University of Nottingham

UK | CHINA | MALAYSIA

Engineering a Mastoparan Peptide Concatemer Prodrug From CircRNA for Cancer Therapy

Declan Grewcock

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

July 2022

Abstract

CircRNAs are covalently closed loops of RNA formed as products of RNA backsplicing in mammalian cells. Engineered circRNAs containing a desired coding sequence have been produced using self-splicing introns. Translatable circRNAs require an internal ribosomal entry site or m^6A methylation site for translation initiation. CircRNAs with a nucleotide length a multiple of three, a start codon, and no stop codon in the same frame have an infinite open reading frame. This project aimed to produce a mastoparan peptide concatemer prodrug from circRNA for treatment in cancer therapeutics. *Anabaena* group I self-splicing introns were used to circularise a mastoparan prodrug containing a metalloproteinase cleavage site for activation (construct named *Anabaena* Mastoparan). RNA circularisation was achieved *in vitro* but not in mammalian cells, indicating that group I *Anabaena* introns do not have the catalytic ability to splice in mammalian cells. Mastoparan peptides were detected *in vitro* and *in vivo* after adding a Flag tag to the *Anabaena* Mastoparan construct. However, only peptides produced from unspliced RNA translation were detected. Mastoparan peptides extracted from *Anabaena* Mastoparan transfected cells caused cytotoxicity when added to the culture medium of MDA-MB-231 and MCF-7 cells. *Anabaena* Mastoparan transfection did not directly lead to cytotoxicity, demonstrating the effectiveness of mastoparan as a prodrug, only being activated by metalloproteinase cleavage in the extracellular environment.

This project aimed to identify endogenous circRNAs that have the coding potential to produce a peptide with a different biological function to their parent gene. Using a Bioinformatics approach, circRNAs containing an ORF through the circular junction were identified. Their ORF through junction peptides were investigated for differences in predicted function to their parent gene using InterProScan and Protein Homology/analogY

Recognition (Phyre2). Using this approach, four candidate circRNAs were identified that encode a predicted peptide with a different biological function to their parent gene. The four candidate circRNAs contain either a predicted m^6A or an internal ribosomal entry site for translation initiation, and have a codon adaption index score (CAI) between 0.781 and 0.821, comparable to the 75th percentile of ORFs through the circular junction (079), and the mean CAI score of coding sequence mRNA. This project demonstrates that the circular junction of circRNAs can provide the coding potential to produce unique peptides with a different function to their parent gene.

Acknowledgements

First, I would like to thank my primary supervisor Keith Spriggs for allowing me to work in his lab. Thanks for always being approachable to offer valuable advice. His support and encouragement motivated me through difficult patches. I'm extremely grateful for his encouragement to learn bioinformatics research approaches during COVID lockdown. Without this encouragement, I would have never explored an area of research that I thoroughly enjoy and wish to continue in future endeavours.

In addition, thanks to my secondary supervisor Hilary Collins for introducing me to cell culture and providing support throughout the project.

Thank you to members of the KAS group, Maksim Ivanov, Olubusola Olaleye and Joe Tomlinson, for their presence in weekly lab meetings, and for asking inquisitive questions relating to my research, and after journal club presentations.

Thank you to all the RNA Biology and Gene Regulation lab members. They have made my time working in the lab a pleasant experience. In particular, a special thanks to my lab buddy Abdulmohsin Alamoudi, an extremely thoughtful individual with whom we have shared many laughs. Thanks to Chris Roberts for showing interest in my project and brainstorming ideas with me.

I want to thank my wife, Paula, for pushing me to achieve and believe in myself. Thanks to my parents for their constant support, I am extremely grateful to have an incredibly happy home life, allowing me to focus on my PhD fully.

Table of Contents

List of Figures

List of Tables

Abbreviations

AF – *Anabaena* FRET

AIRES – Artificial internal ribosomal entry site

AM – *Anabaena* Mastoparan

AMF – *Anabaena* Mastoparan Flag

CAI – Codon adaption index

CELF5 – CUGBP Elav-Like Family Member 5

circFRET – circular RNA FRET

circMast – circular RNA Mastoparan

circSDFLAG – Circular RNA Shine Dalgarno Flag

COX1 - Cytochrome c oxidase subunit I

CSP – Circumsporozoite protein

ECM – Extracellular matrix

EGFP – Enhanced green fluorescent protein

eIF2 – Eukaryotic Initiation Factor 2

eIF4F – Eukaryotic initiation factor 4F

EPDR – Ependymin-related protein

ER – Endoplasmic reticulum

FRET – Fluorescence Resonance Energy Transfer

GFP – Green fluorescent protein

GMP – Guanosine-5'-monophosphate

GPI – glycosylphosphatidylinositol

GTP – Guanosine-5'-triphosphate

HA – Haemagglutinin

His – Histidine

HMMs – Hidden Markov models

HSPGs – Heparan sulfate proteoglycans

IORF – Infinite open reading frame

IRES – Internal Ribosomal Entry Site

ITAFs – IRES trans-acting factors

LB – Luria Bertani

LF – Linear FRET

LM – Linear Mastoparan

lncRNAs – long noncoding RNAs

LTBP1 – Latent-transforming growth factor beta-binding protein 1

m^6A – N6-methyladenosine

MIRES – m^6A-induced ribosome engagement sites

MMP – Matrix Metalloproteinases

ORFTJ – ORF through the junction

PABPI – polyA binding protein I

PANTHER– ANalysis THrough Evolutionary Relationships

PCR – Polymerase chain reaction

Phyre – Protein Homology/analogY Recognition

PIE – Permuted Intron Exon

PLMVd – Peach Latent Mosaic Viroid

pre-mRNA – precursor mRNA

PVDF – Polyvinylidene difluoride

RIMS2 – Regulating Synaptic Membrane Exocytosis 2

rRNA – ribosomal RNA

SD – Shine Dalgarno

SDFLAG – Coding sequence containing a Shine Dalgarno and Flag sequence

SDS – Sodium dodecyl sulfate

SFRP4 – Secreted frizzled-related protein 4

snRNPs – Small nuclear ribonucleoproteins

SRAMP – Sequence-based N6-methyladenosine (m^6A) modification site predictor

SRP – Signal-recognition particle

TRBC2 – T Cell Receptor Beta Constant 2

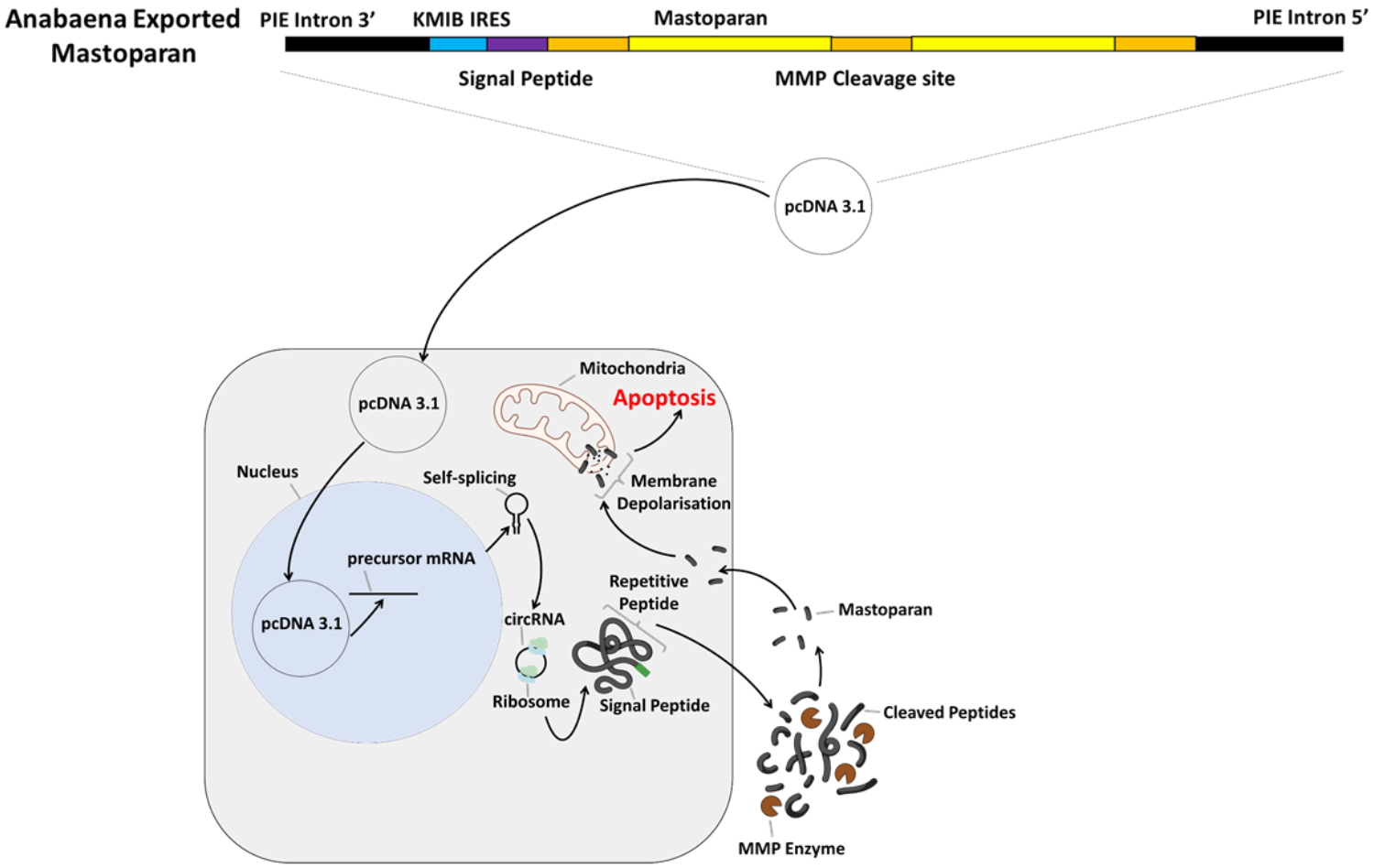
UTR – Untranslated region

YPD – Yeast Extract–Peptone–Dextrose

Chapter One

1.0 Introduction

Graphical Abstract



CircRNAs are covalently closed loops of RNA that form as products of precursor RNA backsplicing in mammalian cells (Jeck et al., 2012). They lack free ends, resulting in greater resistance to exonucleases (Xiao and Wilusz, 2019). Endogenous circRNAs have diverse biological functions with many regulatory roles such as transcription regulation, splicing regulation, and acting as a miRNA sponge (Verduci et al., 2021). Recently, circRNAs have been discovered that are translatable, with over 216 confirmed experimentally (Li et al., 2021). circRNAs lack a 7-methylguanylate cap for cap-dependent translation and must contain either an Internal Ribosomal Entry Site (IRES) or an N6-methyladenosine (m^6A) methylation site for internal ribosomal recruitment to initiate translation (Yang et al., 2017). Designer circRNAs can be engineered to produce circRNAs of a desired sequence and function by rearrangement of 'ribozyme' self-splicing introns flanking the target sequence for circulation.

Self-splicing introns catalyse splicing reactions without spliceosomes (Lambowitz and Zimmerly, 2010). Utilising this system, exogenous translatable circRNAs can be formed in cells after transfecting a vector plasmid (Perriman and Ares, 1998; Wesselhoeft, Kowalski and Anderson, 2018). If translatable circRNAs have a length a multiple of 3, a start codon and lack a stop codon in the same frame, the circRNA has an infinite open reading frame (IORF), resulting in the rolling circle translation in which the ribosome translates around circRNA many times, producing a repetitive peptide concatemer. A large green fluorescent protein (GFP) repetitive peptide concatemer over 300 kDa has been produced from engineered circRNAs with an IORF (Perriman and Ares, 1998). This system has therapeutic applications by producing a peptide concatemer prodrug that requires activation via protease cleavage.

Cancer is the uncontrolled growth of abnormal cells. Cancer is further identified by the tissue from which the abnormal cells originate, e.g., breast cancer (Cooper and Hausman, 2000).

Although the basic principle of uncontrolled cell growth can be applied to any cancer type, the behaviour of cancer at different stages and in different tissue sites can vary vastly. The protease expression of cancer types is one of these many biological variances (Mason and Joyce, 2011). Matrix Metalloproteinases (MMPs) are calcium-dependent zinc-containing endopeptidases that degrade extracellular matrix proteins and process bioactive molecules (Verma and Hansch, 2007). They play a significant role in cancer hallmarks, including migration, invasion, metastasis, and angiogenesis (Foda and Zucker, 2001). Several MMPs are ubiquitously overexpressed across multiple cancer types, giving the potential for pan-cancer treatments where MMPs are the prodrug target (Gobin et al., 2019).

Mastoparan is a 14 aa peptide that acts as a peptide toxin in the defence system of *Vespula lewisii*, Korean yellow jacket wasp. It has been shown to have anti-cancer properties (Hilchie et al., 2016). It induces intrinsic apoptosis in its non-amidated form (COOH) and interacts directly with the membrane, causing cell lysis in its amidated form (NH_2) (de Azevedo et al., 2015; Hilchie et al., 2016).

This project aims to produce a mastoparan peptide concatemer from circRNA formed and translated inside cancer cells, which is then exported into the extracellular environment, activated by MMP cleavage, and causes the cell death of cancer cells.

1.1 Cancer

Cancer is a major global health concern, with an estimated 19.3 million new cases and 10 million deaths in 2020, making it the second leading cause of death globally (World Health Organization [WHO], 2020). In the UK, cancer accounts for around 28% of all deaths

(Parliament. House of Commons, 2023). Lung cancer is the leading cause of cancer death worldwide, followed by liver, stomach, colorectal, and breast cancer (WHO, 2020).

Cancer is a disease that arises due to genetic mutations in a single cell that disrupt its normal function and trigger uncontrolled growth (Hassanpour, 2017). These mutations can be caused by various factors, including environmental exposures, genetic predisposition, and lifestyle factors (Hanahan & Weinberg, 2011). The development of cancer involves several steps, including initiation, promotion, and progression (Compton, 2020). Initiation occurs when genetic mutations occur in a driver gene of an adult stem cell, the mutated cell typically makes the cell more susceptible to mutation (Grizzi et al., 2006). Promotion involves the clonal expansion of the mutated cell. During this stage, cells acquire additional mutations, resulting in the formation of unique subclones (Grizzi et al., 2006). This continually repeats as the nascent cancer grows in a phase of cancer development known as progression.

Cancer cells continue to divide and proliferate, forming tumours and invading nearby tissues and organs. This uncontrolled growth is driven by the acquisition of additional genetic and epigenetic alterations that promote tumour growth and survival, including the activation of oncogenes and the inactivation of tumour suppressor genes (Sadikovic et al., 2008). Oncogenes promote cell growth and division, while tumour suppressor genes inhibit cell growth and promote apoptosis or programmed cell death. Mutations in these genes can disrupt the normal regulation of cell growth and division, leading to uncontrolled proliferation (Sadikovic et al., 2008).

Metastasis occurs when cancer cells spread from the primary tumour to other parts of the body through the bloodstream or lymphatic system (Fares, 2020). Metastasis is a complex process that requires cancer cells to acquire various characteristics that enable them to

survive and grow in new locations. These include the invading surrounding tissues, resisting immune surveillance, and establishing a blood supply to support their growth (Fidler, 2003).

1.1.1 Breast Cancer

Breast cancer is the development of cancer in the breast tissue. Breast cancer is the most diagnosed cancer, accounting for 11.7% of new cancer cases and 6.9% of all cancer deaths in 2020, worldwide (Sung et al., 2021). The incidence of breast cancer varies by region, with the highest incidence rates observed in North America, Europe, and Australia/ New Zealand (Arnold et al., 2022).

Breast cancer is a heterogeneous disease that comprises various subtypes based on histopathological and molecular characteristics (Testa et al., 2020). The most common types of breast cancer are invasive ductal carcinoma (IDC) and invasive lobular carcinoma (ILC), which differ in their anatomic location within the breast tissue (American Cancer Society, 2021). IDC originates in the milk ducts and invades the surrounding breast tissue and accounts for 55% of breast cancer incidences upon diagnosis, ILC constitutes 5%–15% of invasive breast carcinoma and originates in the lobules, the milk-producing glands (Makki, 2015). In addition, immunohistochemistry is used to classify breast cancer based on the expression of cell surface receptors such as estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) (Onitilo et al., 2009). Triple-negative breast cancer (TNBC) is a subtype that lacks expression of these receptors and tends to have a more aggressive clinical behaviour (Lehmann et al., 2011). TNBC is associated with a poorer prognosis and limited treatment options compared to other types of breast cancer (Zerdan, 2021). The different types of breast cancer have varying responses to therapy

and clinical outcomes, highlighting the importance of accurate diagnosis and individualised treatment strategies (Harbeck et al., 2019).

Studies of breast cancer often utilise cell culture models, such as MDA-MB-231 and MCF-7 cell lines (Lacroix and Leclercq, 2004). MDA-MB-231 cells are a TNBC cell line widely used as a model for studying the biology of TNBC and for testing potential therapeutic agents (Chavez, Garimella and Lipkowitz, 2012). MCF-7 cells are ER+ breast cancer cell lines commonly used to study the effects of anti-estrogen therapies (Comşa, Cîmpean and Raica, 2015).

1.1.2 Breast Cancer Treatments

Breast cancer is a complex disease with multiple treatment options surgery, radiation therapy, chemotherapy, targeted therapy, and hormonal therapy (Burguin, Diorio and Durocher, 2021). The choice of treatment depends on several factors, including the stage and type of breast cancer, the patient's overall health, and personal preferences (Burguin, Diorio and Durocher, 2021). Breast cancer treatment is highly individualised, and each patient's treatment plan may differ based on their unique circumstances.

Breast cancer is a highly heterogeneous disease, with intra-tumour (within the same tumour mass) and inter-tumour (between distinct tumours) heterogeneity characterised by a diverse array of phenotypic, transcriptomic, epigenetic and genomic characteristics (Fumagalli and Barberis, 2021). This heterogeneity is an essential factor in determining how cancer behaves and responds to treatment, and its overall prognosis (Liang et al., 2020). Treatment must be tailored each patient’s specific subtype and cancer characteristics. For instance, endocrine therapy may be effective for hormone receptor-positive breast cancers (El Sayed et al., 2019),

while targeted therapy such as trastuzumab, a monoclonal antibody, may be useful for HER2-positive cancers (Derakhshani et al., 2019). The heterogeneity of breast cancer also contributes impact on treatment outcomes, as some breast cancers have a higher risk of recurrence and are more aggressive than others (Ahmad, 2013).

The expense of breast cancer treatments is a significant concern for patients, healthcare systems, and payers. The treatment cost depends on the disease stage, the type of treatment, and the healthcare setting. In 2016, the UK National Health Service (NHS) reported that chemotherapy treatment costs an estimated £1.4 billion a year, almost a tenth of the entire central budget (NHS England, 2016). Other treatments such as targeted therapies and immunotherapies can be even more expensive (Burguin, Diorio and Durocher, 2021), highlighting the need for the development of a low-cost, effective treatment.

Drug resistance remains a major clinical challenge to chemotherapy and targeted therapies (Tang et al., 2016). Resistance can arise from various mechanisms, including genetic alterations in tumour cells, changes in signalling pathways, and alterations in the tumour microenvironment (Tang et al., 2016). There is a pressing need for the development of new therapeutic strategies that can overcome resistance and improve patient outcomes.

Breast cancer treatments such as chemotherapy, radiation therapy, and hormone therapy are associated with varying degrees of toxicity (Lowenthal and Eaton, 1996; Wang and Tepper, 2021; Henry, 2014). The toxicity of breast cancer treatments is largely due to their off-target effects, which can damage healthy cells and tissues (Wolfson et al., 2021). Chemotherapy, for instance, can cause hair loss, nausea, and damage to the gastrointestinal tract (Lowenthal and Eaton, 1996). Radiation therapy can cause skin irritation and damage to the heart and lungs

(Wang and Tepper, 2021). Hormone therapy can lead to bone loss and an increased risk of blood clots (Henry, 2014). Careful consideration of breast cancer treatments' potential benefits and risks is necessary to optimise patient outcomes while minimising treatment-related toxicity.

1.1.3 Proteases and Cancer

Proteases are essential in biological events associated with cancer progression, including tumour angiogenesis, invasion, and metastasis (Koblinski, Ahram and Sloane, 2000). The proteolytic network of cancer microenvironments is complex. Key proteases, cathepsin B, urokinase-type plasminogen activator, calpains, and several matrix metalloproteinases, interact with many signalling pathways in tumour biology (Mason and Joyce, 2011). The expression of proteases in the tumour microenvironment is often tissue-specific, and proteases that are overexpressed in some cancers can be repressed in others (Mason and Joyce, 2011). Immunohistochemistry or mRNA expression levels determine protease levels; however, this may not be directly comparable to the levels of active proteases. This is because the levels of associated inhibitors may also be overexpressed, and activation of proteases can be associated with other cellular conditions (Storr et al., 2015).

1.1.3.1 Matrix Metalloproteinases (MMPs)

Matrix metalloproteinases (MMPs) are calcium-dependent zinc-containing endopeptidases that degrade extracellular matrix proteins and process bioactive molecules (Verma and Hansch, 2007). The dysregulation of MMP expression and activity has been implicated in various stages of cancer progression, including the initiation, promotion, invasion, and metastasis of tumours (Lopez-Otin and Matrisian, 2007). MMPs are involved in the degradation of ECM components,

such as collagen, elastin, laminin, and fibronectin, which provide structural support and signalling cues to cells in the microenvironment. The cleavage of ECM components by MMPs releases bioactive fragments, such as growth factors and cytokines, that promote cancer cell survival, proliferation, and migration. MMPs also participate in the modulation of cell-cell and cell-ECM interactions, which are crucial for tumour invasion and metastasis (Kessenbrock, Plaks and Werb, 2010). MMPs are frequently overexpressed in many types of cancer, including breast, lung, colon, prostate, and melanoma. Their expression levels are often correlated with poor prognosis and increased risk of metastasis (Egeblad and Werb, 2002; Kessenbrock, Plaks and Werb, 2010).

MMPs can be sorted into classes, characterised by their modular domain structure (Al-Alem and Curry, 2015). All MMPs share some common functional domains: a signal peptide, a prodomain, and a catalytic domain with a zinc-binding site (Figure 1.1) (Massova, Kotra, Fridman and Mobashery, 1998). All MMPs with the exception of MMP7, 26, and 23 contain a proline-rich hinge and terminal hemopexin-like domain, which is necessary for substrate specificity and interaction with endogenous inhibitors (Laronha and Caldeira, 2020). The N-terminal signal peptide (or pre-domain) is removed once the enzyme's synthesis is directed to the endoplasmic reticulum. Most MMPs remain in an inactive pro-MMP state inside cells until activated after secretion into the extracellular environment (Hadler-Olsen et al., 2010). Pro-MMPs contain a conserved PRCG(V/N) sequence, in which a catalytic zinc ion forms a covalent bond with the cysteine residue thiol group, known as a cysteine switch (Massova, Kotra, Fridman and Mobashery, 1998). This interaction maintains enzyme latency until an enzyme such as fibrinolytic protease plasmin, serine protease tissue kallikrein, thrombin, trypsin or other MMPs causes proteolysis of the pro-domain or modification of the cysteine thiol group (Hadler-Olsen et al., 2010). Upon activation, the pro-domain is removed, and the

MMP is in its mature form with its catalytic machinery revealed (Al-Alem and Curry, 2015). The majority of the 23 identified human MMP enzymes are secreted. However, seven are 'membrane-type matrix metalloproteinases' containing a transmembrane domain (type I or II) or a glycosylphosphatidylinositol (GPI) membrane anchor expressed as cell surface enzymes (Itoh, 2015). All membrane-bound MMPs and MMPs 23, 2 and 9 contain a furin-cleavage site for activation during secretion (Loffek, Schilling and Franzke, 2010). Some MMPs are almost ubiquitously upregulated across different cancer types, making them a good prodrug target for pan-cancer treatments (Gobin et al., 2019).

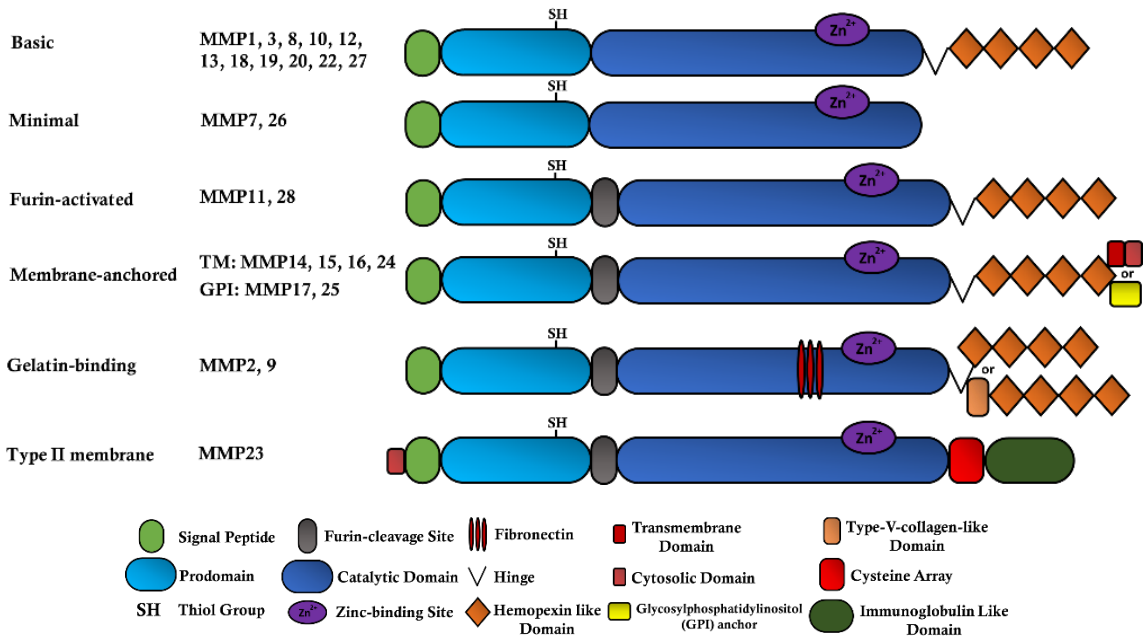


Figure 1.1 The structure of matrix metalloproteinases (MMPs). The MMP family comprises more than 20 members with distinct structural features, substrate specificities, and tissue expression patterns. The general domain structure of MMPs includes a signal peptide, a prodomain, a catalytic domain, a hinge region, and a hemopexin-like domain. The prodomain maintains the enzyme in an inactive state until activation by proteolytic cleavage. The catalytic domain contains the zinc-binding site, essential for enzymatic activity. The hemopexin-like domain is involved in substrate recognition, binding, and interactions with proteins and cell surface receptors. The hinge region connects the catalytic and hemopexin-like domains and contributes to the flexibility of the protein.

1.2 Translation

Translation is the process in which ribosomes in the cytoplasm or endoplasmic reticulum (ER) synthesize proteins from RNA. Translation consists of three stages: initiation, elongation, and termination. Initiation refers to the formation of the 80S ribosome on an mRNA transcript (Merrick and Pavitt, 2018). During elongation, the ribosome reads each codon in the mRNA transcript and adds its corresponding amino acid to the growing chain linked via a peptide bond (Dever and Green, 2012). Termination refers to the ceasing of translation when the ribosome reaches a stop codon (UAA, UAG, and UGA) (Schueren and Thoms, 2016).

1.2.1 Translation Control

Translation regulation is the control of protein synthesis. Protein synthesis takes up a substantial proportion of cellular energy resources. It is estimated that translation accounts for ~30% of energy consumption in differentiating mammalian cells (Buttgereit and Brand, 1995). Protein synthesis, therefore, requires tight regulation to avoid unnecessary energy expenditure. As eukaryotic mRNAs have quite long half-lives (>2 h), rapid changes to protein levels must be regulated by mRNA translation efficiency rates and protein degradation (Raghavan et al. 2002). These changes are demonstrated by the low correlation between cellular mRNA and protein levels (Ghazalpour et al., 2011). Rapid changes in protein expression are particularly important in response to abrupt changes to the cellular microenvironment allowing cellular adaptation and restoring homeostasis (Sonenberg and Hinnebusch, 2009). Ribosomes are the universal molecules required for protein synthesis. The ribosome recruitment to mRNA is often the rate-limiting step of protein synthesis (Aitken and Lorsch, 2012). Therefore, translational control mechanisms are geared toward preventing ribosomal-mRNA binding and initiation (Jackson, Hellen and Pestova, 2010).

1.2.2 Cap-dependent Translation Initiation

Precursor mRNA (pre-mRNA) is formed in the cell nucleus by DNA transcription. pre-mRNA undergoes three main processing steps to be converted into its mature translatable form, capping at the 5' end, adding a poly(A) tail at the 3' end, and splicing to remove introns. During capping of the newly formed pre-mRNA, an N7-methylguanosine cap is added at the 5' terminus (Walters and Thompson, 2016). The translational machinery recognises the cap structure and allows translation initiation to proceed (Scheper and Proud, 2002).

After transcription of pre-mRNA and processing into mRNA, the mRNA is transported through the nuclear envelope and into the cytoplasm, the site of translation (Piñol-Roma and Dreyfuss, 1992). The localisation of these processes results in an uncoupled translation-transcription system. During initiation, Eukaryotic Initiation Factor 2 (eIF2) forms a ternary complex with GTP and the initiator Met-tRNAi (Sokabe and Fraser, 2014). This ternary complex is bound to the 40S ribosomal subunit and forms the 43S preinitiation complex (Figure 1.2). Eukaryotic initiation factor 4F (eIF4F) loads mRNA onto the 43S preinitiation complex. eIF4F is composed of three eukaryotic initiation factors (eIF4A, eIF4G and eIF4E). Each protein plays a distinguishable role in translation initiation: eIF4E recognises the cap structure (Sonenberg et al., 1978), and eIF4A is a DEAD-box RNA helicase that unwinds the mRNA removing its secondary structures (Schmid and Linder, 1991). eIF4G is a scaffold protein that binds to eI4FA, eIF3 (which is bound to the 40S subunit), and polyA binding protein I (PABPI) (which in turn binds the messenger RNA's polyA tail) (Jackson, Hellen and Pestova, 2010).

Three paralogs of eIF4A exist named eIF4AI (eIF4A1), eIF4AII (eIF4A2) and eIF4AIII (eIF4A3) (Cunningham, Chapman and Schatz, 2018). eIF4AI and II are structurally similar, containing an amino acid sequence with 90-95% homology, and both are present in the

cytoplasm with no particular localisation (Bordeleau et al., 2005; Yang et al., 2003). eIF4AI and eIF4AII contribute to translation initiation, forming a complex which binds to eIF4G. Although they have a high degree of similarity (90% amino acid sequence), they are thought to function differently (Nielsen and Trachsel, 1988; Galicia-Vázquez et al., 2012). This is supported by the differential expression of the two genes between tissues and eIF4A1 being essential for cell viability, whereas eIF4A2 is not (Galicia-Vázquez, Chu and Pelletier, 2015). eIF4AIII amino acid sequence is less similar (65% similarity). It functions in RNA metabolism, including coupling mRNA splicing to translation, localisation, and export (Mazloomian et al., 2019).

The 43S preinitiation complex loaded to the mRNA begins scanning unidirectionally from 5′ to 3′ until it reaches a start codon (most commonly AUG) (Hinnebusch, 2014). Upon its encounter, the larger 60S ribosomal subunit binds, forming the 80S ribosome and beginning elongation.

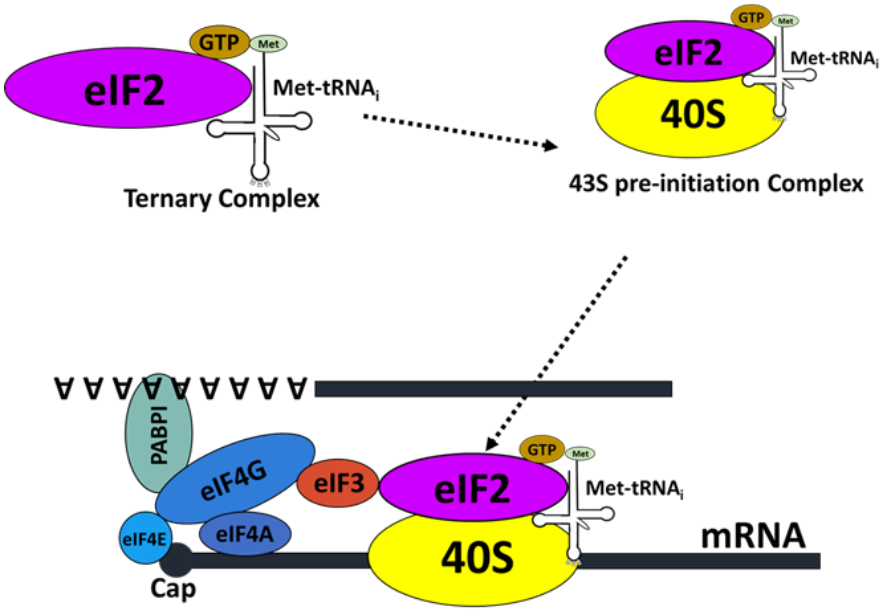


Figure 1.1 Cap-dependent translation initiation. The ternary complex composed of eIF2-GTP, initiator methionyl-tRNA (Met-tRNAi), and eIF2B binds to the 40S ribosomal subunit forming the 43S preinitiation complex (PIC). The 43S PIC then binds to the mRNA, and the small ribosomal subunit scans the mRNA for the start codon. eIF – eukaryotic initiation factor; PABP1 – Poly(A)-binding protein 1; GTP – Guanosine triphosphate; Met – Methionine.

1.2.3 Cap-independent Translation Initiation

Not all translation is driven by the recruitment of the ribosome to the N7-methylguanosine cap. Under cellular stress where cap-dependent translation is inhibited, cap-independent methods become prominent (Thakor and Holcik, 2011). Internal Ribosomal Entry Site (IRES) mediated translation is a common method of cap-independent translation. IRESs are cis-acting RNA elements that recruit the 43S preinitiation complex near or on the start codon without the requirement of the 5′ cap (Rijnbrand et al., 1996). Although efficiencies vary between IRESs, IRES-mediated translation is less efficient than cap-dependent translation (Thoma et al., 2004) (Figure 1.3).

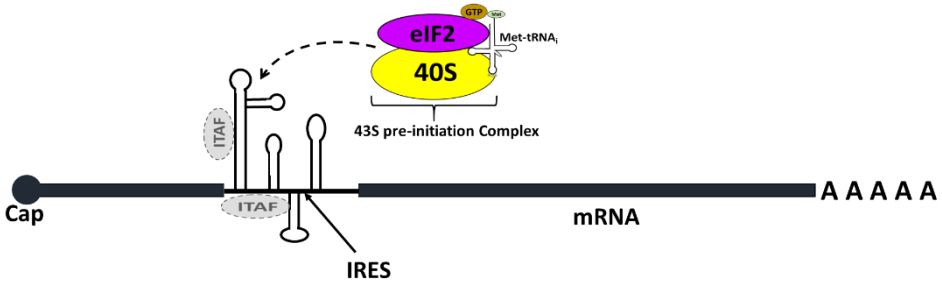


Figure 1.3 IRES mediated translation initiation. The internal ribosome entry site (IRES) present within the mRNA bypasses the need for a 5' cap structure and allows cap-independent translation initiation. The IRES recruits the 43S preinitiation complex with the aid of IRES trans-acting factors (ITAFs) which bind to specific regions of the IRES to facilitate efficient translation initiation. Upon binding to the IRES and ITAFs, the 43S preinitiation complex scans the mRNA for the start codon and initiates protein synthesis. IRES – Internal ribosome entry site; eIF – eukaryotic initiation factor; ITAF – IRES trans-acting factor; GTP – Guanosine triphosphate; Met – Methionine.

Whereas cap-dependent initiation occurs by the same mechanism for all eukaryotes, the mechanism of IRES-dependent translation can differ in the number of IRES trans-acting factors (ITAFs) required to facilitate translation (King, Cobbold and Willis, 2010). ITAFs are proteins that specifically bind to IRES sequences and facilitate ribosome recruitment and translation initiation. They can also interact with other cellular proteins and translation factors to modulate the efficiency of translation initiation (King, Cobbold and Willis, 2010). Generally, the more compact an IRES is, the fewer ITAFs are required for translation initiation (Godet et al., 2019).

Cellular IRESs (cIRESs) are present in the 5' untranslated region (UTR) and have no detectable sequence similarity or predictable secondary structure (Baird et al., 2007), making them extremely difficult to be identified by computational approaches.

It has been estimated that 10% of cellular mRNAs possess IRES activity (Spriggs et al., 2005). IRES-mediated translation has been shown to play a key role in pathology, especially when IRESs are located in mRNAs coding master regulators of cell survival, proliferation, and death (Piccirillo et al., 2014).

1.2.4 Artificial Internal Ribosome Entry Site (AIRES)

Spriggs, Mitchell and Willis, (2005) demonstrated that repeat CCU motifs in stem-loop structures form efficient IRESs in eukaryotes. A functional, minimalistic (110 nt) artificial internal ribosomal entry site (AIRES) was engineered with high translation efficiency in a mammalian expression system. It contained a 30-nucleotide spacer region between the hairpin and the start codon and mismatch base pairs in the (GGA)n half of the hairpin (Figure 1.4).

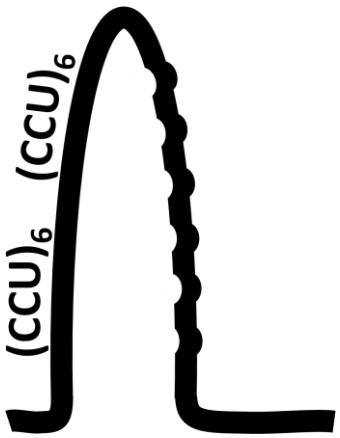


Figure 1.4 KMIB IRES. The KMIB artificial IRES (AIRES) acts as a minimalistic functional IRES and is composed of CUU repeat motifs and GGA mismatch base in a hairpin loop followed by a spacer region at least 30-nucleotide in length before a start codon.

1.2.5 m^6A Methylation

N6-methyladenosine (m^6A) is the most abundant form of internal RNA modification in higher eukaryotes, with an average of one to three m^6A modifications per transcript (Zaccara, Ries and Jaffrey, 2019). m^6A methylation adds a methyl group on the nitrogen-6 position of an adenosine molecule which is catalysed by the methyltransferase complex, composed of METTL3, METTL14, WTAP and KIAA1429 subunits (Schwartz et al., 2014). m6A modification is highly dynamic and is regulated by a group of enzymes called "writers," "erasers," and "readers." Writers are responsible for the addition of the methyl group to the adenosine base (Schwartz et al., 2014), while erasers remove the methyl group (Yang et al., 2018). Readers bind to m6A-modified mRNA and initiate downstream signalling pathways to regulate gene expression (Han et al., 2021). The dynamic nature of this modification plays an essential role in various biological processes, such as RNA processing, stability, and translation (Han et al., 2021). The consensus motif for the m^6A sequence is "RRm^6ACH" (R = G or A; H = A, C, or U) (Csepany, Lin, Baldick and Beemon, 1990). During cap-dependent translation, m^6A readers can facilitate the recruitment of translation initiation factors to the mRNA, thereby promoting ribosome binding and protein synthesis. Specifically, the m6A reader YTHDF3 can enhance cap-independent translation by interacting with the eIF3 complex (Chang et al., 2020). Studies have shown that a single m^6A site is sufficient to initiate translation (Meyer et al., 2015). m^6A sites that can recruit the ribosome directly in a cap-independent manner are named m^6A-induced ribosome engagement sites (MIRES) (Meyer et al., 2015). The mechanism that m^6A sites use to promote cap-independent translation varies depending on the template type; mRNA or circRNA. When present in the 5' UTR of mRNA, they can promote non-AUG translation by slowing the speed of ribosome scanning. m^6A-driven translation initiation is not well understood on circRNAs, but studies show it involves the m^6A reader YTHDF3 and the translation initiation factors eIF4G2 and eIF3A (Yang et al., 2017). m^6A methylation can also

promote cap-independent translation by directly binding to EIF3 and recruiting the 43S preinitiation complex to the transcript mRNA (Meyer et al., 2015). This mechanism has only been found when present in the 5' UTR on an mRNA transcript and not internally (Meyer et al., 2015). Therefore, it is unknown whether this is a mechanism of internal ribosomal recruitment on circRNA.

1.3 Splicing

Splicing refers to the reaction in which intronic regions in pre-mRNA are removed, and the exons are joined together (Figure 1.5). For nuclear-encoded genes, the splicing process occurs after transcription in the nucleus. Splicing is essential to remove the non-coding sequences so that the translated mature mRNA produces a functional protein (Clancy, 2008).

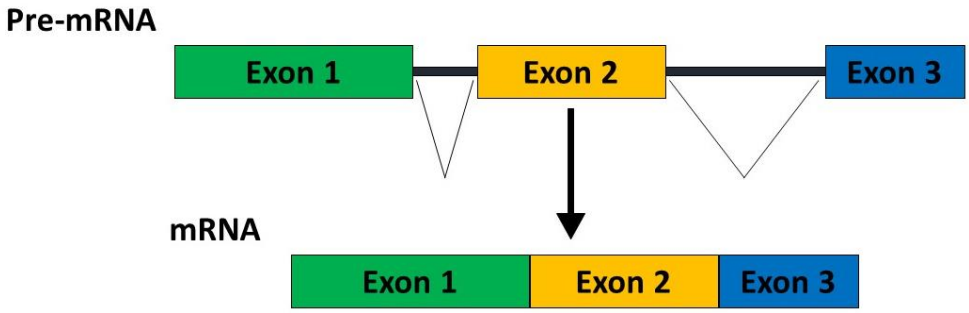


Figure 1.5 Pre-mRNA splicing. The removal of introns from precursor messenger mRNA (Pre-mRNA) and the joining of exons to form mature messenger RNA (mRNA).

1.3.1 Canonical Splicing

Canonical splicing in mammalian cells refers to the mechanism by which introns are removed from the pre-mRNA in a spliceosome-dependent manner. The spliceosome complex comprises five small nuclear ribonucleoproteins (snRNPs) and many other protein-associated factors (Will and Lührmann, 2001). The spliceosome recognises the acceptor and donor splice sites

and catalyses the reaction. All precursor mRNAs have conserved regions that are essential for the splicing reaction. 99% of introns obey a GU-AG rule, in which a dinucleotide pair forms the 5' (GU) donor and 3' acceptor (AG) splice site junctions between the intron-exon boundary (Zheng, 2004). The intron has two other semi-conservative regions, a polypyrimidine tract region composed of C and U repeats upstream of the acceptor site and a branch site required for lariat formation (Coolidge, Seely and Patton, 1997). Pre-mRNA undergoes a two-step transesterification process to excise the intron into its mature form (Padgett et al., 1986). In the first step, the 5' splice site is cleaved, and the first base of the intron forms a lariat with the adenosine branch site. Next, the first base of the 3' exon forms a 2', 5' phosphodiester bond with the last base of the upstream exon, which joins them, excising the intron lariat (Figure 1.6).

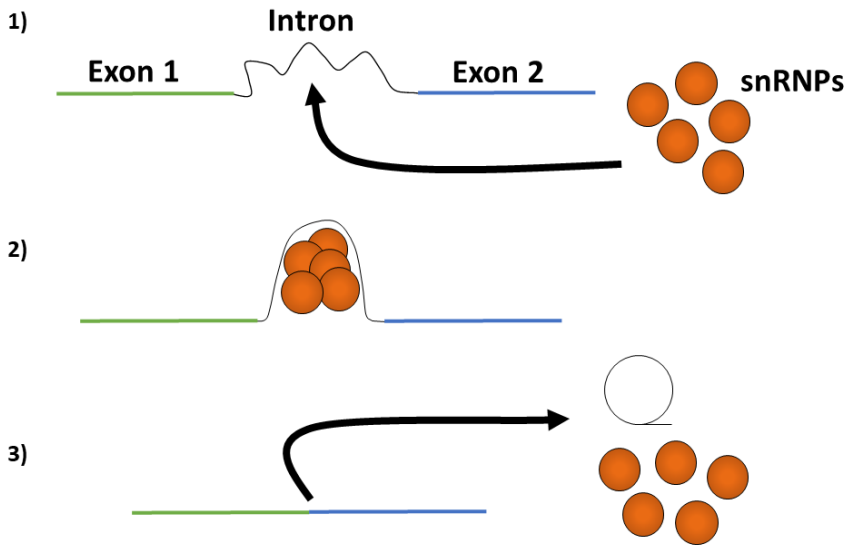


Figure 1.6 Canonical splicing in Eukaryotes. 1) 5 snRNPs form the spliceosome complex which associates with the pre-mRNA 2) The spliceosome-pre-mRNA complex brings the splice sites closer together and catalyses the splicing reaction. 3) The intron splices into a lariat and is excised, and the spliceosome complex is disassembled.

1.3.2 Group I Introns

Not all splicing events are spliceosome dependent. Some introns have ribozyme activity that allow self-splicing without additional proteinaceous components. These ribozymes have been extensively studied as they provide information on DNA genome evolution and spliceosome origin and can be manipulated into producing circular RNAs with desired functions (Lambowitz and Zimmerly, 2010; Perriman and Ares, 1998)

Group I self-splicing introns are widely present in viral, bacterial, bacteriophage, organellar genomes, and ribosomal RNA (rRNA) of eukaryotes (Hausner, Hafez and Edgell, 2014). Group I intron of Tetrahymena was the first discovered intron to display self-splicing ribozyme activity (Kruger et al., 1982). Only Mg^{2+} and guanosine nucleotides are essential components for catalysis; however, in many instances, host factors are recruited to support the reaction in vivo (Nielsen, 2003; Lambowitz and Zimmerly, 2010). The self-splicing occurs via a two-step transesterification reaction shown in figure 1.7. Initially, the 3' hydroxyl of a free guanosine attacks the phosphate of the 5' splice site, excising the upstream exon (step 1). Next, the 3' hydroxyl of the free exon attacks the 3' splice site, excising the intron and ligating the exons (step 2) (Hedberg and Johansen, 2013).

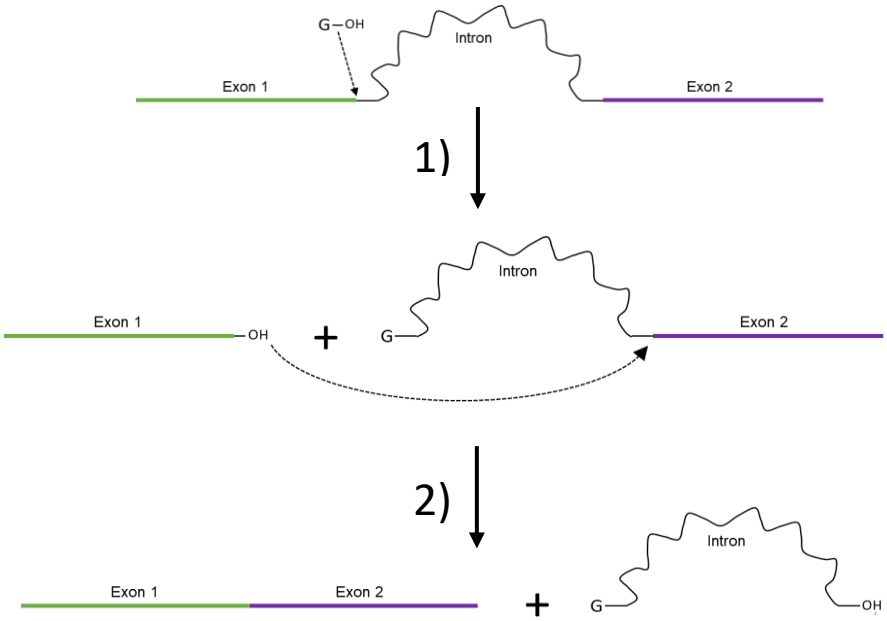


Figure 1.7 Self-Splicing mechanism of group I introns. 1) The 3' hydroxyl of a free guanosine attacks the phosphate of the 5' splice site. 2) Free exon 3' hydroxyl attacks the 3' splice site, the intron is excised, and the exons are ligated. 3) Finally, circular intronic RNA (ciRNA) is formed when the 3' hydroxyl of the excised intron attacks an internal site near the 5' end of the intron. G represents a Guanine nucleotide. Mg^{2+} is also required for the reaction, and host factors may be involved in vivo. All linkages formed are 3', 5'.

1.3.3 Intron Rearrangement and RNA Circularisation

The autocatalytic function of self-splicing introns is not determined by the relative position of the splice sites but by their primary sequence and structure features (Puttaraju and Been, 1992). Therefore, it is possible to reverse the splice sites' order and maintain the intron's autocatalytic capacity while generating unconventional products (Figure 1.8). The rearrangement, known as the permuted intron-exon (PIE) in group I introns, can be engineered to circularise a desired exon sequence (Petkovic and Müller, 2015). The 5' half is downstream of the exon, and the 3' half is upstream. The splice sites and the two-step transesterification reaction in self-splicing introns remain the same as in intron excision. However, the PIE method generates three different products, the second half intron, the first half intron, and a full circRNA (Puttaraju

and Been, 1996). The reaction produces a 3', 5' phosphodiester linkage at the branch site of the circRNA. This is the usual linkage of nucleic acids in the DNA/RNA backbone, making it resistant to debranching (Puttaraju and Been, 1996).

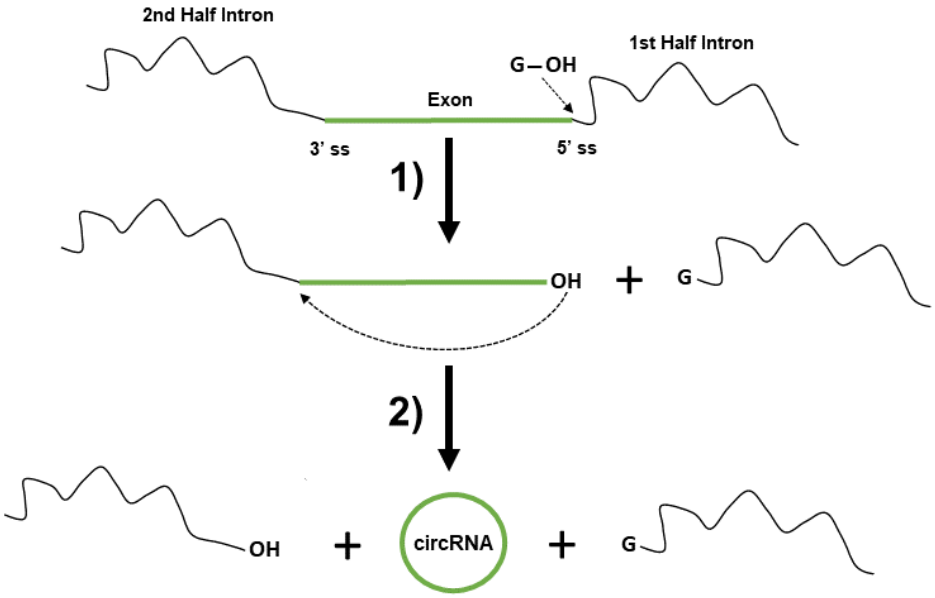


Figure 1.8 Engineering circRNA from PIE method. Group I introns produced a circular human exon with a 3', 5' phosphodiester linkage. Each step indicates a transesterification reaction. 1) The 3' hydroxyl of a free guanosine attacks the phosphate of the 5' splice site. 2) The free hydroxyl at the 3' end of the exon attacks the 3' splice site, generating circRNA and 2 intron RNA productions that are then degraded. Abbreviations: A - bulged adenosine, IVS – Intervening Sequence. Mg^{2+} is also required for the reaction.

Another group of self-splicing introns, group II, can have their intron halves rearranged to produce circRNA in a method known as inversed splicing. Splicing catalysis can occur *in vitro* only with Mg^{2+} as an essential component (Peebles et al., 1995). This method was used to circularise a human exon sequence *in vitro* using group II introns of yeast mitochondrion (Mikheeva et al., 1997). The main differences between the splicing products of group I and group II introns are that 2', 5' phosphodiester bond forms at the adenosine branch site. In addition, the circRNA product that is formed from inverse splicing of group II introns yields a

circular RNA of only the desired sequence, whereas group I introns require insertion of the desired sequence into the naturally occurring exon.

1.4 CircRNAs

CircRNAs are covalently closed loops of RNA formed as the products of backsplicing in mammalian cells. They are predominantly found in the cell cytoplasm and are more resistant to degradation from RNases as they lack free ends (Geng et al., 2020). Endogenous circRNAs were long thought of as non-coding erroneous splicing products. In recent years it has been revealed that circRNAs are abundant and conserved (Jeck et al., 2012). A conservative estimate of 14.4% of linear RNAs also form circRNAs, showing that they contribute significantly to the transcriptome (Salzman et al., 2012). Some circRNAs have greater expression than their linear cognates. The expression profile of circRNA is highly tissue-specific, and aberrant levels can often be associated with disease (Salzman et al., 2013). In cancer, circRNAs carry predictive and prognostic significance; therefore, circRNAs have exciting potential as a non-invasive biomarker for early cancer detection (Li et al., 2015). The biological functions of circRNAs are diverse. Many are involved in gene regulation using a variety of mechanisms (Table 1.1). The classification of endogenous circRNAs as long noncoding RNAs (lncRNAs) has recently been challenged with the discovery of circRNAs capable of being translated (Lei et al., 2020). Designer non-native circRNAs can be engineered to have specific properties such as translation or acting as a miRNA sponge

Table 1.1 Known biological functions of circRNAs.

circRNA Function	Description	Example	Reference
	Interact with RNA Pol II or transcriptional factors	ci-ankrd52 – Positively regulates its parent gene by interacting with an elongation RNA Pol II complex	Zhang et al., 2013
Transcription Regulation	Inducing DNA hypomethylation in the promoter or regulate intronic enhancer	FLI1 exonic circRNA induces DNA hypomethylation in the CpG islands of its promoter	Chen et al., 2018
	cytoplasmic circRNAs regulating the expression of transcriptional factors	circITGA7 suppress transcription factor RREB1 of the Ras pathway to promote transcription of its host gene	Li et al., 2018
Splicing Regulation	circRNAs can affect linear splicing of pre-mRNAs, potentially leading to altered gene expression	circMbl is formed when the MBL protein is in excess, MBL protein binds to the mbl pre-mRNA and causes it to back-splice into circMbl. circMbl then binds to MBL protein so that it can no longer produce circMbl transcripts, consequently lowering its own level	Ashwal-Fluss et al., 2014
miRNA Sponge	Enhances target gene expression by binding to miRNAs and inhibiting their activity	circ-Sirt1 facilitates the expression of host gene SIRT1 by binding to miR-132/212	Kong et al., 2019
Repression of Parental Gene Expression (mRNA trap)	ecircRNA containing the translation start codon in their circular structure which prevents translation of linear transcripts	Formation of circ Formin (FMN1) gene led to decreased expression of the FMN1 protein and affected limb development in mice	Chao, Chan, Kuo and Leder, 1998
Translation Regulation	circRNAs can interact with specific proteins and modulate the translation of their cognate mRNAs	circYap interacts with Yap mRNA, eIF4G and PABP to decrease Yap protein levels in cancer cells	Wu et al., 2019
Post-translational modification Regulation	Competitively interact with enzymes to regulate post-translational modification of the full-length protein	FBXW7-185aa competitively interacts with deubiquitinating enzyme USP28, preventing binding to FBXW7. This suppresses triple-negative breast cancer progression	Ye et al., 2019

1.4.1 Endogenously Translated CircRNAs

216 endogenous circRNAs have been experimentally verified to be translated (Li et al., 2021). Some peptides translated from circRNA have known biological functions and are associated with disease (Wen, Qadir and Yang, 2022). These circRNAs can be identified by ribo-seq when the sequence read spans a unique sequence where the 5' and 3' splice sites are joined, known as the backsplice or circular junction. This technique has identified 1479 mammalian circRNAs translated and present on the riboCIRC database (Li et al., 2021). Although this method identifies translated circRNAs, it does not provide any insight into the functional capabilities of these peptides, which must be experimentally determined. Therefore, most translatable circRNAs identified with a different ORF to their linear form, produce a peptide of an unknown/undiscovered function.

1.4.2 Engineering Translatable CircRNAs

Translatable circRNAs are circRNAs that contain an IRES site or m^6A methylation site, which enables internal ribosomal recruitment and translation (Yang et al., 2017). Engineered circRNAs have been produced inside cells using self-splicing introns. These circRNAs can be of a designer sequence to translate a peptide of a desired function.

A long repetitive protein can be translated from this circRNA with an infinite open reading frame (IORF) characterised by a start codon but no stop codon. Perriman and Ares, (1998) demonstrated the production of a long repetitive protein from circular mRNA (Figure 1.9). They used PIE bacteriophage T4 group I introns to produce a circular mRNA with an IORF encoding Green Fluorescent Protein (GFP). Transcription, self-splicing, and translation all occurred in *E. coli* producing proteins above 300 kDa, far greater than the size of normal GFP

(30 kDa). This shows that the ribosome could cycle at least ten times around the mRNA. A ribosomal binding site was essential to protein expression, as translation was completely abolished in circular mRNA lacking a Shine-Dalgarno sequence (Perriman and Ares, 1998).

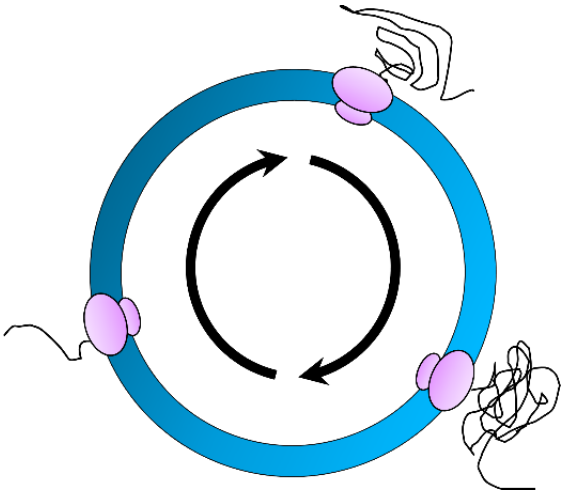


Figure 1.9 Translation of infinite open reading frame (IORF) circRNA. A circRNA with a start codon and no stop codon in the same frame has an IORF. During translation the ribosome cycles around the circRNA many producing a long peptide concatemer.

1.5 Mastoparan

Mastoparan is a 14 aa peptide that acts as a peptide toxin in the defence system of *Vespula lewisii*, Korean yellow jacket wasp. The amphipathic distribution of its residues allows the formation of α-helical structures in an adequate environment (Figure 1.10) (Leite et al., 2010) The three lysine residues give the peptide a net positive charge (net charge 3 at pH 7).

The natural Mastoparan peptide in its amidated form (Mastoparan-NH_2) and non-amidated Mastoparan (Mastoparan-COOH) have been shown to kill mammalian cancer cells (Hilchie et al., 2016). However, Mastoparan-COOH is 8-11 times less potent than Mastoparan-NH_2. The difference in potency of non-amidated and amidated mastoparan is associated with different cell death mechanisms, either directly (lytic) or indirectly (apoptosis-inducing) (de Azevedo et al., 2015). Mastoparan-NH_2 kills cells via direct cell membrane interaction and cell lysis,

whereas Mastoparan-COOH indirectly induces intrinsic apoptosis by causing mitochondrial depolarisation (de Azevedo et al., 2015; Hilchie et al., 2016).

The net surface charge of a cell is generated by the movement of ions across the cell membrane. Normal cells have a net neutral or slightly positive cell surface charge, maintained by ion channels and pumps on the cell surface (Le et al., 2019). Cancer cells have an elevated glycolysis level, leading to up to 30 times greater glucose uptake and lactate secretion than normal cells (Chen et al., 2016). Secreted lactate removes positive ions leaving a net negative charge on the cell surface. The cationic nature of mastoparan causes an electrostatic attraction to cancer cells providing specificity for cancer treatments (Chen et al., 2018). The cytotoxicity to mammalian cells and specificity to cancer cells make mastoparan a good anti-cancer peptide. The potent anti-cancer ability of this peptide has been demonstrated both *in vitro* and *in vivo* and has shown synergy with the chemotherapy agent Gemcitabine (de Azevedo et al., 2015; Hilchie et al., 2016).

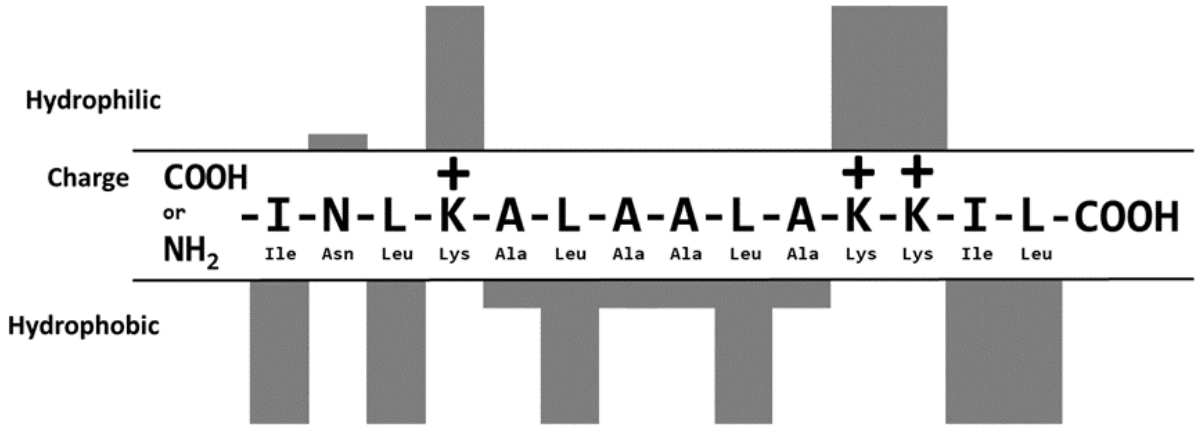


Figure 1.10 The structure of mastoparan. The N-terminus may be aminated (NH_2) or non-aminated (COOH). 3 positively charged amino acid residues give mastoparan an overall net positive charge. The polarity of mastoparan is amphiphilic, with both hydrophilic and hydrophobic regions.

1.6 Fluorescence Resonance Energy Transfer (FRET)

A FRET peptide is two fluorophores with a spectral overlap in the emission wavelength of one (Albertazzi et al., 2009). When FRET peptides are in a proximity of less than 10 nm, excitation of the donor fluorophore is quenched by the acceptor, leading to acceptor fluorescence without donor fluorescence (Hevekerl, Spielmann, Chmyrov and Widengren, 2011). If the distance is greater than 10nm, excitation of the donor leads to donor fluorescence and a reduction in acceptor fluorescence. A FRET-based system can be used *in vivo* to monitor biological processes such as protein-protein interactions (Margineanu et al., 2016).

A protease cleavage site peptide linker can be used in an enzyme-substrate FRET system to determine protease activity with high sensitivity (Zauner et al., 2011). After protease cleavage of the peptide linker, the FRET donor emits fluorescence (Figure 1.11). This system provides a model for assessing protease activity and therefore can be utilised to model the effectiveness of prodrug targets.

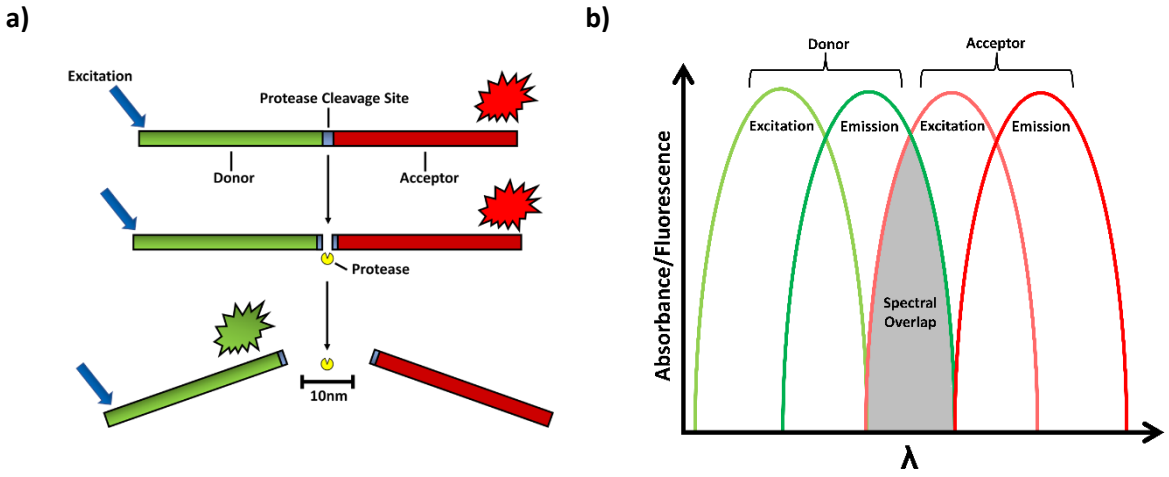


Figure 1.11 The functionality of the FRET peptide system. a) Detection of protease activity by FRET peptide cleavage. A donor fluorophore acceptor quencher pair is separated by a protease recognition sequence. The fluorophore and quencher are in close proximity, causing the fluorescence signal to be quenched. When the protease cleaves the recognition sequence, the peptide is cleaved into two fragments, and the fluorophore and quencher are separated, resulting in an increase in donor fluorescence. b) A graph showing the excitation and emission wavelengths of fluorophores in a FRET system. Spectral overlap (grey) from the emission wavelength of the donor and excitation wavelength of the acceptor results in the quenching of the donor fluorophore.

1.7 Signal Peptide and the Secretory Pathway

Some proteins contain a short peptide sequence (16-30 aa long) at their N-terminus that destines the protein towards the secretory pathway (Nielsen, Engelbrecht, Brunak and von Heijne, 1997). This peptide is known as the signal peptide contains a hydrophilic, usually positively charged N-terminal region, a long stretch of hydrophobic amino acids in its C-terminal region and a cleavage site for signal peptidase (Kapp, Schrempf, Lemberg and Dobberstein, 2009)

In eukaryotes, the secretory pathway occurs co-translationally and prevents any form of cytoplasmic modification (Rutkowski, Ott, Polansky and Lingappa, 2003). Once the signal peptide emerges from the ribosome after translation, the signal-recognition particle (SRP) recognises the signal peptide sequence and halts further translation (Mercier, Holtkamp, Rodnina and Wintermeyer, 2017). Translation only continues when the ribosome/SRP complex encounters the SRP receptor present on the plasma membrane of the Endoplasmic Reticulum. Translation continues, and the signal peptide is pushed through a membrane-spanning channel known as the Sec61 translocon and into the ER lumen (Ge et al., 2014). When a stop codon is reached and translation ceases, the SRP is released from the SRP receptor (Mercier, Holtkamp, Rodnina and Wintermeyer, 2017). An integral membrane protein signal peptidase cleaves the signal peptides converting the precursor protein to its mature form (Auclair, Bhanu and Kendall, 2011). The signal peptide is directed to secretory vesicles, which export the protein out of the cell. Signal peptides can often be utilised in recombinant peptide production to export the desired protein outside of the cell without needing cell lysis for isolation (Figure 1.12).

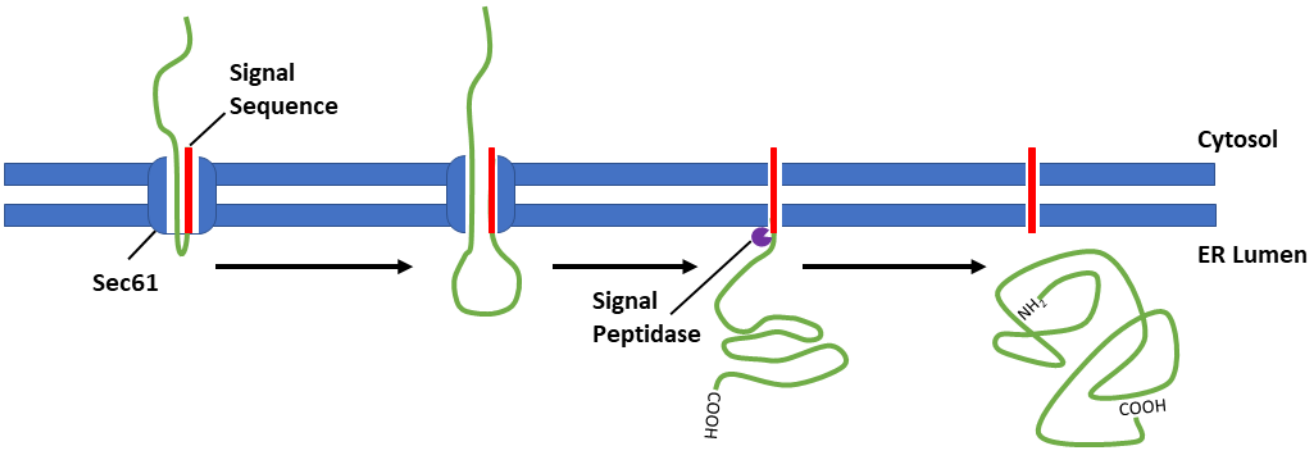


Figure 1.12 Signal peptides and the secretory pathway. Signal peptides are short amino acid sequences found at the N-terminus of newly synthesised proteins that target them to the secretory pathway. The SRP receptor on the ER membrane recognizes the SRP-ribosome-nascent chain complex and facilitates the insertion of the nascent protein into the Sec61 translocon, a protein channel on the ER membrane. The Sec61 translocon allows the nascent protein to cross the ER membrane into the lumen of the ER. The enzyme, signal peptidase, recognises a specific amino acid sequence at the end of the signal peptide and cleaves the peptide into a mature protein free to undergo further processing and trafficking.

1.8 Aims

Breast cancer is the most diagnosed cancer that accounted for 6.9% of all cancer deaths in 2020 worldwide. Due to the significant heterogeneity of breast cancer, the ability to develop drug resistance over time, high cost and toxicity of current treatments, there is a continual need to develop more treatment options. This study aims to develop a novel treatment for breast cancer by utilising engineered circRNA with an infinite open reading frame to produce a cytotoxic peptide prodrug concatemer with specificity to breast cancer cells.

More than 216 endogenous circRNAs have been experimentally verified. However, it is likely that the majority of translatable circRNAs remain unknown. circRNA translation has the potential to produce novel proteins due to the potential of initiation from an alternate start codon and alternative splicing compared to its parent gene. In addition, the sequence

surrounding the circular junction is distinct to circRNAs, further contributing to the possibility of novel protein generation. CircRNA translation can occur from an alternate start codon and circRNAs can be alternatively spliced compared to their parent gene. This project aims to enhance the knowledge of circRNA translation by using a bioinformatics approach to explore the ability of endogenous circRNAs to produce a peptide with a different biological function to their parent gene.

Chapter Two

2.0 Materials and Methods

2.1 Molecular Biology Methods

2.1.1 Yeast DNA Extraction

2.1.1.1 Freeze-Thaw

200 µL of Harju-buffer (2% Triton X-100, 1% SDS, 100 mM NaCl, 10 mM Tris-HCl, pH 8.0, 1 mM EDTA) was added to dry *Saccharomyces cerevisiae* yeast. The tubes were flash-frozen in liquid nitrogen before being transferred to a water bath at 95°C for 1 minute. This freeze-thaw method was repeated two times. The tube was then vortexed for 30 seconds. 200µL of chloroform was added, and the mix vortexed for 2 minutes. The tubes were centrifuged for 3 minutes at room temperature at max speed (13,000 rpm). The upper aqueous phase was transferred to a new microcentrifuge tube containing 400 µL ice-cold 100% ethanol and mixed by inversion. DNA was precipitated by incubation at -20°C for at least 1 hour. Tubes were centrifuged at room temperature for 5 minutes at max speed. The supernatant was removed by vacuum aspiration. The pellet was washed in 0.5 ml 70% ethanol and centrifuged for 5 minutes at room temperature at 13,000 rpm. The supernatant was removed, and the DNA pellet air-dried. The DNA was resuspended in 50 µL nuclease-free water. DNA quality and yield were assessed on a Nanodrop 100 spectrophotometer (Thermo Scientific).

2.1.1.2 Glass Bead

S. cerevisiae yeast was grown on an agar Yeast Extract–Peptone–Dextrose (YPD) plate overnight. Cells were scraped and added to 200 µL of Harju-buffer. ~200µL volume of glass beads (Sigma, Cat# G8772) and 200 µL of phenol/chloroform was added and vortexed at high speed for 10 minutes. 200 µL of TE buffer was added and vortexed briefly. The tubes were spun at max speed for 5 minutes. The supernatant was transferred to a new microcentrifuge

tube. ~400µL of phenol/chloroform was added to the tube and spun at top speed (13,000rpm) for 5 minutes. The upper aqueous layer was transferred to a new tube. 1 ml of 100% ice-cold ethanol was added and mixed. The tube was spun at max speed for 10 minutes, forming a pellet. The supernatant was removed, and the pellet was resuspended in 50 µL nuclease-free water. DNA concentration and yield were assessed by Nanodrop.

2.1.2 PCR

The amplified sequences generated by polymerase chain reaction (PCR) were all generated according to the optimised conditions for Phusion High-Fidelity DNA Polymerase (NEB, Cat# M0530S). The final PCR reaction consists of 1x 5X Phusion HF or GC Buffer, 200 µM dNTPs, 0.5 µM Forward Primer, 0.5 µM Reverse Primer, Template DNA 1 ng, DMSO 3%, Phusion DNA Polymerase 1.0 units/50 µL.

Template concentrations were approximately 100 ng for genomic DNA and 1 ng for plasmids. PCR was also performed with overlapping oligos, in which 1 ng of each oligo was used as a template. Colony screen PCR, a single colony was picked with a pipette tip and mixed with the reaction. For RT-PCR, the first strand synthesis reaction was added to a final volume of 1 % in the PCR reaction mix. Negative controls used in PCR reactions included all components of the PCR reaction other than the template DNA.

The PCR tubes were transferred from ice to a PCR machine with the block preheated to 98°C. The ‘NEB Tm Calculator’ (New England Biolabs) online tool was used to determine the annealing temperature of the forward and reverse primer sets (Table 2.2 – 2.6). In a Techne TC-512 gradient PCR machine, the reaction was performed to conditions as shown in table 2.1

for standard PCR procedure. Following PCR, amplicons were viewed by agarose gel electrophoresis.

Table 2.1 Typical PCR thermocycler conditions.

Step	Temp	Time
Initial Denaturation	98°C	30 seconds
25-35 Cycles:		
Denaturing	98°C	10 seconds
Annealing	X°C	30 seconds
Extension	72°C	15-30 seconds per kb
Final Extension	72°C	10 minutes
Hold	10°C	X

Table 2.2 Primers used for standard PCR amplification.

Primer Name	Sequence (5'-3')
pRSET B Fwd	AATACGCAAACCGCCTCTC
pRSETB R 273	TCAGCTTCCTTTCGGGCTTT
pcDNA Fwd	TAATACGACTCACTATAGGG
pcDNA Rev	AACTAGAAGGCACAGTCG

Table 2.3 Primers used in reverse-transcription PCR.

Primer Name	Sequence (5'-3')
RT Lin Fwd	ACCTTAAACGGTCGTGTGGG
RT Lin Rev	TCTCGCCGGTAACGCATAATAC
RT circ Fwd	GTGTGGGTTCAAGTCCCT
RT circ Rev	GACCGTTTAAGGTCAAACGG

Table 2.4 Primers used for Yeast DNA amplification.

Primer Name	Sequence (5'-3')
MT Fwd	CTATGTTTGCCACCTCGATGTCGACTC
MT Rev	CCCCATAAGGAGTGAGGGACCCCTCCC
GAPDH Fwd	GGGGACAAGTTTGTACAAAAAAGCAGGCTCCATGGTTAGAGTTGCTATT
GAPDH Rev	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTAAGCCTTGGCAACGTG
COX1 Fwd	TGGGTGGTTTAACTGGTGTTG
COX1 Rev	GTGTACAGCTGGTGGAGAAGT
INV I NcoI PmlI Fwd	AGTCCATGGACATCACGTGCGGTCTGAAAGTTATC
INV III BamHI Rev	ATAGAGGATCCACCTATAGTATAAGTTAGC
INV IV XbaI Fwd	CTAACTTATACTTCTAGAGATATGCCTATC
INV VI EcoRV NcoI Rev	CCATTGCAATGGCATGATATCCCGATAGGTAGACC

Table 2.5 Oligos/ primers used for overlapping PCR.

Oligo/Primer Name	Sequence (5'-3')
R1-3 O1	AGTCCATGGACATCACGTGCGGTCTGAAAGTTATCATAAATAATATTTACCATATAATAATGGATAAATTATATTTTTATCAATATAAGTCTAATTACAAGTGTATTAAAATGGTAACAT
R1-3 O2	CACATAATTTTACCATTGTAAAATATGCTAAGCTGTAATGACAAAAGTATCCATATTCTTGACAGTTATATTATAAAAAAAGATGAAGGAACTTTGACTGATCTAATATGCTCAACGAAA
R1-3 O3	TAGATTATACGAGTTGCTTTGTGAATCAAATGTTATAAAATTACTTAGTGTAATTGAAAACCTGTCTGATATTCAATTATTATTTATTATTATATATAATTATATAATAATAAATAAAATGG
R1-3 O4	ATATTATTATTTATTTTACCTTGATGTTATGTATTGGAAATGAGCATACGATAAATCATATAACCATTAGTAATATAATTTGAGAGCTAAGTTAGATATTTACGTATTTATGATAAAACA
R1-3 O5	ATGCATAAATACTATTTTGTGAATAAACCCTATAAATTATTATTATTAATAATAAAAAATAATAATAATACCAATATATATATTATTTAATTTATTATTATTATATTAATAAAATTTAAT
R1-3 O6	AATATAATTATTTTAAATTAATATATTATAAATAATTATTGGATTAAGAAATATAATATTTTATAGAAATTTTCTTTATATTTAGAGGGTAAAAGATTGTATAAAAAGCTAATGCCATAT
R1-3 O7	TATTTTTCGATTACGGTATATGTAATGATATGGATAAGAATTATTATTCTAAAGATGAAAATCTGCTAACTTATACTATAGGTGGATCCTCTAT
R4-6 O1	CTAACTTATACTTCTAGAGATATGCCTATCTTTATTTATATATATATTATTATTATTAATAATAAAAAAAAAAATTAAAAAAAAGATAGGAGGTTTATATATAACTG
R4-6 O2	TCCTCCAAATATATATTGACATAAATATTTATTATATTATTTTTTTTTATAATAAATATTAAAAGATATTGCGTGAGCCGTATGCGATGAAAGTCGCACGTACG
R4-6 O3	GCTACTTTCAGCGTGCATGCGTTCTTACCGGGGGAAAACTTGTAAAGGTCTACCTATCGGGATATCATGCCATTGCCATGGAAGT
R1-3 P1	TCAGGTACCTGTAGTGCACG
R1-3 P2	TTTCGTTGAGCATATTAGAT
R1-3 P3	ATCTAATATGCTCTTCGAAA
R1-3 P4	TGTTTTATCATAAATACGTA
R1-3 P5	TACGTATTTATGATAAAACA
R1-3 P6	ATATGGCATTAGCTTTTTAT
R1-3 P7	ATAAAAAGCTAATGCCATAT
R1-3 P8	ATAGAGGATCCACCTATAGT
R4-6 P1	GATTGAATATGAAGATCTCT
R4-6 P2	CGTACGTGCGACTTTCATCG
R4-6 P3	CGATGAAAGTCGCACGTACG
R4-6 P2	ACTTCCATGGCAATGGCATG

Table 2.6 Oligos Used in Gibson Assembly.

Oligo/Primer Name	Sequence (5'-3')
XbaI + 10 Rev	TCTAGAGGGAAACCGT
BamHI + 10 Fwd	GGATCCGAGCTCGAGA
BamHI + 10 Rev	GGATCCTTATCGTCAT
HindIII + 10 Fwd	AAGCTTGATCCGGCTG
XbaI + 10 Fwd	TCTAGAGATATGCCTA
HindIII + 10 Rev	AAGCTTACCTATAGTA
G1 O1	GGAGACCACAACGGTTTCCCTCTAGAGATATGCCTATCTTTATTTATATATATATTATTA
G1 O2	ATAAATATATATATAATAATAATAATTATTATTTTTTTTTTTAATTTTTTTTCTATCCTC
G1 O3	AATTAAAAAAAAAGATAGGAGGTTTATATATAACTGATAAATATTTATTATATTATTTTTT
G1 O4	ATAAATAATATAATAAAAAAAAAATATTATTTATAATTTTCTATAACGCACTCGGCATACG
G1 O5	ATATTGCGTGAGCCGTATGCGATGAAAGTCGCACGTACGGTTCTTACCGGGGGGAAAACTT
G1 O6	AAGAATGGCCCCCTTTTGAACATTTCCAGATGGATAGCCCTATAGTACGGTACCTGTAGT
G1 O7	ATATCATGCCATGGACATCACGTGCGGTCTGAAAGTTATCATAAATAATATTTACCATAT
G1 O8	TATTTATTATAAATGGTATATTATTACCTATTTACCTAGGCTCGAGCTCTAGACGTCGAC
G2 O1	CTGTACGACGATGACGATAAGGATCCCAATAGTATTTATTATAAATGGTATATTATTACC
G2 O2	TATTTACCATATAATAATGGATAAATTATATTTTTATCAATATAAGTCTAATTACAAGTG
G2 O3	ATATTCAGATTAATGTTCACATAATTTTACCATTGTATTTATACGATTCGACATTACTGT
G2 O4	TATGCTAAGCTGTAATGACAAAAGTATCCATATTCTTGACAGTTATATTATAAAAAAAGA
G2 O5	TCAATATAATATTTTTTTCTACTTCCTTGAAACTGACTAGATTATACGAGTTGCTTTCAC
G2 O6	TAATATGCTCAACGAAAGTGAATCAAATGTTATAAAATTACTTACACCACTAATTGAAAA
G2 O7	GAATGTGGTGATTAACTTTTGGACAGACTATAAGTTAATAATAAATAATAATATATTAAT
G2 O8	TATTTATTATTATATAATTATATAATAATAAATACCTAGGCTCGAGCTCTAGACGTCGAC
G3 O1	CTGTACGACGATGACGATAAGGATCCAATTATTATTTATTATTATATAATTATATAATAA
G3 O2	TAATATATTAATATATTATTATTTATTTTACCAACTACAATACATAACCTTTACTCGTAT
G3 O3	ATGTATTGGAAATGAGCATACGATAAATCATATAACCATTAGTAATATAATTTGAGAGCT
G3 O4	TCATTATATTAAACTCTCGATTCAATCTATAAATGCATAAATACTATTTTGTCTTATTTG
G3 O5	TATGATAAAACAGAATAAACCCTATAAATTATTATTATTAATAATAAAAAATAATAATAA
G3 O6	TATTATTTTTTATTATTATTATGGTTATATATATAATAAATTAAATAATAATAATATAAT
G3 O7	AATTTATTATTATTATATTAATAAAATTTAATATATATTATAAATAATTATTGGATTAAG
G3 O8	ATTTATTAATAACCTAATTCTTTATATTATAAAACCTAGGCTCGAGCTCTAGACGTCGAC
G4 O1	CTGTACGACGATGACGATAAGGATCCATATTATAAATAATTATTGGATTAAGAAATATAA
G4 O2	ATAACCTAATTCTTTATATTATAAAATATCTTTAAAAGAAATATAAATCTCCCATTTTCT
G4 O3	TATATTTAGAGGGTAAAAGATTGTATAAAAAGCTAATGCCATATTGTAATGATATGGATA
G4 O4	TATAACATTACTATACCTATTCTTAATAATAAGATTTCTACTTTTAGACGATTGAATATG
G4 O5	GAAAATCTGCTAACTTATACTATAGGTAAGCTTGATCCGGCTGCTAACAAAGCCCGAAAG

2.1.3 Gibson Assembly Reaction

A Gibson assembly approach was used to construct the Inverse intron domains. Oligos were designed so that after 2 Gibson assembly reactions and PCR, the entire group II inverted yeast intron sequence could be constructed (Figure 2.1). Oligos were designed 60 nucleotides in length and grouped into sets, each containing eight oligos. Each oligo in a set had a 20 bp overlap with another in the same set. The first and last oligos in a set had a 40bp overlap with the pRSET B vector digested with a restriction enzyme.

Each oligo set overlapped by 40bp. The ends of each oligo set contained a restriction enzyme site to release it from the vector after cloning (Figure 2.1). pRSET B vectors were linearised using 3 restriction enzyme sets (XbaI/BamHI, BamHI, BamHI/XbaI) (2.1.10) and purified by gel extraction (2.1.6). The linearised vectors were then PCR amplified (2.1.2) with their specific primer sets shown in table 2.7.

Table 2.7 Restriction enzymes used to linearise vector and their complementary primer.

Restriction Ezymes/s Used to Linearise pRSET B	Forward Primer	Reverse Primer
XbaI and BamHI	XbaI + 10 Rev	BamHI + 10 Fwd
BamHI	BamHI + 10 Rev	BamHI + 10 Fwd
BamHI and HindIII	BamHI + 10 Rev	Hind III + 10 Fwd

The Oligos were pooled in their sets with equal volumes of each oligonucleotide and diluted to a concentration of 180nM per oligo. The pooled oligos were annealed, and 50ng of the linearised vector was added to the appropriate set of pooled oligos. A Gibson reaction was assembled by adding 2x Gibson Assembly Master Mix (NEB, Cat #E2611S) to the vector and oligo mix to a final concentration of 1x with nuclease-free H_2O up to 20 μL. The reaction was then incubated for 2 hours at 50°C before transformation into *E.coli* DH5α cells (2.1.11).

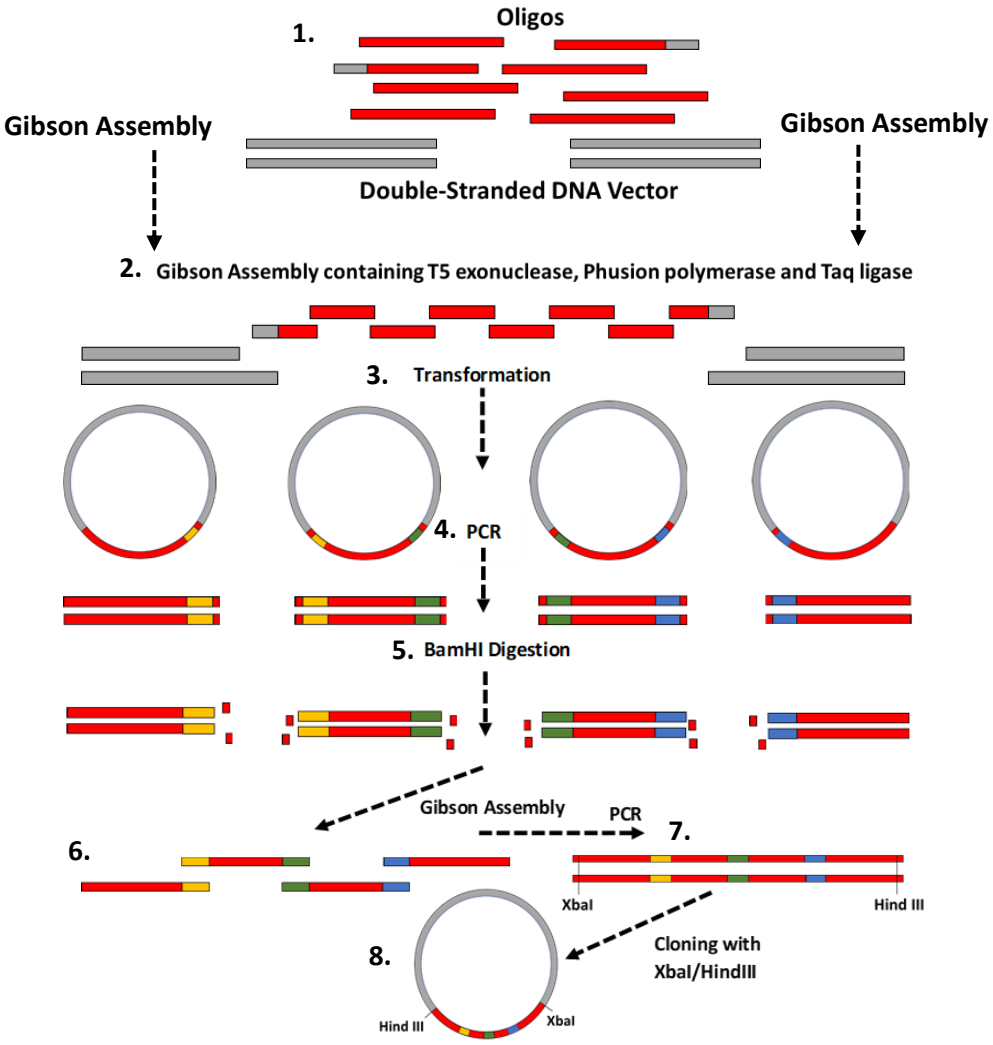


Figure 2.1 Yeast Group II intron construction using Gibson assembly. 1. 8 oligos pooled together 2. Oligos annealed and subject to Gibson assembly with pRSET B 3. Gibson assembly transformed into competent cells. 4. Inserts amplified by PCR 5. Inserts digested with BamHI revealing overlap 6. Inserts annealed together via overlapping region and assembled by Gibson approach 7. Full ai5γ intron assembly amplified by PCR 8. Full length assembly cloned into pRSET B vector using XbaI and HindIII.

2.1.4 Site-Directed Mutagenesis

Site-directed mutagenesis was used to make targeted changes to the DNA sequence. The PCR reaction was the same as described in (2.1.2) but with two separate reactions, one with only the forward primer and the other with only the reverse primer, H_2O was adjusted accordingly (Table 2.7). 200 ng of plasmid DNA was used as the template.

Thermocycler conditions are shown in table 2.8. After 10 cycles, the two PCR reactions were added together.

Table 2.8 Site-directed mutagenesis primers.

Primer Name	Sequence (5'-3')
pRSETB_IORF_Fwd	ATGGCTAGCATGGCTGGTGGACA
pRSETB_IORF_Rev	ACCATGATGATGATGATGATGAGAACC
SDM_Flag_Stop_Fwd	CGACGACAAGTGGCTCGAGAAAA
SDM_Flag_Stop_Rev	TCGTCCTTGTAGTCGATG

Table 2.9 Site-directed mutagenesis thermocycler conditions.

Step	Temp	Time
Initial Denaturation	98°C	3 minutes
25 Cycles:		
Denaturation	98°C	10 seconds
Annealing	X °C	30 seconds
Extension	72°C	20 minutes
Final extension	72°C	10 minutes

After the PCR, the parental methylated DNA template was digested with DpnI (NEB, Cat# R0176S) (2.1.8), selecting the synthesised DNA containing the required mutations. The digestion mix was transformed into competent cells.

2.1.5 Agarose Gel Electrophoresis

DNA was separated according to its molecular weight by agarose (Sigma, Cat# A4718) gel electrophoresis. Typically, 1%-1.2% (w/v) agarose gels were used. Agarose was melted in 1x TAE (40 mM Tris, 40 mM acetic acid, 1 mM EDTA, pH 8.0) and allowed to cool until the flask was safe to hold. The agarose was cast into plastic trays after adding 10 mg/ml Ethidium Bromide (Sigma Cat# E1510) or 1x SYBR Safe DNA gel stain (Thermo Fisher Scientific, Cat# S33102), and the gel comb was put in place. DNA was mixed with 6x loading dye (NEB, Cat# B7025) to a final concentration of 1x and loaded into the gel. The gel was submerged in

1x TAE, and a voltage of 90-120v was applied for 45 minutes to 1 hour. DNA bands were visualised using a UV transilluminator or a Gel Doc XR+ System (BioRad).

2.1.6 DNA Column Purification

Digested plasmids and PCR-amplified DNA was separated on an agarose gel (1%-1.2%). The DNA was purified using Monarch DNA Gel Extraction Kit (NEB, Cat #T1020) to manufacturer's guidelines. DNA fragment excised from the gel and weighed in 1.5 ml Eppendorf. 4 volumes of Gel Dissolving Buffer (e.g., 400 µL buffer per 100 mg agarose). Samples were incubated at 50°C until completely dissolved. For PCR clean-up, the sample was diluted with DNA Cleanup Binding Buffer to a 5:1, buffer: sample ratio. The sample was loaded onto the column placed inside the collection tube and spun at 13,000 rpm for 1 minute and the flow-through discarded. 200 µL wash buffer was added to the column and spun 13,000 rpm for 1 minute. The washing step was repeated. The column was transferred to a clean Eppendorf, and 20 µL elution buffer (preheated to 60°C) added to the centre of the matrix. The sample was incubated for 5 minutes and spun 13,000 rpm for 1 minute.

2.1.7 RNA Purification

RNA was purified using plateclear™ Transcription Clean-Up Kit (Thermo Scientific, Cat# AM1908). Elution solution was added to the RNA sample up to a total volume of 100 µL. 350 µL of Binding Solution Concentrate was added to the sample. 250 µL of 100% ethanol was added to the sample. A filter cartridge was placed in a collection tube, and the sample was added and centrifuged for 1 minute at 13,000 rpm. Flowthrough was discarded, and the column was washed by adding 500 µL wash solution and centrifugation for 1 minute at 13,000 rpm. The wash step was repeated once more. The column was added to a fresh collection tube, and

50 µL of Elution Solution (preheated to 70°C) was applied to the centre of the Filter Cartridge and incubated on a heat block set to 70°C for 5 minutes. The sample was centrifuged for 1 minute at 13,000 rpm.

2.1.8 Plasmid Purification

Plasmid DNA was obtained from *E. coli* DH5α cells using NucleoSpin Plasmid QuickPure (Fisher Scientific, Cat# 11902422) miniprep kit to the manufacturer's recommendations. An isolated colony was picked and grown in 7 mL LB broth + Ampicillin 100 µg/ml incubated overnight at 37˚C with agitation. The culture was centrifuged at 4000 rpm for 10 minutes at 4˚C. The pellet was resuspended in 250 µL Resuspension Buffer A1. 250 µL Lysis Buffer A2 was added, mixed by inversion, and incubated for 5 minutes at room temperature. 300 µL Neutralisation Buffer A3 was added and mixed by inversion. Sample centrifuged for 30 min at 13,000 rpm at room temperature. Supernatant decanted into a column, placed in a collection tube and centrifuged at 13,000 rpm for 1 minute. Flow-through was discarded, and 450 µL Wash Buffer AQ was added and centrifuged at 13,000 rpm for 3 minutes. 50 µL preheated Elution Buffer AE (50 °C) was added to the centre of the column and incubated at 50 °C in a heat block for 5 minutes before the sample was centrifuged at 13,000 rpm for 1 minute. DNA yield and quality assessed with a nanodrop. Plasmid DNA was stored at -20 °C .

2.1.9 Nucleic Acid Ethanol Precipitation

0.1 volumes of 3M sodium acetate and 3 volumes of ice-cold 100% ethanol were added to the DNA sample. The DNA was incubated at -80 °C for 2 hours – overnight. The sample was spun at maximum speed (13000 rpm) at 4 °C for 30 minutes. The pellet was washed twice with 0.5 ml ice-cold 70% ethanol, spinning at maximum speed for 10 minutes at 4°C. Ethanol was

decanted, and the pellet air-dried. The pellet was resuspended in 20-50 μL nuclease-free water. DNA quality and yield were measured with the nanodrop.

2.1.10 Restriction Enzyme Digestion

Restriction enzymes were used in the most complementary buffer and protocol according to manufacturer guidelines (New England Biolabs). Briefly, 1 μL desired restriction enzyme, 5 μL compatible restriction enzyme buffer, 1 μg of DNA and H_2O to 50 μL composed the digestion reaction. The reaction was incubated at 37°C for 1-3 hours. This reaction is scalable to larger volumes. Single digests were subject to vector dephosphorylation by adding Quick Calf Intestinal Alkaline Phosphatase (NEB, Cat #M0525S) to the digestion mix and incubating at 37°C for 30 minutes.

2.1.11 Denaturing Agarose Gel Electrophoresis

Using a MOPs buffer system, denaturing RNA Agarose-Formaldehyde Electrophoresis was performed according to Maniatis et al. (1982). A mixture containing 20 mL 5x MOPS buffer (10 mM $MgSO_4$,0.5 M MOPS, 2.5 M NaCl), 72 mL nuclease-free Water and 1g Agarose was heated to boil and allowed to cool to ~55°C. 17.6 ml of 37% formaldehyde and 5 μL of 10mg/ml ethidium bromide were added to the molten agar. RNA samples were mixed with 2X RNA loading dye to a final concentration of 1X and a total volume of 20-30 μL. Samples were denatured by heating to 80 °C for 10 minutes, then cooled on ice for 2 minutes. The gel was submerged in 1x MOPS, and a voltage of 90-120v was applied for 30 to 45 minutes. RNA bands were visualised using a UV transilluminator or a Gel Doc XR+ System (BioRad)

2.1.12 Competent DH5α *E. coli* Preparation

5 ml Luria Bertani (LB) starter culture was grown overnight at 37 °C with agitation from an isolated *E. coli* colony. 250 ml of sterile LB with 20 mM $MgSO_4$ added was inoculated with the starter culture and incubated at 37 °C with agitation until the optical density reached 0.4-0.6 (OD600). The culture was collected into 5x 50ml Falcon tubes and centrifuged at 4000 rpm for 10 minutes at 4 °C. The resulting pellets were resuspended in 1ml ice cold TBF1 (30 mM KOAc, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 100 mM RbCl, 15% glycerol), combined into a single falcon and incubated on ice for 60 minutes. The cells were then centrifuged at 4000 rpm for 10 minutes at 4 °C and the pellet resuspended in 5 ml TBF2 (10 mM MOPS (pH 6.5), 75 mM $CaCl_2$, 10 mM RbCl, 15% glycerol). Cells were aliquoted (100 µL) and frozen immediately in liquid nitrogen. Aliquots were stored at -80 °C for future use.

2.1.13 Transformation of Competent DH5α *E. coli*

1 ng plasmid DNA or 10µL ligation mix added to 50 µL of competent DH5α *E. coli*. Cells were heat-shocked at 42 °C for 1 minute, followed by incubation on ice for 2 minutes. The cells were added to 700 µL LB and incubated at 37 °C with agitation. The culture was centrifuged at 3000 rpm for 10 minutes. The supernatant was discarded, leaving 150 µL for resuspension. Cells were spread onto an LB plate containing ampicillin at 100 µg/ml and incubated overnight at 37 °C.

2.1.14 DNA Ligation

DNA fragments were ligated into plasmids using T4 DNA ligase (New England Biolabs, Cat# M0202S) using manufacturer guidelines. Briefly, a reaction containing: 2 µL T4 DNA Ligase

Buffer, 1 μL T4 DNA Ligase, insert and vector DNA to a molar ratio of 3:1 (insert: vector), and H2O to 20 μL final volume was assembled and incubated at 4 °C overnight or 3-4 hours at room temperature before transformation.

2.1.15 *In Vitro* Transcription

Linear RNA produced using RiboMAX™ Large Scale RNA Production Kit – T7 (Promega, Cat# P1300). A typical reaction consisted of 10 μL T7 Transcription 5X Buffer, 15 μL rNTPs (25mM ATP, CTP, GTP, UTP), 7.5 μg linear DNA template, 5 μL Enzyme Mix (T7), H_2O to a final volume of 50 μL was assembled. This reaction could be scalable for smaller or larger reaction volumes.

The rNTP mix contained 10-fold more GMP (Sigma, Cat# G0635) than GTP (22.73mM:2.27mM). If the RNA were synthesised would be subject to splint ligation (2.1.9). The reaction was incubated at 37 °C for 4 hours before 1u/μg DNase I (NEB, Cat# M0303S) was added and incubated for another 30 minutes. For circRNA, after DNase treatment additional GTP was added to a final concentration of 2 mM, and then reactions were heated at 55 °C for 15 min. RNA was then column purified (2.1.5).

2.1.16 RNA/DNA Splint Hybridisation

5 μM DNA splint added to 25 μg of purified RNA in a 90 μL reaction. Water boiled in a glass beaker and allowed to cool to 80°C. The tube containing the DNA/RNA mix was added, and the water was allowed to cool to room temperature.

2.1.17 Single-Stranded RNA Ligation

T4 RNA Ligase I Buffer (NEB, Cat# M0204S) was added to the DNA/RNA hybrid to a final concentration of 1X. ATP was added to a final concentration of 1 mM, and 50 U of T4 RNA Ligase I (NEB, Cat# M0204S) was added. Reactions were incubated at 37 °C for 1 hour and then analysed by gel electrophoresis to confirm circular topology.

2.1.18 Ribonuclease R Treatment

Circular RNA purified by Ribonuclease R (Lucigen, RNR07250) treatment. RNase R Reaction Buffer added to a final concentration of 1x. 5U of RNase R was added, and the reaction was incubated at 37 °C for 30 minutes.

2.1.19 Reverse Transcription PCR First-Strand Synthesis

First-strand synthesis of the RNA template produced using SuperScript™ IV Reverse Transcriptase (Thermo Scientific, Cat# 18090010) according to manufacturing guidelines. The following components were assembled in a reaction: 2 μM reverse primer, 2 μL 10 mM dNTP mix, 10 ng total RNA or *in vitro* synthesised RNA, and H_2O to a final volume of 13 μL. The mix was heated to 65°C for 5 minutes and incubated on ice for 1 minute. In another tube, 4 μL 5X SSIV Buffer, 1 μL 100 mM DTT, 0.2 μL Murine RNase R inhibitor (NEB), 1 μL Superscript IV Reverse Transcriptase (200 U/ μL), 0.8 μL H_2O was assembled. The mixtures in both tubes were added and incubated at 50 °C for 30 minutes. The reaction was inactivated by heating to 80 °C for 10 minutes. RT reaction was stored at -20 °C. After first-strand synthesis, PCR was performed as demonstrated in 2.1.1 with 1% of the reverse transcription reaction in the total PCR reaction volume.

2.1.20 *In Vitro* Mammalian Translation

The RNA/DNA template was translated according to TnT® Quick Coupled Transcription/Translation protocol (Promega, L1170). Briefly, a reaction mix containing 40 µL Quick Master Mix, 1 µL Methionine, 1 mM, 2 µg plasmid DNA or 10 µg circRNA, and H_2O to a final volume of 50 µL was assembled. The reaction was incubated at 30°C for 120 minutes. Protein was stored in 5µL aliquots at -20°C.A negative control was produced from this reaction with 10 µL H_2O and lacked any nucleic acid template.

2.1.21 SDS PAGE

Two glass plates with 1 mm separation were placed in holders. A 10-15% resolving gel was prepared. The lower the protein's expected size (kDa), the higher the percentage resolving gel was used. A tricine gel was assembled for very low MW proteins (below 30 kDa). Components of the resolving gel and tricine gels at varying percentages are shown in table 2.10.

Table 2.10 SDS-PAGE and tricine gel compositions. Upper: Composition of an SDS-PAGE resolving gel Lower: Composition of a tricine gel.

Reagent	10% (10 mL)	12% (10 mL)	15% (10 mL)
H_2O	4	3.3	2.3
30% w/v Acrylamide	3.3	4	5
1.5M Tris pH 8.8	2.5	2.5	2.5
10% SDS	0.1	0.1	0.1
10% APS	0.1	0.1	0.1
TEMED	0.004	0.004	0.004

Reagent	Stacking (4%)	Intermediate (10%)	Separation (15%)
38% Glycerol	-	-	1.6 mL
H_2O	1.4 mL	1.2 mL	-
30% w/v Acrylamide	0.3 mL	0.8 mL	2.7 mL
3M Tris HCL pH 6.8	-	1 mL	2.14 mL
1M Tris HCL pH 6.8	0.3 mL	-	-
10% SDS	0.02 mL	0.03 mL	0.06 mL
10% APS	0.02 mL	0.03 mL	0.06 mL
TEMED	0.002 mL	0.003 mL	0.003 mL

This mixture was poured between the glass plates, and a layer of isopropanol was added to the surface. The gel was left to sit at room temperature for twenty minutes. The isopropanol was poured off the gel and dried by gently dabbing with a paper towel between the glass plates. 5% stacking gel (H_20 - 1.4ml, 30% w/v Acrylamide (Protogel, Cat# A2-0072) – 330l, 1.5 M Tris pH 8.8 – 250 µL, 10% SDS – 20 µL, 10% APS – 20 µL, TEMED – 2 µL) was then added to the layer of resolving gel and a 1mm comb placed between the glass plates. For the tricine gel, the intermediate gel was added, first allowed to set, and then stacking gel was added. The stacking gel was left to set at room temperature for 20 minutes. Protein samples were mixed with 4x protein loading dye to a final concentration of 1x. Samples were heated to 95 °C for 5 minutes. Gels were mounted into the tank, and the comb was removed. The inner chamber and at least a third of the outer chamber were filled with 1x running buffer (25 mM Tris pH 8.3, 0.192 M glycine and 1% SDS). 5 µL PageRuler™ Prestained Protein Ladder (ThermoFisher, Cat# 26617) and 20 µL of each sample were loaded into each well. A power pack was connected and ran at 20 mA until the blue dye almost reached the bottom of the gel. Once electrophoresis had finished, the gel was removed from the glass plates and subject to a Western blot.

2.1.22 Western Blot

A polyvinylidene difluoride (PVDF) membrane was activated by soaking in 100% methanol. Foam pads and filter paper were soaked in 1x transfer buffer (25 mM Tris-HCl (pH 7.6), 192 mM glycine, 20% methanol, 0.03% sodium dodecyl sulfate (SDS), and any bubbles in the sponge were removed. Foam pads, filter paper, gel and PVDF membrane were encased in a 'transfer sandwich' configuration with a plastic cassette. The cassette was submerged in a tank containing 1x transfer buffer and an ice block. The transfer was run in the cold room at 100v for 1 hour or overnight.

Once the transfer was complete, the blot was washed with purified water for 3 minutes with gentle agitation. The blot was then blocked with 3% non-fat milk in Tris Buffered Saline (TBS) pH 8.0 for 1 hour at room temperature. The blot was then incubated overnight in the cold room with Monoclonal ANTI-FLAG® M2 antibody (Sigma. Cat# F3165) or Monoclonal Anti-Green Fluorescent Protein (GFP) (Sigma, Cat# G1546) produced in mice at a 1:1,000 dilution in TBST pH 8.0 with 3% non-fat milk. The primary antibody solution was decanted, and the blot was washed with TBS, pH 8.0, for 5 minutes. The blot was then incubated with anti-mouse IgG, HRP-linked antibody 1:2000 (Cell Signalling Technology, Cat# 7076) secondary antibody at a dilution of 1:2000 in TBST with 3% non-fat milk for 1 hour. The blot was washed 3 times (every 5 minutes) in TBS with 0.05% TWEEN 20, pH 8.0. The blot excess liquid dried in the air and was placed on a plastic film. The blots were visualised by chemiluminescence using Pierce ECL western blotting substrate (Thermo Scientific, Cat# 32209), scanned using Fujifilm LAS- 4000 luminescent image analyser and quantified using ImageJ (US National Institutes of Health, Bethesda, Maryland, USA).

2.1.23 Measuring FRET Fluorescence

Protein/media/*in vitro* reaction was diluted in PBS containing 96 plate wells to desired final concentrations. FRET, GFP and mCherry fluorescence was measured at their excitation/ emission wavelengths (Table 2.11) using Synergy HT spectrophotometer.

Table 2.11 The excitation and emission wavelength of FRET Peptide, EGFP and mCherry.

Peptide	Excitation Wavelength (nm)	Emission Wavelength (nm)
FRET	485/20	645/40
EGFP	485/20	528/20
mCherry	590/20	645/40

2.2 Cell Culture

2.2.1 Cell Lines and Maintenance

Breast cancer cell lines used in this project are MDA-MB-231 and MCF-7 cells. Both cell lines were cultured in DMEM (Thermo Fisher, Cat# 10569010). Media was supplemented with 10% foetal bovine serum (Gibco, Cat# 10500064). Cell lines were maintained at 37˚C, 5% CO2. When cells reached 70-80% confluence they were seeded or sub-cultured, using 1x trypsin-0.5 mM EDTA for enzymatic dissociation.

2.2.2 Transfection

Plasmid DNA was transfected into MDA-MB-231 or MCF-7 cells on a 96 well plate or a 100 mm dish according to Lipofectamine™ 3000 Transfection (Thermo Scientific, Cat# L3000001) manufacturer guidelines. For a single well on a 96 well plate containing 10,000 cells per well a transfection mix was assembled containing 5 μL Opti-MEM™ Reduced Serum Medium (Thermo Fisher, Cat# 31985062) and 0.15 μL Lipofectamine™ 3000 Reagent. A mix containing Opti-MEM Medium, 100 ng Plasmid DNA and 0.5 μL P3000™ Reagent was assembled in another tube. The two tube components were mixed and incubated at room temperature for 10 minutes. The transfection mix was then added to cells. A GFP expression plasmid (Lonza) was used as a positive control to ensure successful transfection. A transfection reaction lacking any nucleic acid template served as a negative control. Transfection was also performed in 100 mm plates containing 2.2 million cells with the following concentration changes: 20 μg plasmid DNA, 40 μL P3000™ and 21.7 μL Lipofectamine® 3000 Reagent.

2.2.3 Protein/RNA Extraction

Cells were seeded in a 100 mm dish, transfected and incubated for 24 hours. Media aspirated (or kept for media concentration section 2.2.6) and dish washed with PBS and 1 mL ice-cold Tri reagent (Sigma-Aldrich, Cat # T9424) added. The sample was incubated on ice for 5 minutes. Cells were added to 1.5 mL eppendorf. A cell scraper was used to maximise the number of cells obtained from the dish. 200 μL chloroform was added, shaken vigorously for 15 seconds, and left for 3 minutes at room temperature. Samples spun at 12,000 rpm for 15 minutes. Aqueous (upper) phase aspirated, aliquoted and stored at -80 °C (RNA samples). 0.3 mL of 100% ethanol added to the remaining organic layer, mixed by inversion and incubated at room temperature for 3 minutes. Centrifuged at 2,000 rpm for 5 minutes at 4°C and the supernatant collected into a fresh 2 mL eppendorf. 1.5 mL isopropanol was added to the supernatant, mixed by inversion and left for 10 minutes at room temperature. Samples were centrifuged at 12,000 rpm for 10 minutes at 4 °C, and the supernatant was aspirated and discarded. 2 mL solution containing 0.3 M guanidine hydrochloride in 95 % ethanol was used to wash the pellet x3 times. During each wash cycle, the wash solution was incubated with the pellet at room temperature for 20 minutes before centrifugation at 7,500 rpm for 5 minutes at 4 °C. 2 mL 100% ethanol was added to the pellet incubated for 20 minutes before centrifugation at 7,500g for 5 minutes at 4 °C. Ethanol was aspirated, and the pellet was left to air dry. Pellet resuspended in 1 % SDS and heated to 50 °C if redissolving was difficult. The sample was centrifuged at 10,000 rpm for 10 minutes, and the supernatant was aliquoted and stored at -20 °C. If there is an insoluble pellet remaining, this is dissolved in 8 M urea, aliquoted and stored at -80 °C.

2.2.4 Mastoparan Treatment

10,000 MDA MB-231 cells or MCF-7 cells were seeded into 96-well plates with DMEM media overnight. Mastoparan (Sigma, Cat# M5280) (0 – 40 μM), Mastoparan extracted from MDA MB-231 or MCF-7 cells (0 - 7 ng/μL) (2.2.2), or Mastoparan produced from *in vitro* translation (0 – 2 μL) were added to the cells and incubated at 37°C at 5% CO_2 concentration for 8 or 24 hours.

2.2.5 MTT Assay

An MTT assay was used to measure the cytotoxic activity of mastoparan. Media Aspirated from cells in 96 well plates and discarded. Cells were washed in PBS. 50 μL Gibco Opti-MEM Reduced Serum Medium, and 50 μL 5 mg/mL MTT were added to each well. Plate incubated at 37 °C for 3 hours. After incubation, media was removed containing MTT and 100 μL of DMSO was put into each well. The plate was incubated at 37 °C for 30 minutes and read at absorbance OD=590 nm using Synergy HT spectrophotometer. For amidated mastoparan (0 – 40 μM) treated cells, the IC_{50} was calculated using a nonlinear regression model in GraphPad prism (version 9.3.1).

2.2.6 Concentrating Media

Up to 15 mL culture media from transfected cells on 100 mm dishes, transferred to Amicon® Ultra-15 Centrifugal Filter Unit (Sigma, Cat# UFC901024) with a 10 kDa Molecular Weight cut-off. The filter unit is spun at 3,000 rpm until a 200 μL final volume (15-60 minutes). Concentrated media is then aliquoted and stored at -20 °C.

2.3 Bioinformatics

2.3.1 CircRNA Grouping and Sorting

circAtlas database comprises of a large number of circRNAs identified from high-throughput RNA-seq data sets. The entire circRNA grouping and sorting workflow is represented in figure 2.3. 421,501 unique human circRNAs entries, across 240 samples obtained from 19 tissues have been identified (Wu, Ji and Zhao, 2020). The original sequence file (human_sequence_v2001.txt) was downloaded from circAtlas in a format with the information for each entry on the same line. The file consisted of 578,934 lines of sequences, each with an organism identifier (all human), a circRNA identifier e.g. hsa-MAP3K2_0035, and the circRNA cDNA sequence. All of the python scripts used in this project are original and are publicly available at https://github.com/stxdg8/circRNA-Analysis GitHub repository. Non-nucleotide sequences were removed by running the python script ATGC.py, this acted as a quality control for the entries, removing sequence errors with non-nucleotide sequences or entries with 'partial' written instead of a nucleotide sequence. Any future reference to running a script with an ending of .py, refers to executing a script in python3 (version - 3.6.9) within Ubuntu (version - 18.04) or Jupyter notebook (version - 6.0.3) (Sobell, M.G., 2015). CircRNAs were further sorted into circRNAs that contain an ATG start codon and those that lack an ATG by running ORF.py. CircRNAs that lack an ORF were further sorted into those that contain non-consensus start codons (CTG, GUG and ACG) and those that do not, by running Non_consensus.py. CircRNAs containing an infinite open reading frame (IORF), categorised by a start codon and no stop codon in the same frame were obtained by first running multiple_of_three.py then IORF.py. CircRNAs containing an ORF that reads through the circular/back splice junction were obtained by running ORF_through_junction.py (Figure 2.2).

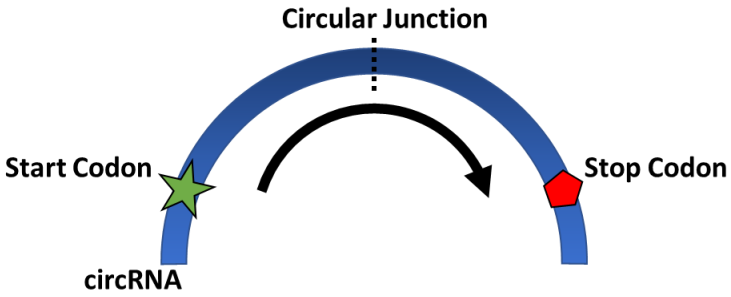


Figure 2.2 ORF through the circular junction. A circRNA with a start codon before the circular junction and a stop codon after it was determined as a circRNA with an ORF though the circular junction.

The pandas (version – 1.0.1) module of python (version - 3.6.9) within the Jupyter notebook (version - 6.0.3) of the anaconda navigator environment was used to organise data in a table format (McKinney., 2010; Van Rossum, G. & Drake, F.L., 2009; Kluyver, T. et al., 2016). A header for each column was added to enable sorting by column characteristics. Any reference to columns being dropped, added, rearranged, and tables merged were all performed in the aforementioned environment.

As the ORF through the junction was of interest, another script was written to extract just this ORF from the full circRNA sequence. The circRNA sequences were first rearranged by running ATG_ rearrangement.py, which rearranged the sequence so that it began with the start codon that closest to the backsplice junction. EMBOSSgetorf was executed in Ubuntu (Rice, Bleasby and Ison, 2011). This generates a file containing every ORF within each sequence. Many sequences contained more than one ORF but the rearrangement with ORF_through_junction_rearrangement.py prior to execution of EMBOSSgetorf, resulted in the first ORF identified for each circRNA sequence being the desired ORF through junction. Only_ORF_through_junction.py script was then run to retrieve only the ORF through junction sequences. Embosstranseq was used to convert the ORF through junction nucleotide sequences into amino acid sequences (Rice, Bleasby and Ison, 2011).

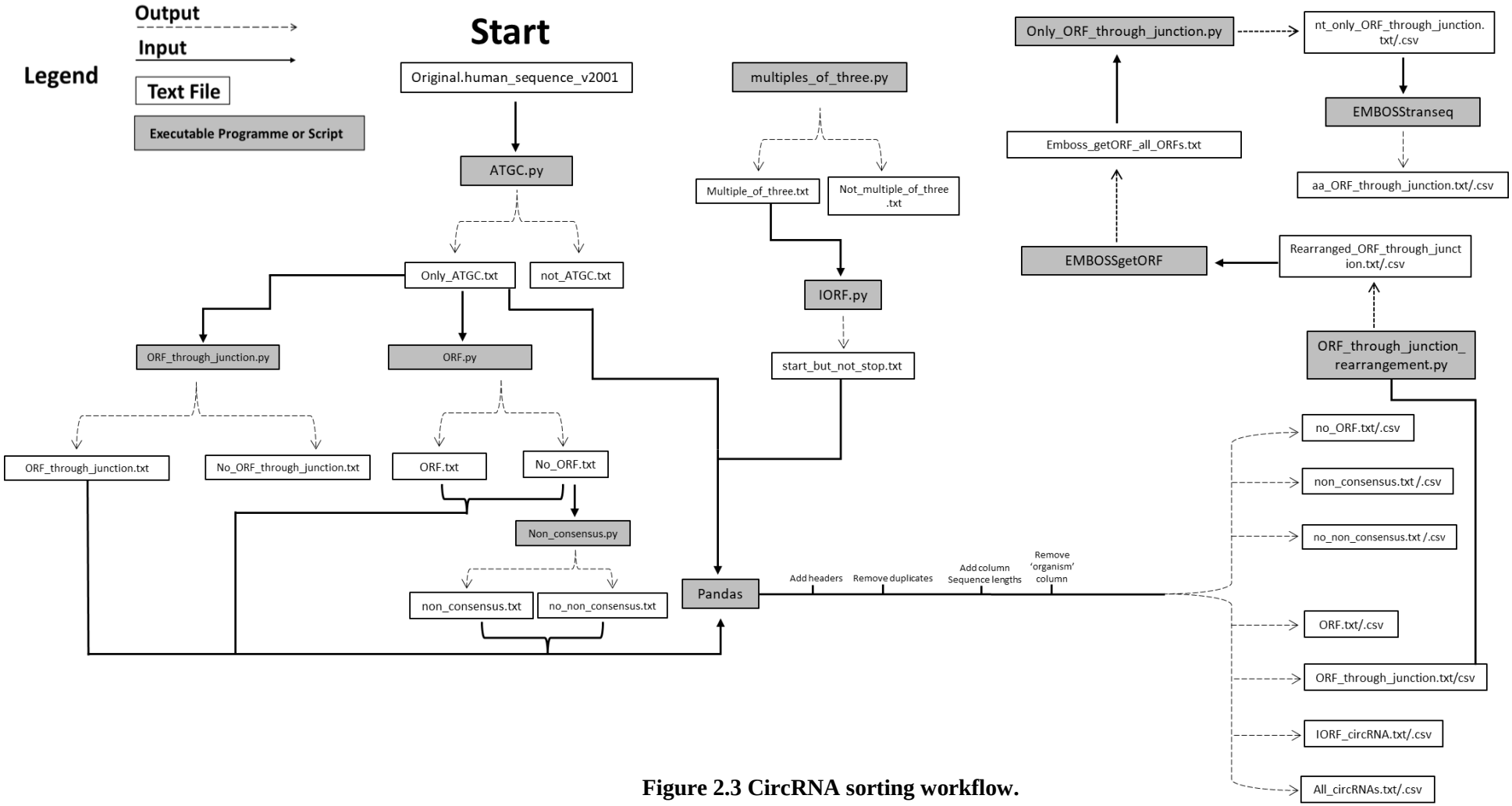


Figure 2.3 CircRNA sorting workflow.

2.3.2 Length Distribution Analysis

Sequence lengths were determined, and this information to the headings of the respective groups using the pandas module of python. Length distribution analysis was performed in GraphPad prism (version 9.3.1) in histogram plots of various bin widths for circRNA abundance and relative frequency.

2.3.3 Codon Adaption Index (CAI)

The Codon adaption index (CAI) gives an indication of how efficiently a sequence is translated by analysing the codon usage and giving a score ranging from 0-1. A score of 0 indicates, every codon is in its rarest possible form and a score of 1 indicates that every codon is the most common. All the individual circRNAs CAI in the groups' ORF through the junction, IORF, not IORF, CDS mRNA and ORF through the junction scrambled were calculated using CAIcal (Puigbò, Bravo and Garcia-Vallve, 2008). Using GraphPad prism (version 9.3.1), the CAI scores were plotted into a box and whiskers chart and then a plotted on a histogram with a bin width of 0.02.

The scrambled sequences of ORF through junction circRNAs were obtained by removing ATG from the beginning of the sequence using Notepad ++ (version 7.8.7), uploading this sequence to Galaxy, and using the shuffleseq tool to randomise the sequences whilst maintaining the same nucleotide composition (Jalili et al., 2020). After shuffling the sequences, the ATG start codon was added back to each sequence to enable the CAIcal program to run correctly.

2.3.4 Peptide Functional Analysis

Multiple programmes and tools were used to identify ORF through the junction circRNAs with a different predicted biological function to their parental gene. The entire workflow is summarised via a schematic in figure 2.4.

A Parent_gene_names_only.py was run to obtain the parental gene names from the circRNA names, this generated a text file of only the parent genes, called parent_genes.txt. The parental gene peptides were obtained using parent_genes.txt in Ensembl Biomart (Cunningham et al., 2022). The parental peptide sequences were obtained and ran in the same peptide analysis programmes as the ORF through the junction peptides, this enabled comparison between the parental and ORF through the junction results, so differential results could be identified.

It is possible my predicted circRNA peptides have been experimentally discovered but are unmapped to the genome. To check this a Blastp (Altchul et al., 1990) was performed with ORF through the junction peptides as the query sequence against the 1,347 unmapped proteins present in the human reference proteome (ID - UP000005640) obtained from Uniprot (UniProt Consortium, 2018). Only Blast results in which the alignment included at least 10 amino acids either side of the backsplice junction were considered, and in which had ≥ 95 % identity, by running over_94_percent_identity.py. This removed any hits that did not run through the junction and hits of very short sequences (< 20aa).

InterPro provides functional analysis of proteins by classifying them into families and predicting domains and important sites (Blum et al., 2021). This was used to provide functional analysis to the circRNA peptides and the parental gene peptides, to enable comparison and see identify differences. An InterProscan was performed for ORF through the junction peptides

and for the gene peptides. The database used for this analysis was PANTHER (version – 16.0). PANTHER is a (Protein ANalysis THrough Evolutionary Relationships) Classification System designed to classify proteins (and their genes) in order to facilitate high-throughput analysis (Mi et al., 2019). PANTHER was used because it reads the peptide sequence and assigns a PANTHER term, therefore this would allow for comparison between parental gene and circRNA peptide results. Some circRNA peptide sequences had multiple results for different regions within the sequence. Like the Blastp alignment only Interproscan results in which the alignment ran at least 10 amino acids either side of the backsplice junction were considered by running important_interproscan_results.py and then Interpro_results_through_junction.py.

The results obtained for ORF through the junction peptides and gene peptides were merged into a single table. The panther terms that differed between the parental gene and ORF through the junction peptides were returned in a single text file after running Panther.py.

The homology between the InterPro result and the parental gene was analysed by running a Blastp alignment (Altchul et al., 1990). A very similar homology would be an indication of biological function similarity. Any result with a Query cover (%) + Percentage Identity (%) over 150 % was disregarded.

Protein Homology/analogY Recognition (Phyre - version 2.0) provides a predicted 3D model of a peptide sequence and indicates confidence and coverage of the query sequence to proteins of known functions (Kelley et al., 2015). Phyre produces accurate (the core of the protein within 2-4Å r.m.s.d from the native) models of about 70% of the domains in a typical genome. The ORF through the junction peptides identified of interest were input into Phyre2 web portal and results obtained (Kelley et al., 2015).

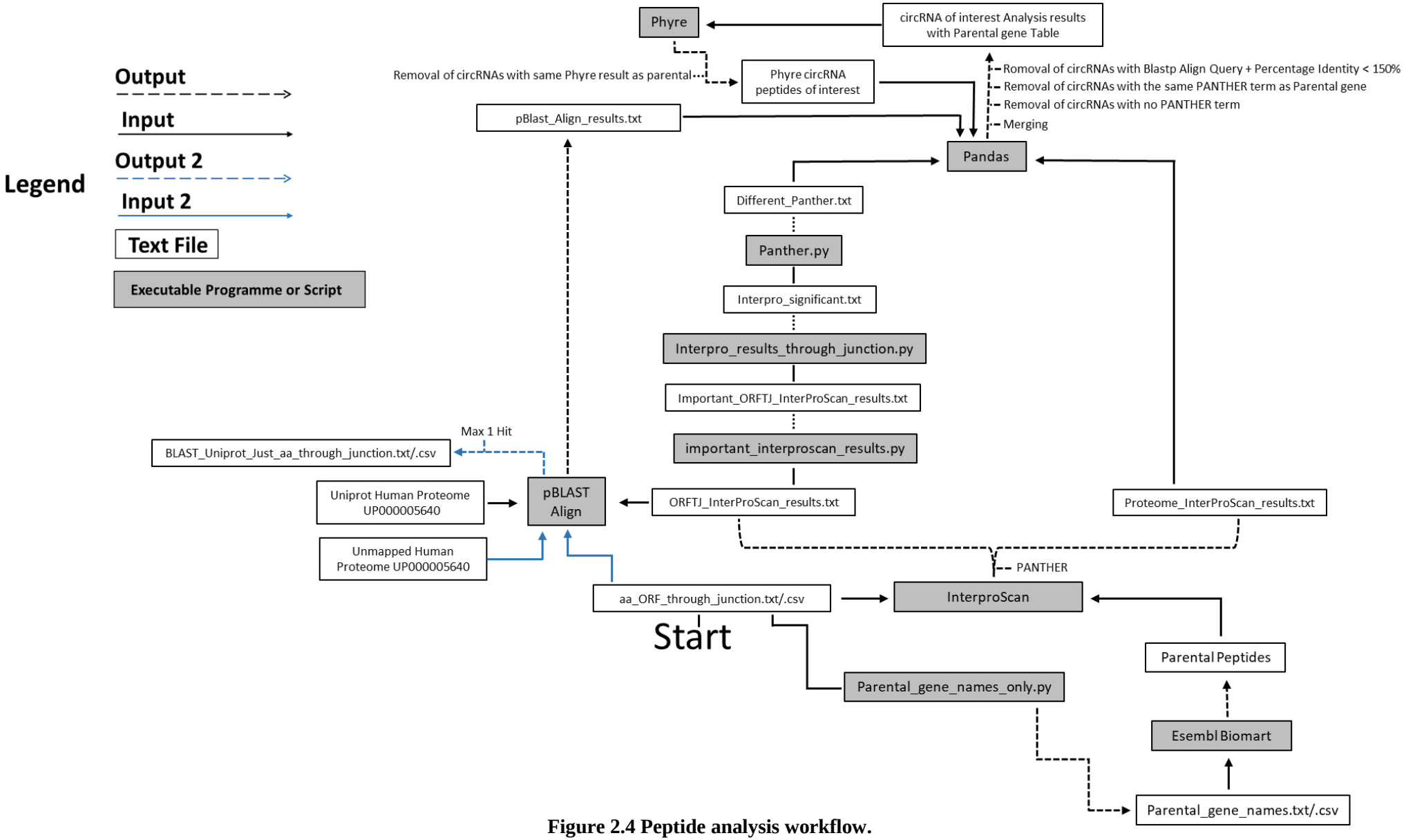


Figure 2.4 Peptide analysis workflow.

2.3.5 CircRNA Translatability

The likelihood of efficient translation from circRNAs was assessed in several ways. The codon adaption index (CAI) score was determined using the graphical interface of CAIcal_v1.4 and returned a value from 0-1 per transcript, depending on the codon usage (Puigbò, Bravo and Garcia-Vallve, 2008).

The individual circRNA profiles in circAtlas database were used to obtain the circRNA expression, location and conservation (Wu, Ji and Zhao, 2020).

SRAMP (A sequence-based N6-methyladenosine (m^6A) modification site predictor) was used to identify the m^6A consensus sites within the circRNA sequence and rank the likelihood of methylation (low confidence to very high confidence) (Zhou, Zeng, YH, Zhang and Cui., 2016).

IRESpy was used to predict potential IRES sites in the circRNA sequences (Wang and Gribskov, 2019). IRESpy uses sequence features such as kmer words, structural features such as QMFE, and sequence/structure hybrid features, to predict IRES probability. The circRNA sequences uploaded in FASTA format to the web server.

RiboCIRC, a database of predicted and known translated circRNAs (Li et al., 2021) was used to determine if circRNAs had previously been identified as translatable.

Chapter Three

3.0 Vector Construction, RNA Circularisation and Translation *In Vitro*

3.1 Introduction

Circular RNAs are covalently closed loops of RNA. The expression of circular RNA molecules has been widely detected across every domain of life (Ebbesen, Hansen and Kjems, 2017). In nature, circRNA sequences are produced from intron sequences via spliceosomal backsplicing and self-splicing of group I and group II introns (Lasda and Parker, 2014). CircRNAs can be engineered to circularise designer sequences with high efficiency.

In vitro RNA circularisation can result by enzymatic ligation of the RNA ends. Ligation of free RNA ends requires a 5' terminal monophosphate (Haugner III and Seelig, 2013). To circumvent the need for de- and re-phosphorylation post transcription, excess GMP can be used in the *in vitro* transcription reaction to obtain 5'-monophosphorylated linear RNA substrates (Abe et al., 2015). A linear oligonucleotide splint improves the circularisation efficacy of enzymatically ligated RNAs by assisting with ligation (Müller and Appel, 2016). The splint anneals to the 5' and 3' RNA ends, bringing them into flanking alignment. Ligase selection is essential to splint design: T4 RNA Ligase 1 requires nucleotide overhang from the DNA-RNA splint, whereas T4 DNA Ligase requires DNA-RNA hybridisation at the distal ends of the RNA precursor (Bain and Switzer, 1992; Moore and Sharp, 1992).

In vivo RNA circularisation strategies utilise biological processes that occur in nature, such as back splicing and self-splicing intron ribozymes (Meganck et al., 2021; Wesselhoeft, Kowalski and Anderson, 2018). Self-splicing intron rearrangement generates circRNAs, in which the 5' intron half is rearranged downstream of the exon and the 3' half upstream (Chen and Lu, 2021) (1.3.3). In group I introns, this method is known as permuted intron-exon (PIE), and in group II introns, the rearrangement is known as inverse splicing (Puttaraju and Been, 1992;

Mikheeva, 1997). The PIE method can generate circRNA in tubes, bacteria, and eukaryotic cells (Umekage and Kikuchi, 2009; Puttaraju and Been, 1992; Ford and Ares, 1994). Group II intron's ability to circularise engineered RNA sequences is less researched due to the formation of a 2' to 5' phosphodiester linkage at the branch site. Inverse splicing has been demonstrated *in vitro*, but it is likely to function in bacterial and eukaryotic cells as both contain endogenous group II introns with autocatalytic function (Jarrell, 1993; Mikheeva, 1997; Lambowitz and Zimmerly, 2010). The main differences between the PIE method and inverse splicing are that PIE creates a 3' to 5' phosphodiester linkage at the branch site and includes part of the native exon sequence in the circRNA. In contrast, inversed splicing produces a 2' to 5' phosphodiester linkage without native exon inclusion (Chen and Lu, 2021).

Several ways to evaluate RNA circularisation include RNase R treatment and RT-PCR. RNase R is an *E. coli* exoribonuclease which exhibits 3' to 5' exonuclease activity, digesting linear RNA, whereas circRNA is resistant due to its lack of free ends (Xiao and Wilusz, 2019). RNase R treatment can purify circRNAs and determine the circularisation efficiency, comparing the intensity of bands after gel electrophoresis before/after treatment (Wesselhoeft, Kowalski and Anderson,2018).

A more cost-effective, sensitive approach to detect circRNAs of a known sequence is RT-PCR, in which a complementary divergent primer set amplifies through the circular junction (Panda and Gorospe, 2018). Amplification is circRNA dependent because the reverse transcriptase must travel through the circular junction during first-strand synthesis for a successful PCR reaction. As this approach uses template amplification, it can be used to sensitively detect circRNAs of a known sequence (Wesselhoeft, Kowalski and Anderson, 2018).

As circRNAs are covalently closed loops, they lack a 5' methylated cap and poly(A) tail (Yang, Fu and Zhou, 2018). Therefore, internal ribosomal recruitment is required for translation initiation. There are two known methods for translation initiation of circRNAs, m^6A methylation sites and internal ribosomal entry sites (IRES) (Lei et al., 2020). The most common approach used for engineered circRNA translation is IRES-driven translation. Cellular IRESs are diverse in their size, differ in cells they have optimal activity, and vary in the IRES trans-acting factors (ITAFs) required to facilitate translation. Over fifty proteins have been described for their ability to specifically regulate cellular IRESs (Godet et al., 2019). Typically, smaller and more compact IRESs require fewer ITAFs for ribosomal recruitment (Yang and Wang, 2019). Spriggs, Mitchell and Willis (2005) engineered an artificial IRES named KMIB that shows activity *in vitro* (rabbit reticulocyte system) and *in vivo* (HeLa cells). KMIB IRES contains CCU repeats in a stem-loop structure and mismatched base pairs in the (GGA)n half of the hairpin.

The ribosome cycles around the circRNA multiple times if there is an infinite open reading frame (IORF) (Perriman and Ares, 1998). KMIB's compact sequence is ideal for translation of IORF circRNA as it causes minimal disruption to translation caused by non-optimal codon usage and lacks a stop codon.

Studying highly repetitive sequences is challenging due to nonspecific binding in PCR-based amplification, resulting in nonspecific products (Scior, Preissler, Koch and Deuerling, 2011). CircRNAs with an IORF have been engineered to express repetitive peptide concatemers in multiple hosts, including *in vitro*, bacterial, and *in vivo* (mammalian cells) (Abe et al., 2015;

Perriman and Ares, 1998; Wesselhoeft., et al., 2019). Overcoming challenges to produce repetitive peptides is beneficial in studying inherited human disorders linked to trinucleotide repeat expansions and producing valuable biomaterials like spider silk.

In vitro synthesis of proteins in a cell-free system is advantageous when the over-expressed product is toxic to the host cell, the product is insoluble or when intracellular proteases degrade the protein in host cells (Shimizu et al., 2006). The cell extracts used for *in vitro* translation contain all the macromolecular components required for mRNA translation, including the 70S or 80S ribosomes, tRNAs, aminoacyl-tRNA synthetases, initiation, elongation, and termination factors (Hammerling, Krüger and Jewett, 2019). The cell-free extracts can theoretically be made from any cells, but the most commonly used are Rabbit Reticulocyte Lysate (Eukaryotic expression system) and *Escherichia coli* (bacterial expression system).

This section describes a system to circularise IORF circRNAs encoding Mastoparan and FRET peptide sequences. *In vitro* experiments are described which establish RNA circularisation and peptide production.

3.2 Results

3.2.1 SDFLAG Design and Cloning

An oligonucleotide sequence was designed to encode epitope tags in both orientations. This allowed multiple copies of the sequence to be ligated together in different orientations while allowing detection of resulting peptide. The oligo named SDFLAGAHSIH encoded a Shine

Dalgarno site and a FLAG tag in one orientation and haemagglutinin (HA) and ^{x6}Histidine (^{x6}His) Tag in the other (Figure 3.1)

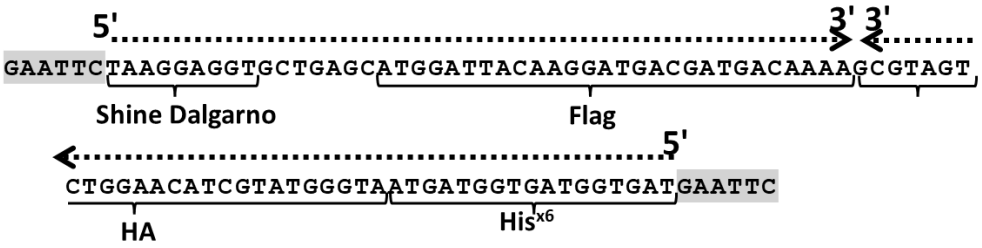


Figure 3.1 SDFLAGAHSIH oligo design. SDFLAGAHSIH nucleotide sequence containing a Shine Dalgarno FLAG sequence in one orientation and haemagglutinin x6His tag in the other orientation. The SDFLAGAHSIH oligo is flanked by *EcoRI* restriction enzyme sites (grey).

The sequence and its complement were annealed together and the double-stranded product was digested with EcoRI. The reaction was cleaned up with a PCR clean-up kit (2.1.4).

The digested oligos were ligated together using T4 DNA Ligase resulting in a string of the oligos ligated together in different orientations. pRSETB vector was also digested with EcoRI, dephosphorylated and purified with a PCR clean-up kit (2.1.6). The vector and oligos were ligated together and transformed into *E. coli* DH5α cells.

The resulting colonies were subject to a PCR colony screen with pRSETB_Fwd and pRSETB _R276 primers. Two colonies contained a double copy of the insert. The double-copy colonies were miniprepped and sequenced. Sequencing revealed that both colonies were in the same orientation of Shine Dalgarno Flag x2 (Figure 3.2)

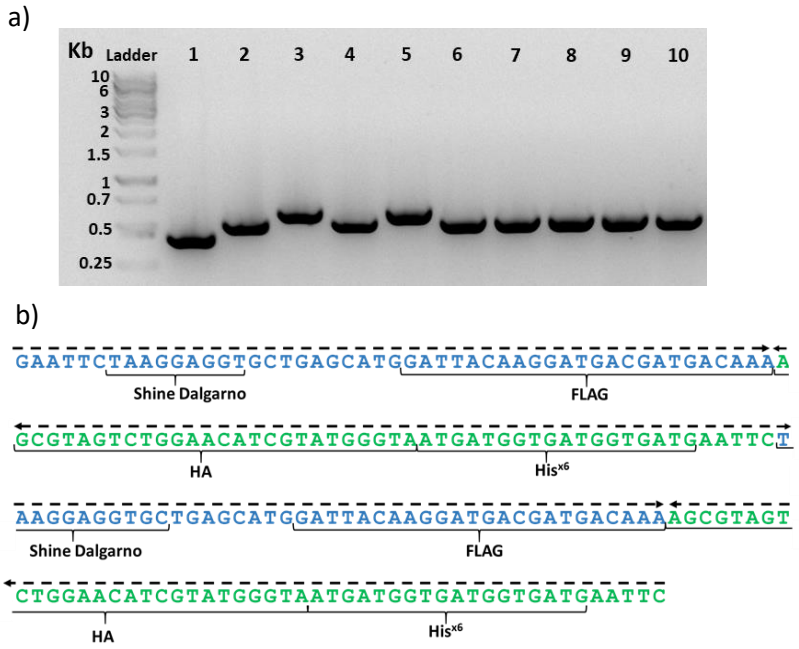


Figure 3.2 SDFLAGAHSIH cloning. a) 1% w/v Agarose Gel Electrophoresis Colony screen PCR for SDFLAGAHSIH in pRSET B. Lane 1 - pRSETB +ve control, Lanes 2,4, 6-10 – pRSETB SDFLAGAHSIH one copy insertion, Lane 3, 5 - pRSETB SDFLAGAHSIH double copy insertion b) The Sequence of the Double Copy SDFLAGAHSIH Insertions – The sequence confirmed to be a double copy Shine Dalgarno Flag, Shine Dalgarno Flag.

The only requirement for the construct was that it contained a Shine Dalgarno sequence for ribosomal recruitment, as these did. The new construct was named pRSETB SDFLAG

For an infinite open reading frame (IORF) to be present in a circularised mRNA, a stop codon was removed by site-directed mutagenesis, changing a TGA to a TGG using primers pRSETB_IORF_Fwd and pRSETB_IORF_Rev (Figure 3.3a). This results in a frame with no stop codons (Figure 3.3b).

a)

B e f o r e

TGG CTA GCA TGA CTG GTG GAC
W L A - L V D

A f t e r

TGG CTA GCA TGG CTG GTG GAC
W L A W L V D

b)

5'3' Frame 2

G D H N G F P L E I I L F N F K K E I Y I C G V L I I I
I I Met V W L A W L V D S K W V G I C T T Met T I R I R
A R D L Q L V P W N S K E V L S Met D Y K D D D D K S V
V W N I V W V Met Met V Met V Met N S K E V L S Met D Y
K D D D D K S V V W N I V W V Met Met V Met V Met N S K
L D P A A N K A R K E A

Figure 3.3 Site-Directed Mutagenesis of SD FLAG and IORF Design. a) Site-Directed Mutagenesis TGA to TGG to remove the stop codon of pRSETB SDF, genereating pRSET SDF IORF plasmid b) Amino Acid Sequence of Protein Expressed from pRSETB - DNA transcribed with T7 Promotor and Translation initiated with Shine Dalgarno Sequence.

After site-directed mutagenesis and Dpn1 treatment, the vector was transformed into DH5α cells. Colonies were selected, miniprepped and the desired mutation confirmed by sequencing. The construct was named pRSET B SDFLAG IORF.

3.2.2 Circularisation of SDFLAG RNA

The pRSET B SDFLAG IORF was linearised by PCR amplification (pRSETB_Fwd and pRSETBR_273) and purified by ethanol precipitation (2.1.7). 5-10 µg of linearised pRSETB SDFLAG IORF DNA template was used for an *in vitro* transcription reaction with 10-fold excess GMP. The reaction was purified using a transcription clean-up kit. A DNA splint was hybridised to both the 5′ and 3′ ends of the RNA. RNA ends were ligated using RNA ligase 1 (2.1.14 - 2.1.15). DNase I was added to remove the splint. Linear RNA contaminants were removed by RNase R treatment. RNA was cleaned up with the in vitro transcription clean-up kit (2.1.5). The RNA circularised with the DNA splint, named circSDFLAG IORF. The efficiency of RNA circularisation was determined by running an agarose gel electrophoresis.

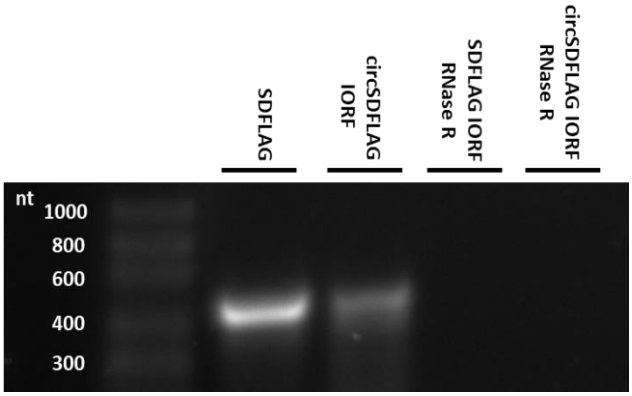


Figure 3.4 SDFLAG RNA circularisation with a DNA splint. 2% w/v Agarose Gel Electrophoresis of Splint Hybridised SDFLAG RNA With RNase R Treatment. Shine-Dalgarno Flag RNA transcribed from pRSET B T7 promoter before circularisation (SDFLAG) and after (circSDFLAG IORF), with or without exonuclease RNase R treatment. SDFLAG RNA expected size 426nt.

The gel electrophoresis shows unsuccessful RNA circularisation. circRNAs migrate slower on agarose gels than linear RNA of the same length (Tabak et al., 1988). Figure 3.4 shows that the SDFLAG RNA migrated the same distance before attempted circularisation and after (circSDFLAG IORF RNA). Therefore, it is unlikely that RNA was circularised. This is confirmed by the absence of an RNA band when treated with RNase R. circRNAs would demonstrate resistance to RNase R, and a band would remain after treatment.

Splint hybridisation was attempted using a variety of DNA splints of different sized homology regions with no success (data not shown). It was concluded that DNA splint hybridisation is not a method that can be used to circularise RNA *in vitro*; alternative methods are necessary.

3.2.3 Vector Construction

3.2.3.1 Amplification of Total Yeast DNA

Group II introns can be engineered to circularise RNA of a desired sequence (1.3.3). The inverse intron sequence is constructed by amplifying group II intron domains I-III and IV-VI from total yeast DNA. The primers amplifying domains I-III (INV I NcoI PmlI Fwd /INV III BamHI Rev) and IV-VI (IV XbaI Fwd/ INV VI EcoRV NcoI Rev) added restriction digest sites to the distal ends of the domains. The reaction enzymes sites allow the construction of the inverse intron sequences flanking a desired coding region and insertion into a plasmid vector (Figure 3.5).

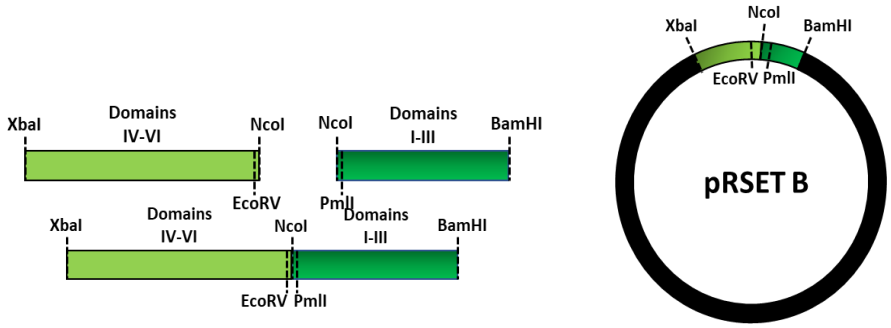


Figure 3.5 Inverse group II intron assembly and insertion into pRSET B. Amplification of domains I-III and IV-VI adds restriction enzyme sites to the ends on the amplified domains. This allows for intron inversion by digestion/ligation, and insertion into pRSET B. Blunt ended digestion and ligation can then be used to insert a desired sequence for circularisation between then end of domains VI and the start of domain I.

Yeast total DNA was extracted from dry yeast by freeze-thaw method (2.1.1). The group II intron domains were amplified by PCR (3.2.3.1). Attempts to amplify the domains using a standard PCR method were unsuccessful, as distinct bands could not be observed on an agarose gel after gel electrophoresis (data not shown). Optimisation by changing various conditions, including temperature, touch-down PCR, magnesium titration, and addition of betaine, also did

not amplify the domains successfully, as visible on an agarose gel. Controls for genomic yeast DNA (GAPDH Fwd, GAPDH Rev) and mitochondrial yeast DNA (MT Fwd, MT Rev) were obtained to check the quality of the extracted DNA.

There are other methods of yeast DNA extraction, including extraction from a live yeast culture and a glass bead approach for lysis rather than a freeze-thaw method.

It was considered that using a freeze-thaw DNA extraction approach may not yield optimal DNA quality and yield, so alternative methods were performed. Genomic and mitochondrial DNA amplification controls were obtained to assess the DNA quality from the various extraction methods. The genomic DNA control amplified a region within the GAPDH gene, and the mitochondrial DNA control amplified a region within the 21S rRNA locus.

A PCR reaction was performed, amplifying the mitochondrial and genomic control regions. DNA was extracted from dry and live yeast using glass beads, and a freeze-thaw approach was used in the reaction. PCR products were viewed on agarose gel electrophoresis (Figure 3.6).

The results demonstrated that DNA extractions from live yeast culture provide the best template for PCR amplification as both genomic and mitochondrial controls amplified more intensely. DNA extracted via the glass bead method yielded the clearest amplification product, so this DNA was used in subsequent amplifications.

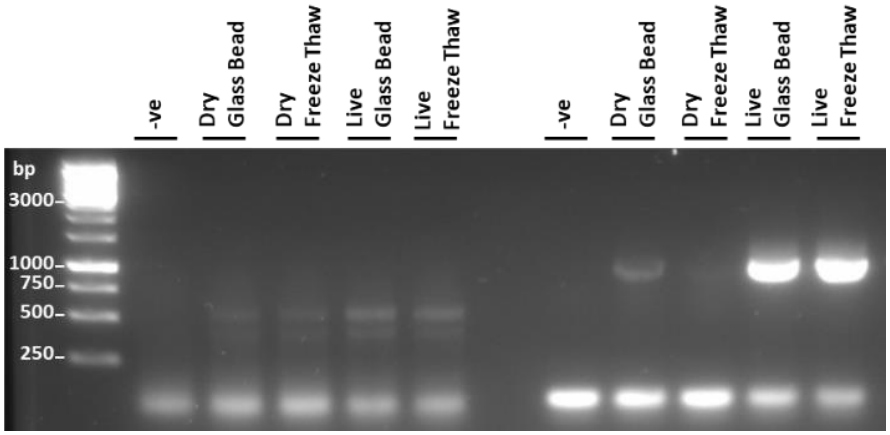


Figure 3.6 Genomic and Mitochondrial DNA yeast amplification using different DNA extraction methods. 1.2% w/v Agarose Gel Electrophoresis. Left – Mitochondrial DNA control amplification (MT Fwd/MT Rev), Right – Genomic DNA control amplification (GAPDH Fwd, GAPDH Rev). Dry – DNA extracted from dry wild yeast; Live – DNA extracted from a live yeast culture; -ve – No DNA template in PCR reaction.

Although the control confirmed successful mitochondrial DNA extraction and integrity of the DNA, subsequent attempts to amplify the group II intron domains were unsuccessful. As the mitochondrial DNA controls were successfully amplified, the group II intron's unsuccessful amplification must be due to a sequence-specific reason, likely due to its highly structured nature. A nested approach was used to amplify a larger region (1375 bp) of the mitochondrial DNA within the Cox 1 gene, containing the desired group II introns sequence within it. This would allow the primers to bind to a less structured region which may aid amplification of the template DNA. The primer set selected for this approach (Cox1 Fwd/Cox 1 Rev) is bound to the nearest coding regions flanking the group II intron sequence. This gave a greater concentrated region for amplification to reduce nonspecific binding primer binding to the genomic/mitochondrial DNA. Using the Cox1 primer set, a single distinct band was amplified (Figure 3.7). However, after purification, subsequent amplification of domains I-III and IV-VI from this region was unsuccessful (data not shown).

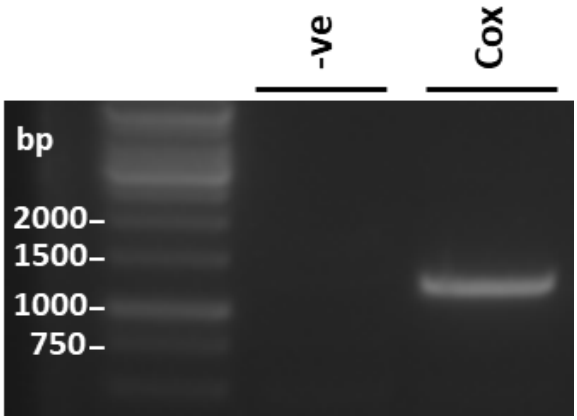


Figure 3.7 Cox1 DNA amplification. 1.2% w/v Agarose Gel Electrophoresis. Cox1 DNA **(**1375 bp**)** amplified with Cox 1 Fwd / Cox 1 Rev. -ve – No DNA template in PCR reaction.

3.2.3.2 Overlapping Oligo PCR

Overlapping PCR was attempted to amplify the yeast group II intron DNA. A set of seven oligos with a 20 bp overlap were designed, which covered the full sequence for domains I-IV, and three oligos were designed with a 20 bp overlap to construct domains IV-VI fully (Figure 3.8). Amplifying in pairs, the full-length sequences could be constructed with 3 rounds of PCR amplification. In the first round of synthesis, oligos 1-2, 3-4, 5-6 and 7 were amplified for domains I-III and oligos 1-2 and 3 alone were amplified for domains IV-VI (Figure 3.9). The primers used to amplify these regions are bound to the beginning of the first oligo (forward) and the end of the other (reverse). In the second round of synthesis, double-stranded DNA for the domains were paired again. For domain I-III synthesis 1,2, and 3,4 DNA was paired, and 5,6 and 7 DNA was paired. For domains, IV-VI synthesis 1,2 and 3 DNA was paired. After the second round, domain IV-VI is fully constructed, but domain I-III requires another round of amplification in which 1-4 and 5-7 DNA are paired.

Figure 3.8 Yeast group I intron construction with 7 overlapping oligos. Each oligo contains a 20 bp overlapping regions indicated by the matching colours.

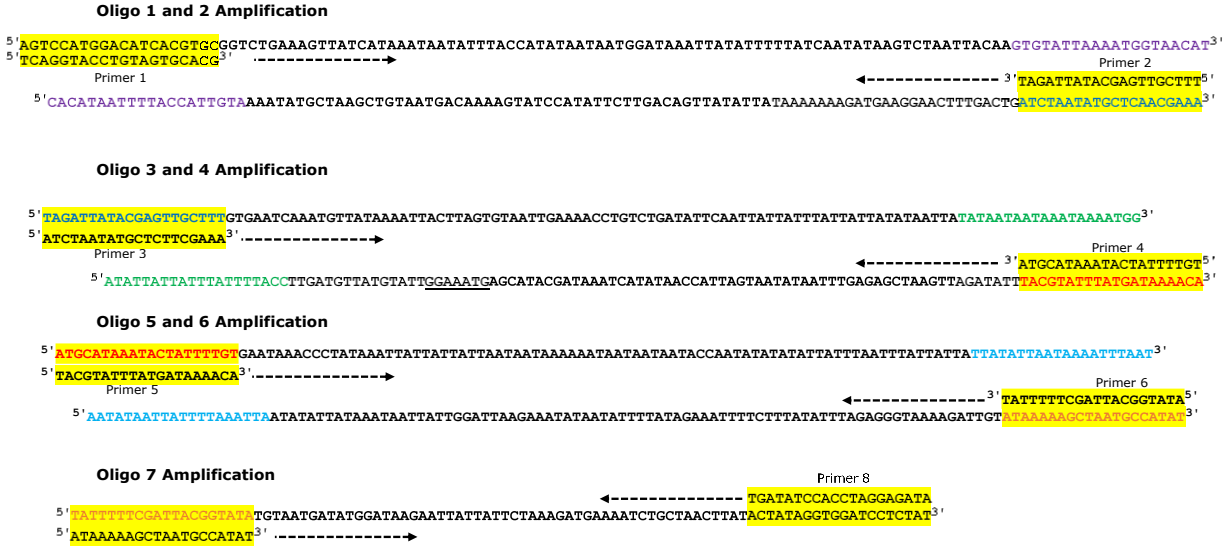


Figure 3.9 First round of domain I-III oligo pair amplification. Primers at the distal ends of each oligo pair allow amplification into double-stranded DNA. The first-round synthesis for domain I-III Oligos pairs are 1-2, 3-4, 5-6 and 7 alone, for domains IV-VI pairs are 1-2 and 3 alone. First round synthesis is followed by another 1 (IV-VI domains) or 2 rounds (I-III domains) of pair amplification to synthesise the entire region.

After two (for domain IV-VI) or three rounds (for domains I-III) of PCR amplifications, the sequencing results revealed that only small segments of the expected sequence had been amplified for both domains, showing that nonspecific annealing was causing problems in amplifying the desired product (Figure 3.10). Therefore, it was observed that overlapping PCR could not be used to amplify the group II intron sequence.

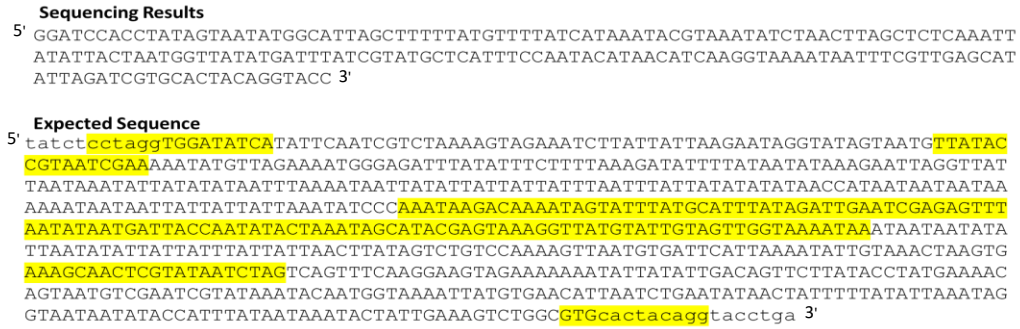


Figure 3.10 Sequencing results for after 3 rounds of domain I-III oligo amplification. The highlighted regions are the sequences of the group II intron I-III domains that were present in the sequencing results.

3.2.3.3 Gibson Assembly

A Gibson assembly approach was used to construct the intron domains as described in methods 2.1.3. After the Gibson reaction and transformation, either no colonies were observed, or the colonies were negative with after a PCR colony screen. This was repeated with ultra-competent cells with the same negative result. Due to the negative result in step 3 (Figure 2.1), the further stages of intron assembly could not be performed.

3.2.4 Self-Splicing *Anabaena* Vector Design and Cloning

An RNA circularisation system was designed (Figure 3.11), adapted from Wesselhoeft et al., 2018 to circularise RNA in mammalian cells. The construct design includes regions of homology arms either side a permuted group I *Anabaena* self-splicing intron, bringing the splice sites closer together by base pairing (Figure 3.11 b). This increases the likelihood of splice-site interaction by generating a splicing bubble, allowing circularisation of large RNA sequences up to 5 kb in length. The design also includes spacer sequences to separate the coding region and IRES from the self-splicing introns, avoiding interference which may reduce the

splice efficiency. The Internal Ribosomal Entry Site (IRES) used in this project is a minimalistic artificial IRES site known as KMIB IRES (Spriggs, Mitchell and Willis, 2005). KMIB IRES is compact lacks any potential stop codons. A minimal sequence is advantageous during translation to reduce likelihood of stalling and disruption caused by rare codons. The spliced circRNA contains a start codon with no stop codon, this gives the circRNA an infinite open reading frame (IORF) for continuous translation, cycling many times around the circRNA producing a long repetitive peptide (Figure 3.11 c).

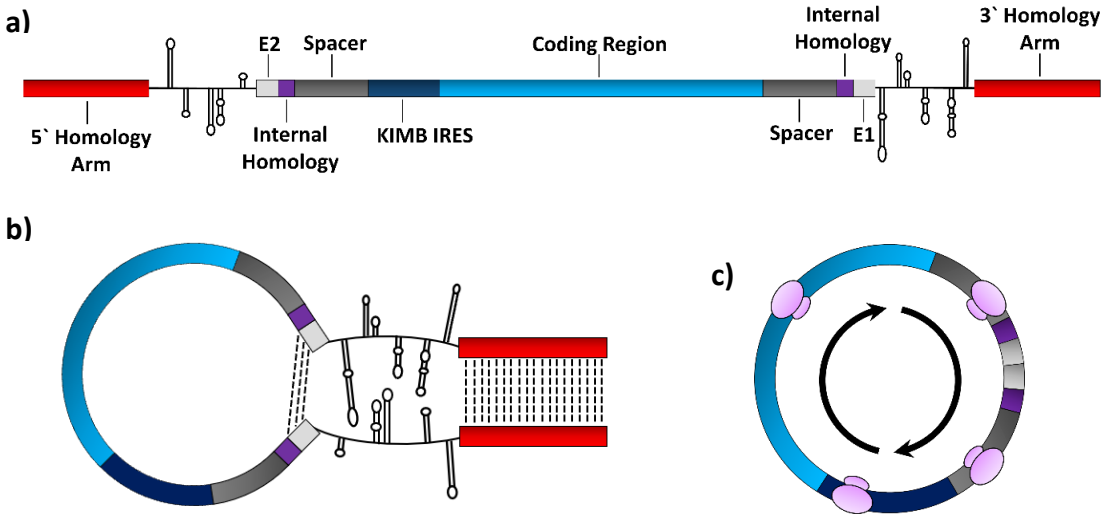


Figure 3.11 Engineered permuted group I *Anabaena* introns. a) Schematic of the Engineered Permuted Group I *Anabaena* Intron b) The splicing bubble generated by the homology regions before and after the splice sites c) Continuous translation of circRNA with an infinite open reading frame (IORF).

3.2.5 Mastoparan/FRET Vector Design and Cloning

Two coding sequences were inserted into the engineered *Anabaena* sequence for circularisation. The two coding sequences were designed, one containing a repeat mastoparan sequence (Figure 3.12) and the other containing a FRET peptide sequence (Figure 3.13). At the N-terminus of both the FRET and mastoparan coding sequence is a signal peptide for

peptide secretion outside of the cell. Between each mastoparan unit or EGFP/mCherry (FRET construct) was a broad MMP cleavage site PLGVRGC. The FRET system was designed to monitor the process from peptide production to activation in. A successful system of peptide production, exportation and MMP activation would lead to EGFP fluorescence in the media in the extracellular environment *in vivo*.

Figure 3.12 FRET peptide design and activation by MMP cleavage. a) Schematic of uncleaved FRET peptide b) Cleavage of the FRET peptide by MMPs resulting in the emission of EGFP.

The mastoparan coding sequence was designed to contain double mastoparan peptide sequence, separated by MMP cleavage sites. It was important that more than one repeat of mastoparan was included, because MMP cleavage would still be required for peptide activation even if circularisation was not successful and translation occurred from linear mRNA. Full LF cleavage at the 3 MMP cleavage sites would release two active mastoparan peptides (Figure 3.13b).

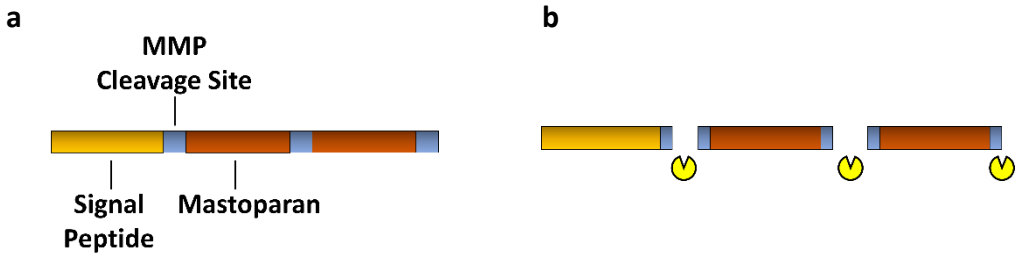


Figure 3.13 Mastoparan peptide design and functionality as a prodrug. a) Schematic of the uncleaved Mastoparan peptide b) Activation of Mastoparan by MMP cleavage.

A summary of the events that take place from transfection to peptide activation *in vivo* are demonstrated in figure 3.14, the events include transfection, transcription, self-splicing into circRNA, IRES-driven translation, peptide secretion, MMP cleavage and activation.

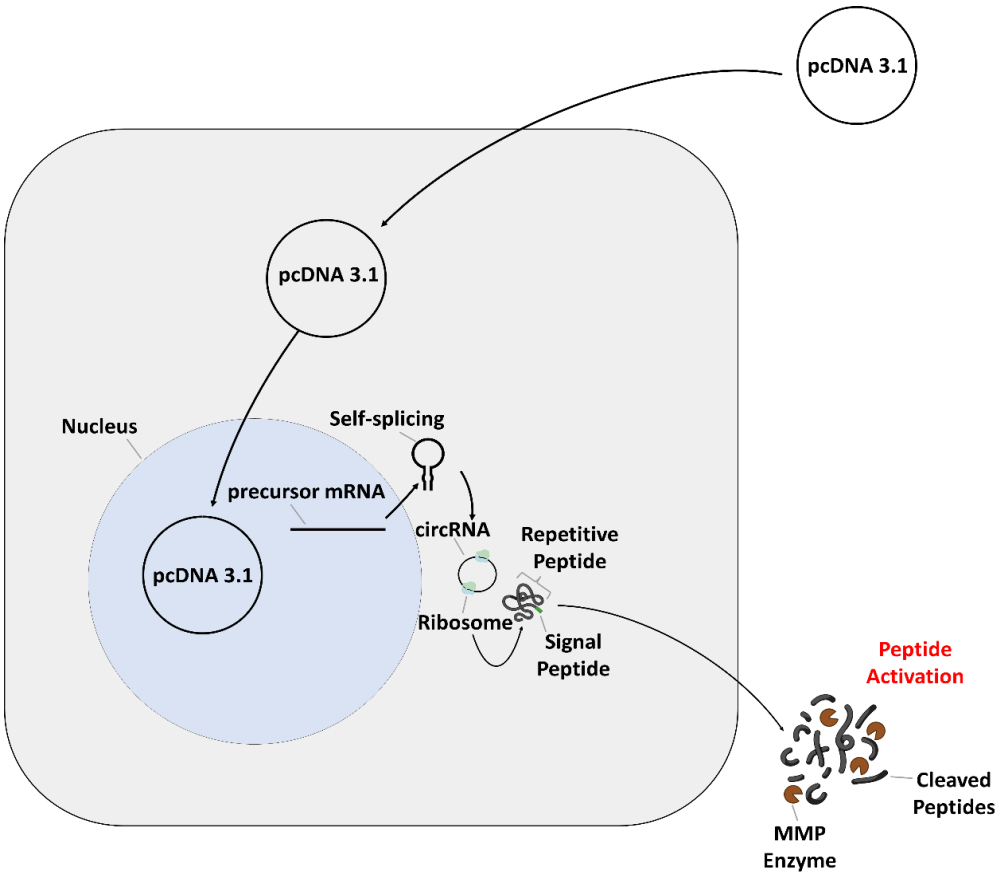


Figure 3.14 Producing a peptide concatemer inside mammalian cells. The full pathway from transfection to peptide cleavage and activation. Transfection of vector into cell, transcription, and self-splicing into circRNA, translation initiation and peptide concatemer production, exportation out of the cell via the signal peptide, activation of the peptide by MMP cleavage.

Controls were designed which contained the same FRET/Mastoparan sequence in the absence of the KMIB IRES and the PIE *Anabaena* self-splicing intron sequence. These controls do not form circRNA and translation is initiated in a cap-dependent manner. All the sequences designed are shown in figure 3.15.

These 4 constructs are designed to produce different peptides, the features and functions are summarised in table 3.1. *Anabaena* FRET construct can self-splice into circRNA (circFRET) which produces a repeating FRET peptide concatemer with a signal peptide. Upon secretion and MMP cleavage, the eGFP will no longer be quenched by the mCherry peptide green fluorescence will result.

Anabaena Mastoparan construct can self-splice into circRNA (circMast) which produces a repeating mastoparan peptide concatemer with a signal peptide. Upon secretion and MMP cleavage, mastoparan will be activated by MMP cleavage, causing resulting in cell death of the cancerous cells.

Linear FRET contains the same coding sequence and signal peptide as Anabaena FRET construct however it lacks the ability to self-splice and translation is initiated in a cap-dependent manner producing a single FRET peptide unit. This FRET peptide is designed to be exported and cleaved resulting in fluorescence.

Linear Mastoparan contains the same coding sequence and signal peptide as *Anabaena* Mastoparan construct however it lacks the ability to self-splice and translation is initiated in a cap-dependent manner producing 2x mastoparan peptides separated by MMP cleavage sites. This mastoparan repeat peptide is designed to be exported and cleaved resulting in activation and ultimately cell death.

The sequences were inserted into PCDNA 3.1 by ligation into HindIII and XbaI sites as shown in figure 3.15. All clones were successfully constructed as show by PCR amplification (primer

set pcDNA Fwd / pcDNA Rev) and agarose gel electrophoresis (Figure 3.16), and confirmed by sequencing.

Table 3.1 Features and functions summary of the four mastoparan/ FRET designed constructs.

	Features and Function							
Construct	Encoded Peptide	Engineered *Anabaena* Introns	KMIB Sequence	Signal Peptide	Circularises	Translation Initiation Mechanism	Number of peptides	Effect After MMP Activation
LF	FRET	✖	✖	✓	✖	N7-methylguanosine Cap	1 FRET Peptide	EGFP Fluorescence
AF	FRET	✓	✓	✓	✓ (circFRET)	IRES	Multiple FRET (Concatemer)	EGFP Fluorescence
LM	Mastoparan	✖	✖	✓	✖	N7-methylguanosine Cap	2 x Mastoparan	Cancer Cell Cytotoxicity
AM	Mastoparan	✓	✓	✓	✓ (circMast)	IRES	Multiple Mastoparan (Concatemer)	Cancer Cell Cytotoxicity

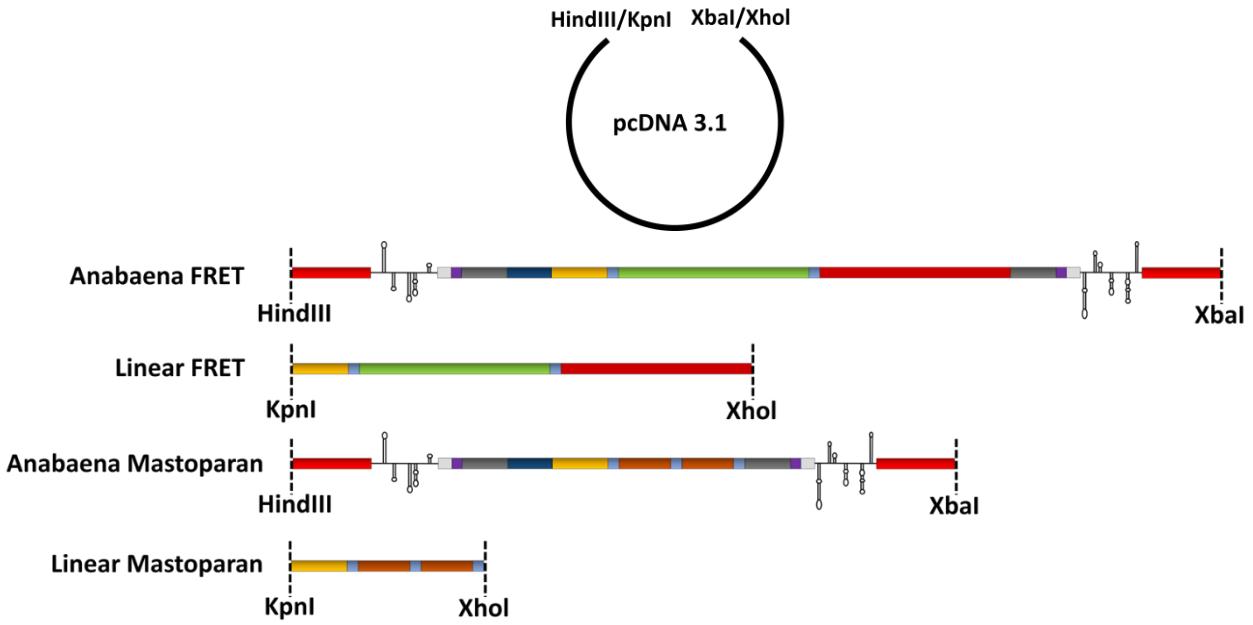


Figure 3.15 Cloning of the coding sequences into pcDNA 3.1. Some coding sequences are contained within PIE *Anabaena* group I intron sequence (*Anabaena* FRET and *Anabaena* Mast) for precursor mRNA self-splicing and circularisation, others (Exported FRET and Exported Mast) act as linear controls that do not self-splice. *Anabaena* containing constructs are cloned using HindIII and XbaI restriction enzymes and non-*Anabaena* constructs cloned using KpnI and Xhol restriction enzymes.

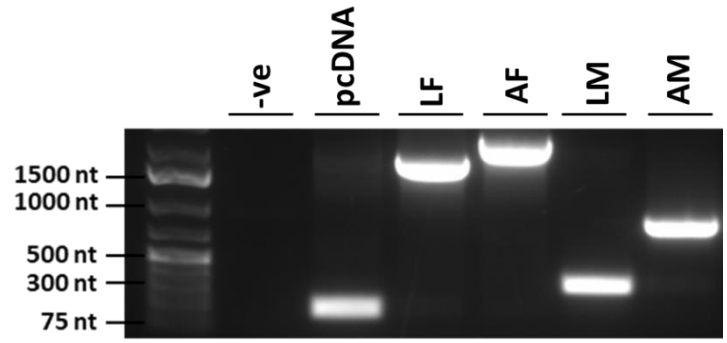


Figure 3.16 All constructs cloned into pcDNA3.1. 1% agarose gel. Electrophoresis ran after PCR amplification using primer set pcDNA Fwd / pcDNA Rev. -ve – PCR reaction lacking DNA template; Band sizes: pcDNA – 180 nt; Linear FRET (LF) – 1648 nt; *Anabaena* FRET (AF) – 2180 nt; Linear Mastoparan (LM) – 337 nt, *Anabaena* Mastoparan (AM) – 871 bp.

3.2.6 Self-splicing CircRNA Formation and Confirmation

To determine the self-splicing ability of both mastoparan and FRET sequences inserted into PIE *Anabaena, in vitro* self-splicing was performed by adding guanosine nucleotides and heating to 52^{o}C (2.1.15). Figure 3.17 and 3.18 shows *Anabaena* Mastoparan and *Anabaena* FRET, preceding transcription (before self-splicing), and after being subject to self-splicing condition. Lanes with 'R' indicate samples treated with RNase R, an exonuclease that degrades linear RNA. The expected product sizes after *Anabaena* Mastoparan self-splicing correspond to the bands in figure 3.17a: unspliced RNA – 878 nt (top), circMast – 474 nt (middle), and the two introns both of lengths 187 nt (bottom). The expected product sizes after *Anabaena* FRET self-splicing correspond to the bands in figure 3.17b, unspliced RNA – 2164 nt (top), circFRET – 1788 nt (top), two introns both of length 187 nt. The positive control is RNA (1800nt) produced from the transcription kit control vector. Only two bands can be seen in this gel as the unspliced RNA spliced circRNA product cannot be distinguished between. The results show that even before the *Anabaena* sequences are treated under self-splicing conditions, self-splicing still occurs, as there is a band/bands before self-splicing for the spliced products. The band remaining after RNase R treatment confirms the presence of circRNA. After *Anabaena* sequences are treated under self-splicing conditions, the splicing products are increased.

RT-PCR was performed to confirm the presence of circMast RNA and circFRET. The results confirm the presence of circRNA both before and after self-splicing. The expected product sizes for circMast are 450 nt for the linear control primers (convergent primers), and 474 nt for the circRNA primers (divergent primers) (Figure 3.18a). The expected product sizes for circFRET are 1774nt for linear control primers and 1788 nt for the circRNA primers (Figure 3.18b). As expected, the vector control lacks a band for the circRNA primers set, as the extension time of 30 seconds was not enough for vector amplification at a rate 1kb per 15-30 seconds.

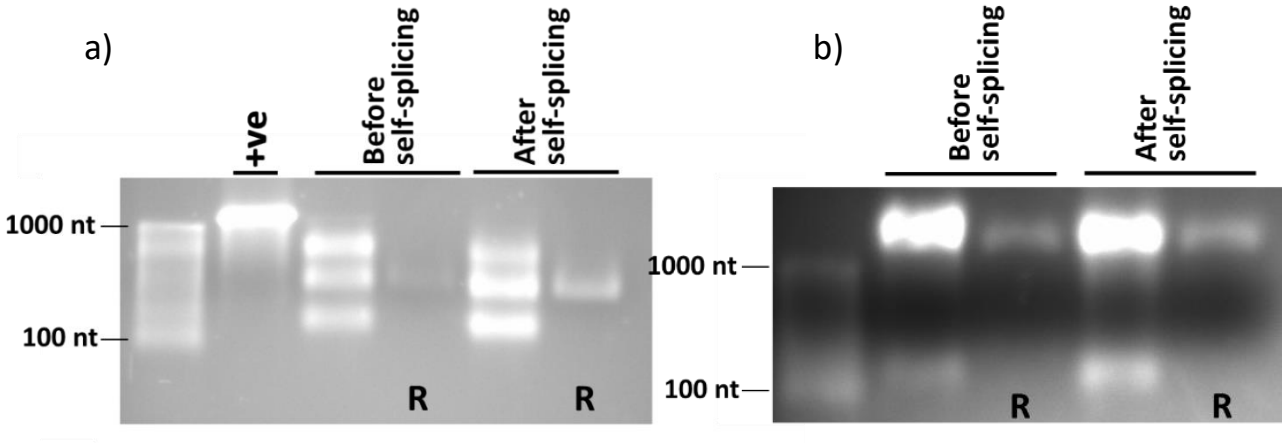


Figure 3.17 Self splicing of *Anabaena* Mastoparan and *Anabaena* FRET. 1% denaturing agarose/formaldehyde gel**.** a) *Anabaena* Mastoparan b) *Anabaena* FRET. R indicates wells with samples treated with RNase R. After self-splicing is after addition of guanosine nucleotides and incubation at 52°c and for 30 minutes a) +ve – Transcription kit control positive (1800nt).

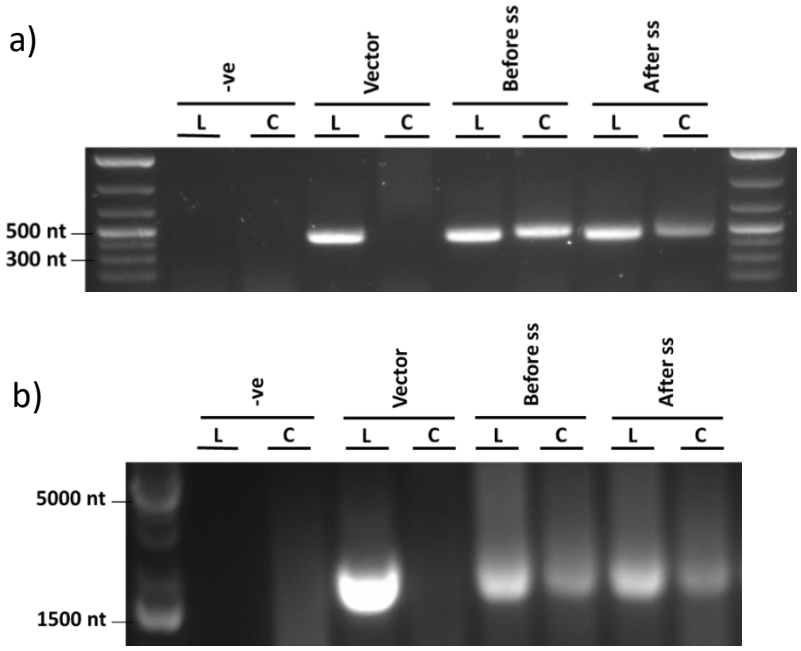


Figure 3.18 RT-PCR of *Anabaena* FRET and *Anabaena* Mastoparan RNA produced *in vitro*. a) *Anabaena* Mastoparan b) *Anabaena* FRET. -ve – PCR reaction lacking a nucleic acid template; L – Convergent linear RNA control primers; C – Divergent circRNA primers; ss – self-splicing.

3.2.7 Protein Analysis of FRET Peptide

Once confirmation of circRNA formation was obtained, it was important to establish protein translation from the circRNA. An *in vitro* coupled transcription/translation reaction was performed with FRET vector constructs/circFRET followed by protein analysis via western blot (Figure 3.19). The results indicate that that a peptide was been produced and detected for Linear FRET (58.08 kDa) and for *Anabaena* FRET where the RNA is unspliced (67.36kDa). The negative control is from the *in vitro* coupled transcription/translation reaction without any template nucleic acid added to the reaction. The signal is not very strong for the unspliced *Anabaena* FRET band so there are 2 images of the same gel at different exposures. A band could not be seen for circFRET.

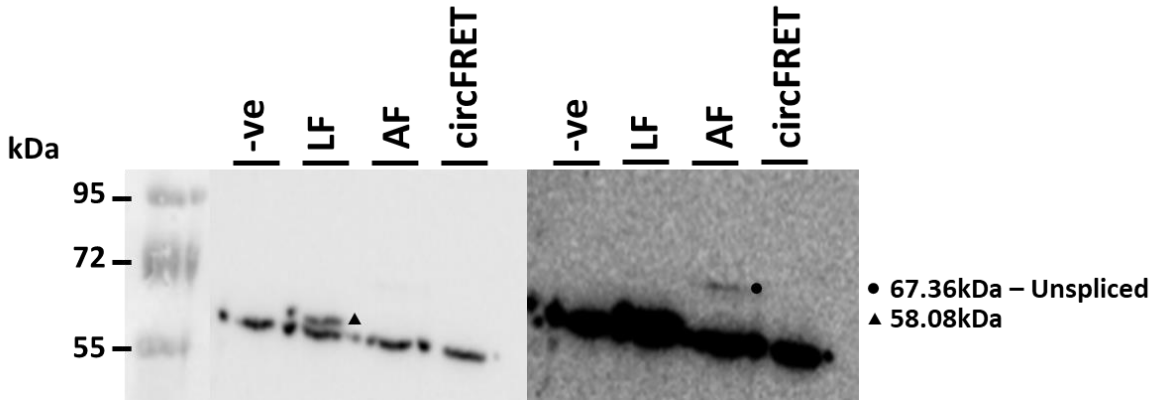


Figure 3.19 Western blot results of *in vitro* synthesised FRET peptides. Left low exposure, right high exposure. The -ve is the transcription/translation reaction lacking a nucleic acid template. LF – Linear FRET; AF – *Anabaena* FRET.

3.2.8 Detection of FRET Fluorescence

A Synergy HT spectrophotometer (2.1.23) was used to analyse the fluorescence of the *in vitro* produced FRET peptides diluted in PBS (10 - 40% sample volume in PBS) (2.1.23) (Figure 3.20). The FRET peptide would be detected by excitation of EGFP leading to mCherry emission. This provided no significant fluorescence for any of the samples at any concentration, compared with the control. This was also true for both EGFP (Figure 3.20b) and mCherry (3.20c) independent excitation, emission. There was a significant mCherry detection at 40% concentration for Linear FRET ($P = 0.0446$), however considering there was not a positive result for FRET detection, and other concentrations did not yield a positive result, the presence of functional mCherry remains unconfirmed. A functional FRET peptide would be detected, in addition independent mCherry excitation would cause mCherry fluorescence (3.20b). The control is from the *in vitro* coupled transcription/translation reaction without any template nucleic acid added to the reaction. No fluoresce was observed for the control as expected. Overall, the results indicate that a functional FRET peptide was not produced *in vitro*.

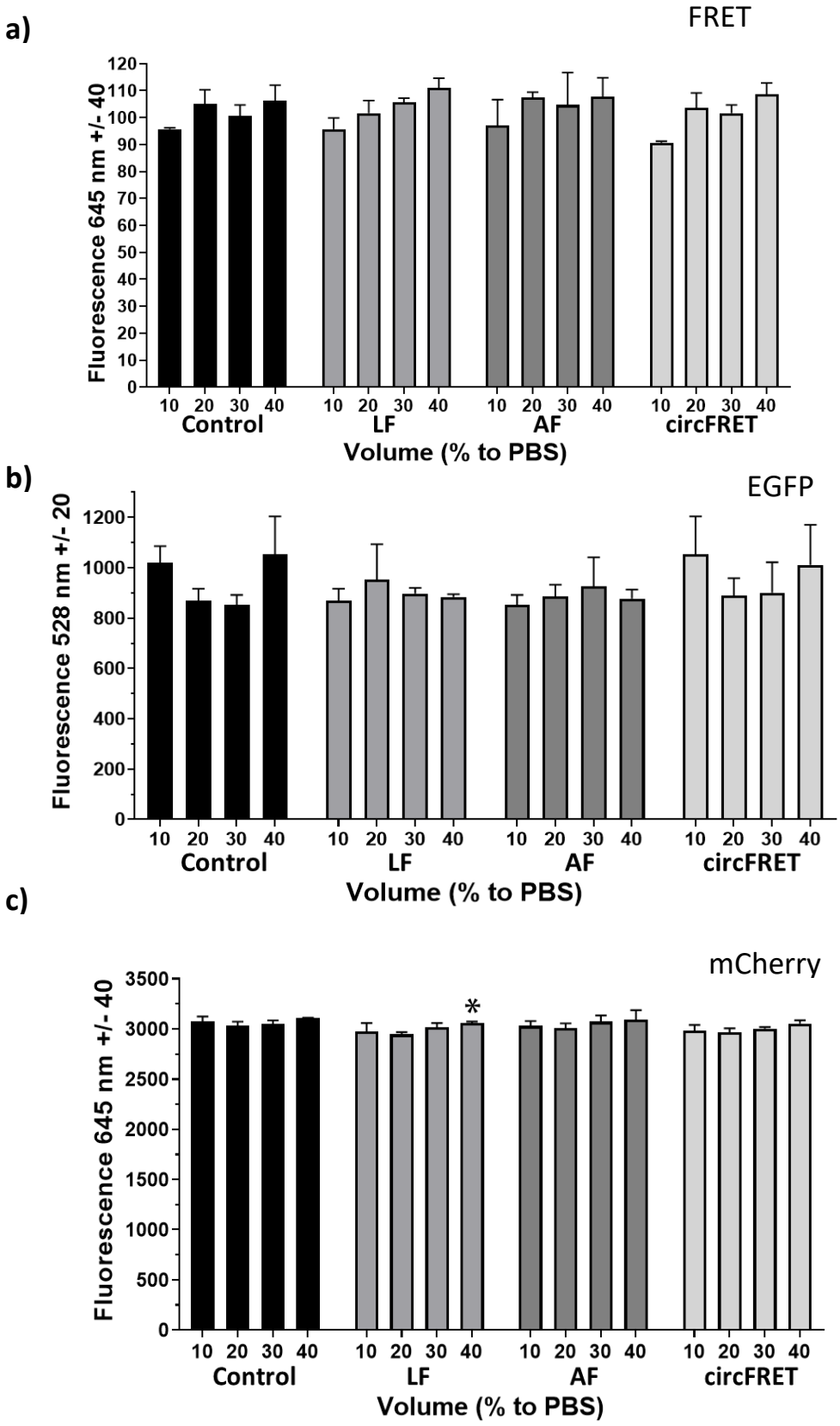


Figure 3.20 Fluorescence Analysis of *in vitro* produced FRET. a) FRET peptide detection – Excitation 485/20 nm, emission 645/40 nm b) EGFP detection – Excitation 485/20 nm, emission 528/20 nm c) mCherry detection – 590/20 nm, emission 645/40 nm. Control – Transcription/translation reaction lacking a nucleic acid template; LF – Linear FRET; AF – *Anabaena* FRET. The error bars represent the standard deviation. n = 3. Data is derived from a single experiment.

3.2.9 Designing Mastoparan Flag Constructs

A Flag tag was added to the end of the mastoparan construct to allow for the detection of the peptide. The Flag DNA sequence was cloned into pRSET B Mastoparan with a Xhol single digest (Figure 3.21). This created a construct with a stop codon at the end of the FLAG sequence (named *Anabaena* Mastoparan Flag). Another construct was made via site-directed mutagenesis (2.1.4) (SDM_Flag_Stop_Fwd/ SDM_Flag_Stop_Rev), removing the stop codon of *Anabaena* Mastoparan Flag to produce an infinite open reading frame that would generate a peptide concatemer when translated, this construct was named *Anabaena* Mastoparan Flag IORF (Figure 3.21c).

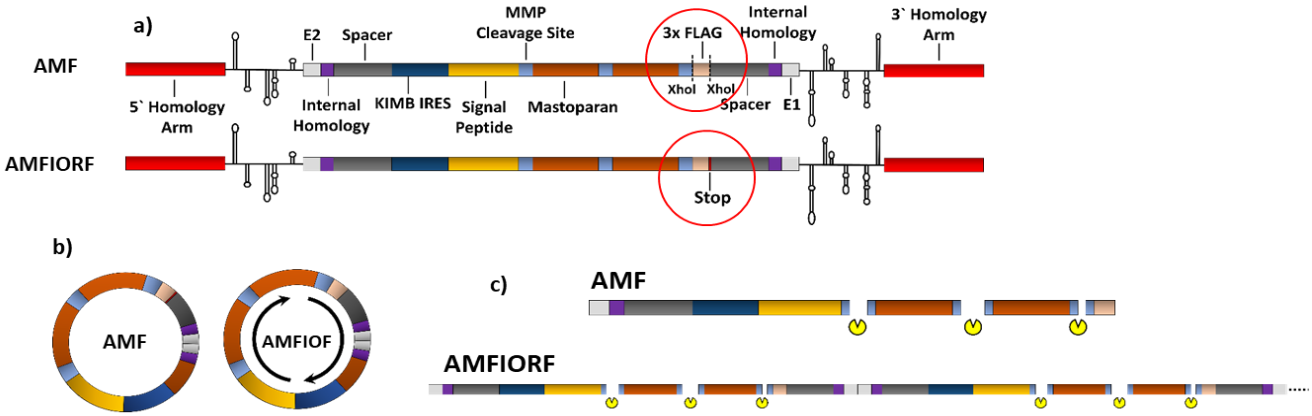


Figure 3.21 Cloning a Flag tag into *Anabaena* Mastoparan. a) Upper – Schematic of all the features of *Anabaena* Mastoparan Flag with the Xhol restriction site's location for FLAG cloning, middle – *Anabaena* Mastoparan Flag with a stop codon location, lower – *Anabaena* Mastoparan Flag IORF produced after site directed mutagenesis of *Anabaena* Mastoparan Flag stop codon. Red circles indicate location of the cloned Flag peptide. b) circRNAs produces after transcription of *Anabaena* Mastoparan Flag (left) or *Anabaena* Mastoparan Flag IORF (right) c) The peptides produce from *Anabaena* Mastoparan Flag circRNA (left) translation or *Anabaena* Mastoparan Flag IORF circRNA translation (right). AMF – *Anabaena* Mastoparan Flag; AMFIORF – *Anabaena* Mastoparan Flag IORF; E1 – Exon 1; E2 – Exon 2.

3.2.10 Peptide Analysis of *Anabaena* Mastoparan Flag and *Anabaena* Mastoparan Flag IORF

Peptides were expressed from the new Mastoparan Flag constructs with an *in vitro* coupled transcription/translation reaction and analysed on a western blot (Figure 3.22). A band was visible for unspliced *Anabaena* Mastoparan Flag (18kDa), but no visible band could be detected for *Anabaena* Mastoparan Flag IORF or circMast Flag (data not shown). In addition, the positive control Flag peptide (100 ng) was not visible.

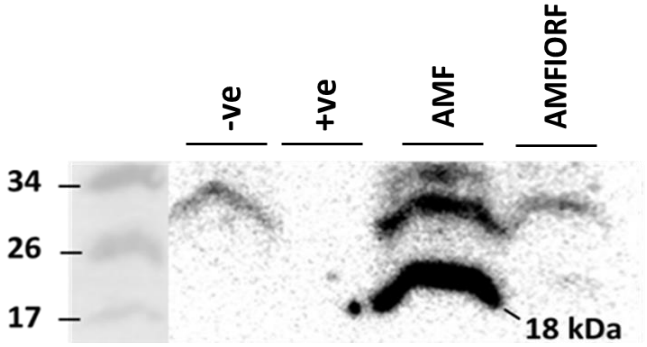


Figure 3.22 Western blot results of *Anabaena* Mastoparan Flag and *Anabaena* Mastoparan Flag IORF. Band present at 18kDa represents a peptide produced from unspliced *Anabaena* Mastoparan Flag RNA. The -ve is the transcription/translation reaction lacking a nucleic acid template. The +ve control was Flag peptide in solution (1:1,000 dilution). AMF – *Anabaena* Mastoparan Flag; AMFIORF – *Anabaena* Mastoparan Flag IORF.

3.3 Discussion

RNA circularisation models have been established *in vitro* using group I and II introns. (Puttaraju and Been, 1992; Mikheeva, 1997). Translatable circRNAs have been engineered to produce a biologically functional peptides (Wesselhoeft, Kowalski and Anderson, 2018). circRNAs with an infinite open reading frame have demonstrated the ability to produce a long repetitive peptide concatemers through many cycles of ribosomal translation around the circRNA (Perriman and Ares, 1998). This section aimed to design an RNA circularisation

model to produce translatable circRNA with an infinite open reading frame, capable of translating a mastoparan or FRET peptide concatemer *in vitro*.

The first part of this chapter aimed to design an IORF coding sequence containing a Shine-Dalgarno sequence and Flag tag (SDFLAGAHSIH) (3.2.1). The SDFLAGAHSIH was designed and successfully constructed but attempts of RNA circularisation using an enzymatic ligation approach were unsuccessful (3.2.2).

Next, the project focused on designing and constructing self-splicing introns in a reversed order for RNA circularisation. Initially, the focus was on constructing yeast inverse group II introns, however multiple construction strategies, including amplification from yeast DNA (3.2.1.3), overlapping oligo PCR (3.2.3.2) and Gibson assembly were unsuccessful (3.2.3.3). Attention switched to designing engineered *Anabaena* group I introns, using a model based on Wesselhoeft et al., 2018. This model was successfully designed and constructed along with linear controls (3.2.4 – 3.2.5).

After construction, the *Anabaena* mastoparan and *Anabaena* FRET constructs were circularised *in vitro.* RNA circularisation was confirmed by RNase R treatment and RT-PCR (3.2.5).

Detection of the mastoparan peptide was achieved by adding an epitope Flag tag to into the *Anabaena* Mastoparan construct (3.2.8). Two constructs were made, the first with a stop codon following the mastoparan sequence (*Anabaena* Mastoparan Flag) and another without a stop codon (*Anabaena* Mastoparan Flag IORF).

Finally, the constructs were subject to a coupled *in vitro* transcription/translation reaction. Results indicated some peptide translation, although the peptides produced were from unspliced, non-circularised RNA (3.2.9). The FRET peptide could not be detection using a spectrophotometer (3.2.7).

The *in vitro* work described in this section demonstrated that RNA circularisation occurred *in vitro*, but subsequent circRNA translation to produce a peptide concatemer was unsuccessful. However, both FRET and mastoparan peptides were detected from unspliced RNA on a western blot. The constructs described in this section can be used for analysis of circRNA formation, peptide production and ultimately, mastoparan anti-cancer activity *in vivo*.

3.3.1 SDFLAG Cloning and Circularisation

An engineered IORF circRNA has been shown to produce a long repetitive GFP protein in a bacterial expression system, using a Shine Dalgarno sequence for internal translation initiation (Perriman and Ares, 1998). In this section a double repeat of Shine Dalgarno Flag was constructed in pRSET B. Site-directed mutagenesis was required on a region between the T7 promoter sequence and the SDFLAG sequence to mutate a stop codon for an IORF after RNA circularisation.

The best 3′-hydroxyl acceptor for *in vitro* RNA circularisation is adenosine (Petkovic and Müller, 2015). The reverse primer (pRSETB_R_273) for template linearisation was designed so that the RNA transcript produced after *in vitro* transcription was a multiple of 3 (423 nt), and the terminal base was an adenosine nucleotide. This provided the best conditions for RNA circulation whilst maintaining an infinite open reading frame (IORF).

A 5’-monophosphate is required for RNA circularisation via enzymatic ligation. The transcription reaction was primed by including 10-fold excess guanosine-5’-monophosphate (GMP) to guanosine-5’-triphosphate (GTP) (Wesselhoeft et al., 2019). This method provides the 5’-monophosphate without the need for enzymatic modification (Schindewolf, Braun and Domdey, 1996). In general, generating circRNA from linear precursors is challenging due to the negative entropy of the circularisation reaction. Circularisation efficiency is significantly increased with a DNA splint, as this brings the ends of the hybridised mRNA ends closer together (Müller and Appel, 2016). The splint was 40 nt, with 20 bp of hybridisation to each end of the linear precursor RNA (Kurschat, Müller, Wombacher and Helm, 2005), and 3 nucleotides overhang at each distal RNA end for ligation with T4 RNA Ligase I (Müller and Appel, 2016). CircRNA was not obtained via splint hybridisation and ligation. Using the splint method and RNA ligase I, Wesselhoeft et al. 2019 achieved a circularisation efficiency of 40% with a sequence length of 1.5 kB, indicating the problem with the sequence was not size related. Although unsuccessful, enzymatic ligation was not a utilisable method for RNA circularisation inside cells, which is an aim of the project. Therefore, circularisation methods that could function both *in vitro* and *in vivo* were the future focus.

3.3.2 Inversed Group II Intron Construction

Although inversed group II introns are less commonly used for intron circularisation due to their 2’- 5’ phosphodiester bond, they were selected for the following reasons. Inversed group II introns produce a full circle RNA without any native intron inclusion. It is currently unknown whether an RNA polymerase II can traverse the 2’, 5’ phosphodiester linkage, therefore it would be an interesting discovery had inverse intron construction been successful. Group II introns would likely be the self-splicing intron of choice for RNA circularisation if high circularisation and debranching resistance is shown. Primer extension assays have displayed

continual elongation by reverse transcriptase through the 2', 5' phosphodiester linkage of Peach Latent Mosaic Viroid (PLMVd) RNA (Côté, Lévesque and Perreault, 2001) . This provides an indication that RNA polymerase II will translate through the junction.

Attempts to amplify/synthesise the yeast group II intron sequence were unsuccessful. A generalist view of the cause of the problems is that the intron sequence is highly structured and contains many repeat elements causing self-annealing, non-specific primer binding problems during amplification. Amplification of the intron sequence by standard PCR was not successful for domains I-III or IV-VI. Various conditions were altered to obtain a clean template amplification, including temperature gradient PCR, touch down PCR, magnesium titration, addition of betaine, and adjusting DMSO concentrations. DNA extraction from a live yeast culture with a glass bead approach provided the best DNA template for PCR amplification, as shown by the genomic and mitochondrial DNA controls (Figure 3.6). This extraction method has previously been shown to obtain high yields of mitochondrial DNA with high sensitivity in PCR-based applications (Yamada et al., 2002). A nested PCR approach amplified a large region (3kB) within the Cox1 gene of the yeast mitochondrial genome. The forward (Cox1 Fwd) and reverse (Cox1 Rev) primers were designed to bind to coding regions in the Cox1 gene. These regions were selected as the annealing sites would be less structured in coding regions allowing for a greater chance of successful primer annealing and amplification. The Cox1 region was successfully amplified (Figure 3.7) but subsequent amplifications from the purified Cox1 product were unsuccessful. The DNA polymerase would have read through the problematic intron sequence for the DNA to amplify the sizeable Cox1 region successfully. This demonstrates that the problem is primer binding dependent and not caused by the inability of polymerase readthrough.

A different approach was used to synthesise the intron sequence using 20 nucleotide overlapping oligonucleotide pairs. After the third round of amplification, the fragments were ran on a gel, extracted and sequenced. The sequencing reads showed the template/primers had non-specifically annealed with most of the expected sequence missing (Figure 3.10). This non-specific annealing is likely due to the repetitive nature of the intron sequence. A shorter primer design may prevent this by allowing targeting of less repetitive regions, but a relatively long 20bp overlap was chosen due to the difficulty of primer annealing in the standard PCR reaction.

Finally, a Gibson assembly approach was attempted to synthesise the intron sequence. This approach was based on a study where the entire mouse mitochondrial genome was synthesised (Gibson et al., 2010). The genome was constructed using non-discriminatory selection of the overlapping regions, demonstrating the flexibility of the approach. However, using this method either no colonies were observed, or the colonies were negative even after transforming ultra-competent cells.

3.3.3 *Anabaena* Mastoparan and FRET Design and Construction

The engineered *Anabaena* group I intron sequence was designed based on the sequence produced in Wesselhoeft et al., 2018, with a few changes to generate an infinite open reading frame (IORF). An additional T was added into the *Anabaena* sequence to cause a frameshift that removed two stop codons same frame as the coding sequence. The location of the additional nucleotide impacts the RNA structure the least, as predicted by RNAfold. The additional nucleotide maintained an IORF in circMast, and circFRET as both circRNA nt lengths were a multiple of three (*Anabaena* Mastoparan – 477 nt, *Anabaena* FRET– 1788 nt,

respectively). A minor substitution was made to the original sequence in the 5' internal homology region. A G was substituted with a C to remove a stop codon (TAG to TAC). The corresponding C in the original 3' internal homology region that bound to the G was replaced by a G to maintain complete complementarity in the internal homology regions.

The largest change from the sequence designed by Wesselhoeft et al., 2018, was the replacement of the CV3B IRES with a KMIB IRES. This change was necessary because the CV3B IRES is less compact than KMIB and contains multiple stop codons in all 3 frames, making it incompatible with designing circRNA with an IORF. The minimal KMIB IRES sequence is also advantageous during IORF translation as ribosomal stalling and disruption to translation is associated with rare codons that an IRES will likely contain.

The linear mastoparan control was designed to encode a double mastoparan peptide sequence separated by an MMP cleavage site. This double sequence would require cleavage for activation and therefore function similarly to mastoparan expressed as a peptide concatemer from circMast. All the constructs were designed to contain an N-terminus signal peptide. Lacking a signal peptide, the production of long repetitive proteins would likely be toxic to the cell, causing cell death. It is thought that secretion via the signal peptide would allow for continuous production of a large repetitive protein whilst maintaining a viable cell. In addition, secretion is essential for MMP peptide activation due to active MMPs being almost exclusively present in the extracellular environment (1.5.1).

The FRET constructs containing EGFP and mCherry were designed to view the process from peptide production to activation *in vivo.* With everything functioning correctly, the FRET peptide would be produced inside cells providing red fluorescence under UV light, through

mCherry quenching of eGFP. FRET peptide secretion out of the cell causes red fluorescence in the media until it is cleaved by MMPs, releasing EGFP from mCherry quenching, resulting in green fluorescence.

The original mastoparan sequence lacked an epitope tag for peptide detection so Flag tag was added via cloning. Two mastoparan Flag constructs were made, one with a stop codon and another maintaining the infinite open reading frame. The construct with a stop codon was included in the case that the peptide concatemer produced from circRNA caused cell death.

3.3.4 RNA Circularisation and Peptide Analysis

Both *Anabaena* FRET and *Anabaena* Mastoparan were successfully circularised *in vitro* as confirmed by RNase R treatment and RT-PCR. The circularisation efficiency, however, was less than reported by Wesselhoeft et al., 2018. This may be due to the slight changes in *the Anabaena* sequence necessary to maintain the infinite open reading frame, the most disruptive likely being the additional T nucleotide.

Precursor-RNA self-splicing occurred even before treatment under self-splicing conditions. This indicates that the *in vitro* transcription reaction conditions were suitable for some circRNA formation. This is unsurprising since the only cofactors required for group I intron self-splicing are Mg^{2+} and guanosine nucleotides, which are already present in the transcription reaction.

Peptides were detected from *Anabaena* FRET and *Anabaena* Mastoparan Flag. However, they appeared to be from unspliced RNA. A band could not be seen from the circFRET, *Anabaena*

Mastoparan Flag IORF or the positive controls for either anti-GFP or anti-Flag blots. The positive control 3x Flag Peptide (Sigma) has a very low molecular weight (2.86 kDa), so it may be difficult to detect on a western even when using a high 15 % SDS PAGE gel or a tricine gel. The GFP control protein at 28 kDa is much larger and should have no issue being detected. The loading concentration for both control proteins was 100 ng, which should be enough for detection. Western blots were repeated numerous times, but the control proteins were still undetected. Peptides were detected at sizes corresponding to unspliced RNA for both *Anabaena* FRET and *Anabaena* Mastoparan Flag. *Anabaena* Mastoparan Flag showed very high expression levels, whereas *Anabaena* FRET was very low. No peptide was detected for circFRET, a sample which used circRNA as a template, compared to the other samples which used DNA plasmids. This indicates that the peptide expression from the circRNA was an issue. KMIB IRES has demonstrated activity in a rabbit reticulocyte system. The TNT Quick coupled Transcription/Translation (Promega) system used in this reaction uses rabbit reticulocyte solution (Spriggs, Mitchell and Willis, 2005), so circRNA peptide detection would be expected.

FRET fluorescence was not detected, which is in contrast to the result of the western. As there are no MMPs in the rabbit reticulocyte system, no FRET cleavage was expected, resulting in mCherry fluorescence (emission detection wavelength 645/40 nm) when EGFP is excited (485/20 nm). When exciting the EGFP and mCherry fluorophores individually and measuring emission, neither showed fluorescence (apart from mCherry at one concentration). This indicates an issue with FRET peptide production from the in vitro synthesis kit. The TnT® Quick Coupled Transcription/Translation kit includes a luciferase control plasmid. To check the kit's functionality, a luciferase assay could have been performed as a troubleshooting step. A positive control GFP protein was used to test the spectrophotometer, giving strong signals.

A control mCherry protein would have been beneficial to ensure all excitation/ emission filters required for FRET detection were functional.

Chapter Four

4.0 RNA Circularisation in Mammalian Cells and Mastoparan Cytotoxicity

4.1 Introduction

Strategies to circularise designer RNA sequences in mammalian cells aim to maximise circularisation efficiency whilst not impacting the biological function of the circRNA. Wesselhoeft et al. used an engineered PIE *Anabaena* group I intron sequence to circularise RNA with near 100% efficiency. The PIE sequence contained homology arms flanking the splice sites, creating a splicing bubble that increased the chance of interaction. Coding regions of proteins ranging from 1289 nt to 4934 nt in size were circularised with no loss of efficiency using this method. The length of most circRNAs in human cells is < 1000 nt, with a median value of ~ 500 nt, therefore, this circularisation method greatly increases the length of engineered circRNAs beyond those that are likely formed in human cells (Ding et al., 2018).

Cancer is a leading cause of death worldwide, accounting for nearly 10 million deaths in 2020 (World Health Organisation, 2022). Cancer occurs when abnormal cells divide in an uncontrolled way. Chemotherapeutic strategies target the cell cycle and the cell's ability to proliferate, using the rationale that cancer cells are more likely to be replicating than normal cells (Sun, Liu, Ma and Hu, 2021). This poor specificity causes significant toxic effects as normal cells are damaged along with the cancer cells. There is an ongoing pursuit in cancer research to produce new targeted cancer treatments that are effective, inexpensive and limit toxicity from off-target effects.

Mastoparan is a 14 aa peptide toxin in the defence system of *Vespula lewisii*, Korean yellowjacket wasp. Mastoparan can also act as an anti-cancer peptide, demonstrating toxicity and specificity toward cancer cells in cell culture conditions and mice (Hilchie et al., 2016). The natural mastoparan peptide is capped with a C-terminal amide (NH_2) which confers 8-11

times more potency than the C-terminal carboxyl group end (COOH). In addition to potency, the different C-termini demonstrate variation in their mechanism of cancer cell killing. mastoparan-NH_2 kills cells by a lytic mechanism, and mastoparan-COOH induces apoptosis via the intrinsic mitochondrial pathway, causing mitochondrial membrane disruption and increased ROS generation (de Azevedo et al., 2015). Mastoparan net positive charge provides selectivity to cancer cells via electrostatic attraction (Hilchie et al., 2016). Cancer cells have a net negative charge due to elevated levels of glycolysis, increased glucose uptake and lactate secretion compared to normal cells (Chen et al., 2016). Mastoparan has shown synergistic effects with chemotherapy drug Gemcitabine in breast cancer *in vivo* and can be conjugated to a drug Fluvastatin for enhanced cytotoxicity of lung cancer cells *in vitro* (Alhakamy, Ahmed, Md and Fahmy, 2021).

Two different breast cancer cell lines used in this project are MDA-MB-231 cells and MCF-7 cells. Both are ductal/breast carcinoma cells but have distinct phenotypic differences. MDA-MB-231 cells are highly invasive and aggressive, whereas MCF-7 cells are poorly invasive and less aggressive (Gest et al., 2013). MMPs are endopeptidases that degrade extracellular matrix proteins and process bioactive molecules (Verma and Hansch, 2007). In cancer, MMPs are associated with disease progression, angiogenesis, invasion, metastasis, and avoidance of immune surveillance (Foda and Zucker, 2001). Highly invasive, aggressive cancers like MDA-MB-231 have greater MMP expression and activity levels than MCF-7 (Figueira et al., 2009).

Protease activity can be monitored using a FRET peptide system incorporating two fluorophores with a spectral overlap in the emission wavelength of one fluorophore (donor) and the excitation wavelength of the other (acceptor) (Bajar et al., 2016). When FRET peptides are < 10nm in proximity, donor quenching occurs by the acceptor fluorophore (Schaap,

Hancock, Wilderspin and Wells, 2013). Incorporating an MMP cleavage site between the fluorophores can determine protease activity to a high sensitivity as cleavage results in measurable donor fluorescence (Lu et al., 2013).

The MTT assay is a colorimetric assay for measuring cell metabolic activity (Stockert, Horobin, Colombo and Blázquez-Castro, 2018). The assay functions by nicotinamide adenine dinucleotide phosphate (NADPH)-dependent cellular oxidoreductase enzymes reducing tetrazolium dye MTT to purple coloured insoluble formazan (Stockert, Horobin, Colombo and Blázquez-Castro, 2018). Dehydrogenases located mostly in the mitochondria of living cells convert the tetrazolium compound to water insoluble formazan crystals. Formazan can be measured by a spectrophotometer between 500 – 600 nm. The amount of reduction is proportionate to the number of viable cells, therefore MTT assays are commonly used to analyse the effectiveness of drug treatments on cells (Adan, Kiraz and Baran, 2016).

MTT assays is the gold standard for cytotoxicity testing, due to its low cost, easy execution, and rapid results. However, it should be considered that the conversion to formazan crystals depends on metabolic rate of the mitochondria which has external interferences, including D-glucose, NADH or NADPH concentration in the culture medium (Lotze and Thomson, 2005). Also, cells in early apoptosis may still have a functionally intact mitochondrion and can still reduce the tetrazolium compound.

This section focuses on work *in vivo* using MDA-MB-231 and MCF-7 cells, including RNA circularisation, FRET and mastoparan peptide analysis after transfection and the effect of non-amidated mastoparan on MDA-MB-231 and MCF-7 cell viability.

4.2 Results

4.2.1 CircRNA Confirmation

The mastoparan/Flag plasmid constructs were transfected into MDA-231 cells, and after 24 hours, the total RNA and protein were extracted via the trizol extraction method (2.2.3). RT-PCR of total RNA was performed for mastoparan and FRET constructs (2.1.19). Expected results were similar to *in vitro* (3.2.5), with the convergent primer set (RT Lin Fwd/ RT Lin Rev) amplifying both linear and circular RNA and the divergent primer set (RT circ FWD/ RT circ Rev) only amplifying circRNA. As expected, no amplification was detected for Linear Mastoparan and Linear FRET constructs, as these do not contain *Anabaena* introns for template binding and amplification (Figure 4.1). A band was detected for *Anabaena* Mastoparan (450 nt) and *Anabaena* FRET (1774 nt) for the convergent primers, indicating successful RNA production in cells. However, amplification of the divergent primer set was not successful. Bands were expected to run slightly higher than in the convergent primer set produced *in vitro* (3.2.5). The lack of amplification for the circRNA primer set indicates that the *Anabaena* group I introns were not splicing *in vitro.* There is a contamination band in all lanes around 200 nt, but this does not impact the test results.

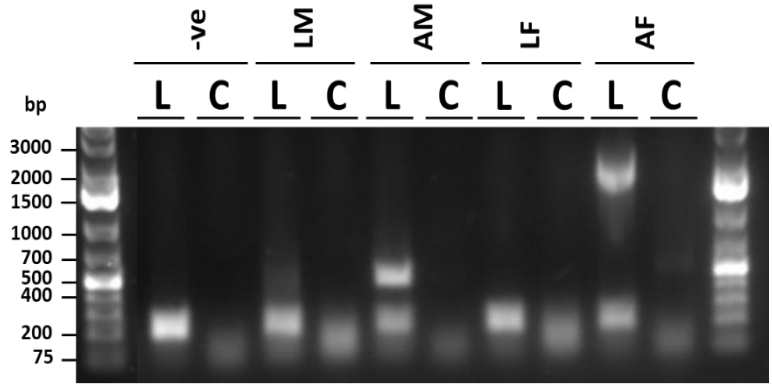


Figure 4.1 RT-PCR of FRET and Mastoparan RNA purified from cells. 1% Agarose Gel. RNA extracted from MDA-MB-231 cells. L – Convergent linear RNA control primers; C –Divergent circRNA primers; ve – PCR reaction lacking a nucleic acid template.

4.2.2 Peptide Analysis

A western blot was performed on FRET peptides contained in the culture media of transfected cells (2.1.22). This was to detect any FRET peptides secreted from cells via its signal peptide. The media proteins were first concentrated by spinning down to 200 μL from 12 mL (for 100 mm plate transfections) or 5 mL (for 60 mm plate transfections) in Amicon® Ultra-15 Centrifugal Filter Units (2.2.6). The western blot result was completely negative with no peptide detection in the media of Linear FRET , *Anabaena* FRET or circFRET transfected cells or for the positive control peptide (data not shown).

Western blot analysis was performed for the cell lysis of transfected MDA-MB-231 cells. Although the expression levels were low, bands were visible for *Anabaena* Mastoparan Flag for translated peptides at approximately 10 kDa and 18 kDa MW (Figure 4.2). The band present around 18 kDa is a peptide produced from unspliced *Anabaena* Mastoparan Flag RNA (18.03 kDa). The band around 10 kDa could represent a peptide translated from *Anabaena* Mastoparan Flag circRNA (10.63 kDa), but as the RT-PCR results indicated no *Anabaena* self-splicing in cells, this band is more likely to be present for IRES translation of the linear transcript. A peptide was detected from *Anabaena* Mastoparan Flag IORF unspliced RNA at 20.84 kDa. There is a possible peptide band at 31.89 kDa, representing translation after 3 cycles around circ*Anabaena* Mastoparan Flag IORF. Similarly to the *Anabaena* Mastoparan Flag band present at around 10 kDa, there is doubt whether this band is from circ*Anabaena* Mastoparan Flag IORF translation due to negative RT-PCR results. In addition, no other bands are present for peptides produced for other cycles of circRNA translation (1 cycle – 10.63 kDa, 2 cycles – 21.26 kDa, 4 cycles – 42.52 kDa). A FLAG peptide positive control could not be detected.

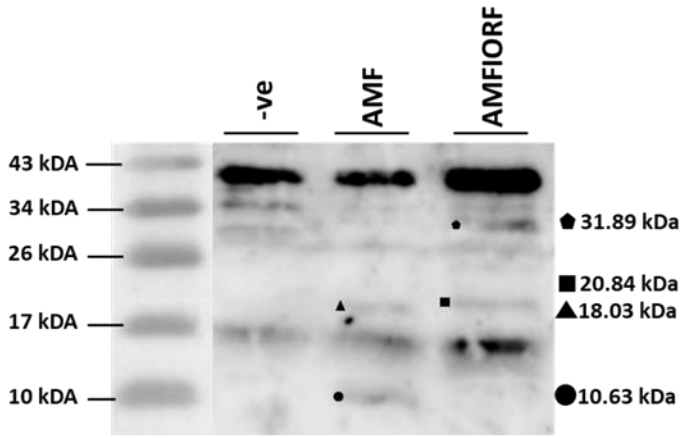


Figure 4.2 Western blot results for *Anabaena* Mastoparan Flag and *Anabaena* Mastoparan Flag IORF peptides. Protein purified from MDA-MB-231 cells transfected with *Anabaena* constructs. ● – *Anabaena* Mastoparan Flag peptide from spliced circRNA; ▲ – *Anabaena* Mastoparan Flag peptide from unspliced mRNA; ■ – *Anabaena* Mastoparan Flag IORF peptide from unspliced mRNA; ⬟ – *Anabaena* Mastoparan Flag IORF peptide from 3x cycles around circRNA; -ve control – Protein purified from non-transfected cells.

A western blot was performed using concentrated culture media from cells transfected with *Anabaena* Mastoparan Flag, and *Anabaena* Mastoparan Flag IORF constructs to detect mastoparan Flag peptides secreted out of the cell via the signal peptide. This result was negative with no Flag peptides detected or positive control (data not shown).

4.2.3 Detection of FRET Fluorescence

FRET (Excitation 485/20 nm, emission 645/40 nm), EGFP (Excitation 485/20 nm, emission 528/20 nm) and mCherry (590/20 nm, emission 645/40 nm) fluorescence was analysed in multiple sample types (2.1.23).

The first sample type tested was the medium of Linear FRET and *Anabaena* FRETMDA-MB-231 transfected cells (Figure 4.3). The expected results for a functional FRET peptide exported into the medium and cleaved by MMPs would be 1) Some FRET fluorescence from uncleaved

FRET in the medium (the greater MMP activity, the less FRET fluorescence) 2) EGFP fluorescence, caused by cleaved FRET in the media (the greater MMP activity, the more EGFP fluorescence) 3) mCherry fluorescence, as mCherry emission wavelength does not overlap with EGFP excitation wavelength and so is unaffected by its proximity. The results obtained did not match the expected outcome. There was no significant FRET or GFP detection in the Linear FRET or *Anabaena* FRET samples. The only detected fluorescence was mCherry at 20 % ($P = 0.0157$), 30 % ($P < 0.0001$) and 40 % ($P < 0.0001$) Linear FRET medium in PBS. mCherry fluorescence in the media indicates that the FRET peptide was being exported into the extracellular environment for the Linear FRET construct. However, as described in the expected results, FRET fluorescence and/or (depending on MMP activity) EGFP fluorescence would be expected, which is not seen in the Linear FRET sample.

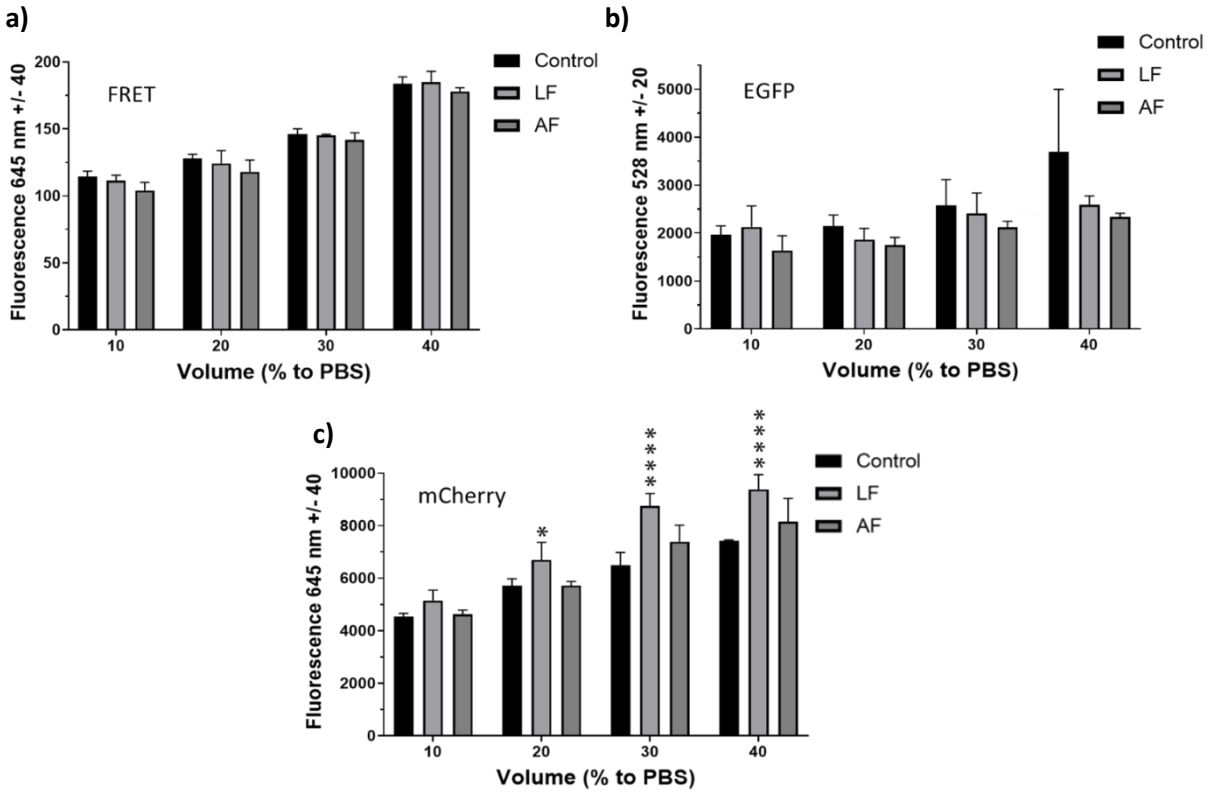


Figure 4.3 Fluorescence detected in the medium of Linear FRET and *Anabaena* FRET transfected MDA-MB-231 cells. a) FRET fluorescence – Excitation 485/20 nm, emission 645/40 nm b) EGFP fluorescence – Excitation 485/20 nm, emission 528/20 nm c) mCherry fluorescence – 590/20 nm, emission 645/40 nm Medium extracted 24 hours post transfection (or without transfected for control). The error bars represent the standard deviation. n = 3. Data is derived from a single experiment. The control is medium from non-transfected cells. Two-way ANOVA with Dunnett's test between samples and GFP. * $P < 0.05$, **** $P < 0.0001$. LF – Linear FRET; AF – *Anabaena* FRET.

The total proteins extracted from FRET transfected MDA-MB-231 cells (2.2.3) were diluted in PBS to concentrations ranging from 0.025 – 0.1 µg/µL, and fluorescence was observed (Figure 4.4). Expected results for Linear FRET and *Anabaena* FRET samples were FRET and mCherry detectable fluorescence. GFP fluorescence would not be expected as the FRET protein should remain uncleaved until exported out of the cell. Neither FRET, EGFP and mCherry fluorescence were detected significantly in Linear FRET or *Anabaena* FRET samples. Only the positive control GFP peptide, diluted to the same concentrations as the samples (0.025 – 0.1 µg/µL), demonstrated GFP and FRET fluorescence. Slight background FRET fluorescence is observed as measuring mCherry emission at 645/40 nm can detect background GFP at a low intensity. The most unexpected result was from proteins extracted from cells transfected with a GFP control plasmid (pmaxGFP, Lonza), as no significant GFP fluorescence was detected. Transfection of the GFP plasmid was successful as fluorescence was observed under a Fluorescence microscope (data not shown). This result may indicate that the proteins had degraded or protein extraction from cells was unsuccessful.

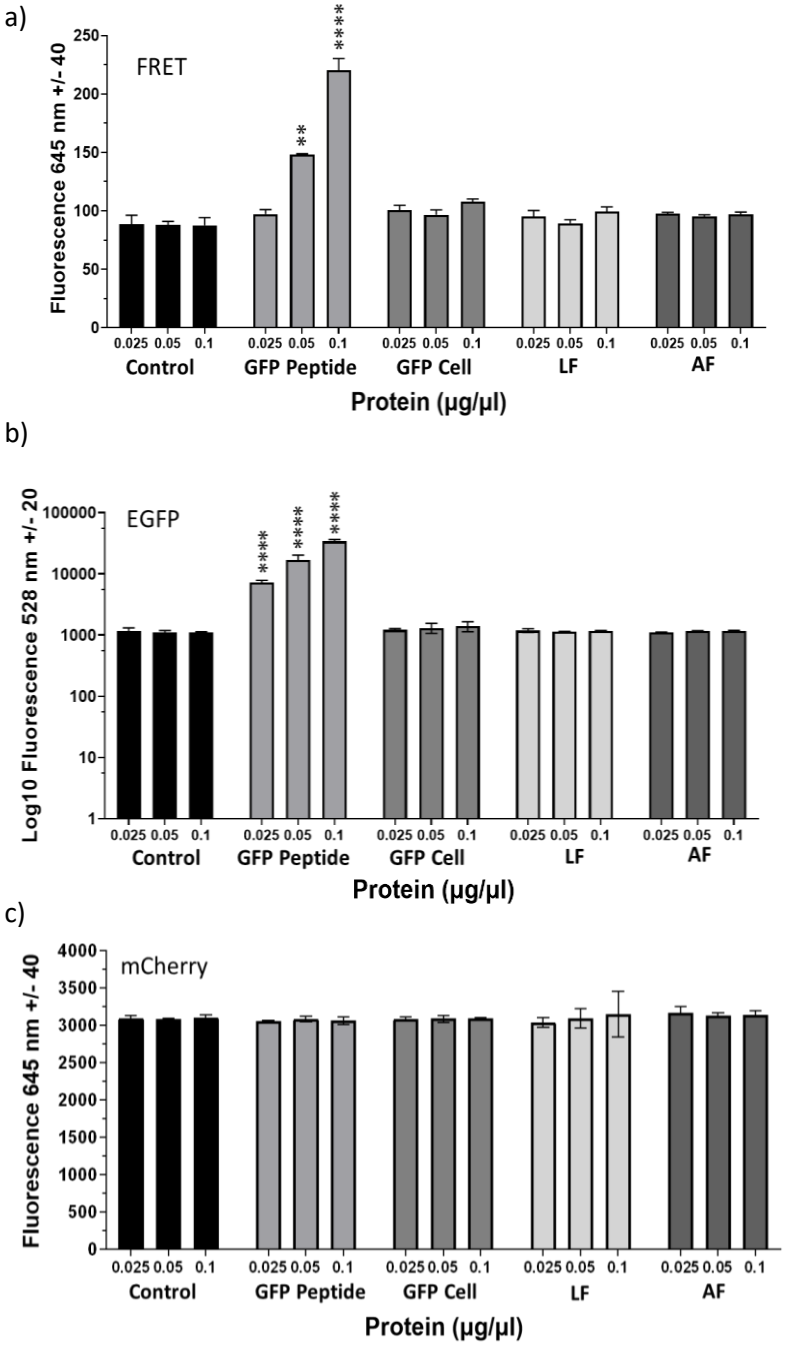


Figure 4.4 Fluorescence detected from proteins extracted from Linear FRET and *Anabaena* FRET. a) FRET fluorescence – Excitation 485/20 nm, emission 645/40 nm b) EGFP fluorescence – Excitation 485/20 nm, emission 528/20 nm c) mCherry fluorescence – 590/20 nm, emission 645/40 nm Proteins extracted 24 hours post transfection. GFP peptide is a purified peptide in solution, and GFP cell is GFP extracted from GFP plasmid transfected cells. The error bars represent the standard deviation. n = 3. Data is derived from a single experiment. The control is from proteins extracted from non-transfected cells. Two-way ANOVA with Dunnett's test between samples and GFP. * $P < 0.05$, **** $P < 0.0001$. Protein extracted from Linear FRET and AF transfected MDA-MB-231 cells. LF – Linear FRET; AF – *Anabaena* FRET.

4.2.4 Mastoparan Cytotoxicity

To demonstrate mastoparan cytotoxicity to MDA-MB-231 and form a basis of comparison for mastoparan produced in a vector expression system, MDA-MB-231 cells were incubated with amidated mastoparan (0-40 µM) for 24 hours, and an MTT Assay performed. As seen in figure (Figure 4.5), greater concentrations of mastoparan resulted in greater cell death until around 25 µM, in which the difference from the blank was < 0.1. The IC_{50} was 12.06 µM µM (95% CI, 8.14 – 17.45), lower to concentrations reported in previous literature for MDA-MB-231 cells (20-24 µM) (Hilchie et al., 2016).

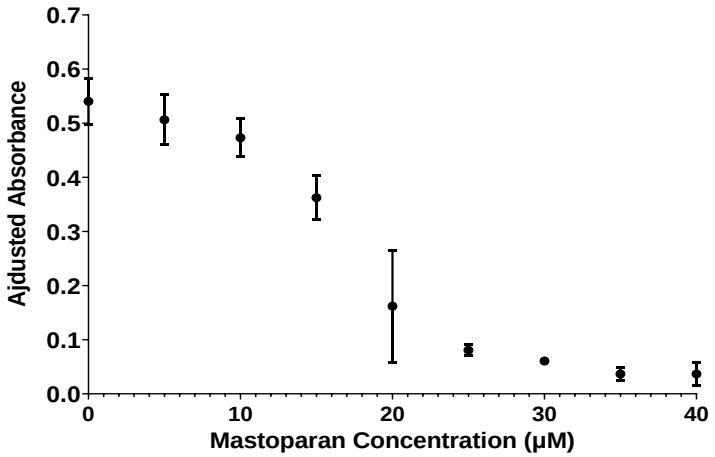


Figure 4.5 Effects of amidated mastoparan (0-40 µM) on MBA-MB-231 cells. MTT assays results after incubating mastoparan of various concentrations from 0-40 µM with MDA-MB-231 cells for 24 hours. Amidated mastoparan power was dissolved in PBS and added to cells containing DMEM media to the final concentrations from 0-40 µM. These results are the adjusted absorbances with a DMSO blank subtracted from each data point. Data is derived from a single experiment. The error bars represent the standard deviation. n = 3. The IC_{50} 12.06 was calculated using a nonlinear regression model.

Mastoparan peptides produced *in vitro* in a coupled transcription/translation reaction from *Anabaena* Mastoparan and *Anabaena* Mastoparan Flag constructs were added to the media of MDA-MB-231 or MCF7 cells at varying volumes (0 - 2 µL) in 96 well plates in a total volume

of 100 µL. The MTT assay performed after 24 hours demonstrated that mastoparan produced *in vitro* did not affect cell death after incubation for 8 hours or 24 hours (Figure 4.6). This was unexpected considering that *in vitro* mastoparan had a strong signal on the western blot (3.2.9).

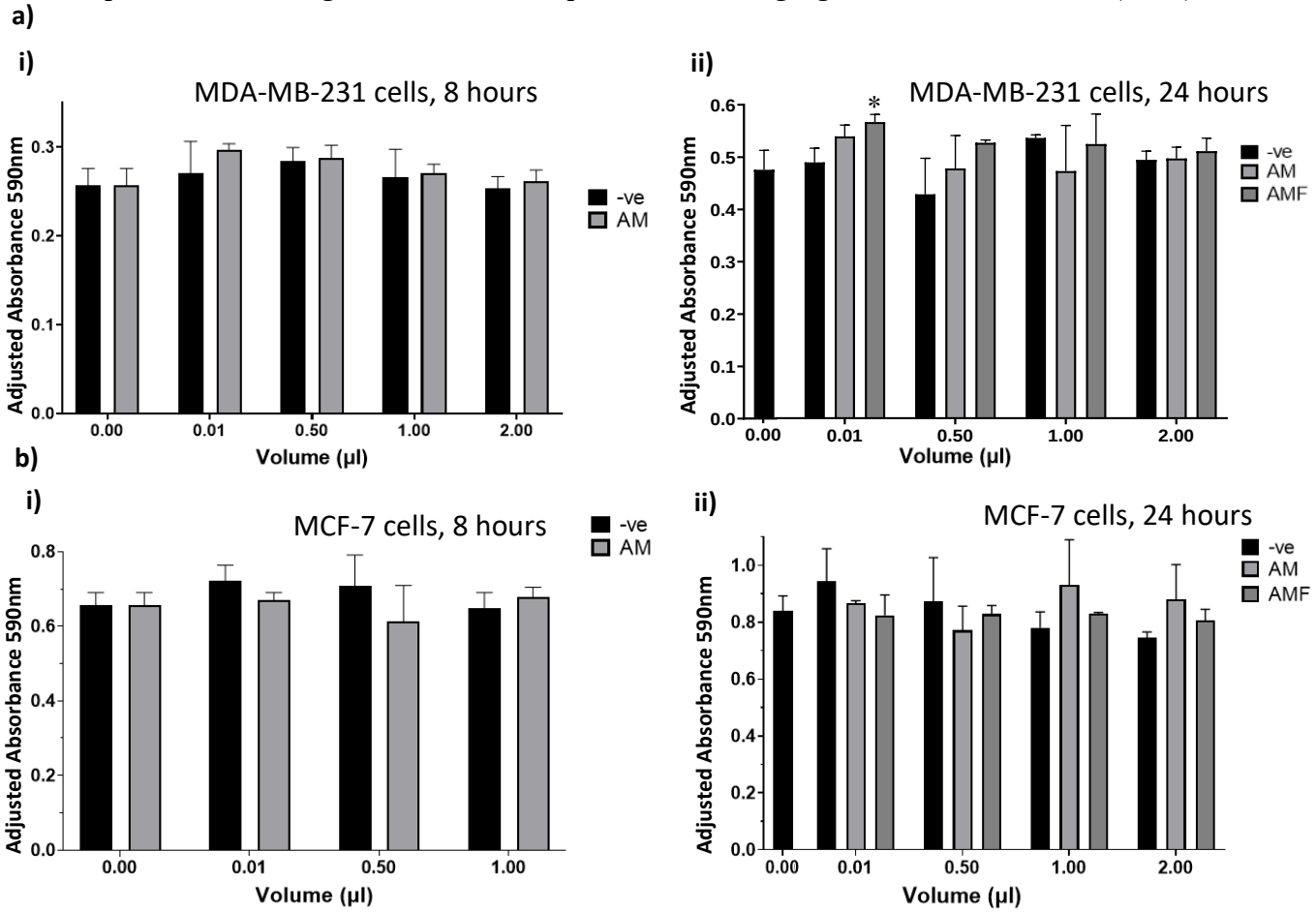


Figure 4.6 Treating MDA-MB-231 and MCF-7 cells with mastoparan produced *in vitro*. a) MDA-MB-231 treated cells b) MCF-7 treated cells i) MTT assay 8 hours post treatment ii) MTT assay 24 hours post treatment. The -ve control is the translation reaction, without adding template nucleic acid. These results are the adjusted absorbances with a DMSO blank subtracted from each data point. Data is derived from a single experiment. The error bars represent the standard deviation. n = 3. -ve – An *in vitro* coupled transcription/translation reaction with no plasmid template. i) Two-way ANOVA with Šídák's multiple comparisons test ii) Two-way ANOVA with Dunnett's multiple comparisons test between samples and -ve. * $P < 0.05$. AM – *Anabaena* Mastoparan; AMF – *Anabaena* Mastoparan Flag.

The effect mastoparan expressed inside cells had on cell death was tested by transfecting mastoparan containing constructs into MDA-MB-231 cells or MCF-7 cells (Figure 4.7). Linear Mastoparan transfection caused significant cell death compared to the negative control

(pmaxGFP, Lonza) in MDA-231-cells (P = 0.0024), but the *Anabaena* Mastoparan construct did not. Neither Linear Mastoparan nor *Anabaena* Mastoparan caused cell death in MCF-7 cells (Figure 4.7b).

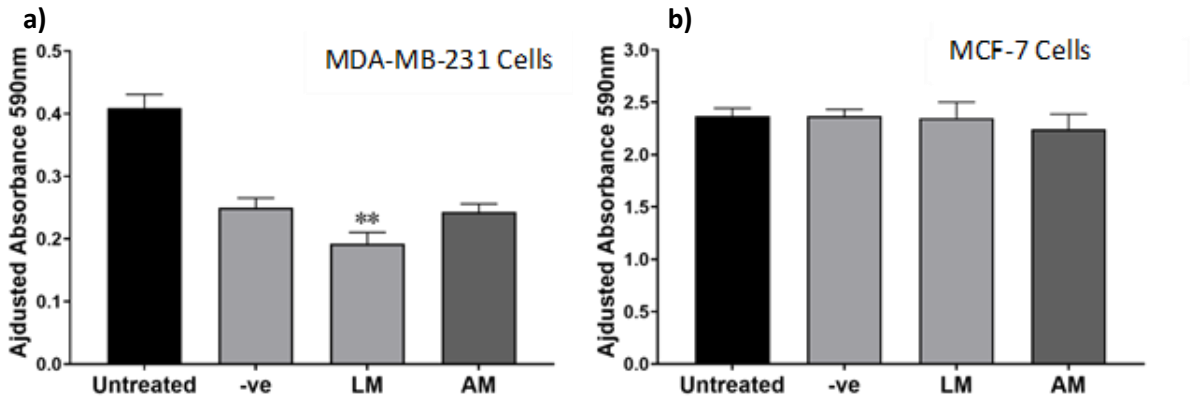


Figure 4.7 MTT assay 24 hours after transfection of mastoparan containing constructs. a) MDA-MB-231 transfected cells b) MCF-7 transfected cells. These results are the adjusted absorbances with a DMSO blank subtracted from each data point. The -ve control is cells transfected with a GFP containing plasmid. Data is derived from a single experiment. The error bars represent the standard deviation. n = 3. Ordinary one-way ANOVA with Dunnett's post hoc test between samples and -ve. ** P < 0.01. LM – Linear Mastoparan; AM – *Anabaena* Mastoparan.

Proteins were purified from the cells using trizol extraction (2.2.3). These extracted proteins were then added to the growth media of MDA-MB-231 cells and incubated for 24 hours before an MTT assay was performed (2.2.5). The *Anabaena* Mastoparan sample caused significant cell death compared to the GFP control at every concentration (2.5 – 5 ng/ μL). Linear Mastoparan caused significant cell death at 5 ng/ μL concentration (Figure 4.8)

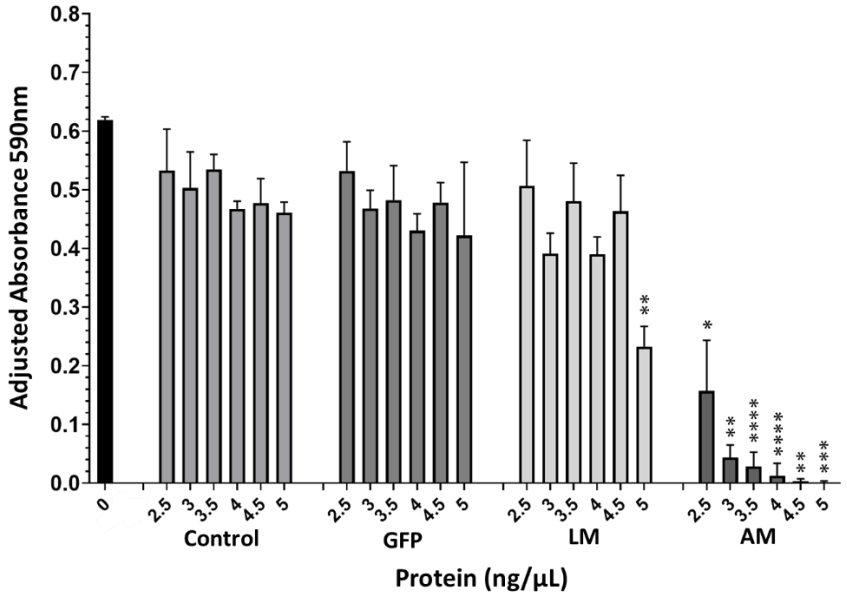


Figure 4.8 MTT assay 24 hours after treatment mastoparan extracted peptides (0 – 5 ng/µL). Proteins extracted 24 hours post transfection (or without transfection for -ve) with GFP control or mastoparan containing plasmids. Cells treated for 24 hours with varying protein concentrations (2.5 ng/ µL - 5 ng/ µL) before an MTT assay. **A)** MDA-MB-231 cells **B)** MCF-7 cells. The error bars represent the standard deviation. n = 3. These results are the adjusted absorbance with a DMSO blank subtracted from each data point. Data is derived from a single experiment. Two-way ANOVA with Dunnett's post hoc test between samples and GFP. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$. LM – Linear Mastoparan; AM – *Anabaena* Mastoparan.

Another experiment was performed, using the same purified proteins as before, but varying the concentration ranges and incubating the proteins with another cell line, MCF-7 cells (Figure 4.9). Both cell lines had very similar results. At higher concentrations, Linear Mastoparan did induce cell death. The concentration of *Anabaena* Mastoparan proteins had to reduce to 1 ng/ µL for the cell death to be similar to the control. The GFP control also caused cell death at higher concentrations which were unexpected.

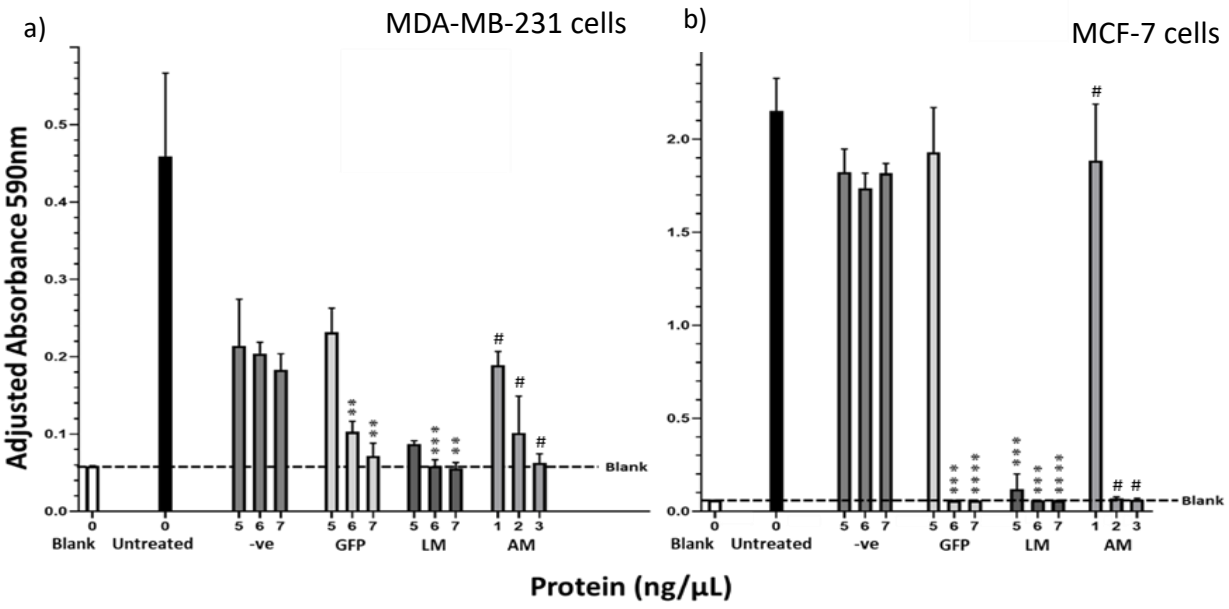


Figure 4.9 MTT assay 24 hours after treatment mastoparan extracted peptides (0 – 7 ng/µL). Proteins extracted 24 hours post transfection (or without transfection for -ve) with GFP or mastoparan containing plasmids. Cells treated for 24 hours with varying protein concentrations (1 ng/µL - 7 ng/µL) before MTT assay. a) MDA-MB-231 cells b) MCF-7 cells. The error bars represent the standard deviation n = 3. These results are the adjusted absorbance with a DMSO blank subtracted from each data point. Absorbance measured at 590 nm. Data is derived from a single experiment. Two-way ANOVA with Dunnett's multiple comparison test between samples and -ve control. ** P < 0.01, *** P < 0.001, **** P < 0.0001. # No statistical test possible for *Anabaena* Mastoparan, as no equal negative control concentration available. Proteins were extracted from MDA-MB-231 cells. LM – Linear Mastoparan; AM – *Anabaena* Mastoparan.

To provide further insight as to whether active mastoparan was present in the cell culture media, the medium from Linear Mastoparan and *Anabaena* Mastoparan transfected cells was concentrated at 200 µL (2.2.6) and then added to MDA-MB-231 cells at varying volumes (0.63 µL – 10 µL) and incubated for 24 hours (Figure 4.10). The results show that *Anabaena* Mastoparan medium at volumes 0.63 µL and 2.5 µL caused significant cell death. Linear Mastoparan did not cause cell death at any concentration. This was unexpected, considering that *Anabaena* Mastoparan transfection had no significant cytotoxicity, so it's peculiar why the media of those cells would cause cell death. As Linear Mastoparan transfection causes significant

cell death, it would be expected that the medium of Linear Mastoparan transfected cells also causes cell death, given that mastoparan requires exportation into the media for activation.

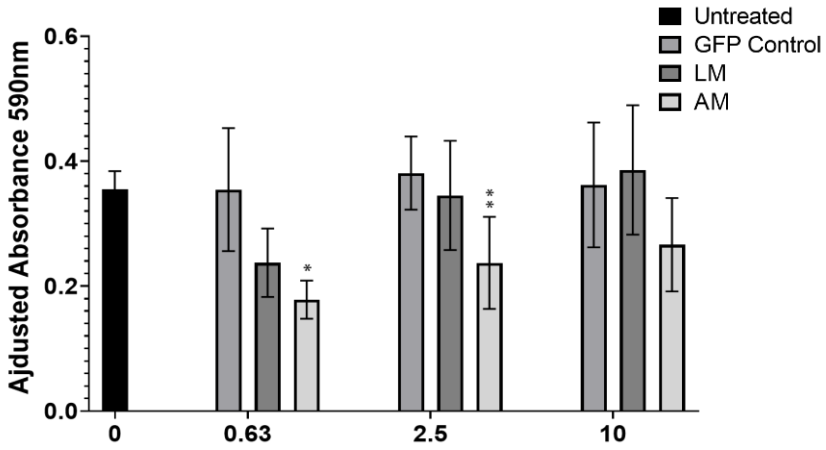


Figure 4.10 Treating MDA-MB-231 cells with mastoparan transfected cells culture medium. Medium extracted 24 hours post transfection. Cells treated for 24 hours with concentrated culture medium (0.63 μL - 10 μL) to a final volume of 100 μL, before an MTT assay is performed. GFP control is cells treated with culture medium of GFP transfected cells. The error bars represent the standard deviation n = 3. These results are the adjusted absorbance with a DMSO blank subtracted from each data point. Data is derived from a single experiment. Two-way ANOVA with Dunnett's post hoc test between samples and GFP Control. * $P < 0.05$, ** $P < 0.01$. LM – Linear Mastoparan; AM – *Anabaena* Mastoparan.

4.3 Discussion

Engineered circRNAs can produce peptides inside mammalian cells. CircRNAs with an infinite open reading frame produces long repetitive peptide concatemers by translating for many cycles around the circRNA. This project attempts to produce peptide concatemer from circRNA for therapeutic applications. This section describes the *in vivo* work in MDA-MB-231 and MCF-7 breast cancer cell lines, including RNA circularisation, peptide analysis and genetically encoded mastoparan effect on cancer cells.

The first part of this section uses RT-PCR to detect circRNA formation inside cells. The results indicate that the linear RNA precursor was not self-splicing inside cells to produce circRNA (Table 4.1).

Table 4.1 RNA Circularisation in MDA-MB-231 cells results summary.

	Linear (convergent)	circRNA (divergent)
AM	✓	×
AF	✓	×

✓- A positive band after RT-PCR
× - No band after RT-PCR

Next, the peptides produced from cells transfected with mastoparan Flag and FRET containing constructs were analysed by western blot. The results supported that precursor RNA was not self-splicing in cells as both *Anabaena* Mastoparan Flag and *Anabaena* Mastoparan Flag IORF constructs detected bands at the unspliced RNA sizes. Neither *Anabaena* Mastoparan Flag or *Anabaena* Mastoparan Flag IORF peptides were detected in the culture medium of transfected cells, indicating the unspliced *Anabaena* constructs were not being exported out of the cell via the signal peptide. Linear FRET peptides could not be detected in the cell lysate or culture medium by anti-GFP western blot analysis, and positive controls were still undetectable, as described in 3.2.7. However, red fluorescence was detected in the media of Linear FRET transfected cells.

Table 4.2 Western blot FRET and mastoparan Flag results summary.

	Culture Medium	Cell Lysate
LF	×	×
AF	×	×
AMF	×	✓*
AMFIORF	×	✓*

✓ Band detected at correct MW
× No band detected
* Peptide produced from unspliced RNA

Table 4.3 Fluorescence detection of FRET peptides results summary.

		Culture Medium	Cell Proteins
LF	FRET	×	×
	EGFP	×	×
	mCherry	✓	×
AF	FRET	×	×
	EGFP	×	×
	mCherry	×	×

✓ Fluorescence detected
× No fluorescence detected

Finally, numerous MTT assays demonstrated the effect of genetically encoded mastoparan on cell death. Initially, amidated mastoparan was used to treat MDA-MB-231 cells and demonstrated a dose-dependent response. MTT assays (2.2.5) were performed on transfected cells, *in vitro* produced mastoparan, media from transfected cells and extracted proteins of mastoparan transfected cells.

Table 4.4 MTT assay results summary.

	In vitro	Transfection	Cell Extracted Proteins	Media
LM	-	✓ MDA-MB-231 × MCF-7	✓	×
AM	×	×	✓	✓
AMF	×	-	-	-

✓ Cytotoxicity
× No Cytotoxicity

The MTT assays indicated that genetically encoded mastoparan causes cytotoxicity in both MDA-MB-231 and MCF-7 cells. Treatment with extracted proteins of Linear Mastoparan and *Anabaena* Mastoparan transfected cells caused cell death. Transfecting cells with Linear Mastoparan caused cell death but *Anabaena* Mastoparan transfection did not. Finally, only treatment with *Anabaena* Mastoparan culture medium from transfected cells caused cell death. The *in vitro* translated *Anabaena* Mastoparan Flag peptide displayed the strongest signal on the western. However, treating MDA-MB-231 and MCF-7 cells with the translation reaction did not affect the viability of cells.

4.3.1 RNA Circularisation

The RT-PCR reaction's convergent and divergent primer sets bind to the *Anabaena* intron sequence. Therefore, only visible bands were expected for the *Anabaena* constructs. The linear control primers bind to 5' and 3' *Anabaena* intron regions that are present after self-splicing and amplify in a convergent manner. This results in the amplification of unspliced *Anabaena* containing sequences and circRNA sequences. The circRNA primer set binds to the 5' *Anabaena* intron region and amplifies in a divergent manner through the circular junction, amplifying only circRNA. The results demonstrate transcription of the RNA precursor in MDA-MB-231 cells, but circRNA was not detected.

The initial self-splicing intron selection for RNA circularisation was yeast group II introns. Although group II intron self-splicing capabilities are yet to be evaluated in mammalian cells,

given that they are present and functional in eukaryotes, it is hypothesised that they would. Inverse yeast group II intron construction difficulties forced RNA circularisation from alternative methods. It was hypothesised that the *Anabaena* group I introns would be capable of self-splicing in mammalian cells as *Tetrahymena* group I introns have demonstrated this ability, albeit to a low efficiency (Shin, Sullenger and Lee, 2004).

The only cofactors for group I intron self-splicing are ~1 mM Mg^{2+} and micromolar concentrations of guanosine or GTP, which are readily available in mammalian cells. Insertion of the *Tetrahymena* group I introns into *E. coli* 23S rRNA produced similar self-splicing efficiencies to its natural *Tetrahymena* location, whereas insertion into an mRNA region in *E.coli* resulted in much lower efficiency. This implies that the folding capability of the catalytic RNA structure is the primary factor in self-splicing efficiency, and there is no requirement for a species-specific protein (Zhang et al., 1995). Therefore, the difficulty of *Anabaena* intron splicing in mammalian cells is likely caused by cellular conditions preventing its folding capability.

The engineered *Anabaena* sequence generates a splicing bubble, which aids self-splicing increasing the efficiency. Therefore, it was thought that the engineered *Anabaena* sequence would provide a better model for RNA circularisation in mammalian cells than the *Tetrahymena* group I introns. *Anabaena* self-splicing was not successful in mammalian cells (4.2.1) Designing a *Tetrahymena* group I intron sequence with homology arms similar to the *Anabaena* construct may allow for efficient circularisation in mammalian cells, which is of interest for future investigation.

4.3.2 Peptide Analysis

Peptides were not detected in the culture media of transfected cells for any of the constructs used (4.2.2). This indicates that exportation via the N-terminus signal peptide is not occurring. There is a start codon in the 5' *Anabaena* intron immediately preceding the splice site. When the RNA remains unspliced, translation is initiated from the upstream start codon, generating a peptide that starts 66 aa upstream of circMastoparan Flag and circFRET and lacks an N-terminus signal peptide. Therefore, it is unsurprising that no peptides were detected in the culture media for *Anabaena* Mastoparan or *Anabaena* FRET by western blotting or the fluorescent assays.

The engineered *Anabaena* introns used by Wesselheoft et al. 2018, that this project was based on, did not translate circRNA with an infinite open reading frame. Therefore, the ribosome did translate through the *Anabaena* sequence. It was unknown how much disruption it would cause to translation in this project. Therefore, the linear controls were designed to circumvent this problem and still produce a function FRET/ mastoparan concatemer if the *Anabaena* sequence completely abolished translation. The flag control peptide was not detected. However, peptide bands were present at the correct size for unspliced *Anabaena* Mastoparan Flag (18.03 kDa) and AMIORF (20.84 kDa), albeit with low expression.

The *Anabaena* regions contain non-optimal codon usage, including a rare Leucine (TTA) at the end of the 5' and 3' *Anabaena* sequences. The unspliced RNA generates a peptide that contains only two unique additional amino acids generated from circMastoparan Flag IORF, one aa before the 5' splice site and one after the 3' splice site. Therefore, the low expression seen in cells for unspliced *Anabaena* Mastoparan Flag and *Anabaena* Mastoparan Flag IORF

would also result in circMastoparan Flag and circMastoparan Flag IORF. IORF circRNAs lack a stop codon for translation termination. Thus, a ribosome cycling around the circRNA that encounters a series of rare codons would be more prone to ribosome stalling and stacking, which would disrupt translation.

The peptide produced from unspliced *Anabaena* Mastoparan Flag and *Anabaena* Mastoparan Flag IORF still contains a double mastoparan repeat, with an MMP cleavage site linker separating them. Therefore, the proteins from *Anabaena* Mastoparan Flag and *Anabaena* Mastoparan Flag IORF transfected cells can be extracted from the cells and subsequently added to the culture medium to see the effect of genetically encoded mastoparan cytotoxic effect on cancer cells. A band was present for spliced *Anabaena* Mastoparan Flag around 10 kDa, which could represent the peptide produced from spliced mastoparan at 10.63 kDa. However, since circRNA was not detected via the very sensitive RT-PCR approach in cells, this band was likely the result of IRES-mediated translation of the linear transcript.

The FRET constructs containing EGFP and mCherry fluorophores were designed for viewing the peptide production process, exportation and extracellular MMP cleavage (3.3.4). FRET peptide expression in cells could not be detected, as extracted proteins showed no significant fluorescence (4.2.3). The GFP control protein extracted from cells transfected with a GFP encoding plasmid (pmaxGFP, Lonza) also did not demonstrate GFP fluorescence. It is possible that the extraction failed for that sample, or the proteins were degraded at the time of detection, or more concerning, it may indicate an extraction problem. A future approach that would circumvent any possible protein extraction issue is measuring the fluorescence of growing viable cells using the bottom optics of the spectrophotometer. If the FRET peptide were functioning as designed, the peptide would be detectable in the culture media via signal peptide

exportation and after *in vitro* translation (3.2.7). The only positive fluorescence result was mCherry detection in the media of Linear FRET transfected cells (590/20 nm, emission 645/40 nm). This would indicate expression of FRET in the cell and exportation via the signal peptide if either FRET (Excitation 485/20 nm, emission 645/40 nm) or EGFP (Excitation 485/20 nm, emission 528/20 nm) filters led to fluorescence. The former would indicate uncleaved FRET in the medium, and the latter, cleaved FRET releasing the fluorophores. Therefore, mCherry fluorescence in the media alone does not provide much biological relevance.

4.3.3 Mastoparan Cytotoxicity

The effect the genetically encoded mastoparan prodrug had on cancer cells was assessed. The natural mastoparan peptide, mastoparan-NH_2, was initially used on MDA-MB-231 cells to demonstrate cytotoxicity. The results showed a lower IC50 (12.06 µM) to that reported in previous literature (20-24 µM) (Hilchie et al., 2016). This was likely due to standard deviation in the biological replicates at 20 µM mastoparan concentration, affecting the nonlinear regression model. The purpose of the experiment was to demonstrate cytotoxicity to MDA-MB 231 cells, which was successful and therefore further optimisation was not required. Amidated mastoparan (mastoparan-NH_2) has much greater potency than non-amidated mastoparan mastoparan-COOH) produced in this project. Mastoparan- NH_2 causes cell death via penetrating the cell's plasma membrane in a direct-acting lytic mechanism (de Azevedo et al., 2015). This mechanism leads to an inflammatory response, whereas mastoparan-COOH causes cell death permeating the mitochondrial membrane, leading to controlled cell death by intrinsic apoptosis. Therefore, mastoparan-COOH may be preferential for therapeutic uses, although it is less potent than its amidated form.

Transfection of MDA-MB-231 and MCF-7 cells with *Anabaena* Mastoparan resulted in no cell death. This is not surprising considering that the peptides produced from unspliced *Anabaena* Mastoparan RNA do not contain N-terminal signal peptides for exportation, and activation of the mastoparan peptide occurs via MMP cleavage in the tumour microenvironment. Transfection of MDA-MB-231 with Linear Mastoparan resulted in cell death compared to the -ve GFP control ($P = 0.0024$), indicating that the linear mastoparan control was functional.

In vitro translated mastoparan from *Anabaena* Mastoparan unspliced RNA generated a peptide with the highest expression. However, subsequent treatment using the *in vitro* reaction on MDA-MB-231 and MCF-7 cells showed no effect.

Proteins were extracted from Linear Mastoparan and *Anabaena* Mastoparan transfected MDA-MB-231 cells, and these proteins were subsequently added to the media of MDA-MB-231 and MCF-7 cells to see their effect on cell viability. Initially, MDA-MB-231 cells were treated with 2.5 - 5 ng/μL of cell extracted proteins. *Anabaena* Mastoparan caused significant cell death at all concentrations, increasing cell death dose-dependently (Figure 4.8). Only Linear Mastoparan cell proteins at the highest quantity (5 ng/μL) caused any other significant death. The protein concentration ranges were then increased/decreased depending on the initial results, with *Anabaena* Mastoparan decreasing (1 - 3 ng/μL) and control, GFP and Linear Mastoparan concentrations increasing (5 - 7 ng/μL). MCF-7 cells were also treated. These results showed even 1 ng/μL of *Anabaena* Mastoparan proteins was sufficient to cause cell death. Linear Mastoparan also caused cell death at the increased concentration. *Anabaena* FRET transfection does not cause cytotoxicity to either cell line. However, its extraction from the cell lysate and subsequent use as a treatment for cells results in significant cell death. This

indicates that the prodrug design is functional (Figure 3.15), as mastoparan only becomes active and cytotoxic in the extracellular environment where MMPs are active.

MDA-MB-231 cells are highly invasive, aggressive cancer cell lines, whereas MCF-7 is less invasive and aggressive. MDA-MB-231 cells display invasiveness by mediating the proteolytic degradation of the extracellular matrix (ECM) and have much greater amounts of active MMPs in their tumour microenvironment (Gobin et al., 2019). The mastoparan constructs used in this project were designed to be activated by MMPs. Thus, it would be expected that MCF-7 cells would have less cell death relative to concentration. However, the results were relatively similar between the two cell lines, with 2 ng/μL and 3 ng/μL causing significant cell death compared with the control.

The results demonstrated that treatment with the cell extracted proteins caused cell death, even without containing mastoparan. This is seen in the absorbance difference between the control and the untreated. It is unknown why this effect occurs, and it appears to have much greatest effect on cell death in MDA-MB-231 cells and not MCF-7 cells (Figure 4.9). It is also shown that the GFP control causes significant cell death compared to the control at concentrations > 5 ng/ μL, this may be because transfection with a plasmid causes cell stress and a greater amount of cells undergo apoptosis than the control (Figure 4.7). The stress response proteins produced in the cell may have had a cytotoxic effect when added to the media.

The media from concentrated cells was concentrated and then used to treat MDA-MB-231 cells (Figure 4.10). This provided interesting results considering that it had previously been established that mastoparan was not being secreted out of the cell, given that the precursor-RNA was not self-splicing. The likely cause is that dead cells remain after concentration to 200

μL. When added to the media at greater concentrations, there is a greater quantity of dead cells added, having a negative effect on cell viability.

Chapter Five

5.0 Novel Peptides from CircRNA Translation

5.1 Introduction

Endogenous circRNAs were long thought of as non-coding erroneous splicing products (Xu and Zhang, 2021). In recent years it has been revealed that circRNAs are abundant and conserved (Jeck et al., 2012). A conservative estimate is that 14.4% of linear RNAs also form circRNAs, showing that they contribute significantly to the transcriptome (Salzman et al., 2012). Aberrant levels of circRNAs are a feature of many cancers and carry predictive and prognostic significance (Wu, Li, Wu and Liu, 2019). Therefore, circRNAs have great potential as a non-invasive biomarker for early cancer detection (Li et al., 2015). The biological functions of circRNAs are diverse, and many are involved in gene regulation using a variety of mechanisms (Table 1.1). The classification of endogenous circRNAs as lncRNAs has recently been challenged with the discovery of circRNAs capable of being translated (Lei et al., 2020). Translatable circRNAs can be identified by ribo-seq when the sequence read spans the unique sequence where the 5' and 3' splice sites are joined, known as the backsplice or circular junction. This technique has identified 1479 mammalian circRNAs recorded on the riboCIRC database (Li et al., 2021). Although this method identifies translated circRNAs, it does not provide insight into these peptides' functional capabilities, which must be experimentally determined. Therefore, most translatable circRNAs identified producing a unique peptide to their parent gene have an unknown/undiscovered function. Examples of circRNAs that produce a peptide of experimentally characterised function are shown in table 5.1.

Table 5.1 Functionally characterised translatable circRNAs. Examples translatable circRNAs and their functionally characterised peptide.

CircRNA Name	Description	Reference
circ-ZNF609	Circ-ZNF609 contains a 753-nt ORF and expresses a protein that plays an important role in myoblast differentiation.	Legnini et al., 2017
circ-SHPRH	Circ-SHPRH expresses an isoform of histone-linker PHD and RING finger domain-containing helicase (SHPRH), the 17 kDa protein SHPRH-144aa, which has a role in tumour suppression.	Zhang et al., 2018
circFPXW7	circFPXW7 produces a 10 kDa protein which suppress glioblastoma cell proliferation	Yang et al., 2017
circPINTexon2	circPINTexon2 produces a 26 kDa protein which suppress glioblastoma cell proliferation	Zhang et al., 2018
circβ-catenin	circβ-catenin produces β-catenin-370aa, a β-catenin isoform of 50 kDa. It's role protects β-catenin from ubiquitination and degradation mediated by the kinase GS3Kβ.	Liang et al., 2019

Translatable circRNAs require internal translation initiation as they are covalently closed loops and therefore lack free 5' ends for cap-dependent translation. Internal ribosomal entry sites and m^6A methylation sites are the two known mechanisms for ribosomal recruitment on circRNAs. Endogenous IRES-dependent circRNA translation functions similarly to those on artificial circRNA constructs (see chapter 1). N6-methyladenosine (m^6A is the most abundant form of internal RNA modification, with an average of one to three m^6A modifications per transcript (Zaccara, Ries and Jaffrey, 2019). m^6A methylation is characterised by methylation of adenosine at the nitrogen-6 position. The m^6A methylation profile is different in mRNAs under different cell stresses, and circRNA methylation patterns are different from their linear mRNA form (Zhou et al., 2017). The consensus m^6A methylation sequence is DRACH, where D = A/G/U, R = A/G, H = A/C/U. However, only 10% of the consensus sequence adenosines are methylated (Wang and Wang, 2020). A change in methylation pattern, either from mRNA to circRNA or under different cellular stresses, can result in the translation of unique peptides

(Zhou et al., 2017). CircMAP3k4 is highly expressed in Hepatocarcinoma (HCC). CircMAP3k4 produces a 455-aa long peptide from m^6A-mediated translation that causes chemoresistance to cisplatin (Duan et al., 2022).

Translatable circRNAs are difficult to detect by computational approaches because IRESs do not have a conserved primary, secondary or tertiary sequence (Koirala et al., 2019). Computational approaches for predicting IRESs from circRNAs provide a predictability score, but definitive proof of activity requires experimental work. m^6A methylation sites occur very frequently in circRNAs (Zhou et al., 2017). However, not all m^6A methylation sites result in internal translation initiation (Zhou et al., 2017). Looking through peptide databases alone, it is impossible to determine whether a peptide has been produced from a circRNA or mRNA unless the translation occurs through the circular junction (Li et al., 2021). There is currently not a reliable high-throughput experimental method to determine translatable circRNAs, therefore is difficult to predict how circRNA translation contributes to the proteome.

Peptides from circRNAs can differ from those translated by linear mRNA in several ways (Figure 5.1). As translation initiation is cap-independent on circRNAs, translation can initiate at an alternative start codon to the corresponding mRNA. When translated, the unique amino acid sequence at the back splice junction offers greater variance in sequence. In addition, if the sequence of the circRNA is not a multiple of three, then a frameshift occurs through the junction giving rise to peptides translated from multiple codon frames from the circRNA. The variance in the amino acid sequence of circRNAs can result in peptides with varying functions. This was previously observed in a protein encoded by circFNDC3B named circFNDC3B-218aa, which initiates translation via an IRES and has an ORF through the circular junction (Pan et al., 2020). This protein inhibited tumour progression by increasing the metabolic switch

from glycolysis to oxidative phosphorylation by inhibiting Snail-FBP-1 signalling, which suppresses EMT progression in colon cancer cells, a feature that FNDC3B did not possess (Pan et al., 2020).

Translation can also be initiated from near cognate start codons. This is particularly apparent in pathological conditions such as cancer, where non-AUG translation events can promote or inhibit cancer progression (Créancier, Morello, Mercier and Prats, 2000). Fibroblast Growth Factor 2 (FGF2) controls cell proliferation, differentiation, and angiogenesis. The canonical AUG start codon produces a protein localised in the cytoplasm, whereas four upstream in-frame CUG codons generate longer isoforms in transformed cells that localise in the nucleus (Créancier, Morello, Mercier and Prats, 2000). Non-cognate start codons also add to the complexity of predicting proteins expressed from circRNAs as non-AUG translation events are typically much less efficient than AUG initiated translation, so including them when identifying circRNA ORFs would lead to overestimation of frequency (Kozak, 1989). Cap-dependent translation is more efficient than internal forms of translation initiation, although internal translation initiation events occur more frequently under stress conditions and/or when cap-dependent translation is suppressed (Thoma et al., 2004). Therefore, if an IRES or m^6A site capable of internal translation initiation is present downstream of the cap, translation may be repressed from these internal sites until circularisation of the RNA occurs, and cap-dependent translation is abolished.

The work described in this chapter uses a bioinformatics approach to attempt to further the knowledge of peptide production from circRNA translation by investigating the translational capacity and peptide functionality of grouped circRNAs and reveals circRNAs with the potential to produce peptides with an alternative function to the gene their parent genes.

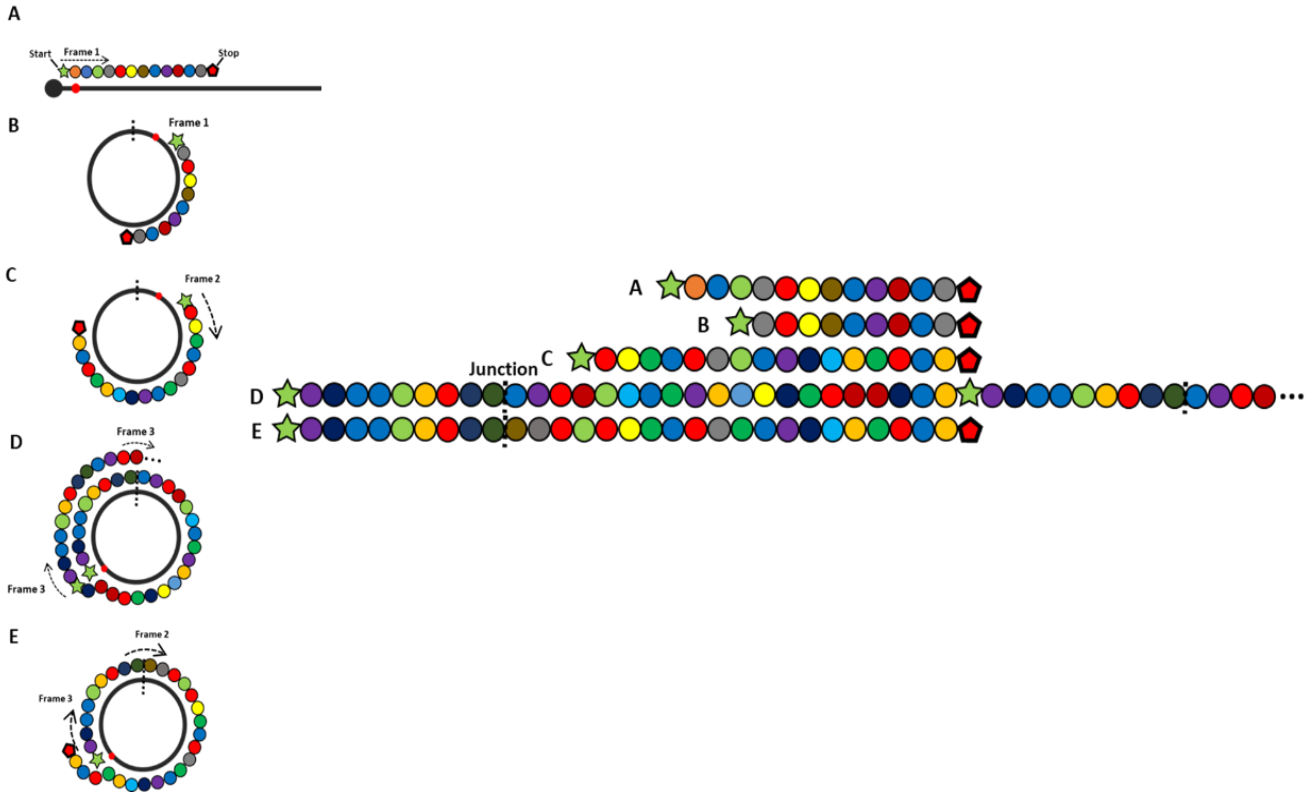


Figure 5.1 Peptide variation from circRNA translation. Left – mRNA/circRNA with aa sequence mapped to transcript location, Right – Aligned amino acids sequences from transcripts A-D. A – Cap-dependent translation on mRNA initiating at the common start codon in frame 1 . B – m^6A-driven translation on circRNA initiating at a start codon downstream of transcript A start codon, resulting in a shorter peptide (right). C - m^6A-driven translation on circRNA initiating at a start codon in a different frame to transcript A and B (frame 2), resulting in a new peptide sequence. D – Translation of a long repetitive peptide from Infinite open reading frame (IORF) circRNA. E - m^6A-driven translation on circRNA with a different m^6A methylation position to the mRNA. Translation initiation results in a new peptide that starts translation in frame 3, then changes frame through the back splice junction resulting in a peptide with a novel N-terminus and a conserved C-terminus to peptide C.

5.2 Results

5.2.1 CircRNA Sorting and Length Distribution Analysis

The circRNAs were obtained from circAtlas database and sorted into groups ORF, no ORF, IORF, ORF through junction, and all circRNAs using python scripts (Table 5.2). The entire workflow is shown in figure 2.3. The length distributions were plotted as a histogram plot for the sorted circRNAs with a 9nt interval ranging from 27 nt to 900 nt in size (Figure 5.2.)

Table 5.2 Sorted circRNAs and their discriminatory feature.

Group	Full Name	Discriminatory Feature
ORF	Open Reading Frame	Contains a start codon
no ORF	No Open Reading Frame	Lacks a start codon
IORF	Infinite Open Reading Frame	Contains a start codon and lacks a stop codon in the same frame
ORF through junction	Open Reading Frame through junction	Contains a start codon and not a stop codon before the end of the sequence is reached
All circRNAs	-	No Discriminatory Feature

For all circRNA groups apart from no ORF, 144nt bin is the peak number of circRNAs, meaning that most circRNAs are between 141-148nt in length (9nt bin width). No ORF circRNA lengths were more skewed towards shorter length circRNAs. All circRNAs show a step up at 144nt, where the abundance of circRNAs increases more than expected from the trend of the graph. This is easily seen when representing the circRNA abundance as relative frequency (%) (Figure 5.2 aii). This step-up appears to be of the same degree uniformly across all the circRNA groups with their relative abundance.

The no ORF circRNA group was characterised by circRNAs that did not contain an ATG start codon in any frame. Python script non_consensus_ORF.py was run, which retrieved circRNAs from the ‘no ORF’ group that contained non-consensus start codons, CTG, GTG, and ACG. The length distributions were plotted in the same way as before for non-consensus circRNAs (contain CTG, GTG or ACG) and no non-consensus ORF circRNAs (do not contain ATG, CTG, GTG or ACG) (Figure 5.2 b). The number of circRNAs now determined as no ORF was very low and heavily skewed towards the smaller circRNAs. However, it was clear that the step up at 144 nt remained in all circRNA groups.

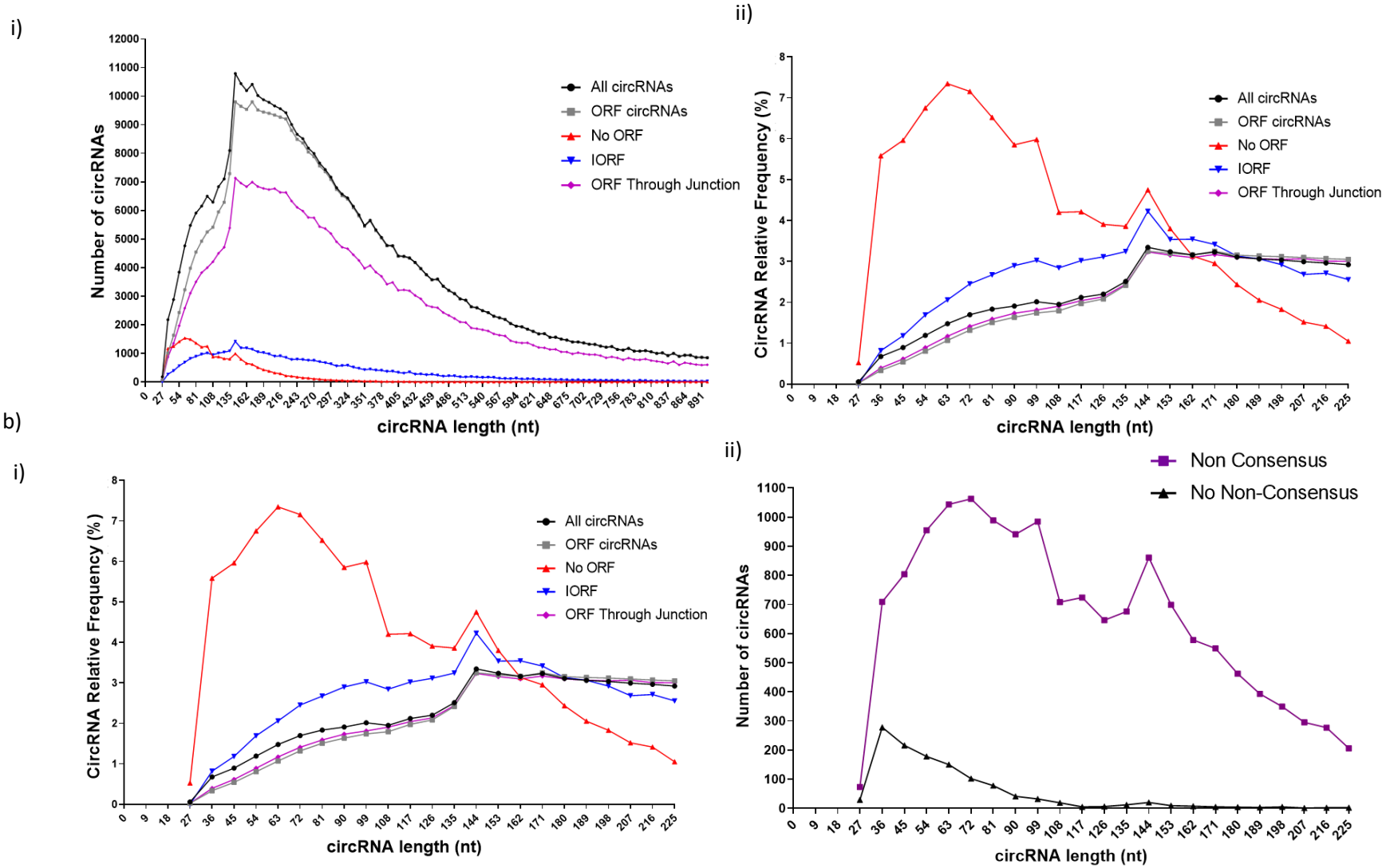


Figure 5.2 The length distribution analysis circRNAs. a) Length distribution of grouped circRNAs b) Length distribution of circRNAs with non-consensus start codons i) Number of circRNAs ii) circRNA relative frequency. Bin width 9nt for all graphs. All circRNAs — n = 463110; ORF circRNAs — n = 442,109; No ORF circRNAs — n = 21,001; IORF circRNAs — 39,574; ORF through junction circRNAs — n = 321,473, Non consensus start codon circRNAs — n = 16,505; No non consensus circRNAs — n = 1207.

5.2.2 Codon Adaption Index (CAI)

The codon adaption index (CAI) for the open reading frame through the junction circRNAs was calculated using CAIcal. The CAI for the parent gene coding sequence (CDS) and a scrambled ORF through junction sequence was also calculated (Figure 5.3). All groups were statistically significant from each other ($P < 0.0001$), with mean (Standard deviation) CAI values of 0.79 (0.045) for CDS mRNA, 0.76 (0.060) for ORFs through the junction, and 0.70 (0.056) for the scrambled sequences. The median (25^{th} percentile ─ 75^{th} percentile) was 0.79 (0.75 ─ 0.82), 0.76 (0.72 ─ 0.79) and 0.70 (0.67 ─ 0.73) for CDS mRNAs, the open reading frames through the junction, and scrambled sequences, respectively. The open reading frame through junction sequences, CDS mRNA and scrambled sequences demonstrated a similar CAI distribution pattern, with the mean, median, 25^{th} percentile and 75^{th} percentile all 0.03 higher in CDS mRNA than in the open reading frame through junction sequences, and 0.06 higher than in the scrambled sequences.

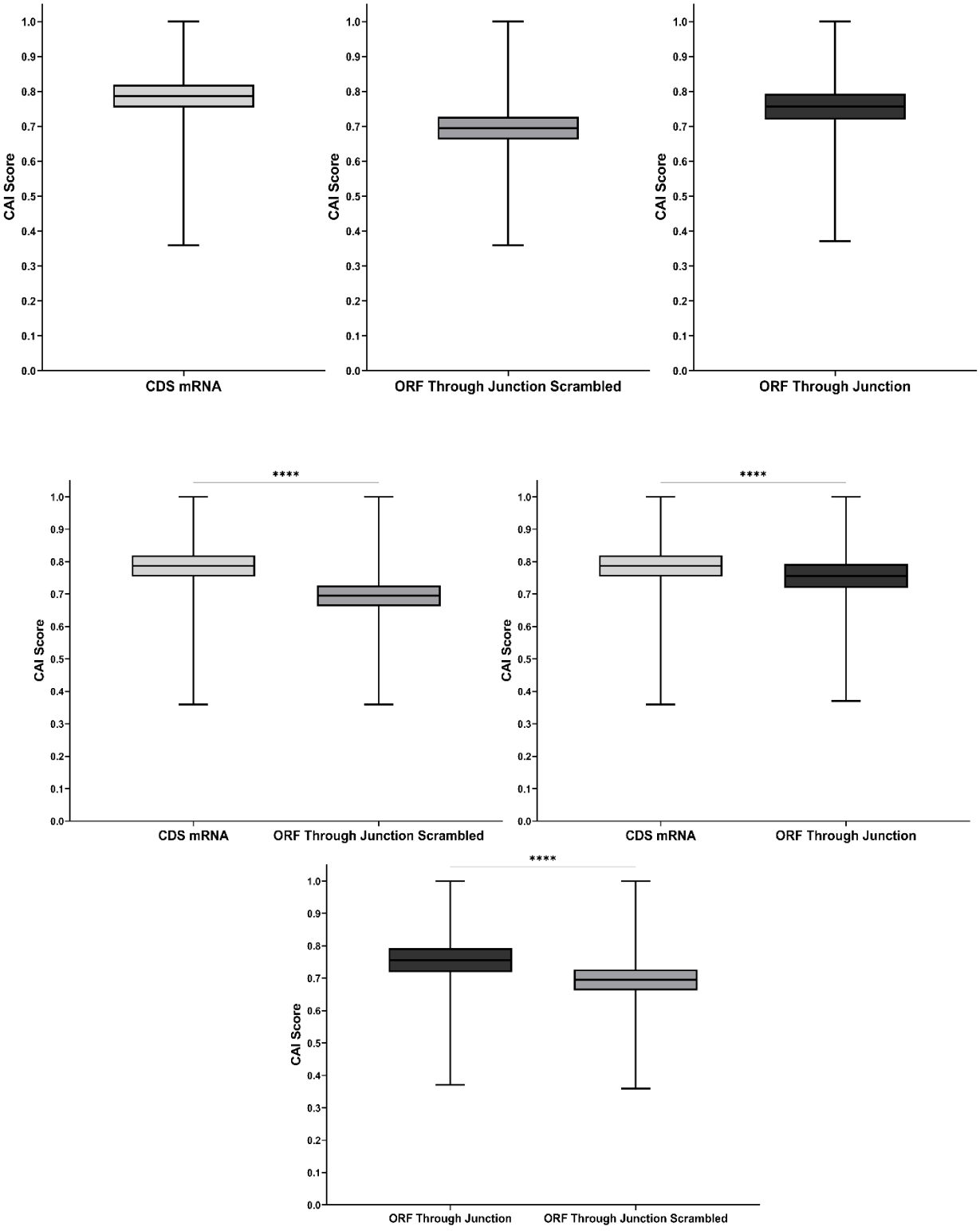


Figure 5.3 The CAI score distribution of open reading frames through circular junction compared with scrambled sequences and CDS mRNA. Within each box, the inner horizontal line denotes median value; boxes extend from the 25th to the 75th percentile of each group's distribution of values; vertical extending lines denote the minimum and maximum values of each group. Unpaired t-test performed between groups. **** P < 0.000. ORF Through Junction n = 429,414; ORF Through Junction scrambled n = 429,414; CDS mRNA n = 99,030.

The CAI scores towards the extreme high (95th and 99th) and low (5th and 1st) percentiles were then investigated. CDS mRNA had a CAI score 0.86 and 0.88 for their 95th and 99th percentile compared with 0.85 and 0.90 for open reading frames through the junction and 0.79 and 0.84 for scrambled sequences. For both the 5th and 1st percentiles, the CAI scores were 0.72 and 0.68 for CDS mRNA, 0.66 and 0.60 for open reading frames through the junction, and 0.60 and 0.55 for the scrambled sequences. With the values at the extremities of the percentiles, the range in CAI becomes larger for ORFs through the junction and scrambled sequences compared to the CDS mRNA sequence. A relative CAI frequency histogram chart was plotted to better visualise the percentile differences.

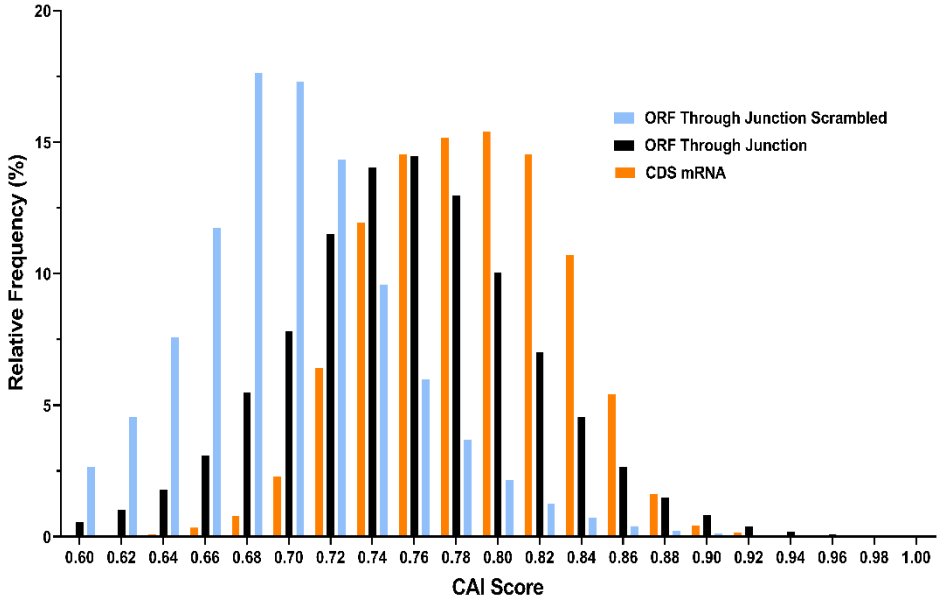


Figure 5.4 Codon Adaption Index (CAI) relative frequency score distributions of CDS mRNA, ORFs through the circular junction and scrambled sequences. Histogram plot with bin widths 0.02. ORF Through Junction n = 429,414; ORF Through Junction scrambled n = 429,414; CDS mRNA n = 99,030.

The histogram plot demonstrates the trend that the scrambled sequences generally had a lower CAI score than the other two groups, with the CDS mRNAs being the most skewed towards a higher score. A bell curve was observed in all three groups with the ORF through

the junction and scrambled sequences having a greater proportion of circRNAs towards the ends of the distribution.

5.2.3 Peptide Analysis

Using a series of python scripts and Emboss tools (Rice, Bleasby and Ison, 2011) (2.3.1), amino acid sequences were obtained from the ORFs that run through the circular junction. Multiple programmes/tools were run to identify homology and function. A summary of how each was used in the project is shown in table 5.3. The peptide analysis workflow can be followed in figure 2.3.

Table 5.3 Programme / tools used for peptide analysis.

Programme/Tool	Function in Project	Reference
Embosstranseq	Converts nucleotide to amino acids sequences	Rice, Bleasby and Ison, 2011
BioMart - Ensembl	To obtain peptide sequences for parental genes	Cunningham et al., 2022
Blastp	1) Search for unmapped circRNA peptides 2) Aligned InterProScan results to parental peptides	Altchul et al., 1990
InterPro	Used with PANTHER database to give functional analysis of circRNA and parental peptides.	Blum et al., 2021
Phyre2	Predicted function of circRNA and parental peptides based alignment to known protein structures	Kelley et al., 2015

Some proteins present in reference proteomes are not mapped to the genome. In the reference proteome used in this project (UP000005640) obtained from Uniprot, there are 80,027 proteins, with 1347 unplaced. This constitutes 1.68% of the reference proteome (UniProt Consortium, 2018). It was investigated whether the ORF through the junction peptides could be aligned to

any of these sequences by performing a Blastp Alignment (2.3.4). None of the BLAST results that ran through at least 10 amino acids on either side of the circular junction was ≥ 95% identical, indicating no ORF through the junction peptides were unplaced.

The protein family domains within the peptide sequences were then analysed using the PANTHER (protein analysis through evolutionary relationships) database integrated with the InterPro software. The corresponding parental coding sequences to the ORF through the junction circRNAs were also obtained from Ensembl Biomart (Cunningham et al., 2022) and analysed with InterPro to compare with the IORF through the junction results. The results obtained were PANTHER terms for parental and ORF through the junction peptides. IORF through the junction circRNAs were disregarded with peptides that did not have an assigned PANTHER term or had PANTHER terms that matched the parental.

The InterPro result found domains with the amino acid sequence to assign a PANTHER term. Therefore the entire length of the ORF through the junction peptide sequence was used to determine the PANTHER term. A Blastp alignment was performed between the sequence in which the InterproScan result was used and the parent protein. A high homology would indicate a similar biological function; therefore, the circRNA would be disregarded as of interest. Sequences with greater Query cover (%) + Identity (%) than 150 were removed.

At this stage, there were 77 unique circRNAs (+32 circRNA isoforms with the same PANTHER term) of interest. The protein for the parental gene for each of the identified circRNAs, was manually searched in Uniprot. Uniprot annotates all the domains within the parental sequence. It assessed whether any domains have the same or similar function to the

PANTHER term corresponding ORF through the junction peptide. If this was true, then the circRNA was disregarded. After this filtering step, 33 identified ORF through the junction peptides with a different biological function to their parent gene were identified.

To check that each circRNA had a correctly assigned parent gene, the full circRNA sequence was subject to a BLASTn. Some results showed 100% similarity to the parent gene of the InterProScan result, indicating an incorrectly assigned circRNA name. An example is the circRNA named hsa-TRBC2_0002 in circAtlas database with the parent gene T Cell Receptor Beta Constant 2 (TRBC2). InterProScan prediction was for serine protease Trypsin-2. The BLASTn result returned 100% similarity to Trypsin 2, demonstrating an incorrectly assigned circRNA identifier in the circAtlas database.

The ORF through the junction peptides were analysed in Phyre. This programme returns a 3D model of a peptide based upon alignment to known protein structures. Some results were the same or similar to the PANTHER terms returned from the InterproScan, validating the InterproScan results (Table 5.4). ORF through the junction peptides Phyre results that returned the same biological function as the parent gene were filtered out, leaving 4 remaining circRNAs of interest. Some peptides had a high degree of coverage and confidence to a known peptide. These peptides are likely to have a biological function related to the known peptide of high similarity.

5.2.4 CircRNA Expression Profile and Translatability

To see if the identified circRNAs were overrepresented in a tissue, their expression profiles were viewed by manually searching for each circRNA in circAtlas. The expression profile of the circRNAs was diverse, with expression found in the brain, testis, spinal cord, uterus, spleen, and bone marrow (Table 5.4). It would be expected to find the greater overrepresentation of filtered circRNAs in the brain tissue, as circRNA enrichment in the brain is already a known feature. There are around 300,000 circRNA sequences identified from brain tissue samples, and the next tissue of most abundance is the liver, with 150,000 circRNAs. 3 of the 4 circRNAs identified are expressed in the brain, which fits the expected overrepresentation. However, the sample size (4) is too low to comment on the significance.

The likelihood of translation initiation was assessed on the 4 circRNAs by predicting m^6A methylation or IRES sites using SRAMP and IRESpy (2.3.5) (Zhou, Zeng, YH, Zhang and Cui., 2016; Wang and Gribskov, 2019). All 4 circRNAs had an m^6A or an IRES site.

The m^6A sites had varying degrees of confidence; 1 high (LTBP1_0013) and 2 moderate (RIMS2_0163 and SFRP4_0005). The 4 circRNAs were investigated to predict potential IRES sites using IRESpy (2.3.5). The Prediction P cut-off value used was 0.1, the same as Wang and Gribskov, 2019 used to predict IRES human IRES present in 5'UTRs. The results revealed a potential IRES in CELF5_0014 circRNA (IRES probability 0.142). The other 3 circRNAs did not show any predicted IRES sites, with probability values ranging from 0.019 - 0.073 (Table 5.4)

The 4 circRNAs were checked to see if translation had been reported in Ribo-seq data or experimentally determined by searching them individually in ribocirc, a database of translatable circRNAs (Li et al., 2021). None of the 4 identified was present in the database.

5.2.5 Final CircRNA Peptides of Interest

Four candidate circRNAs were identified that translate a peptide with a different biological function to their parent gene. A summary of their circRNA and peptide characteristics is below. Table 5.4 details the key characteristics of each candidate circRNA and can be compared to the parent gene in Table 5.5. The functions of the peptides translated from the circRNA are described in this section, along with the function of their parent gene.

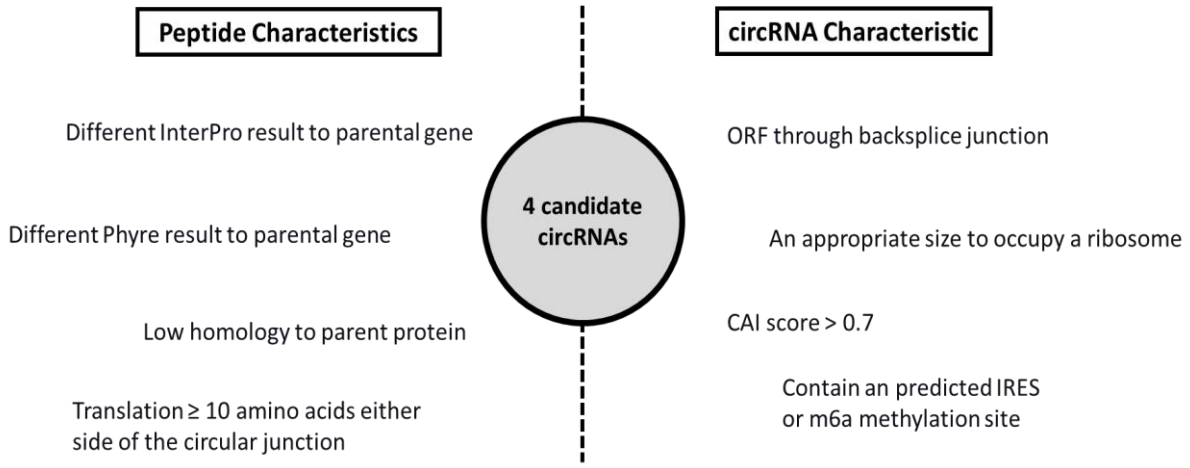


Figure 5.5 Characterises of candidate circRNAs and their peptides.

Table 5.4 Open reading frame through junction peptide

CircRNA Name	circRNA Size (nt)	ORF Length (nt)	CAI	Expression	Conservation	IRES	m6a	InterproScan	Phyre Result	Phyre Confidence (%)	Phyre Coverage (%)
CELF5_0014	470	318	0.794	Brain	Humans	IRES Predicted 0.142	N/A	SPOT2-Related	Plasmodium falciparum recombinant shortened csp	99.3	85
RIMS2_0163	241	183	0.781	Testis	Humans	No IRES Predicted 0.019	1 m6a: Moderate confidence	40S Ribosomal Protein S6	40S Ribosomal Protein S6	99.9	90
SFRP4_0005	440	330	0.831	Brain, Uterus, Spleen, Spinal Cord and the Bone Marrow	Humans	No IRES Predicted 0.064	1 m6a: Moderate confidence	Ependymin-Related Protein 1	Ependymin-Related Protein 2	99.9	88
LTBP1_0013	480	IORF	0.821	Testis, Heart, Brain and the Placenta	Humans, Monkey, Mouse and Pig	No IRES Predicted 0.073	1 m6a: High confidence	Laminin G-Like Domain-Containing Protein	Low-density lipoprotein receptor-related protein 4	100	91

Table 5.5 Parental gene peptides features.

Parental Gene Name	Protein	CDS Size (nt)	Expression
CELF5	CUGBP Elav-Like Family Member 5	1455	Tissue enriched - Brain
RIMS2	Regulating Synaptic Membrane Exocytosis 2	4233	Tissue enriched - Adrenal gland, Brain
SFRP4	Secreted frizzled-related protein 4 (SFRP4)	1038	Tissue enriched - Cervix, Uterine, Endometrium, Retina
LTBP1	Latent Transforming Growth Factor Beta Binding Protein 1	5163	Low tissue specificity

CELF5_0014 circRNA is 470 nt in length with a 318 nt (106 aa) ORF. The predicted peptide for InterproScan and Phyre was SPOT2-Related protein and *Plasmodium falciparum* recombinant shortened circumsporozoite protein (CSP). SPOT2 is an uncharacterised protein. CSP in the vertebrate host binds to highly sulphated heparan sulfate proteoglycans (HSPGs) on the surface of host hepatocytes (Rathore, Sacci, de la Vega and McCutchan, 2002). The Phyre result had an 85% coverage and 99.3% confidence. CELF5_0014 circRNA has a CAI score of 0.794 and, had no m^6A methylation sites, it did contain a potential IRES. The brain was the only tissue found to express CELF5 circRNA, the same as the parental gene. The circRNA is not conserved in any organism other than humans. RIMS2_0163 circRNA is 241 nt in length with a 183 nt (61 aa) ORF. The predicted peptide for InterproScan and Phyre was 40S Ribosomal Protein S6 (249 aa). This protein plays an important role in controlling cell growth and proliferation through the selective translation of particular classes of mRNA. The region of greatest homology between the RIMS2_0163 ORF through the junction and 40S Ribosomal Protein S6 is between positions 1-39 of the RIMS2_0163 peptide and 129 – 167 of the 40S Ribosomal Protein S6. 32/39 amino acids in the region are identical (83%), 34/39 (87%) positive with 0 gaps (Figure 5.6).

```
Query 1    MPHHLGPKRASRIHKLFNLSKEDDVHQYVVRKPLNKEDQ 39
           +P  LGPKRASRI KLFNLSKEDDV QYVVRKPLNKE +
Sbjct 129 VPRRLGPKRASRIRKLFNLSKEDDVRQYVVRKPLNKEGK 167
```

Figure 5.6 RIMS2_0163 peptide homology to Eukaryotic 40S Ribosomal Protein S6. Red – Differences in aa sequence, + – Positive (similar biochemical properties), Query – RIMS2_0163 peptide, (subject) – Eukaryotic 40S Ribosomal Protein S6.

The parental gene RIMS2 codes for Regulating Synaptic Membrane Exocytosis 2 presynaptic protein interacts with RAB3, a protein important for normal neurotransmitter release. RIMS2 circRNA has a CAI score of 0.781 and had a 1 m^6A methylation site of moderate confidence. The only tissue that has been found to express RIMS2_0163 circRNA is the testis, whereas the gene is highly expressed in the adrenal gland and the brain. The circRNA is not conserved in any organism other than humans.

SFRP4_0005 circRNA is 440nt in length with a 330 nt (110 aa) ORF. The predicted peptide result for InterproScan and Phyre returned Ependymin-Related Protein 1/2 (457 aa). The Phyre result returned a 99.9% confidence and 88% coverage. This protein's exact function is unknown, but its structure suggests it binds to hydrophobic molecules like fatty acids (Park et al., 2019). The region of greatest homology between the SFRP4_0005 peptide and Ependymin-Related Protein 1 is between positions 32-104 of the SFRP4_0005 peptide and 90-162 of the Ependymin-Related Protein. 71/73 residues are identical (97%), 73/73 (100%) positive with 0 gaps (Figure 5.7).

```
Query  32   RLFEYILLYKDGVMFQIDQATKQCSKMTLTQPWDPLDIPQNSTFEDQYSIGGPQEQITVQEWSDRKSARSFES   104
            RLFEYILLYKDGVMFQIDQATKQCSKMTLTQPWDPLDIPQNSTFEDQYSIGGPQEQITVQEWSDRKSARS+E+
Sbjct  90   RLFEYILLYKDGVMFQIDQATKQCSKMTLTQPWDPLDIPQNSTFEDQYSIGGPQEQITVQEWSDRKSARSYET   162
```

Figure 5.7 SFRP4_0005 peptide with homology to Ependymin-Related Protein 1. Red – Differences in aa sequence, + – Positive (similar biochemical properties), Query – SFRP4_0005 peptide, Sbjct (subject) – Ependymin-Related Protein 1 peptide.

SFRP4 parental gene encodes Secreted frizzled-related protein 4, this functions as a modulator of Wnt signalling through direct interaction with Wnts (Muley et al., 2010). SFRP4_0005 had a CAI score of 0.831, and one m^6A methylation site had moderate confidence of methylation. SFRP4_0005 circRNA is expressed in the brain, uterus, spleen, spinal cord, and bone marrow,

whereas the parental gene is expressed in the cervix, uterine, endometrium and retina. The circRNA is not conserved in any organism other than humans.

LTBP1_0013 circRNA is 480nt in length and an IORF. The predicted peptide result for InterproScan was Laminin G-Like Domain-Containing Protein. The Laminin G domain is critical for heparin-binding and cell attachment activity (Okazaki et al., 2002). Phyre analysis returned the results of low-density lipoprotein receptor-related protein 4, a single-pass transmembrane protein that carries out a variety of functions, including signal transduction and receptor-mediated endocytosis. The confidence of the Phyre result was 100%, and the coverage was 91%. The parent gene encodes Latent-transforming growth factor beta-binding protein 1. The function of this is unknown. However, similarities and colocalization of the LTBPs and fibrillins suggest a structural role for LTBP1 in the extracellular matrix (ECM). LTBP1_0013 ORF had a CAI score of 0.821 and contained an m^6A methylation site that had high confidence of methylation. LTBP1_0013 circRNA is expressed in the testis, heart, brain and placenta. The parental gene LTBP1 has very low tissue specificity. The circRNA is conserved in monkeys, mice and pigs.

5.3 Discussion

CircRNAs were sorted into desired characteristics of ORF, no ORF, IORF, ORF through junction and all circRNAs using python scripts (Table 5.2). Length distributions were plotted on a histogram with the data representing actual numbers and relative frequency (5.2 a). This revealed a step-up in every sorted circRNA group around 140 nt, which may indicate a selection for circRNAs to be a translatable size, although the step-up was present for circRNAs that did

not contain an ORF. To investigate whether the step-up size around 140 nt for the no ORF circRNA group was due to translation of near cognate start codons, the non ORF group was further sorted into circRNAs that contain a common non-consensus start codon and those that do not (5.2 b). The step-up was still seen in the no non-consensus start codon group, indicating the step up in size is due to a translation independent factor.

The codon adaption index (CAI) was calculated for circRNAs that contained an ORF through the circular junction and compared to a scrambled sequence and coding sequence mRNA (Figure 5.3). The results showed a significant difference ($P < 0.0001$) between all groups, with CDS mRNA mean CAI the highest (0.787) followed by ORFs through junction (0.756) and then scrambled sequences (0.695).

The OFTJ nucleotide sequences were translated into amino acid sequences. The ORF through the junction peptides were aligned to the unplaced peptides in the human reference proteome (UP000005640), but no hits were obtained for OTFTJ peptides that ran 10 amino acids on either side of the circular junction with > 94% identity (5.3).

The biological function of the ORF through the junction peptides was investigated using InterProScan with the PANTHER database and compared to their parent gene. The PANTHER terms that were different for ORF through the junction peptides were identified. A BLASTp alignment of the region spanning the InterProScan result and the parent gene peptide was performed, and any result above 150% similarity (Query cover + Identity) was removed.

The sorting continued by individual analysis of the ORF through the junction peptides by

searching the parent gene in the Uniprot database for domains within the parent protein that can explain the PANTHER term of the ORF through the junction peptide. For quality control, each of the ORF through the junction circRNA full sequences was subject to a BLASTn alignment to their parent gene, this identified circRNAs in the circAtlas database with an incorrect identifier name in which the parent gene was incorrectly assigned.

ORF through the junction peptides were analysed using Phyre. Predictions that returned the same or similar result to the parent gene were removed. After all these sorting steps, 4 circRNAs were identified with the coding capacity to produce a peptide with a different biological function to their parent gene.

The expression profile of the circRNAs was obtained from circAtlas, and the translatability of the circRNAs was then assessed by using SRAMP and IRESpy to identify m^6A methylation and IRES sites. All of the 4 candidate circRNAs contained a predicted m^6A methylation or IRES site. The 4 circRNAs were searched in ribocirc, a database of translatable circRNAs, but none of the candidate circRNAs was present.

Finally, a summary of the key characteristics for each candidate circRNA was made, detailing the biological function of the peptide predictions for Phyre and InterProScan (5.2.5).

5.3.1 CircRNA Length Distribution and Codon Adaption Index (CAI)

The circRNA length distribution analysis revealed a step up in size around 140 nt (Figure 5.2). Abe et al., 2013 reported a minimal circRNA threshold size of around 126 nt for efficient translation in Prokaryotes. As Eukaryotes have larger ribosomes, it can be presumed that this threshold size for circRNA translation is slightly larger. This step-up may be a selection for circRNAs to be of a translatable size. This would indicate that translation is a significant feature of circRNAs, and circRNA peptides may contribute significantly to the proteome. If the spike in circRNA size were due to the minimum threshold for translation, it would be expected that circRNAs without an ORF lack this spike around 140 nt. However, the circRNA relative frequency graph shows that the spike in the percentage of circRNAs around 140nt is very similar in circRNAs that lack an ORF to those circRNAs that contain an ORF. The circRNAs that contain an ORF were found using a python script that searched for a consensus ATG start codon anywhere in the sequence. The spike in abundance in circRNAs that did not contain an open reading frame may be explained by non-consensus start codons that cause translation. Using a python script, the most common non-consensus start codons, CTG, GTG, and ACG, were searched for in non ORF circRNAs (2.3.5). The length distribution showed that non-consensus start codons also contained a step up around 140nt. Therefore, this is a strong indication that the step up is of circRNAs around 140nt in length due to a reason other than translation. This step-up has not been reported as a feature of circRNAs before, and it is uncertain why this exists if not due to the minimum threshold size required for translation

The CAI compares the codons in the RNA sequence to all those in the codon usage table and gives a score based upon the commonality that the codons occur in the sequence. The score

ranges from 0-1. A score of 0 demonstrates the rarest codon usage for every codon, and a score of 1 signifies a sequence that contains only the most common codons. The CAI scores revealed that all groups were highly significant from each other ($P < 0.0001$) (Figure 5.3). Considering the vast quantities of CAI scores in each group, 429,414 for both open reading frame through junction and scrambled sequence scores and 99,030 for CDS mRNA, it was expected that even minor differences in the mean CAI scores would result in the observed significance. The results demonstrate that 25% of the ORF through the junction of circRNAs had a greater or equal CAI score to the mean CDS of mRNA, demonstrating the vast quantity of ORF through the circular junction that have the potential to be efficiently translated. Whereas the scrambled sequences 75^{th} percentile (0.73) still did not reach the 25^{th} percentile of CDS mRNAs (0.75), demonstrating that when there is no selection for translation, as with the scrambled sequences, there is a considerably lower CAI distribution.

The distribution pattern between the 25^{th} percentile and 75^{th} percentile was similar in all groups, however the CAI scores for CDS mRNA and ORFs through the circular junction were skewed higher by 0.9 and 0.6, respectively. If most of the ORF through junction sequences were non-biologically relevant, it would be expected that their score distribution would more closely resemble the scrambled sequences rather than the known translated CDS sequences. As observed the ORF through junction sequences more closely resembled the CDS sequences, which may indicate a translation selection for ORF through junction sequences. The CAI scores in the higher percentiles in the ORF through the junction sequences represent the circRNAs that if translated, would have the greatest efficiency, these were very similar in both the ORF through junction sequences with 95^{th} percentiles of 0.85 and 0.86, for ORF through the junction and CDS mRNA sequences, respectively, supporting a selection for translation..

The CAI score is useful when looking at the 4 identified circRNAs ORFs CAI and comparing with the general trend of ORF through the junction circRNAs. The 4 identified circRNAs had a range of CAI values of 0.794-0.831, all above the 75th percentile of ORF through the junction sequences and both the mean and median of CDS mRNA. This indicates a possible selection of the identified circRNAs to be more efficiently translated than the average circRNA with an ORF through the circular junction.

5.3.2 Peptide Analysis

This project searched for candidate circRNAs that translated peptides with a different function from their parent gene. The ORF through the junction circRNAs were identified because the sequence translated through the junction will always be unique to circRNAs. It was important for comparison between linear gene functions that the peptide sequences expressed from parental genes were obtained, and all the protein analyses performed on circRNA peptides were also performed on the parental genes as a control.

When obtaining the ORF circRNAs, EMBOSSgetORF was used after the ATG rearrangement. EMBOSSgetORF allows for a parameter that inputs the nucleotide sequence from a circRNA (2.3.1). Therefore, when translating the circRNA sequence, if the end of the sequence is reached before a stop codon, translation cycles back to the start of the sequence, as would be the case in circRNA translation. This continues until 1) A stop codon is reached or 2) The circRNA has an infinite open reading frame, in which case the sequence is repeated 3 times, revealing the 3 potential frames of translation (if the sequence length is not a multiple of 3) or maintaining the same frame (if the sequence length is a multiple of 3). It is important that EMBOSSgetORF loops translation at the end of the sequence, as the circular junction can a

frameshift resulting in a different amino acid sequence being translated compared to the first translation cycle around the circRNA (Figure 5.1).

The focus was then shifted to whether any predicted circRNA peptides could carry a different function to the parental gene by looking at the predicted biological functions of the protein domains within the peptides. The software used for this analysis was InterproScan, and the database for peptide analysis was PANTHER. PANTHER contains many protein families divided by function into related subfamilies. PANTHER focuses greater on the functional relationship than other databases and diverges specific functions within protein families (Mi et al., 2019). The functional identification of unique sequences is acquired using Hidden Markov models (HMMs), built for each family and subfamily. PANTHER was used over other peptide domain databases because it considers large amounts of the peptide sequence, returning a prediction of function in a single result. Therefore, the PANTHER term for the parental gene and the circRNA peptides can be directly compared, revealing peptides of interest. Not all peptides run through InterproScan are given a PANTHER term. For the circRNA peptides, only 2291 circRNAs were given a PANTHER term from 378,626 circRNA peptide sequences. It is expected that the vast majority do not have a predicted function, as all the ORF that run through the junction are used as the input without considering the actual likelihood that they are translated and the number of circRNAs that have an ATG before the backsplice junction by chance, without biological significance.

The most prominent function identified in endogenous circRNAs is non-coding regulatory RNAs (ncRNAs), regulating transcription and translation. CircRNAs can directly associate with RNA Polymerase II (RNA Pol II) to stimulate transcription, act as miRNA sponges to prevent miRNA regulation and act as protein decoys by sequestering proteins (Zhang et al.,

2013; Hansen et al., 2013). CircRNAs with an infinite open reading frame translate a repeating peptide concatemer. Considering circRNAs many known regulatory functions and the relatively recent discovery of endogenous circRNA translation, it is likely that regulation would occur from circRNA translated peptides. A peptide concatemer would be an effective regulator, containing multiple binding sites for the same protein, giving the potential for effective protein sequestering. 1 of the 4 identified circRNAs (LTBP1_0013) has an IORF. The InterProScan prediction was Laminin G-Like Domain-Containing Protein. Laminin G is around 180 amino acid long domain found in a large and diverse set of extracellular proteins, critical for heparin-binding and cell attachment (Okazaki et al., 2002). Phyre also predicted a protein with an extracellular function, low-density lipoprotein receptor-related protein 4. LTBP1_0013 peptide lacks an N-terminus signal sequence for exportation. Therefore it is unknown what function the peptide would have inside cells and does not support the likelihood of LTBP1_0013 translation.

Not all the peptide predictions are for human proteins. Phyre predicted peptide for CELF5_0014 circRNA is *Plasmodium falciparum* circumsporozoite protein (CSP). Ordinarily, this would be evidence that this peptide is not translated in CELF5_0014 circRNA. However, CSP is functional in humans during *Plasmodium falciparum* infection. CSP facilitates *Plasmodium* entry into cells by binding to hepatocytes (Rathore, Sacci, de la Vega and McCutchan, 2002). However, the CELF5_0014 peptide does not contain a signal sequence for exportation, making it unlikely that the circRNA is translated as it is unlikely to have a biological function inside cells.

The peptide prediction RIMS2_0163 circRNA was 40S Ribosomal Protein S6 for Phyre and InterProScan. Ribosomal protein S6 is a component of the 40S ribosomal subunit. It is

conserved across all eukaryotes and is fundamental to translation. Its deficiency results in p53-dependent cell-cycle arrest (Panić, Montagne, Cokarić and Volarević, 2007). Various protein kinases can phosphorylate ribosomal protein S6 via five C-terminal serine residues (Ruvinsky et al., 2005). Phosphorylated ribosomal protein S6 controls translation of a subset of mRNAs but phosphorylation is not essential for overall protein synthesis (Puighermanal et al., 2017). The RIMS2_0163 peptide does not contain the C-terminal serine residues for phosphorylation. During growth and proliferation, a substantial amount of the total energy resources are allocated to synthesising new ribosomes (Warner, 1999). Given that RIMS2_0163 has only been detected in the testis tissue, a site of constant proliferation due to spermatogenesis, this further supports that RIMS2_0163 produces Ribosomal protein S6 peptide.

The peptide prediction SFRP4_0005 circRNA was Ependymin-related protein (EPDR) 1 or 2 for InterproScan and Phyre. EPDRs are conserved across species, including fish, mice and humans, although details of their function are unknown (Suárez-Castillo and García-Arrarás, 2006). EPDR1 binds to hydrophobic molecules like fatty acids and interacts with insulin-like growth factor 2 and flotillin proteins, known to be involved in protein and vesicle translocation (Park et al., 2019). The SFRP4_0005 expression has been detected in a diverse range of tissues. Even though the function is unknown, the high level of conservation of this Ependymin-related protein and the SFRP4_0005 circRNAs diverse expression may support SFRP4_0005 translation and its biological importance.

The secondary structure of the circRNA peptides was analysed using Phyre2; this software predicts a potential function of a peptide based on its predicted structure similarity to known peptides. Using two different approaches, InterproScan for primary structure and Phyre for the

secondary structure analysis, provides validation of the same or similar results (RIMS2_0162 and SFRP4_0005) (Table 5.5).

5.3.3 CircRNA Translatability

This Bioinformatics approach has identified circRNAs that have the coding potential to function differently from their parent gene. Translatability was assumed for the identification of these circRNAs. This was then investigated post identification. SRAMP was used to consider the predicted m^6A methylation sites, considering the secondary structure. SRAMP finds all the methylation sites within the circRNA transcript and returns a score for each methylation site based on the likelihood of methylation, as not all m^6A sites on RNA transcripts are methylated. This score is determined by the primary sequence of the methylation site and the % of false positives that occur after cross-validation testing (Zhou, Zeng, YH, Zhang and Cui., 2016; Wang and Gribskov, 2019). A 15% false-positive rate returned a low confidence m^6A site, 10% moderate, 5% high confidence, and 1% very high confidence. There is currently no computational prediction to calculate the likelihood that an m^6A site recruits the ribosome. Little is understood why some m^6A sites are capable of ribosomal recruitment, and others lack this ability. An estimated 10% of cellular and viral mRNA use IRES have a functional IRES capable of initiating translation (Stoneley and Willis, 2004). IRES-mediated translation is difficult to predict computationally due to the lack of primary/secondary/tertiary structure across known IRES sequences. IRESpy uses a machine learning strategy for IRES prediction independent of sequence length using global kmer and structural feature analysis (Wang and Gribskov, 2019. IRESpy predicted only CELF5_0014 circRNA to contain an IRES with a threshold of 0.1. However, although IRESpy is the most accurate computational method for IRES prediction, IRES activity still requires experimental determination.

This bioinformatics approach used very stringent filters to identify 4 candidate circRNAs that have the coding potential to produce a peptide with a different biological function than their parent gene. Due to the stringency of the filters, this is likely a vast underestimation of circRNAs with unique functioning peptides. However, this approach was implemented to identify a small enough pool of candidate circRNAs, for the experimental investigation of translatability, peptide structure, and function.

Chapter Six

6.0 Discussion

This project was undertaken to produce a mastoparan peptide concatemer from circRNA for cancer therapy. A system was constructed for RNA circularisation into IORF circRNA and translation of FRET and mastoparan peptide concatemers. FRET peptide production and activity were assessed by western blot analysis and fluorescence. Mastoparan production and activity was assessed by western blot analysis and its effect on MDA-MB-231 cells and MCF-7 cells.

Another focus of this project was identifying candidate circRNAs that produce a peptide with a different function to that from the linear transcript of their parent gene using a bioinformatics approach. First, circRNAs were sorted from the circAtlas database into those that contain an ORF through the circular junction. The functional potential of these ORTJ peptides was assessed using InterProScan and Phyre. Differences in peptide predictions identified candidate circRNAs. Finally, translatability of the candidate circRNAs was explored with SRAMP and IRESpy, predicting m^6A methylation and IRES sites.

6.1 RNA Circularisation

The first aim of this project was to produce an RNA circularisation system capable of RNA circularisation *in vitro* and *in vivo.* The results show that engineered PIE *Anabaena* introns can self-splice *in vitro*. This supports earlier research in which an engineered *Anabaena* intron sequence was used to circularise RNA *in vitro* to near 100% efficiency (Wesselhoeft, Kowalski and Anderson, 2018). This high-efficiency rate was not reached in this project, potentially due to slight adaptations to the *Anabaena* sequence required to design a circRNA with an infinite open reading frame (IORF).

In vivo circularisation was not achieved in mammalian cells, demonstrating that *Anabaena* introns should not be used for RNA circularisation in mammalian cells. Group I introns of *Tetrahymena* had previously been shown to self-splice in mammalian cells, albeit with low efficiency (Long and Sullenger, 1999). Therefore, it can be proposed that different group I introns have different splicing capacities in cells, likely due to cell conditions impacting the folding capability of the catalytic structure. Using the *Tetrahymena* introns in an engineered approach to generating a splicing bubble like the *Anabaena* introns used in this project may increase the *Tetrahymena* circularisation efficiency in cells by bringing the splice sites closer in proximity.

Given the complexity of group II introns and the inability or low efficiency of group I introns to self-splice in mammalian cells, a backsplicing approach can be adopted for circRNA formation in mammalian cells, utilising the native method of circRNA formation in humans. Meganck et al., 2021 achieved synthetic circularisation and translation of RNA in cell culture conditions and mice.

6.2 Mastoparan and FRET Translation

This project aimed to express a mastoparan and FRET peptide concatemer inside cells, which would be exported out of the cell and activated by MMP cleavage in the extracellular environment. The success of this system would result in 1) A detectable peptide concatemer inside cells and 2) A detectable cleaved peptide in the extracellular environment.

FRET detection was measured by an anti-GFP western blot and measuring fluorescence. Neither fluorescence nor detection by western blot was successful for the Linear FRET or *Anabaena* FRET constructs after purification from cells or in the culture medium. Linear FRET and *Anabaena* FRET peptides were detected on a western after *in vitro* translation. However,

fluorescence was not observed. This indicates that the FRET peptide was not functional. EGFP and mCherry fluorophores have been successfully used in solution and live cells (Albertazzi et al., 2009). It is unknown why the peptide was detectable by western but not fluorescent.

Anabaena Mastoparan Flag and *Anabaena* Mastoparan Flag IORF peptides were detected from unspliced RNA in cells but not in the culture medium. The *Anabaena* Mastoparan Flag peptide was detected with a western blot after *in vitro* translation. As described previously, the RNA circularisation was not achieved in cells. *Anabaena* Mastoparan Flag and *Anabaena* Mastoparan Flag IORF peptides were not expected to be present in the culture medium as translation begins from a start codon upstream of the signal peptide.

The problems of no/low expression may result from *Anabaena* intron translation which contains a non-optimal codons at both the 5' and 3' end. Although circularisation was not achieved, the difficulty of translating the linear RNA would also occur with circRNA translation, likely to an even greater degree of difficulty. IORF circRNA translation of stretches of non-optimal codons would result in ribosomal stalling, queuing and potentially abolishment of translation (Yang et al., 2019).

This finding supports alternative strategies of IORF circRNA translation, using methods that do not require native exon inclusion in the circRNA sequence like group I introns. As previously mentioned in this chapter, inverse group II or a backsplicing approach should be used to circularise RNA with no native exon inclusion.

6.3 Mastoparan Peptide Concatemer Cancer Therapeutics

The main aim of this project was to produce a mastoparan concatemer prodrug for cancer therapeutics, in which mastoparan would be activated in the extracellular environment by MMP cleavage, causing cancer cell cytotoxicity. Overall, the results are not consistent, making it difficult to understand how each construct is functioning biologically.

The Linear Mastoparan control induced significant cell death upon transfection in MDA-MB-231 cells, and after treatment of cell lysate purified Linear Mastoparan peptide. This indicates that the mastoparan production in cells, export and activation in the culture medium was successfully occurring. However, treating cells with the culture medium did not result in cell death. An epitope Flag should be added to this construct for detection to better understand the discrepancies in the results.

Anabaena Mastoparan transfection resulted in no significant cell death, but adding purified *Anabaena* Mastoparan peptide from cell lysate and concentrated culture medium of *Anabaena* Mastoparan transfected cells did. It was expected that transfection and culture medium treatments results would be the same, given that the mastoparan prodrug mechanism of action involves activation in the culture medium.

The clearest result was treatment of cells with *Anabaena* Mastoparan cell lysate. It caused significant cell death, and the presence of cell extracted *Anabaena* Mastoparan Flag peptide was confirmed by western blot. Treatment with cell extracted GFP protein also caused cell death, which was unexpected, but only at a concentration 3x higher than *Anabaena* Mastoparan. *Anabaena* Mastoparan peptides purified from the cell lysate causing cell death is enough to determine that a mastoparan prodrug does result in cancer cell death. However, the

desired system of mastoparan prodrug translation in cells, export and activation in the media requires more work. Due to greater levels of MMP activity in MDA-MB-231 to MCF-7 cells, it was expected that the mastoparan concatemer would have greater cytotoxicity in MDA-MB-231 cells; however, both Linear Mastoparan and *Anabaena* Mastoparan exhibited similar MTT assay results in both cell lines.

Future work should focus on peptide detection and MTT assays to gain a better understanding of what it is occurring biologically. This could be achieved by repeat experiments, obtaining higher peptide expression using an alternative RNA circularisation method and optimising the western blot for positive control detection.

Utilising mastoparan as a prodrug introduces a double layer of specificity to target cancer cells. Firstly, by linking each mastoparan peptide in a concatemer with an MMP cleavage site, the prodrug is cleaved into its individual functional mastoparan units exclusively in the presence of activated MMPs in the extracellular environment. MMP overexpression and increased activity is frequently observed in highly aggressive and invasive cancers such as triple negative breast cancer (Egeblad and Werb, 2002), ensuring that peptide activation occurs solely in the tumour microenvironment. Furthermore, mastoparan features three lysine residues, which confer a net positive charge to the peptide (Le et al., 2019). Cancer cells exhibit elevated glycolysis levels, leading to up to 30-fold greater glucose uptake and lactate secretion compared to normal cells (Chen et al., 2016). The secretion of lactate anions causes a loss of cations from the cancer cell surface and leaving behind a negative charge (Chen et al., 2016). The combination of the positive charge of mastoparan and the negative charge on the surface of cancer cells generates an electrostatic attraction, thus enhancing mastoparan's specificity to cancer cells.

The activation of mastoparan in the tumour microenvironment and the electrostatic attraction of mastoparan to cancer cells provide two layers of specificity. By minimising off-target effects to healthy cells and thus drug toxicity, this approach could significantly improve patient's quality of life during treatment by limiting side effects compared to current treatments. This project demonstrates the prodrug capabilities of a mastoparan concatemer. Transfection of *Anabaena* Mastoparan and Linear Mastoparan constructs caused minimal cell death, whereas purification of the *Anabaena* Mastoparan transfected peptides from cells and addition to the culture medium led to significant cell death. This indicates that transfection resulted in the production of inactive mastoparan, then when purified and added to the culture medium with MMPs in the extracellular environment, the concatemer was activated by cleavage, causing cytotoxicity.

A benefit of producing a mastoparan peptide prodrug from self-spliced circRNA, is that the self-splicing and peptide production occurs in the host cells. The manufacturing cost to produce drugs on a large scale is expensive, therefore, utilising the host’s cellular machinery to produce the mastoparan prodrug lowers the cost of treatment, providing more accessibility to patients. In this project, although mastoparan was produced inside cells, as confirmed by western blot, this did not lead to cytotoxicity, likely as the mastoparan prodrug was not exported via the signal peptide for activation in the extracellular environment.

6.4 Novel Peptides from CircRNA Translation

Another aim of this project was to identify circRNAs with the coding potential to produce a peptide with a different biological function to their parent gene. This project identified 4 candidate circRNAs that fit this aim, corroborated using two programmes (InterproScan and Phyre) for functional peptide prediction.

These CAI score distributions found that a large proportion of ORF through the junction sequences of circRNAs had the coding potential to be efficiently translated. To investigate this further, and explore the likelihood of translation, the presence of predicted m^6a methylation or IRES sites in circRNAs with a CAI score above the 75^{th} percentile could be compared to CDS mRNA and the circRNAs lower than the 25^{th} percentile to see if there is a correlation between increased CAI score and the likelihood of translation.

The sorting steps to identify candidate circRNAs were very stringent, as the approach was to identify a small enough pool for future experimental analysis. This workflow likely underestimates the number of circRNAs with different coding potential to their parent gene.

The most interesting finding, given the vast effects the peptide may have and its well-characterised functions, is RIMS2_0163 circRNA which produced a 61 aa long peptide with a prediction of 40S Ribosomal Protein S6. This would be the priority for further investigation experimentally.

Continuation of this work would involve identifying whether the candidate circRNAs are translatable. To achieve this, a reporter gene should be inserted into the circRNA sequences in the same frame as the ORF of interest to detect internal translation initiation. If peptide translation is confirmed, then its 3D structure can be confirmed with NMR or X-ray crystallography. A similar actual structure to the structure of the predicted peptide would support that it has a similar function. Additionally, proteomics datasets could be interrogated

to identify proteins that could only be derived from backspliced mRNAs. Work towards this is ongoing in the group.

circRNAs can be overexpressed in cancer, and peptides translated from circRNAs have been identified to contribute to the pathogenesis of the disease (Patop and Kadener, 2018; Duan et al., 2022). Future work of interest is identifying circRNAs with the coding potential to produce a peptide with a different biological function to their parent gene in cancer. CircRNAs can be identified from RNA-seq cancer tissue data using circRNA identifier algorithms such as CIR2 (Gao, Wang and Zhao, 2015). once the circRNA sequences are obtained, the workflow will remain the same as in this project.

6.5 Outlook and Conclusion

This research project hypothesised that the production of a mastoparan peptide concatemer prodrug from circRNA is effective in cancer therapeutics. In this system, a mastoparan peptide concatemer is produced inside cells from circRNA, exported out of the cell and subsequently activated by MMP cleavage, causing cytotoxicity to cancer cells. The results determined that *Anabaena* introns are not inappropriate for RNA circularisation and peptide concatemer production in mammalian cells. The mastoparan prodrug purified from cell lysate demonstrated cytotoxicity to both MDA-MB-231 and MCF-7 cells, supporting the use of mastoparan oligomers as an effective prodrug. For the proposed system to function correctly, it is clear that alternative means of circRNA production are necessary for mammalian cells, with the best candidate being a backsplicing approach. The mastoparan prodrug designed in this study provides two stages of specificity. This project evaluates the use of one of these stages, prodrug activation by MMP cleavage. The specificity to cancer cells caused by the electrostatic

attraction of mastoparan produced inside cells warrants investigation. The non-malignant cell line MCF-10A cells are commonly used to model the function of normal breast tissue, these cells could be used to assess the cytotoxicity of mastoparan to normal cells compared to breast cancer cells. As MMP overexpression and increased activity is a feature of many cancers, particularly those that are highly invasive and aggressive, and all cancer cells carry a negative surface charge and, therefore, mastoparan specificity. Given the cytotoxicity of the mastoparan prodrug to breast cancer cell lines demonstrated in this study, the use of mastoparan as a prodrug for the treatment of other cancer types should be investigated.

Another aim was to identify candidate circRNAs that have the coding potential to produce a peptide with a different biological function to their parent gene. The results demonstrate that this is the case, with 4 candidate circRNAs identified. The bioinformatics approach identified circRNAs that have 1) An ORF through the backsplice junction, 2) Are of an appropriate size to occupy the ribosome, 3) Have a codon adaption index (CAI) score > 0.7, 4) Contain a predicted IRES or m^6A site for internal translation initiation. The peptides translated from the candidate circRNAs ORFs also had distinct characteristics, including 1) A Different Interproscan and Phyre result to the parent gene, 2) Low homology to the parent protein, 3) Translation ≥ 10 amino acids through each side of the backsplice junction. This project laid the foundation for future experimental research identifying whether the candidate circRNA ORFs are translated in cells and whether the peptide produced has a similar function to the Phyre and Interpro predictions. The study offers a comprehensive workflow and a set of Python scripts that are openly accessible in a GitHub repository to identify circRNAs that possess the coding potential to produce a peptide with a different biological function compared to their parent gene. This approach can be effectively replicated for the analysis of circRNAs in pathological conditions. Given that circRNAs have been linked to cancer progression and their abundance

is known to increase in cancer, it is intriguing to apply this workflow to identify cancer specific circRNA peptides that may have a novel function different from their parent gene.

7.0 References

1. Abe, N., Matsumoto, K., Nishihara, M., Nakano, Y., Shibata, A., Maruyama, H., Shuto, S., Matsuda, A., Yoshida, M., Ito, Y. and Abe, H. (2015) Rolling Circle Translation of Circular RNA in Living Human Cells. *Scientific Reports*, 5(1).
2. Adan, A., Kiraz, Y. and Baran, Y. (2016) Cell Proliferation and Cytotoxicity Assays. *Current Pharmaceutical Biotechnology*, 17(14), pp.1213-1221.
3. Ahmad, A. (2013) Pathways to breast cancer recurrence. *ISRN Oncology*, *2013*, 1–16.
4. Aitken, C. and Lorsch, J. (2012) A mechanistic overview of translation initiation in eukaryotes. *Nature Structural & Molecular Biology*, 19(6), pp.568-576.
5. Al-Alem, L. and Curry, T. (2015) Ovarian cancer: involvement of the matrix metalloproteinases. *Reproduction*, 150(2), pp.R55-R64.
6. Albertazzi, L., Arosio, D., Marchetti, L., Ricci, F. and Beltram, F. (2009) Quantitative FRET Analysis With the E^0GFP-mCherry Fluorescent Protein Pair. *Photochemistry and Photobiology*, 85(1), pp.287-297.
7. Alhakamy, N., Ahmed, O., Md, S. and Fahmy, U. (2021) Mastoparan, a Peptide Toxin from Wasp Venom Conjugated Fluvastatin Nanocomplex for Suppression of Lung Cancer Cell Growth. *Polymers*, 13(23), p.4225.
8. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *Journal of molecular biology*, *215*(3), pp.403-410.
9. American Cancer Society. (2021, March 25) *Breast Cancer Types*. Available at: https://www.cancer.org/cancer/breast-cancer/understanding-a-breast-cancer-diagnosis/types-of-breast-cancer.html (Accessed: 18th February 2023).
10. Arnold, M., Morgan, E., Rumgay, H., Mafra, A., Singh, D., Laversanne, M., Vignat, J., Gralow, J. R., Cardoso, F., Siesling, S., & Soerjomataram, I. (2022) Current and future burden of breast cancer: Global Statistics for 2020 and 2040. *The Breast*, *66*, 15–23.
11. Ashwal-Fluss, R., Meyer, M., Pamudurti, N., Ivanov, A., Bartok, O., Hanan, M., Evantal, N., Memczak, S., Rajewsky, N. and Kadener, S. (2014) circRNA Biogenesis Competes with Pre-mRNA Splicing. *Molecular Cell*, 56(1), pp.55-66.
12. Auclair, S., Bhanu, M. and Kendall, D. (2011) Signal peptidase I: Cleaving the way to mature proteins. *Protein Science*, 21(1), pp.13-25.

13. Bain, J. and Switzer, C. (1992) Regioselective ligation of oligoribonucleotides using DNA splints. *Nucleic Acids Research*, 20(16), pp.4372-4372.
14. Baird, S.D., Lewis, S.M., Turcotte, M. and Holcik, M. (2007) A search for structurally similar cellular internal ribosome entry sites. *Nucleic acids research*, *35*(14), pp.4664-4677.
15. Bajar, B., Wang, E., Zhang, S., Lin, M. and Chu, J. (2016) A Guide to Fluorescent Protein FRET Pairs. *Sensors*, 16(9), p.1488.
16. Blum, M., Chang, H.Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S. and Richardson, L. (2021) The InterPro protein families and domains database: 20 years on. *Nucleic acids research*, *49*(D1), pp.D344-D354.
17. Bordeleau, M., Matthews, J., Wojnar, J., Lindqvist, L., Novac, O., Jankowsky, E., Sonenberg, N., Northcote, P., Teesdale-Spittle, P. and Pelletier, J. (2005) Stimulation of mammalian translation initiation factor eIF4A activity by a small molecule inhibitor of eukaryotic translation. *Proceedings of the National Academy of Sciences*, 102(30), pp.10460-10465.
18. Bou Zerdan, M., Ghorayeb, T., Saliba, F., Allam, S., Bou Zerdan, M., Yaghi, M., Bilani, N., Jaafar, R. and Nahleh, Z. (2022) Triple negative breast cancer: Updates on classification and treatment in 2021. *Cancers*, *14*(5), p.1253.
19. Burguin, A., Diorio, C., and Durocher, F. (2021) Breast cancer treatments: Updates and new challenges. *Journal of Personalized Medicine*, *11*(8), 808.
20. Buttgereit, F. and Brand, M. (1995) A hierarchy of ATP-consuming processes in mammalian cells. *Biochemical Journal*, 312(1), pp.163-167.
21. Chang, G., Shi, L., Ye, Y., Shi, H., Zeng, L., Tiwary, S., Huse, J. T., Huo, L., Ma, L., Ma, Y., Zhang, S., Zhu, J., Xie, V., Li, P., Han, L., He, C., and Huang, S. (2020) YTHDF3 induces the translation of M6A-enriched gene transcripts to promote breast cancer brain metastasis. *Cancer Cell*, *38*(6).
22. Chao, C., Chan, D., Kuo, A. and Leder, P. (1998) The Mouse formin (Fmn) Gene: Abundant Circular RNA Transcripts and Gene-Targeted Deletion Analysis. *Molecular Medicine*, 4(9), pp.614-628.
23. Chavez, K. J., Garimella, S. V., and Lipkowitz, S. (2011) Triple negative breast cancer cell lines: One tool in the search for better treatment of triple negative breast cancer. *Breast Disease*, *32*(1-2), 35–48.

24. Chen, B., Le, W., Wang, Y., Li, Z., Wang, D., Lin, L., Cui, S., Hu, J., Hu, Y., Yang, P., Ewing, R., Shi, D. and Cui, Z. (2016) Targeting Negative Surface Charges of Cancer Cells by Multifunctional Nanoprobes. *Theranostics*, 6(11), pp.1887-1898.
25. Chen, N., Zhao, G., Yan, X., Lv, Z., Yin, H., Zhang, S., Song, W., Li, X., Li, L., Du, Z., Jia, L., Zhou, L., Li, W., Hoffman, A., Hu, J. and Cui, J. (2018) A novel FLI1 exonic circular RNA promotes metastasis in breast cancer by coordinately regulating TET1 and DNMT1. *Genome Biology*, 19(1).
26. Chen, X. and Lu, Y. (2021) Circular RNA: Biosynthesis in vitro. *Frontiers in Bioengineering and Biotechnology*, 9.
27. Clancy, S. (2008) RNA splicing: introns, exons and spliceosome. *Nature Education*, *1*(1), p.31.
28. Compton, C. (2020) Cancer initiation, promotion, and progression and the acquisition of key behavioral traits. *Cancer: The Enemy from Within*, pp. 25–48.
29. Comşa, Ş., Cimpean, A. M., & Raica, M. (2015) The story of MCF-7 breast cancer cell line: 40 years of experience in research. *Anticancer research*, *35*(6), 3147-3154.
30. Coolidge, C., Seely, R. and Patton, J. (1997) Functional analysis of the polypyrimidine tract in pre-mRNA splicing. *Nucleic Acids Research*, 25(4), pp.888-896.
31. Cooper, G. and Hausman, R. (2000) *The Cell: A Molecular Approach*. 2nd ed. Washington, D.C.: ASM Press.
32. Côté, F., Lévesque, D. and Perreault, J. (2001) Natural 2′,5′-Phosphodiester Bonds Found at the Ligation Sites of Peach Latent Mosaic Viroid. *Journal of Virology*, 75(1), pp.19-25.
33. Créancier, L., Morello, D., Mercier, P. and Prats, A. (2000) Fibroblast Growth Factor 2 Internal Ribosome Entry Site (Ires) Activity Ex Vivo and in Transgenic Mice Reveals a Stringent Tissue-Specific Regulation. *Journal of Cell Biology*, 150(1), pp.275-281.
34. Csepany, T., Lin, A., Baldick, C. and Beemon, K. (1990) Sequence specificity of mRNA N6-adenosine methyltransferase. *Journal of Biological Chemistry*, 265(33), pp.20117-20122.
35. Cunningham, F., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M., Armean, I., Austine-Orimoloye, O., Azov, A., Barnes, I., Bennett, R., Berry, A., Bhai, J., Bignell, A., Billis, K., Boddu, S., Brooks, L., Charkhchi, M., Cummins, C., Da Rin Fioretto, L., Davidson, C., Dodiya, K., Donaldson, S., El Houdaigui, B., El Naboulsi, T., Fatima, R., Giron, C., Genez, T., Martinez, J., Guijarro-Clarke, C., Gymer, A., Hardy, M., Hollis, Z., Hourlier, T., Hunt, T., Juettemann, T., Kaikala, V., Kay, M., Lavidas, I., Le,

T., Lemos, D., Marugán, J., Mohanan, S., Mushtaq, A., Naven, M., Ogeh, D., Parker, A., Parton, A., Perry, M., Piližota, I., Prosovetskaia, I., Sakthivel, M., Salam, A., Schmitt, B., Schuilenburg, H., Sheppard, D., Pérez-Silva, J., Stark, W., Steed, E., Sutinen, K., Sukumaran, R., Sumathipala, D., Suner, M., Szpak, M., Thormann, A., Tricomi, F., Urbina-Gómez, D., Veidenberg, A., Walsh, T., Walts, B., Willhoft, N., Winterbottom, A., Wass, E., Chakiachvili, M., Flint, B., Frankish, A., Giorgetti, S., Haggerty, L., Hunt, S., IIsley, G., Loveland, J., Martin, F., Moore, B., Mudge, J., Muffato, M., Perry, E., Ruffier, M., Tate, J., Thybert, D., Trevanion, S., Dyer, S., Harrison, P., Howe, K., Yates, A., Zerbino, D. and Flicek, P. (2021) Ensembl 2022. *Nucleic Acids Research*, 50(D1), pp.D988-D995.

36. Cunningham, T., Chapman, E. and Schatz, J. (2018) eIF4A inhibition: ready for primetime?. *Oncotarget*, 9(85), pp.35515-35516.

37. de Azevedo, R.A., Figueiredo, C.R., Ferreira, A.K., Matsuo, A.L., Massaoka, M.H., Girola, N., Auada, A.V., Farias, C.F., Pasqualoto, K.F., Rodrigues, C.P. and Barbuto, J.A. (2015) Mastoparan induces apoptosis in B16F10-Nex2 melanoma cells via the intrinsic mitochondrial pathway and displays antitumor activity in vivo. Peptides, 68, pp.113-119.Nex2 melanoma cells via the intrinsic mitochondrial pathway and displays antitumor activity in vivo. *Peptides*, 68, pp.113-119.

38. Derakhshani, A., Rezaei, Z., Safarpour, H., Sabri, M., Mir, A., Sanati, M. A., Vahidian, F., Gholamiyan Moghadam, A., Aghadoukht, A., Hajiasgharzadeh, K., and Baradaran, B. (2019) Overcoming trastuzumab resistance in her2-positive breast cancer using combination therapy. *Journal of Cellular Physiology*, *235*(4), 3142–3156.

39. Dever, T. and Green, R. (2012) The Elongation, Termination, and Recycling Phases of Translation in Eukaryotes. *Cold Spring Harbor Perspectives in Biology*, 4(7), pp.a013706-a013706.

40. Ding, X., Zhang, S., Li, X., Feng, C., Huang, Q., Wang, S., Wang, S., Xia, W., Yang, F., Yin, R., Xu, L., Qiu, M., Li, M. and Wang, J. (2018) Profiling expression of coding genes, long noncoding RNA, and circular RNA in lung adenocarcinoma by ribosomal RNA-depleted RNA sequencing. *FEBS Open Bio*, 8(4), pp.544-555.

41. Duan, J., Chen, W., Xie, J., Zhang, M., Nie, R., Liang, H., Mei, J., Han, K., Xiang, Z., Wang, F., Teng, K., Chen, R., Deng, M., Yin, Y., Zhang, N., Xie, D. and Cai, M. (2022) A novel peptide encoded by N6-methyladenosine modified circMAP3K4 prevents apoptosis in hepatocellular carcinoma. *Molecular Cancer*, 21(1).

42. Ebbesen, K., Hansen, T. and Kjems, J. (2017) Insights into circular RNA biology. *RNA Biology*, 14(8), pp.1035-1045.
43. Egeblad, M., & Werb, Z. (2002) New functions for the matrix metalloproteinases in cancer progression. *Nature Reviews Cancer*, *2*(3), pp.161–174.
44. Egeblad, M., and Werb, Z. (2002). New functions for the matrix metalloproteinases in cancer progression. *Nature Reviews Cancer*, *2*(3), 161–174.
45. El Sayed, R., El Jamal, L., El Iskandarani, S., Kort, J., Abdel Salam, M., and Assi, H. (2019) Endocrine and targeted therapy for hormone-receptor-positive, HER2-negative advanced breast cancer: Insights to sequencing treatment and overcoming resistance based on clinical trials. *Frontiers in Oncology*, *9*.
46. Fares, J., Fares, M. Y., Khachfe, H. H., Salhab, H. A., and Fares, Y. (2020) Molecular principles of metastasis: a hallmark of cancer revisited. *Signal transduction and targeted therapy*, *5*(1), 28.
47. Fidler, I. J. (2003) The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. *Nature Reviews Cancer*, 3(6), 453-458.
48. Figueira, R., Gomes, L., Neto, J., Silva, F., Silva, I. and Sogayar, M. (2009) Correlation between MMPs and their inhibitors in breast cancer tumor tissue specimens and in cell lines with different metastatic potential. *BMC Cancer*, 9(1).
49. Foda, H. and Zucker, S. (2001) Matrix metalloproteinases in cancer invasion, metastasis and angiogenesis. *Drug Discovery Today*, 6(9), pp.478-482.
50. Ford, E. and Ares, M. (1994) Synthesis of circular RNA in bacteria and yeast using RNA cyclase ribozymes derived from a group I intron of phage T4. *Proceedings of the National Academy of Sciences*, 91(8), pp.3117-3121.
51. Fumagalli, C., and Barberis, M. (2021) Breast cancer heterogeneity. *Diagnostics*, *11*(9), 1555.
52. Galicia-Vázquez, G., Cencic, R., Robert, F., Agenor, A. and Pelletier, J. (2012) A cellular response linking eIF4AI activity to eIF4AII transcription. *RNA*, 18(7), pp.1373-1384.
53. Galicia-Vázquez, G., Chu, J. and Pelletier, J. (2015) eIF4AII is dispensable for miRNA-mediated gene silencing. *RNA*, 21(10), pp.1826-1833.
54. Gao, Y., Wang, J. and Zhao, F. (2015) CIRI: an efficient and unbiased algorithm for de novo circular RNA identification. *Genome Biology*, 16(1).

55. Ge, Y., Draycheva, A., Bornemann, T., Rodnina, M. and Wintermeyer, W. (2014) Lateral opening of the bacterial translocon on ribosome binding and signal peptide insertion. *Nature Communications*, 5(1).

56. Geng, X., Jia, Y., Zhang, Y., Shi, L., Li, Q., Zang, A. and Wang, H. (2020) Circular RNA: biogenesis, degradation, functions and potential roles in mediating resistance to anticarcinogens. *Epigenomics*, 12(3), pp.267-283.

57. Gest, C., Joimel, U., Huang, L., Pritchard, L., Petit, A., Dulong, C., Buquet, C., Hu, C., Mirshahi, P., Laurent, M., Fauvel-Lafève, F., Cazin, L., Vannier, J., Lu, H., Soria, J., Li, H., Varin, R. and Soria, C. (2013) Rac3 induces a molecular pathway triggering breast cancer cell aggressiveness: differences in MDA-MB-231 and MCF-7 breast cancer cell lines. *BMC Cancer*, 13(1).

58. Ghazalpour, A., Bennett, B., Petyuk, V., Orozco, L., Hagopian, R., Mungrue, I., Farber, C., Sinsheimer, J., Kang, H., Furlotte, N., Park, C., Wen, P., Brewer, H., Weitz, K., Camp, D., Pan, C., Yordanova, R., Neuhaus, I., Tilford, C., Siemers, N., Gargalovic, P., Eskin, E., Kirchgessner, T., Smith, D., Smith, R. and Lusis, A. (2011) Comparative Analysis of Proteome and Transcriptome Variation in Mouse. *PLoS Genetics*, 7(6), p.e1001393.

59. Gibson, D., Smith, H., Hutchison, C., Venter, J. and Merryman, C. (2010) Chemical synthesis of the mouse mitochondrial genome. *Nature Methods*, 7(11), pp.901-903.

60. Gobin, E., Bagwell, K., Wagner, J., Mysona, D., Sandirasegarane, S., Smith, N., Bai, S., Sharma, A., Schleifer, R. and She, J. (2019) A pan-cancer perspective of matrix metalloproteases (MMP) gene expression profile and their diagnostic/prognostic potential. *BMC Cancer*, 19(1).

61. Godet, A., David, F., Hantelys, F., Tatin, F., Lacazette, E., Garmy-Susini, B. and Prats, A. (2019) IRES Trans-Acting Factors, Key Actors of the Stress Response. *International Journal of Molecular Sciences*, 20(4), p.924.

62. Grizzi, F., Di Ieva, A., Russo, C., Frezza, E. E., Cobos, E., Muzzio, P. C., and Chiriva-Internati, M. (2006) Cancer initiation and progression: an unsimplifiable complexity. *Theoretical biology & medical modelling*, *3*, 37.

63. Hadler-Olsen, E., Fadnes, B., Sylte, I., Uhlin-Hansen, L. and Winberg, J. (2010) Regulation of matrix metalloproteinase activity in health and disease. *FEBS Journal*, 278(1), pp.28-45.

64. Hammerling, M., Krüger, A. and Jewett, M. (2019) Strategies for in vitro engineering of the translation machinery. *Nucleic Acids Research*, 48(3), pp.1068-1083.

65. Han, B., Wei, S., Li, F., Zhang, J., Li, Z., and Gao, X. (2021) Decoding M6A mRNA methylation by reader proteins in cancer. *Cancer Letters*, *518*, 256–265.
66. Hanahan, D., and Weinberg, R.A. (2011) Hallmarks of cancer: The next generation. *Cell*, 144(5), pp. 646–674.
67. Hansen, T., Jensen, T., Clausen, B., Bramsen, J., Finsen, B., Damgaard, C. and Kjems, J. (2013) Natural RNA circles function as efficient microRNA sponges. *Nature*, 495(7441), pp.384-388.
68. Harbeck, N., Penault-Llorca, F., Cortes, J., Gnant, M., Houssami, N., Poortmans, P., Ruddy, K., Tsang, J., and Cardoso, F. (2019) Breast cancer. *Nature Reviews Disease Primers*, *5*(1).
69. Hassanpour, S.H., and Dehghani, M. (2017) Review of cancer from perspective of molecular. *Journal of Cancer Research and Practice*, 4(4), pp. 127–129.
70. Haugner III, J. and Seelig, B. (2013) Universal labeling of 5′-triphosphate RNAs by artificial RNA ligase enzyme with broad substrate specificity. *Chemical Communications*, 49(66), p.7322.
71. Hausner, G., Hafez, M. and Edgell, D. (2014) Bacterial group I introns: mobile RNA catalysts. *Mobile DNA*, 5(1).
72. Hedberg, A. and Johansen, S. (2013) Nuclear group I introns in self-splicing and beyond. *Mobile DNA*, 4(1), p.17.
73. Henry, N. L. (2014) Endocrine therapy toxicity: Management options. *American Society of Clinical Oncology Educational Book*, (34).
74. Hevekerl, H., Spielmann, T., Chmyrov, A. and Widengren, J. (2011) Förster Resonance Energy Transfer beyond 10 nm: Exploiting the Triplet State Kinetics of Organic Fluorophores. *The Journal of Physical Chemistry B*, 115(45), pp.13360-13370.
75. Hilchie, A., Sharon, A., Haney, E., Hoskin, D., Bally, M., Franco, O., Corcoran, J. and Hancock, R. (2016) Mastoparan is a membranolytic anti-cancer peptide that works synergistically with gemcitabine in a mouse model of mammary carcinoma. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1858(12), pp.3195-3204.
76. Hinnebusch, A. (2014) The Scanning Mechanism of Eukaryotic Translation Initiation. *Annual Review of Biochemistry*, 83(1), pp.779-812.
77. Itoh, Y. (2015) Membrane-type matrix metalloproteinases: Their functions and regulations. *Matrix Biology*, 44-46, pp.207-223.

78. Jackson, R., Hellen, C. and Pestova, T. (2010) The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Reviews Molecular Cell Biology*, 11(2), pp.113-127.

79. Jalili, V., Afgan, E., Gu, Q., Clements, D., Blankenberg, D., Goecks, J., Taylor, J. and Nekrutenko, A. (2020) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update. *Nucleic Acids Research*, 48(W1), pp.W395-W402.

80. Jarrell, K. (1993) Inverse splicing of a group II intron. *Proceedings of the National Academy of Sciences*, 90(18), pp.8624-8627.

81. Jeck, W., Sorrentino, J., Wang, K., Slevin, M., Burd, C., Liu, J., Marzluff, W. and Sharpless, N. (2012) Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA*, 19(2), pp.141-157.

82. Kapp, K., Schrempf, S., Lemberg, M.K. and Dobberstein, B. (2009) Post-targeting functions of signal peptides.

83. Kelley, L., Mezulis, S., Yates, C., Wass, M. and Sternberg, M. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 10(6), pp.845-858.

84. Kessenbrock, K., Plaks, V., and Werb, Z. (2010) Matrix metalloproteinases: Regulators of the tumor microenvironment. *Cell*, *141*(1), pp.52–67.

85. King, H., Cobbold, L. and Willis, A. (2010) The role of IREStrans-acting factors in regulating translation initiation. *Biochemical Society Transactions*, 38(6), pp.1581-1586.

86. Koblinski, J., Ahram, M. and Sloane, B. (2000) Unraveling the role of proteases in cancer. *Clinica Chimica Acta*, 291(2), pp.113-135.

87. Koirala, D., Shao, Y., Koldobskaya, Y., Fuller, J., Watkins, A., Shelke, S., Pilipenko, E., Das, R., Rice, P. and Piccirilli, J. (2019) A conserved RNA structural motif for organizing topology within picornaviral internal ribosome entry sites. *Nature Communications*, 10(1).

88. Kong, P., Yu, Y., Wang, L., Dou, Y., Zhang, X., Cui, Y., Wang, H., Yong, Y., Liu, Y., Hu, H., Cui, W., Sun, S., Li, B., Zhang, F. and Han, M. (2019) circ-Sirt1 controls NF-κB activation via sequence-specific interaction and enhancement of SIRT1 expression by binding to miR-132/212 in vascular smooth muscle cells. *Nucleic Acids Research*, 47(7), pp.3580-3593.

89. Kozak, M. (1989) Circumstances and mechanisms of inhibition of translation by secondary structure in eucaryotic mRNAs. *Molecular and Cellular Biology*, 9(11), pp.5134-5142.
90. Kruger, K., Grabowski, P., Zaug, A., Sands, J., Gottschling, D. and Cech, T. (1982) Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell*, 31(1), pp.147-157.
91. Kurschat, W., Müller, J., Wombacher, R. and Helm, M. (2005) Optimizing splinted ligation of highly structured small RNAs. *RNA*, 11(12), pp.1909-1914.
92. Lacroix, M. and Leclercq, G. (2004) Relevance of breast cancer cell lines as models for breast tumours: An update. *Breast Cancer Research and Treatment*, 83(3), pp. 249–289.
93. Lambowitz, A. and Zimmerly, S. (2010) Group II Introns: Mobile Ribozymes that Invade DNA. *Cold Spring Harbor Perspectives in Biology*, 3(8), pp.a003616-a003616.
94. Laronha, H. and Caldeira, J. (2020) Structure and Function of Human Matrix Metalloproteinases. *Cells*, 9(5), p.1076.
95. Lasda, E. and Parker, R. (2014) Circular RNAs: diversity of form and function. *RNA*, 20(12), pp.1829-1842.
96. Le, W., Chen, B., Cui, Z., Liu, Z. and Shi, D. (2019) Detection of cancer cells based on glycolytic-regulated surface electrical charges. *Biophysics Reports*, 5(1), pp.10-18.
97. Legnini, I., Di Timoteo, G., Rossi, F., Morlando, M., Briganti, F., Sthandier, O., Fatica, A., Santini, T., Andronache, A., Wade, M., Laneve, P., Rajewsky, N. and Bozzoni, I. (2017) Circ-ZNF609 Is a Circular RNA that Can Be Translated and Functions in Myogenesis. *Molecular Cell*, 66(1), pp.22-37.e9.
98. Lehmann, B. D., Bauer, J. A., Chen, X., Sanders, M. E., Chakravarthy, A. B., Shyr, Y., & Pietenpol, J. A. (2011). Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *Journal of Clinical Investigation*, *121*(7), 2750–2767.
99. Lei, M., Zheng, G., Ning, Q., Zheng, J. and Dong, D. (2020) Translation and functional roles of circular RNAs in human cancer. *Molecular Cancer*, 19(1).
100. Leite, N., da Costa, L., dos Santos Alvares, D., dos Santos Cabrera, M., de Souza, B., Palma, M. and Ruggiero Neto, J. (2010) The effect of acidic residues and amphipathicity on the lytic activities of mastoparan peptides studied by fluorescence and CD spectroscopy. Amino Acids, 40(1), pp.91-100.

101. Li, H., Xie, M., Wang, Y., Yang, L., Xie, Z. and Wang, H. (2021) riboCIRC: a comprehensive database of translatable circRNAs. *Genome Biology*, 22(1).

102. Li, X., Wang, J., Zhang, C., Lin, C., Zhang, J., Zhang, W., Zhang, W., Lu, Y., Zheng, L. and Li, X. (2018) Circular RNA circITGA7 inhibits colorectal cancer growth and metastasis by modulating the Ras pathway and upregulating transcription of its host geneITGA7. *The Journal of Pathology*, 246(2), pp.166-179.

103. Li, Y., Zheng, Q., Bao, C., Li, S., Guo, W., Zhao, J., Chen, D., Gu, J., He, X. and Huang, S. (2015) Circular RNA is enriched and stable in exosomes: a promising biomarker for cancer diagnosis. *Cell Research*, 25(8), pp.981-984.

104. Liang, W., Wong, C., Liang, P., Shi, M., Cao, Y., Rao, S., Tsui, S., Waye, M., Zhang, Q., Fu, W. and Zhang, J. (2019) Translation of the circular RNA circβ-catenin promotes liver cancer cell growth through activation of the Wnt pathway. *Genome Biology*, 20(1).

105. Liang, Y., Zhang, H., Song, X., and Yang, Q. (2020) Metastatic heterogeneity of breast cancer: Molecular mechanism and potential therapeutic targets. *Seminars in Cancer Biology*, *60*, 14–27.

106. Loffek, S., Schilling, O. and Franzke, C. (2010) Biological role of matrix metalloproteinases: a critical balance. *European Respiratory Journal*, 38(1), pp.191-208.

107. Long, M. and Sullenger, B. (1999) Evaluating Group I Intron Catalytic Efficiency in Mammalian Cells. *Molecular and Cellular Biology*, 19(10), pp.6479-6487.

108. López-Otín, C., & Matrisian, L. M. (2007) Emerging roles of proteases in tumour suppression. *Nature Reviews Cancer*, *7*(10), pp.800–808.

109. Lotze, M. and Thomson, A. (2005) *Measuring immunity: Basic Biology and Clinical Assessment*. Academic Press, pp.343-349.

110. Lowenthal, R. M., and Eaton, K. (1996) Toxicity of chemotherapy. *Hematology/Oncology Clinics of North America*, *10*(4), 967–990.

111. Lu, S., Wang, Y., Huang, H., Pan, Y., Chaney, E., Boppart, S., Ozer, H., Strongin, A. and Wang, Y. (2013) Quantitative FRET Imaging to Visualize the Invasiveness of Live Breast Cancer Cells. *PLoS ONE*, 8(3), p.e58569.

112. Makki, J. (2015) Diversity of breast carcinoma: Histological subtypes and clinical relevance. *Clinical Medicine Insights: Pathology*, 8.

113. Maniatis, T., E.F. Fritsch and J. Sambrook (1982) Molecular Cloning a Laboratory Manual. *Cold Spring Harbor Laboratory*, pp: 6.

114. Margineanu, A., Chan, J., Kelly, D., Warren, S., Flatters, D., Kumar, S., Katan, M., Dunsby, C. and French, P. (2016) Screening for protein-protein interactions using Förster resonance energy transfer (FRET) and fluorescence lifetime imaging microscopy (FLIM). *Scientific Reports*, 6(1).

115. Mason, S.D. and Joyce, J.A. (2011) Proteolytic networks in cancer. *Trends in cell biology*, *21*(4), pp.228-237.

116. Massova, I., Kotra, L., Fridman, R. and Mobashery, S. (1998) Matrix metalloproteinases: structures, evolution, and diversification. *The FASEB Journal*, 12(12), pp.1075-1095.

117. Mazloomian, A., Araki, S., Ohori, M., El-Naggar, A.M., Yap, D., Bashashati, A., Nakao, S., Sorensen, P.H., Nakanishi, A., Shah, S. and Aparicio, S. (2019) Pharmacological systems analysis defines EIF4A3 functions in cell-cycle and RNA stress granule formation. *Communications biology*, *2*(1), pp.1-15.

118. Meganck, R., Liu, J., Hale, A., Simon, K., Fanous, M., Vincent, H., Wilusz, J., Moorman, N., Marzluff, W. and Asokan, A. (2021) Engineering highly efficient backsplicing and translation of synthetic circRNAs. *Molecular Therapy - Nucleic Acids*, 23, pp.821-834.

119. Mercier, E., Holtkamp, W., Rodnina, M. and Wintermeyer, W. (2017) Signal recognition particle binds to translating ribosomes before emergence of a signal anchor sequence. *Nucleic Acids Research*, 45(20), pp.11858-11866.

120. Merrick, W. and Pavitt, G. (2018) Protein Synthesis Initiation in Eukaryotic Cells. *Cold Spring Harbor Perspectives in Biology*, 10(12), p.a033092.

121. Meyer, K., Patil, D., Zhou, J., Zinoviev, A., Skabkin, M., Elemento, O., Pestova, T., Qian, S. and Jaffrey, S. (2015) 5′ UTR m^6A Promotes Cap-Independent Translation. *Cell*, 163(4), pp.999-1010.

122. Mi, H., Ebert, D. and Muruganujan, A. (2021) Caitlin Mills, Laurent-Philippe Albou, Tremayne Mushayamaha, and Paul D. Thomas. 2021."PANTHER Version 16: A Revised Family Classification, Tree-Based Classification Tool, Enhancer Regions and Extensive API.". *Nucleic acids research, 49*, pp.D394-403.

123. Mi, H., Muruganujan, A., Huang, X., Ebert, D., Mills, C., Guo, X. and Thomas, P. (2019) Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nature Protocols*, 14(3), pp.703-721.

124. Mikheeva, S. (1997) Use of an engineered ribozyme to produce a circular human exon. *Nucleic Acids Research*, 25(24), pp.5085-5094.

125. Moore, M. and Sharp, P. (1992) Site-specific modification of pre-mRNA: the 2'-hydroxyl groups at the splice sites. *Science*, 256(5059), pp.992-997.

126. Muley, A., Majumder, S., Kolluru, G., Parkinson, S., Viola, H., Hool, L., Arfuso, F., Ganss, R., Dharmarajan, A. and Chatterjee, S. (2010) Secreted Frizzled-Related Protein 4. *The American Journal of Pathology*, 176(3), pp.1505-1516.

127. Müller, S. and Appel, B. (2016) In vitro circularization of RNA. *RNA Biology*, 14(8), pp.1018-1027.

128. NHS England. (2016, May 23) *Chemo drug optimisation to improve patient experience of cancer treatment*. Chemo drug optimisation. Available at: https://www.england.nhs.uk/2016/05/chemo-drug-optimisation/ (Accessed: 18th February 2023).

129. Nielsen, H. (2003) The ability to form full-length intron RNA circles is a general property of nuclear group I introns. *RNA*, 9(12), pp.1464-1475.

130. Nielsen, H., Engelbrecht, J., Brunak, S. and von Heijne, G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering Design and Selection*, 10(1), pp.1-6.

131. Nielsen, P.J. and Trachsel, H. (1988) The mouse protein synthesis initiation factor 4A gene family includes two related functional genes which are differentially expressed. *The EMBO journal*, *7*(7), pp.2097-2105.

132. Okazaki, I., Suzuki, N., Nishi, N., Utani, A., Matsuura, H., Shinkai, H., Yamashita, H., Kitagawa, Y. and Nomizu, M. (2002) Identification of Biologically Active Sequences in the Laminin α4 Chain G Domain. *Journal of Biological Chemistry*, 277(40), pp.37070-37078.

133. Onitilo, A. A., Engel, J. M., Greenlee, R. T., & Mukesh, B. N. (2009) Breast cancer subtypes based on ER/PR and HER2 expression: Comparison of clinicopathologic features and survival. *Clinical Medicine & Research*, *7*(1-2), 4–13.

134. Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S. and Sharp, P.A. (1986) Splicing of messenger RNA precursors. *Annual review of biochemistry*, *55*(1), pp.1119-1150.

135. Pan, Z., Cai, J., Lin, J., Zhou, H., Peng, J., Liang, J., Xia, L., Yin, Q., Zou, B., Zheng, J., Qiao, L. and Zhang, L. (2020) A novel protein encoded by circFNDC3B inhibits tumor progression and EMT through regulating Snail in colon cancer. *Molecular Cancer*, 19(1).

136. Panda, A. and Gorospe, M. (2018) Detection and Analysis of Circular RNAs by RT-PCR. *BIO-PROTOCOL*, 8(6).

137. Panić, L., Montagne, J., Cokarić, M. and Volarević, S. (2007) S6-Haploinsufficiency Activates the p53 Tumor Suppressor. *Cell Cycle*, 6(1), pp.20-24.

138. Park, J., Kim, K., Sim, Y., Kim, Y., Kim, J., Lee, C., Han, J., Kim, C., Lee, J. and Park, S. (2019) Structures of three ependymin-related proteins suggest their function as a hydrophobic molecule binder. *IUCrJ*, 6(4), pp.729-739.

139. Parliament. House of Commons (2023). *Cancer statistics for England*. (HC 06877). London: The Stationery Office.

140. Patop, I. and Kadener, S. (2018) circRNAs in Cancer. *Current Opinion in Genetics & Development*, 48, pp.121-127.

141. Peebles, C., Zhang, M., Perlman, P. and Franzen, J. (1995) Catalytically critical nucleotide in domain 5 of a group II intron. *Proceedings of the National Academy of Sciences*, 92(10), pp.4422-4426.

142. Perriman, R. and Ares, M. (1998) Circular mRNA can direct translation of extremely long repeating-sequence proteins in vivo. *RNA*, 4(9), pp.1047-1054.

143. Petkovic, S. and Müller, S. (2015) RNA circularization strategies in vivo and in vitro. *Nucleic Acids Research*, 43(4), pp.2454-2465.

144. Piccirillo, C., Bjur, E., Topisirovic, I., Sonenberg, N. and Larsson, O. (2014) Translational control of immune responses: from transcripts to translatomes. *Nature Immunology*, 15(6), pp.503-511.

145. Piñol-Roma, S. and Dreyfuss, G. (1992) Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm. *Nature*, 355(6362), pp.730-732.

146. Puigbò, P., Bravo, I. and Garcia-Vallve, S. (2008) CAIcal: A combined set of tools to assess codon usage adaptation. *Biology Direct*, 3(1), p.38.

147. Puighermanal, E., Biever, A., Pascoli, V., Melser, S., Pratlong, M., Cutando, L., Rialle, S., Severac, D., Boubaker-Vitre, J., Meyuhas, O., Marsicano, G., Lüscher, C. and Valjent, E. (2017) Ribosomal Protein S6 Phosphorylation Is Involved in Novelty-Induced Locomotion, Synaptic Plasticity and mRNA Translation. *Frontiers in Molecular Neuroscience*, 10.

148. Puttaraju, M. and Been, M. (1992) Group I permuted intron-exon (PIE) sequences self-splice to produce circular exons. *Nucleic Acids Research*, 20(20), pp.5357-5364.

149. Puttaraju, M. and Been, M. (1996) Circular Ribozymes Generated in Escherichia coli Using Group I Self-splicing Permuted Intron-Exon Sequences. *Journal of Biological Chemistry*, 271(42), pp.26081-26087.

150. Raghavan, A., Ogilvie, R., Reilly, C., Abelson, M., Raghavan, S., Vasdewani, J., Krathwohl, M. and Bohjanen, P. (2002) Genome-wide analysis of mRNA decay in resting and activated primary human T lymphocytes. *Nucleic Acids Research*, 30(24), pp.5529-5538.

151. Rathore, D., Sacci, J., de la Vega, P. and McCutchan, T. (2002) Binding and Invasion of Liver Cells by Plasmodium falciparum Sporozoites. *Journal of Biological Chemistry*, 277(9), pp.7092-7098.

152. Rice, P., Bleasby, A. and Ison, J. (2011) *EMBOSS user's guide*. Cambridge: Cambridge University Press.

153. Rijnbrand, R., Abbink, T., Haasnoot, P., Spaan, W. and Bredenbeek, P. (1996) The Influence of AUG Codons in the Hepatitis C Virus 5′ Nontranslated Region on Translation and Mapping of the Translation Initiation Window. *Virology*, 226(1), pp.47-56.

154. Rutkowski, D., Ott, C., Polansky, J. and Lingappa, V. (2003) Signal Sequences Initiate the Pathway of Maturation in the Endoplasmic Reticulum Lumen. *Journal of Biological Chemistry*, 278(32), pp.30365-30372.

155. Ruvinsky, I., Sharon, N., Lerer, T., Cohen, H., Stolovich-Rain, M., Nir, T., Dor, Y., Zisman, P. and Meyuhas, O. (2005) Ribosomal protein S6 phosphorylation is a determinant of cell size and glucose homeostasis. *Genes & Development*, 19(18), pp.2199-2211.

156. Sadikovic, B., Al-Romaih, K., Squire, J., and Zielenska, M. (2008) Cause and consequences of genetic and epigenetic alterations in human cancer. *Current Genomics*, *9*(6), 394–408.

157. Salzman, J., Chen, R., Olsen, M., Wang, P. and Brown, P. (2013) Cell-Type Specific Features of Circular RNA Expression. *PLoS Genetics*, 9(9), p.e1003777.

158. Salzman, J., Gawad, C., Wang, P., Lacayo, N. and Brown, P. (2012) Circular RNAs Are the Predominant Transcript Isoform from Hundreds of Human Genes in Diverse Cell Types. *PLoS ONE*, 7(2), p.e30733.

159. Schaap, M., Hancock, R., Wilderspin, A. and Wells, G. (2013) Development of a steady-state FRET-based assay to identify inhibitors of the Keap1-Nrf2 protein-protein interaction. *Protein Science*, 22(12), pp.1812-1819.

160. Scheper, G. and Proud, C. (2002) Does phosphorylation of the cap-binding protein eIF4E play a role in translation initiation?. *European Journal of Biochemistry*, 269(22), pp.5350-5359.

161. Schindewolf, C., Braun, S. and Domdey, H. (1996) In vitro Generation of a Circular Exon from a Linear Pre-mRNA Transcript. *Nucleic Acids Research*, 24(7), pp.1260-1266.

162. Schmid, S. and Linder, P. (1991) Translation initiation factor 4A from Saccharomyces cerevisiae: analysis of residues conserved in the DEAD family of RNA helicases. *Molecular and Cellular Biology*, 11(7), pp.3463-3471.

163. Schueren, F. and Thoms, S. (2016) Functional Translational Readthrough: A Systems Biology Perspective. *PLOS Genetics*, 12(8), p.e1006196.

164. Schwartz, S., Mumbach, M. R., Jovanovic, M., Wang, T., Maciag, K., Bushkin, G. G., Mertins, P., Ter-Ovanesyan, D., Habib, N., Cacchiarelli, D., Sanjana, N. E., Freinkman, E., Pacold, M. E., Satija, R., Mikkelsen, T. S., Hacohen, N., Zhang, F., Carr, S. A., Lander, E. S., and Regev, A. (2014) Perturbation of M^6A writers reveals two distinct classes of mRNA methylation at internal and 5′ sites. *Cell Reports*, *8*(1), 284–296.

165. Scior, A., Preissler, S., Koch, M. and Deuerling, E. (2011) Directed PCR-free engineering of highly repetitive DNA sequences. *BMC Biotechnology*, 11(1).

166. Shimizu, Y., Kuruma, Y., Ying, B., Umekage, S. and Ueda, T. (2006) Cell-free translation systems for protein engineering. *FEBS Journal*, 273(18), pp.4133-4140.

167. Shin, K., Sullenger, B. and Lee, S. (2004) Ribozyme-Mediated Induction of Apoptosis in Human Cancer Cells by Targeted Repair of Mutant p53 RNA. *Molecular Therapy*, 10(2), pp.365-372.

168. Sokabe, M. and Fraser, C. (2014) Human Eukaryotic Initiation Factor 2 (eIF2)-GTP-Met-tRNAi Ternary Complex and eIF3 Stabilize the 43 S Preinitiation Complex. *Journal of Biological Chemistry*, 289(46), pp.31827-31836.

169. Sonenberg, N. and Hinnebusch, A. (2009) Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets. *Cell*, 136(4), pp.731-745.

170. Sonenberg, N., Morgan, M., Merrick, W. and Shatkin, A. (1978) A polypeptide in eukaryotic initiation factors that crosslinks specifically to the 5'-terminal cap in mRNA. *Proceedings of the National Academy of Sciences*, 75(10), pp.4843-4847.

171. Spriggs, K., Bushell, M., Mitchell, S. and Willis, A. (2005) Internal ribosome entry segment-mediated translation during apoptosis: the role of IRES-trans-acting factors. *Cell Death & Differentiation*, 12(6), pp.585-591.

172. Spriggs, K., Mitchell, S. and Willis, A. (2005) Investigation of interactions of polypyrimidine tract-binding protein with artificial internal ribosome entry segments. *Biochemical Society Transactions*, 33(6), pp.1483-1486.

173. Stockert, J., Horobin, R., Colombo, L. and Blázquez-Castro, A. (2018) Tetrazolium salts and formazan products in Cell Biology: Viability assessment, fluorescence imaging, and labeling perspectives. *Acta Histochemica*, 120(3), pp.159-167.

174. Stoneley, M. and Willis, A. (2004) Cellular internal ribosome entry segments: structures, trans-acting factors and regulation of gene expression. *Oncogene*, 23(18), pp.3200-3207.

175. Storr, S., Thompson, N., Pu, X., Zhang, Y. and Martin, S. (2015) Calpain in Breast Cancer: Role in Disease Progression and Treatment Response. *Pathobiology*, 82(3-4), pp.133-141.

176. Suárez-Castillo, E. and García-Arrarás, J. (2006) Deciphering the molecular evolution of the ependymin protein family. *Developmental Biology*, 295(1), p.424.

177. Sun, Y., Liu, Y., Ma, X. and Hu, H. (2021) The Influence of Cell Cycle Regulation on Chemotherapy. *International Journal of Molecular Sciences*, 22(13), p.6923.

178. Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021) Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209–249.

179. Tabak, H., Van der Horst, G., Smit, J., Winter, A., Mul, Y. and Koerkamp, G. (1988) Discrimination between RNA circles, interlocked RNA circles and lariats using two-dimensional polyacrylamide gel electrophoresis. *Nucleic Acids Research*, 16(14), pp.6597-6605.

180. Tang, Y., Wang, Y., Kiani, M. F., and Wang, B. (2016) Classification, treatment strategy, and associated drug resistance in breast cancer. *Clinical Breast Cancer*, *16*(5), 335–343.

181. Testa, U., Castelli, G. and Pelosi, E. (2020) Breast cancer: A molecularly heterogenous disease needing subtype-specific treatments. *Medical Sciences*, 8(1), p. 18.

182. Thakor, N. and Holcik, M. (2011) IRES-mediated translation of cellular messenger RNA operates in eIF2α- independent manner during stress. *Nucleic Acids Research*, 40(2), pp.541-552.

183. Thoma, C., Bergamini, G., Galy, B., Hundsdoerfer, P. and Hentze, M. (2004) Enhancement of IRES-Mediated Translation of the c-myc and BiP mRNAs by the Poly(A) Tail Is Independent of Intact eIF4G and PABP. *Molecular Cell*, 15(6), pp.925-935.

184. Umekage, S. and Kikuchi, Y. (2009) In vitro and in vivo production and purification of circular RNA aptamer. *Journal of Biotechnology*, 139(4), pp.265-272.

185. Verduci, L., Tarcitano, E., Strano, S., Yarden, Y. and Blandino, G. (2021) CircRNAs: role in human diseases and potential use as biomarkers. *Cell Death & Disease*, 12(5).

186. Verma, R. and Hansch, C. (2007) Matrix Metalloproteinases (MMPs): Chemical-Biological Functions and (Q)SARs. *ChemInform*, 38(24).

187. Walters, B. and Thompson, S. (2016) Cap-Independent Translational Control of Carcinogenesis. *Frontiers in Oncology*, 6.

188. Wang, J. and Gribskov, M. (2019) IRESpy: an XGBoost model for prediction of internal ribosome entry sites. *BMC Bioinformatics*, 20(1).

189. Wang, J. and Wang, L. (2020) Deep analysis of RNA N6-adenosine methylation (m^6A) patterns in human cells. *NAR Genomics and Bioinformatics*, 2(1).

190. Wang, K., and Tepper, J. E. (2021) Radiation therapy-associated toxicity: Etiology, management, and prevention. *CA: A Cancer Journal for Clinicians*, *71*(5), 437–454.

191. Warner, J. (1999) The economics of ribosome biosynthesis in yeast. *Trends in Biochemical Sciences*, 24(11), pp.437-440.

192. Wen, S., Qadir, J. and Yang, B. (2022) Circular RNA translation: novel protein isoforms and clinical significance. *Trends in Molecular Medicine*, 28(5), pp.405-420.

193. Wesselhoeft, R., Kowalski, P. and Anderson, D. (2018) Engineering circular RNA for potent and stable translation in eukaryotic cells. *Nature Communications*, 9(1).

194. Wesselhoeft, R., Kowalski, P., Parker-Hale, F., Huang, Y., Bisaria, N. and Anderson, D. (2019) RNA Circularization Diminishes Immunogenicity and Can Extend Translation Duration In Vivo. *Molecular Cell*, 74(3), pp.508-520.e4.

195. WHO International Agency for Research on Cancer (2020). Latest global cancer data: Cancer burden rises to 19.3 million new cases and 10.0 million cancer deaths in 2020. Press release 292. Available at: https://www.iarc.who.int/wp-content/uploads/2020/12/pr292_E.pdf.

196. Who.int. (2022) *Cancer*. [online] Available at: <https://www.who.int/news-room/fact-sheets/detail/cancer> [Accessed 13 January 2022].

197. Will, C.L. and Lührmann, R. (2001) Spliceosomal UsnRNP biogenesis, structure and function. *Current opinion in cell biology*, *13*(3), pp.290-301.

198. Wolfson, B., Padget, M. R., Schlom, J., and Hodge, J. W. (2021) Exploiting off-target effects of estrogen deprivation to sensitize estrogen receptor negative breast cancer to immune killing. *Journal for ImmunoTherapy of Cancer*, *9*(7).

199. Wu, N., Yuan, Z., Du, K., Fang, L., Lyu, J., Zhang, C., He, A., Eshaghi, E., Zeng, K., Ma, J., Du, W. and Yang, B. (2019) Translation of yes-associated protein (YAP) was antagonized by its circular RNA via suppressing the assembly of the translation initiation machinery. *Cell Death & Differentiation*, 26(12), pp.2758-2773.

200. Wu, Q., Li, P., Wu, M. and Liu, Q. (2019) Deregulation of Circular RNAs in Cancer From the Perspectives of Aberrant Biogenesis, Transport and Removal. *Frontiers in Genetics*, 10.

201. Wu, W., Ji, P. and Zhao, F. (2020) CircAtlas: an integrated resource of one million highly accurate circular RNAs from 1070 vertebrate transcriptomes. *Genome Biology*, 21(1).

202. Xiao, M. and Wilusz, J. (2019) An improved method for circular RNA purification using RNase R that efficiently removes linear RNAs containing G-quadruplexes or structured 3′ ends. *Nucleic Acids Research*, 47(16), pp.8755-8769.

203. Xu, C. and Zhang, J. (2021) Mammalian circular RNAs result largely from splicing errors. *Cell Reports*, 36(4), p.109439.

204. Yamada, Y., Makimura, K., Merhendi, H., Ueda, K., Nishiyama, Y., Yamaguchi, H. and Osumi, M. (2002) Comparison of different methods for extraction of mitochondrial DNA from human pathogenic yeasts. *Japanese journal of infectious diseases*, *55*(4), pp.122-125.

205. Yang, H.S., Jansen, A.P., Komar, A.A., Zheng, X., Merrick, W.C., Costes, S., Lockett, S.J., Sonenberg, N. and Colburn, N.H. (2003) The transformation suppressor Pdcd4 is a novel eukaryotic translation initiation factor 4A binding protein that inhibits translation. *Molecular and cellular biology*, *23*(1), pp.26-37.

206. Yang, L., Fu, J. and Zhou, Y. (2018) Circular RNAs and Their Emerging Roles in Immune Regulation. *Frontiers in Immunology*, 9.

207. Yang, Q., Yu, C., Zhao, F., Dang, Y., Wu, C., Xie, P., Sachs, M. and Liu, Y. (2019) eRF1 mediates codon usage effects on mRNA translation efficiency through premature termination at rare codons. *Nucleic Acids Research*, 47(17), pp.9243-9258.

208. Yang, Y. and Wang, Z. (2019) IRES-mediated cap-independent translation, a path leading to hidden proteome. *Journal of Molecular Cell Biology*, 11(10), pp.911-919.

209. Yang, Y., Fan, X., Mao, M., Song, X., Wu, P., Zhang, Y., Jin, Y., Yang, Y., Chen, L., Wang, Y., Wong, C., Xiao, X. and Wang, Z. (2017) Extensive translation of circular RNAs driven by N6-methyladenosine. *Cell Research*, 27(5), pp.626-641.

210. Yang, Y., Gao, X., Zhang, M., Yan, S., Sun, C., Xiao, F., Huang, N., Yang, X., Zhao, K., Zhou, H., Huang, S., Xie, B. and Zhang, N. (2017) Novel Role of FBXW7 Circular RNA in Repressing Glioma Tumorigenesis. *JNCI: Journal of the National Cancer Institute*, 110(3), pp.304-315.

211. Ye, F., Gao, G., Zou, Y., Zheng, S., Zhang, L., Ou, X., Xie, X. and Tang, H. (2019) circFBXW7 Inhibits Malignant Progression by Sponging miR-197-3p and Encoding a 185-aa Protein in Triple-Negative Breast Cancer. *Molecular Therapy - Nucleic Acids*, 18, pp.88-98.

212. Zaccara, S., Ries, R. and Jaffrey, S. (2019) Reading, writing and erasing mRNA methylation. *Nature Reviews Molecular Cell Biology*, 20(10), pp.608-624.

213. Zauner, T., Berger-Hoffmann, R., Müller, K., Hoffmann, R. and Zuchner, T. (2011) Highly Adaptable and Sensitive Protease Assay Based on Fluorescence Resonance Energy Transfer. *Analytical Chemistry*, 83(19), pp.7356-7363.

214. Zhang, F., Ramsay, E.S. and Woodson, S.A. (1995) In vivo facilitation of Tetrahymena group I intron splicing in Escherichia coli pre-ribosomal RNA. *Rna*, *1*(3), p.284.

215. Zhang, M., Huang, N., Yang, X., Luo, J., Yan, S., Xiao, F., Chen, W., Gao, X., Zhao, K., Zhou, H., Li, Z., Ming, L., Xie, B. and Zhang, N. (2018) A novel protein encoded by the circular form of the SHPRH gene suppresses glioma tumorigenesis. *Oncogene*, 37(13), pp.1805-1814.

216. Zhang, Y., Zhang, X., Chen, T., Xiang, J., Yin, Q., Xing, Y., Zhu, S., Yang, L. and Chen, L. (2013) Circular Intronic Long Noncoding RNAs. *Molecular Cell*, 51(6), pp.792-806.

217. Zheng, Z. (2004) Regulation of Alternative RNA Splicing by Exon Definition and Exon Sequences in Viral and Mammalian Gene Expression. *Journal of Biomedical Science*, 11(3), pp.278-294.

218. Zhou, C., Molinie, B., Daneshvar, K., Pondick, J., Wang, J., Van Wittenberghe, N., Xing, Y., Giallourakis, C. and Mullen, A. (2017) Genome-Wide Maps of m^6A circRNAs Identify Widespread and Cell-Type-Specific Methylation Patterns that Are Distinct from mRNAs. *Cell Reports*, 20(9), pp.2262-2276.

219. Zhou, Y., Zeng, P., Li, Y.H., Zhang, Z. and Cui, Q. (2016) SRAMP: prediction of mammalian N6-methyladenosine (m^6A) sites based on sequence-derived features. *Nucleic acids research*, *44*(10), pp.e91-e91.