025719	0.023053	0.477771	0.0034356	1

Table 7.3: P-Values comparing samples found in figure 4.9B.

	Native	Opt.	Dis.	Fast	Middling	Swap
Native	1	0.06762	0.598424	0.761149	0.0467839	0.512283
Opt.	0.06762	1	0.167986	0.092393	0.0067226	0.041479
Dis.	0.598424	0.167986	1	0.733207	0.469398	0.808436
Fast	0.761149	0.092393	0.733207	1	0.1247738	0.825638
Middling	0.046784	0.006723	0.469398	0.124774	1	0.066789
Swap	0.512283	0.041479	0.808436	0.825638	0.066789	1

Table 7.4: P-Values comparing samples found in figure 4.10B.

	Native	Opt.	Dis.	Fast	Middling	Swap
Native	1	0.046741	0.179159	0.025351	0.1001108	0.009471
Opt.	0.046741	1	0.094296	0.068836	0.0421949	0.244278
Dis.	0.179159	0.094296	1	0.701873	0.128412	0.099547
Fast	0.025351	0.068836	0.701873	1	0.0197821	0.022752
Middling	0.100111	0.042195	0.128412	0.019782	1	0.008572
Swap	0.009471	0.244278	0.099547	0.022752	0.0085721	1

Table 7.5: P-Values comparing samples found in figure 4.13B.

	Native	Opt.	Dis.	Fast	Middling	Swap
Native	1	0.12655	0.054274	0.301464	0.077296	0.073799
Opt.	0.12655	1	0.160669	0.128996	0.4242908	0.211788
Dis.	0.054274	0.160669	1	0.074489	0.3096123	0.995586
Fast	0.301464	0.128996	0.074489	1	0.1009135	0.081608
Middling	0.077296	0.424291	0.309612	0.100913	1	0.377723
Swap	0.073799	0.211788	0.995586	0.081608	0.377723	1

Table 7.6: P-Values comparing samples found in figure 4.14B.

	Native	Opt.	Dis.	Fast	Middling	Swap
Native	1	0.598676	0.222212	0.192711	0.0898315	0.209538
Opt.	0.598676	1	0.126475	0.105299	0.0315606	0.120591
Dis.	0.222212	0.126475	1	0.344325	0.0123699	0.604332
Fast	0.192711	0.105299	0.344325	1	0.0344007	0.656098
Middling	0.089832	0.031561	0.01237	0.034401	1	0.025832
Swap	0.209538	0.120591	0.604332	0.656098	0.0258324	1

Table 7.7: P-Values comparing samples found in figure 4.15B.

	Native	Opt.	Dis.	Fast	Middling	Swap
Native	1	0.991038	0.228223	0.055802	0.9962849	0.210474
Opt.	0.991038	1	0.209559	0.038233	0.9926609	0.175565
Dis.	0.228223	0.209559	1	0.583421	0.2019766	0.722315
Fast	0.055802	0.038233	0.583421	1	0.0300803	0.235626
Middling	0.996285	0.992661	0.201977	0.03008	1	0.159681
Swap	0.210474	0.175565	0.722315	0.235626	0.1596813	1

Table 7.8: P-Values comparing samples found in figure 4.16B.

	Native	Opt.	Dis.	Fast	Middling	Swap
Native	1	0.553716	0.052345	0.700606	0.0836374	0.141447
Opt.	0.553716	1	0.113893	0.737405	0.1908935	0.370732
Dis.	0.052345	0.113893	1	0.038404	0.1584277	0.016117
Fast	0.700606	0.737405	0.038404	1	0.0714206	0.134514
Middling	0.083637	0.190894	0.158428	0.071421	1	0.134702
Swap	0.141447	0.370732	0.016117	0.134514	0.1347017	1

Table 7.9: P-Values comparing samples found in figure 4.19B.

	Native	Opt.	Dis.	Fast	Middling	Swap
Native	1	0.029619	0.083394	0.032269	0.280268	0.740283
Opt.	0.029619	1	0.726653	0.096522	0.0243928	0.033201
Dis.	0.083394	0.726653	1	0.269158	0.067543	0.083149
Fast	0.032269	0.096522	0.269158	1	0.0221485	0.051624
Middling	0.280268	0.024393	0.067543	0.022148	1	0.601716
Swap	0.740283	0.033201	0.083149	0.051624	0.601716	1

Table 7.10: P-Values comparing samples found in figure 4.20B.

	Native	Opt.	Dis.	Fast	Middling	Swap
Native	1	0.159024	0.049687	0.091972	0.1550595	0.225466
Opt.	0.159024	1	0.171298	0.651409	0.7846727	0.392746
Dis.	0.049687	0.171298	1	0.083297	0.0691309	0.028253
Fast	0.091972	0.651409	0.083297	1	0.336581	0.104924
Middling	0.155059	0.784673	0.069131	0.336581	1	0.407679
Swap	0.225466	0.392746	0.028253	0.104924	0.4076789	1

Table 7.11: P-Values comparing samples found in figure 4.21B.

	Native	Opt.	Dis.	Fast	Middling	Swap
Native	1	0.257135	0.117587	0.506621	0.4828063	0.055779
Opt.	0.257135	1	0.003165	0.096461	0.0351527	0.001206
Dis.	0.117587	0.003165	1	0.01754	0.005038	0.0008
Fast	0.506621	0.096461	0.01754	1	0.9086782	0.003878
Middling	0.482806	0.035153	0.005038	0.908678	1	0.002076
Swap	0.055779	0.001206	0.0008	0.003878	0.0020764	1

Table 7.12: P-Values comparing samples found in figure 4.22B.

	Native	Opt.	Dis.	Fast	Middling	Swap
Native	1	0.112056	0.084923698	0.008963	0.796637479	0.168107
Opt.	0.1120564	1	0.055961916	0.056701	0.333498603	0.418138
Dis.	0.0849237	0.055962	1	0.35992	0.275957796	0.063848
Fast	0.0089627	0.056701	0.359920342	1	0.376213573	0.058388
Middling	0.7966375	0.333499	0.275957796	0.376214	1	0.590225
Swap	0.168107	0.418138	0.063848106	0.058388	0.590225402	1

Table 7.13: P-Values comparing samples found in figure 4.25B.

	Native	Opt.	Dis.	Fast	Middling	Swap
Native	1	0.970236	0.055364	0.395341	0.0095414	0.037993
Opt.	0.970236	1	0.052926	0.394039	0.0070749	0.032462
Dis.	0.055364	0.052926	1	0.232461	0.0121693	0.021353
Fast	0.395341	0.394039	0.232461	1	0.0415139	0.079859
Middling	0.009541	0.007075	0.012169	0.041514	1	0.112639
Swap	0.037993	0.032462	0.021353	0.079859	0.1126393	1

Table 7.14: P-Values comparing samples found in figure 4.26B.

	Native	Opt.	Dis.	Fast	Middling	Swap
Native	1	0.125453	0.044772	0.40586	0.012315184	0.445166
Opt.	0.1254532	1	0.034126	0.575871	0.000622183	0.017871
Dis.	0.044772	0.034126	1	0.146201	0.000983251	0.009833
Fast	0.4058603	0.575871	0.146201	1	0.014663584	0.185071
Middling	0.0123152	0.000622	0.000983	0.014664	1	0.003757
Swap	0.4451658	0.017871	0.009833	0.185071	0.003757075	1

Table 7.15: P-Values comparing samples found in figure 5.1B.

	Nat24	Opt24	Dis24	Nat48	Opt48	Dis48
Nat24	1	0.3389245	0.7183938	0.0177006	0.4864792	0.1569959
Opt24	0.3389245	1	0.3743728	0.0604035	0.2749322	0.0896061
Dis24	0.7183938	0.3743728	1	0.0052549	0.4041409	0.0736698
Nat48	0.0177006	0.0604035	0.0052549	1	0.0696982	0.0064355
Opt48	0.4864792	0.2749322	0.4041409	0.0696982	1	0.9107369
Dis48	0.1569959	0.0896061	0.0736698	0.0064355	0.9107369	1

Table 7.16: P-Values comparing samples found in figure 5.2B.

	Nat24	Opt24	Dis24	Nat48	Opt48	Dis48
Nat24	1	0.04662434	0.13855381	0.0495119	0.99386933	0.05795592
Opt24	0.04662434	1	0.03633824	0.9406843	0.09183202	0.00953591
Dis24	0.13855381	0.03633824	1	0.04932994	0.26365447	0.0127097
Nat48	0.0495119	0.9406843	0.04932994	1	0.0932418	0.01050719
Opt48	0.99386933	0.09183202	0.26365447	0.0932418	1	0.09690886
Dis48	0.05795592	0.00953591	0.0127097	0.01050719	0.09690886	1

Table 7.17: P-Values comparing samples found in figure 5.3B.

	Opt24	Dis24	Nat48	Opt48	Dis48
Opt24	1	0.722776	0.373562	0.599769	0.972094
Dis24	0.722776	1	0.798406	0.905003	0.704621
Nat48	0.373562	0.798406	1	0.415554	0.236769
Opt48	0.599769	0.905003	0.415554	1	0.123084
Dis48	0.972094	0.704621	0.236769	0.123084	1

Table 7.18: P-Values comparing samples found in figure 5.4B.

	Nat24	Opt24	Dis24	Nat48	Opt48	Dis48
Nat24	1	0.3816501	0.735932134	0.013585	0.012173	0.060818
Opt24	0.3816501	1	0.25363933	0.0074293	0.0090371	0.0423265
Dis24	0.7359321	0.2536393	1	0.0150295	0.0129653	0.0682743
Nat48	0.013585	0.0074293	0.015029522	1	0.0895537	0.319093
Opt48	0.012173	0.0090371	0.012965292	0.0895537	1	0.0746528
Dis48	0.060818	0.0423265	0.068274322	0.319093	0.0746528	1

Table 7.19: P-Values comparing samples found in figure 5.5B.

	Nat-Lys	Nat-SN	Opt-Lys	Opt-SN	Dis-Lys	Dis-SN
Nat-Lys	1	0.064842	0.23873	0.014100434	0.122783	0.01430719
Nat-SN	0.064842	1	0.282612	0.668973967	0.283613	0.41572291
Opt-Lys	0.23873	0.282612	1	0.142266999	0.828782	0.108198726
Opt-SN	0.0141	0.668974	0.142267	1	0.104862	0.397999269
Dis-Lys	0.122783	0.283613	0.828782	0.10486245	1	0.078951272
Dis-SN	0.014307	0.415723	0.108199	0.397999269	0.078951	1

Table 7.20: P-Values comparing samples found in figure 5.6B.

	Nat-Lys	Nat-SN	Opt-Lys	Opt-SN	Dis-Lys	Dis-SN
Nat-Lys	1	0.105992	0.007933	0.343664849	0.031449	0.100827571
Nat-SN	0.105992	1	0.002399	0.245269228	0.038999	0.168802319
Opt-Lys	0.007933	0.002399	1	0.544186097	0.00327	0.016926468
Opt-SN	0.343665	0.245269	0.544186	1	0.153258	0.147349806
Dis-Lys	0.031449	0.038999	0.00327	0.15325762	1	0.653513054
Dis-SN	0.100828	0.168802	0.016926	0.147349806	0.653513	1

Table 7.21: P-Values comparing samples found in figure 5.8B.

	NOGO-Lys	NOGD-Lys	NOGS-Lys	sNOGO-SN	sNOGD-SN	sNOGS-SN
NOGO-Lys	1	0.06392447	0.920232265	0.00821697	0.47684884	0.13694333
NOGD-Lys	0.063924472	1	0.037421276	5.75208E-05	0.00014691	0.00030693
NOGS-Lys	0.920232265	0.03742127	1	0.00423385	0.41860109	0.07183101
sNOGO-SN	0.008216972	5.75208E-05	0.004233858	1	8.08888E-05	0.00029746
sNOGD-SN	0.476848844	0.00014691	0.418601093	8.08888E-05	1	0.000977373
sNOGS-SN	0.136943333	0.00030693	0.071831019	0.00029746	0.00097737	1

Table 7.22: P-Values comparing samples found in figure 5.9B.

	NOGO-Lys	NOGD-Lys	NOGS-Lys	sNOGO-SN	sNOGD-SN	sNOGS-SN
NOGO-Lys	1	0.01177017	0.040225955	0.00131071	0.02220894	0.01030005
NOGD-Lys	0.01177017	1	0.001008125	3.59E-05	0.01933508	3.2425E-05
NOGS-Lys	0.040225955	0.00100812	1	4.6744E-05	0.02401199	5.8639E-05
sNOGO-SN	0.001310712	3.59E-05	4.67484E-05	1	0.000109988	0.00012980
sNOGD-SN	0.022208947	0.01933508	0.024011998	0.00010998	1	0.00034734
sNOGS-SN	0.010300055	3.242E-05	5.86394E-05	0.00012980	0.00034734	1

Table 7.23: P-Values comparing samples found in figure 5.10B.

	NOGO-Lys	NOGD-Lys	NOGS-Lys	sNOGO-SN	sNOGD-SN	sNOGS-SN
NOGO-Lys	1	0.31159438	0.473223402	0.06208976	0.32862389	0.1552034
NOGD-Lys	0.31159438	1	0.079060147	0.01774267	0.19057945	0.0003905
NOGS-Lys	0.47322340	0.07906014	1	0.02095428	0.09714332	0.0022369
sNOGO-SN	0.06208976	0.01774267	0.020954283	1	0.01805813	0.0829590
sNOGD-SN	0.32862389	0.19057945	0.097143329	0.01805813	1	0.0003259
sNOGS-SN	0.15520349	0.000390592	0.002236925	0.082959079	0.000325957	1

Table 7.24: P-Values comparing samples found in figure 5.11B.

	NOGO-Lys	NOGD-Lys	NOGS-Lys	sNOGO-SN	sNOGD-SN	sNOGS-SN
NOGO-Lys	1	0.13185707	0.12501667	0.009370885	0.295601282	0.010640246
NOGD-Lys	0.13185707	1	0.17656961	0.003927038	0.057420336	1.17069E-06
NOGS-Lys	0.125016672	0.176569612	1	0.003895284	0.050710929	1.13403E-05
sNOGO-SN	0.009370885	0.003927038	0.00389528	1	0.004641663	0.035323246
sNOGD-SN	0.295601282	0.057420336	0.05071093	0.004641663	1	0.000538038
sNOGS-SN	0.010640246	1.17069E-06	1.134E-05	0.035323246	0.000538038	1