IS SPLICING ESSENTIAL IN *CYANIDIOSCHYZON MEROLAE*?

by

María Begoña Zumárraga Rivera

B.Sc., Universidad San Francisco de Quito-Ecuador, 2019

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
BIOCHEMISTRY

UNIVERSITY OF NORTHERN BRITISH COLUMBIA

December 2024

Abstract

Protein-coding sequences of eukaryotic genes are often interrupted by non-coding sequences called introns. Introns must be removed from mRNA precursors, and retained segments known as exons are then ligated together to form mature messenger RNA. This essential process in eukaryotic gene expression is called mRNA splicing. Splicing is carried out by a large complex that contains small nuclear RNAs (snRNAs) and many proteins. Due to the complexity of the spliceosomal machinery in other eukaryotes, studying splicing has been very challenging. *Cyanidioschyzon merolae* is a unicellular red alga with only 39 introns in its genome and a much simpler set of splicing machinery than in humans. It has been estimated that the ancestral red alga contained ~1700 introns, from which we can infer that *C. merolae* has lost almost all of its introns. This raises the possibility that splicing is no longer essential for this organism.

I addressed whether this vestigial splicing system is biologically important by inhibiting splicing in various ways. If splicing is not required for the survival of this organism, inhibiting this process should not impact cell survival. In contrast, if splicing is essential, a deleterious phenotype and cell death are expected upon inhibition. I attempted to inhibit splicing using antisense RNA and degron techniques. In the first approach, which seeks to silence a target by base pairing the transcript to inactivate it, I transiently expressed the antisense version of three essential splicing factors under the control of a nitrate inducible promoter by transforming an engineered plasmid with a selectable marker into the cells. Additionally, I integrated antisense versions of two splicing factors in the genome under the control of the same inducible promoter. The antisense RNA should bind the target RNAs in both cases, leading to their degradation or sequestration. The nitrogen source for *C. merolae* in rich media is ammonium, where the

antisense promoter will be off. By shifting cells to nitrate media, I activated antisense expression, after which I expected splicing to be inhibited and cell death to occur. Control experiments showed that the inducible promoter works, but I could not demonstrate the antisense strand induction.

In the second approach, I implemented an inducible degron system to degrade splicing proteins. Degrons are motifs that target proteins for degradation, and they can be fused to target genes to allow the corresponding protein to be degraded by adding rapamycin. This small molecule activates the degradation system. I targeted Prp8 and Clf1, both core spliceosomal proteins. The degron results were consistent with protein degradation and splicing inhibition, but for technical reasons, I cannot conclude whether splicing is essential for *C. merolae* . Despite numerous attempts with morpholino oligos, gene deletion, splicing inhibitors, *etc., the only experiment* in our lab to date consistent with Cm splicing being essential was our failure to delete the gene for the splicing protein Cef1.

Key Words: Splicing, pre-mRNA, inhibition, antisense RNA, degron, northern blot, western blot, RT-qPCR.

Table of Contents

List of Figures

List of Tables

Acknowledgement

First of all, I would like to thank my supervisor, Dr. Stephen Rader, for giving me the opportunity to perform my graduate research in his lab and for all his help and support since the very beginning of my application process to UNBC. I want to thank Dr. Martha Stark for all her guidance and support in developing my research skills and for teaching me all the techniques I have been using throughout my degree, which will be extremely useful for me in the future. I want to thank my Committee Members, Dr. Sarah Gray, Dr. Andrea Gorrell, and Dr. June Garcia-Becerra, for all their support and feedback throughout my graduate work. Thank you, Committee Members, for all your recommendations, guidance, and help and for cheering me up when things did not seem to work appropriately. I want to thank my labmates, especially Maryam, Fafa, and Raha, for their feedback, help and support throughout this process and for helping me finish my experiments due to my restricted schedule.

Finally, I would like to thank my family, my son, my daughter, and my husband, for all their support and love and for always encouraging me to follow my dreams and keep going with my studies.

I dedicate this work to my loved family, Jose Maria, Sara and Sebastian.

Abbreviation Glossary

PCR: Polymerase Chain Reaction

RT-qPCR: Reverse Transcription- Quantitative Polymerase Chain Reaction

NRT: Non- Reverse Transcriptase control

NTC: Non-Template Control or Nineteen Complex

Pre-mRNA: pre-messenger RNA

mRNA: messenger RNA

5'SS:5' splice site

3'SS: 3' splice site

BPS: Branch Point Sequence

snRNA: small nuclear RNA

sisRNA: Stable intronic sequence RNA

FRB: FKBP12 binding protein

FKBP12: FK506 binding protein 12

TOR: Target of Rapamycin

HRA: Homology Recombination Arms

DMSO: Dimethyl sulfoxide

PS: Parent strain

NRTp: nitrite transporter promoter

NRp: nitrate reductase promoter

NIRp: nitrite reductase promoter

Gene nomenclature (*C. merolae*)

C. merolae genes are named serially, starting with the species CM (*Cyanidioschyzon merolae*), then the chromosome letter (A, B, C… T), then the serial number of the gene along the chromosome, ending with a letter designating whether the gene is protein-coding (C), transcribed (T), or a hypothetical transcript (Z). To simplify the gene references and make it easier to follow this work, I refer to the genes below using only the chromosome letter and serial number, as shown in Table 0-1. If an intron is deleted from a gene, the gene name is followed by Δi. CSR (*C. merolae* from Stephen Rader's lab) represents the name for biological replicates of the intron deletion or other modified strains in the Rader lab. Each biological replicate (CSR) has a specific number, as Table 0-2 shows.

Table 1: Gene nomenclature for *C. merolae* genes used in this project.

Gene	Gene nomenclature
NRp (nitrate reductase promoter)	(CM)G019C
NIRp (nitrite reductase promoter)	G021C
UBQ (ubiquitin) 3' UTR (untranslated region) Terminator	K296C
Cef1	R098C
Dib1	S018C
Clf1	K252C
Prp8	H168C
Neutral locus	D184/185

Table 2: CSR numbers for *C. merolae* strains used in this project.

Strain (Biological replicate number)	**CSR number**
Clf1-1	100
Clf1-2	101
Prp8-1	102
Prp8-2	103
mVenus-FRB-1	90
HA-FKBP-SKP1 (parent strain)	87

1. Chapter One- Introduction

1.1. Precursor Messenger RNA Splicing

Eukaryotes possess "genes in pieces," in which protein-coding sequences are often interrupted by non-coding sequences called introns (Koonin *et al*., 2012). Introns must be removed from mRNA precursors (pre-mRNAs), and protein-coding segments, known as exons, must be ligated together to form mature messenger RNAs. This essential process in eukaryotic gene expression is called pre-mRNA splicing (Wilkinson *et al*., 2020).

Decades of biochemistry and genetic studies, combined with recent studies of the spliceosome structure, have produced a detailed view of the splicing mechanism (Wilkinson *et al*., 2020). Biochemical characterization of splicing intermediates has elucidated a two-step phosphorylation transfer mechanism of splicing. Introns are defined by three important sites: the 5' splice site (5'SS), the branch point (BP) adenosine, and the 3' splice site (3'SS) (Figure 1a). In the first reaction, called branching, the 2' hydroxyl group of the BP adenosine attacks the phosphodiester group at the 5'SS, producing a cleaved 5' exon and a lariat-intron-3'exon intermediate in which the 5' phosphate of the first intron nucleotide is linked to the 2' oxygen of the BP adenosine (Wilkinson *et al*., 2020) (Figure 1b and c). In the second step (exon ligation), the newly exposed 3' hydroxyl group of the 5' exon attacks the phosphodiester group at the 3'SS, ligating the 5' and 3' exons to form mRNA and releasing the lariat intron. The spliceosome mediates these two reactions.

The 5' splice site (5'SS), the branch point (BP) adenosine, and the 3' splice site (3'SS) are all short, conserved sequences in introns. In yeast, for example, the 5'SS is followed by a highly conserved sequence (GUAUGU), and the 3'SS is preceded by a YAG trinucleotide (Figure 1a). The BP adenosine is commonly located 18-40 nucleotides upstream of the 3'SS in a highly conserved sequence (UACUAAC), in which the last A of the sequence denotes the BP adenosine

(Wilkinson *et al*., 2020). In humans, the sequences surrounding the 5'SS and the BP adenosine are less stringently conserved, and the 3'SS trinucleotide is preceded by a poly-pyrimidine tract (PPT).

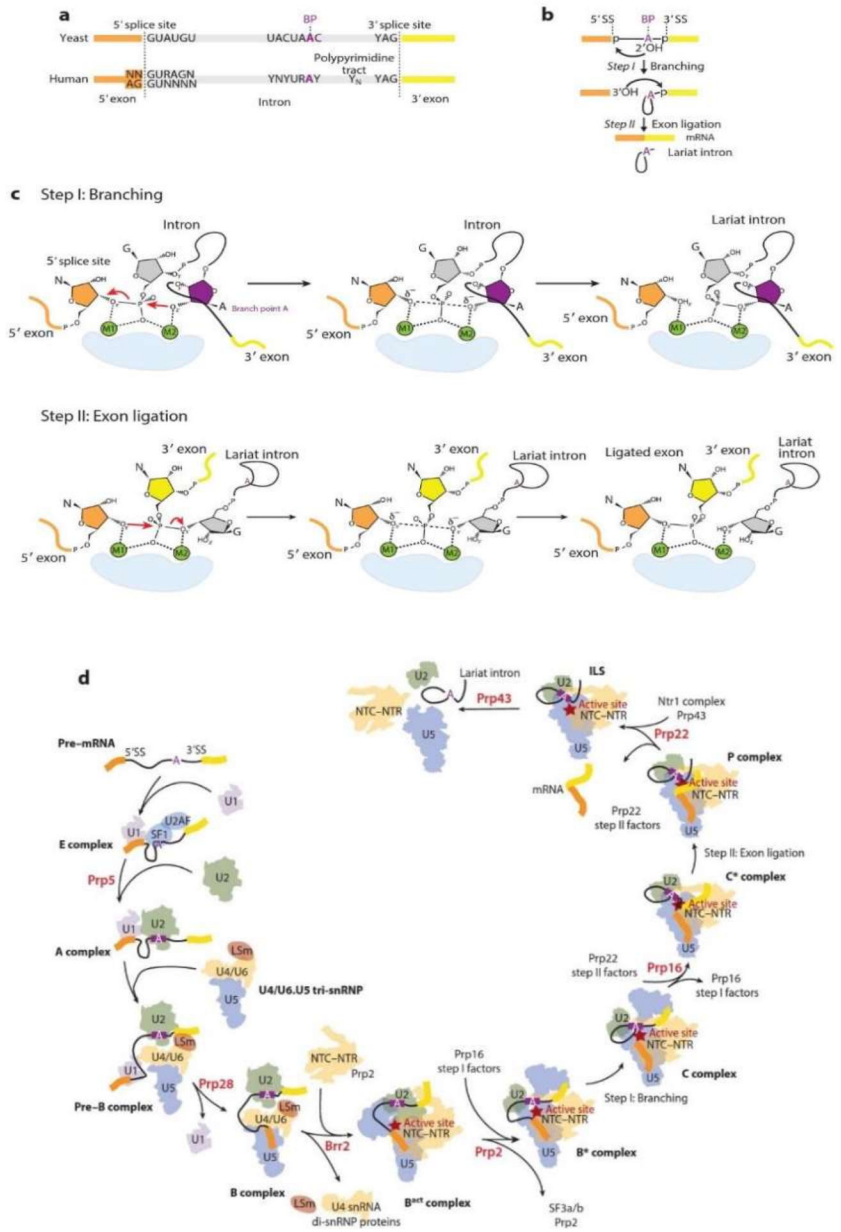


Figure 1-1: Chemistry of splicing (adapted Wilkinson *et al*., 2020). a) Introns are characterized by three conserved sequences: the 5'SS, BP sequence, and 3'SS. Branch Point adenosine is highlighted in purple. Human and yeast intron-conserved sequences are shown. b) and c) Representation of the two transesterification reactions involved in splicing. d)The spliceosome is assembled in a highly ordered and stepwise manner, activated to form the active site, and extensively remodelled to perform branching and exon ligation (Permission to use the figure, cropped from original).

The spliceosome is not a preassembled enzyme (holoenzyme) but is assembled *de novo* for each splicing reaction from 5 small nuclear RNAs (snRNAs – U1, U2, U4, U5, and U6) and approximately 100-300 proteins (Wilkinson *et al*., 2020). During spliceosome assembly, activation, catalysis, and disassembly, large ribonucleoprotein (RNP) complexes associate and dissociate in an ordered and stepwise manner (Hegele *et al*., 2012) (Figure 1d). Seven homologous Sm proteins assemble into a ring around the U-rich sequence at the 3' end of U1, U2, U4, and U5 snRNAs, and seven paralogous LSm proteins (LSm2-8) forming a preassembled ring attach to a U-rich sequence at the 3' end of the U6 snRNA (Wilkinson *et al*., 2019). These snRNAs bind a specific set of additional proteins and form a small nuclear ribonucleoprotein (snRNP). Several other non-snRNP-associated proteins and protein complexes, including splicing factors and eight ATP-dependent helicases, are also involved in the splicing process.

Within the spliceosome, the snRNAs perform the essential roles of substrate recognition and catalysis (Wilkinson *et al*., 2020). The U1 and U2 snRNPs recognize the 5'SS and the BP sequence, respectively, and form the pre-spliceosome or A complex. The A complex then associates with the preassembled U4/U6-U5 tri-snRNP to generate the B complex in a reaction catalyzed by the Prp28 ATP-dependent helicase (Matera & Wang, 2014) (Figure 1d). The tri-snRNP structure remains unchanged when joining the pre-spliceosome to form the pre-B complex. During pre-spliceosome formation, binding of U2 to the BPS releases the 5' end of U2 snRNA, enabling it to pair with an exposed sequence at the 3' end of U6 snRNA. This forms the U2/U6 snRNA stem II within the pre-B complex. The Prp28 DEAD-box helicase releases the 5'SS from the U1 snRNP and transfers it to the ACAGAGA box in the U6 snRNA. Formation of the 5'SS/U6 duplex induces extensive conformational changes and remodelling of the

spliceosome. The RNA helicase Brr2 separates U4 snRNA from U6 snRNA on its U4 snRNA loading site, which was base paired with U6 to form stem III, making it single-stranded and exposed as the U6 half of stem III anneals with the 5'SS. This allows Brr2 to get ready to unwind the U4/U6 duplex for spliceosome active site formation,

The B complex is activated in a subsequent step involving large rearrangements in which the U1 and U4 snRNPs are destabilized or released, yielding the B_{act} complex (Hegele *et al*., 2012). U6 snRNA is extensively base-paired with the U4 snRNA within the tri-snRNP and folds to form the active site of the spliceosome. U4/U6 unwinding allows the U6 snRNA sequence adjacent to the 5'SS-bound ACAGAGA box to fold and associate with part of U2 snRNA to yield the active site harbouring two catalytic metal ions (M1 and M2) (Wilkinson *et al*., 2019). The 5'SS is positioned at the M1 metal ion. When the BP adenosine is joined to the active site, the branching reaction produces the cleaved 5' exon and the lariat-intron intermediate. The 5' exon remains in the active site, but the BP adenosine must leave the active site for the 3'SS to join and the exon-ligation reaction to occur. Finally, the 5' and 3' exons are ligated, and the resulting mature mRNA is released from the active site (Wilkinson *et al*, 2019) (Figure 1b and c).

1.2. Intron density and spliceosomal complexity

Evolutionary reconstructions using maximum likelihood methods suggest intron-rich ancestors for each major group of eukaryotes. For the last common ancestor of animals, the highest intron density of all extant and extinct eukaryotes was inferred at 120-130% of the human intron density (Koonin *et al*., 2012). Furthermore, an intron density within 53-74% of the human values was inferred for the last eukaryotic common ancestor (LECA) (Koonin *et al*., 2012). According to this, the evolution of eukaryotic genes in all lines of descent involved primarily intron loss, with a remarkable gain only at the base of several branches, including

plants and animals. The biological implication of these conclusions is that the common ancestor of all modern eukaryotes was a complex organism with a gene architecture resembling that of multicellular organisms.

That said, intron density differs widely between eukaryotic lineages, from 6 to 8 introns per kilobase (kb) of the coding sequence in vertebrates, some invertebrates, and green plants to only a few introns across the entire genome in many unicellular eukaryotes (Koonin *et al*., 2012). In the yeast *Saccharomyces cerevisiae,* there are approximately 5,000 protein-coding genes, of which approximately 300 contain an intron with an average length between 100 and 400 nucleotides (Parenteau *et al*., 2008). On the other hand, in humans, there are 233,785 exons and 207,344 introns in the 26,564 annotated genes in the human genome (Sakharkar *et al*., 2004). Of the 26,564 genes, 25,877 contain introns, representing more than 97.4% of all genes. On average, there are 8.8 exons and 7.8 introns per gene (Sakharkar *et al*., 2004). Only 1.1% of the human genome is spanned by exons, whereas about 24% is spanned by introns, and 75% is intergenic (Venter *et al*., 2001). About 80% of the exons on each human chromosome are less than 200 bp long, less than 0.01% of the introns are less than 20 bp long, and less than 10% are more than 11 kb long. These numbers demonstrate the challenge for the splicing machinery to splice out very long or short introns (Sakharkar *et al*., 2004). The cell faces a massive challenge in correctly identifying the exons within an ocean of intron sequences and correctly ligating them together (Wilkinson *et al*., 2020).

In addition to this challenge, most of the human introns (95%) are not spliced in the same way every time but alternatively spliced, which means that a single gene can produce multiple protein isoforms either by including or skipping exons or by choosing an alternative exon (Wilkinson *et al*., 2020). This enormously increases the proteome derived from a limited number

of genes, contributing to the incredible complexity of metazoans (Wilkinson *et al*., 2020).

The spliceosome is a protein-rich molecular machine. In humans, more than 200 proteins are associated at one or more stages with spliceosomes assembled on pre-mRNAs. Based on their known function in splicing and the fact that they always copurify together, 141 of these proteins are designated core components of the human spliceosome (Agafonov *et al*., 2011). A set of more than 100 non-core proteins, including mRNA binding and regulatory proteins, is also part of the human spliceosome.

Mass spectrometry studies have revealed that more than 50 snRNP core spliceosomal proteins and more than 100 non-snRNP proteins are contained within the human spliceosome (Fabrizio *et al*., 2009). These studies have shown that the spliceosome's protein complement varies substantially between one splicing stage and another. Under similar conditions, the human and *Drosophila* spliceosomal B complexes contain ~110 proteins, whereas the yeast B complex comprises ~60 proteins (Fabrizio *et al*., 2009). The C complex contains ~110 proteins in most metazoans versus only ~50 in yeast (Fabrizio *et al*., 2009). The total number of proteins per spliceosomal complex in yeast is ~90, representing less than half of the core spliceosomal proteins in humans and other metazoans (Fabrizio *et al*., 2009). Although new proteins are still being found (Lipinski *et al*., 2023), the total number of proteins involved in splicing will probably be less than 100 in yeast, in contrast to the ~170 in humans and ~190 in *Drosophila* (Fabrizio *et al*., 2009). In addition, most of the proteins with critical regulatory roles in alternative splicing are absent in yeast, yielding the conclusion that yeast rarely undergo alternative splicing (Hudson *et al*., 2015).

Hagele *et al*., 2012 performed a study that systematically investigated the protein-protein interactions (PPIs) among 224 human spliceosomal core and non-core proteins. A total of 632

interactions were found, 242 of which were interactions between human spliceosomal core proteins. The splicing process is very complex and dynamic, and what dictates the protein-protein interactions is the large number of components involved. Many non-core human spliceosomal proteins and their interactions are required for alternative splicing. Therefore, these proteins or their interactions are not found in organisms like yeast that do not alternatively splice because they have almost exclusively single intron-containing genes.

Besides this, the *S. cerevisiae* genome contains only 296 introns in 283 genes (out of more than 6,725), accounting for only 5% of yeast genes. As mentioned above, in the vast majority of cases, only one intron per gene is found in yeast, and it is thought that both intron distribution and splicing regulation seem to be much simpler in yeast than in higher eukaryotes. The length of introns in yeast is about 100 - 400 nucleotides compared to the more than 10,000 nucleotide introns found in some human genes (Parenteau *et al*., 2008). All of these attributes make yeast a very suitable organism for studying splicing. Nevertheless, the yeast spliceosome is still a highly complex system for more incisive experiments. Even though the yeast spliceosome has fewer spliceosomal proteins than in humans, it remains complex, with ~90 core proteins and another ~100 associated proteins (Stark *et al*., 2015).

The data presented earlier sharply differ from the organism studied in this thesis. This organism contains only 39 introns in 38 genes out of 5,331 genes (Matsuzaki *et al*., 2004), and there is no evidence of alternative splicing. Furthermore, a significantly reduced spliceosomal machinery has been identified (Stark *et al*., 2015).

1.3. Is splicing essential?

The removal of introns from pre-mRNA and its regulation by alternative splicing are key processes for eukaryotic gene expression and cellular function, as evidenced by the numerous

pathologies induced or caused by splicing alterations. Splicing and alternative splicing play essential roles in all aspects of human physiology (*e.g.* mutations that affect the splicing process of T-cell factors contribute to abnormal T-cell function and the development of autoimmune diseases) (Banerjee *et al*., 2023)

Mutations affecting splicing significantly contribute to the development of rare diseases. Variants such as stop-gained and frameshift gene mutations lead to spliceosomal misfunction, including changes to the canonical splice sites (Lord & Baralle, 2021). Non-coding point mutations causing splicing defects account for approximately 13.5% of hereditary disease alleles. Alterations in splice site recognition are associated with a wide range of human diseases (Fredericks *et al*., 2015). The highly conserved GU/AG motifs mark the beginning and end of 99% of introns. Mutations in these motifs prevent the interaction of core spliceosomal proteins and allow non-coding regions to be transcribed, thereby producing truncated proteins (Fredericks *et al*., 2015). Mutations in RNA species that increase interaction stability with core spliceosomal components have been shown to cause diseases, particularly neurological and muscular degenerative disorders (Fredericks *et al*., 2015). Mutations of splicing factors also contribute to the development of splicing-related diseases. In neurons, the combination of two splicing factors regulates alternative events, and the mutation or loss of either leads to severe pathogenesis (Fredericks *et al*., 2015). Additionally, RBFOX1, a neuron-specific splicing factor, serves as a crucial splicing regulator in the early development of neurons, and its loss negatively affects most neurodevelopmental pathways (Fredericks *et al*., 2015). *In vitro* splicing analyses have demonstrated that the depletion of IKAP, a component of the neuronal Elongator complex, due to mutations in the intron 20 splice sites, which causes its skipping, leads to neurological dysfunction, resulting in Familial Dysautonomia (FD) (Ward & Cooper, 2010).

Several techniques developed to study the effects of missplicing on the development of various diseases have supported the pathobiology of splicing. Computational software such as SpliceAI has been used to predict the impact of variants on splicing trained on pre-mRNA sequence alone. These findings have been supported by experimental methods such as RT-qPCR and mini gene-based confirmations. RNA-sequencing has played a big role in splicing-related disease diagnostics and it has helped clarify the effects of splicing in disease development. A study performed by Cummins *et al.* (2017) showed the effect of splicing in the development of muscle disorders. In 25 patients with muscular dystrophy, RNA-seq methods established a diagnostic for 36% of the cases, confirming that the disease was caused by splicing conditions in all of them (Gonorazky *et al*., 2019).

Although the evolutionary eukaryotic trend goes towards intron-loss, some eukaryotic organisms have retained a small number of introns in specific gene groups, such as the ribosomal protein (RP) genes, despite losing most of their introns. This preservation of introns and the complete splicing machinery required to remove them could be attributed to several factors: (i) the role of introns in regulating gene expression, (ii) the presence of crucial non-coding elements within the introns (e.g., snoRNAs, sisRNAs), and (iii) the constraints of the intron loss process, with the second reason being particularly relevant to this research project. Recent studies in yeast have demonstrated that most of the introns in RP genes have been preserved throughout evolution because they regulate responses to changes in environmental and growth conditions (Parenteau & Abou Elela, 2019). Intron retention occurs in yeast under stress conditions (osmotic stress), suggesting that splicing regulation is essential in yeast (Lukacisin *et al*., 2022).

Studies in yeast revealed that the accumulation of introns helps the organism to cope with nutrient deprivation (Parenteau *et al*., 2019). In yeast, two major categories of upregulated genes

can be identified when intron-deleted strains are exposed to nutrient depletion. One category contains translation-related genes, including RP genes, while the other contains genes associated with cellular respiration (Parenteau *et al*., 2019). The deleted introns specifically upregulate a set of genes associated with ribosome production, which are in part regulated by the nutrient-sensitive target of the rapamycin (TOR) pathway (Parenteau *et al*., 2019). Excised introns enhance the repression of TOR–dependent RP genes during starvation. The introns' effects include inhibiting TOR pathways for repressing ribosomal biosynthesis, which is required for translation in response to nutrient concentration changes (Parenteau *et al*., 2019). Accumulated introns sensitize the TOR nutrient-sensing pathway, which sends the appropriate repression signals to its RP-regulated genes. These examples suggest why introns must be preserved in every eukaryote in at least one gene, and splicing is an essential biochemical process.

For the starvation-dependent repression of ribosome biogenesis to work, only a few genes need to be spliced, and it is not relevant to the gene to which the accumulated introns belong. The only requirement is the accumulation of just one intron in a sufficient amount during the starvation period to sequester the spliceosome and repress the splicing of RP genes, thereby preventing their translation (Parenteau & Abou Elela, 2019). The expression of a group of highly expressed RP genes is enabled by the reduction of the splicing efficiency under starvation conditions in an intron-dependent manner (Parenteau *et al*., 2019). This suggests that splicing is essential for yeast cells to survive under starvation conditions regardless of the significant metabolic cost to a cell of transcribing, excising and degrading introns.

In yeast, introns in ribosomal protein genes resolve the evolutionary conflict between precise expression control and environmental responsiveness. The primary source of debate is whether introns represent a burden or an evolutionary advantage to eukaryotic organisms. Both

sides of the hypothesis have been tested. In exponentially growing cell cultures in rich media, introns confer a high energetic cost and, for instance, reduce the cell growth rate significantly (96% intron deleted strains outcompeted the WT in rich media). Conversely, introns provide a net advantage under stress and starvation conditions (31% of yeast introns enhance growth under stress conditions) (Parenteau & Abou Elela, 2019), highlighting the importance of splicing for cell survival under these conditions.

The information discussed above suggests that splicing is essential for organisms such as humans and yeast. However, considering that some eukaryotes have lost all of their introns, along with the splicing machinery associated with their removal, the question remains whether splicing is essential for organisms with a handful of introns in their genomes and a highly reduced splicing machinery, such as *Cyanidioschyzon merolae*. It has been shown that intron accumulation occurs under stress conditions (high temperature) in *C. merolae*, but the function of those introns still needs to be elucidated. If intron accumulation under high-temperature stress occurs as a regulatory measurement for handling cell survival, then splicing would also be an essential process for the alga. Several other reasons why splicing might be essential for this organism will be discussed below in section 1.5.

1.4. *Cyanidioschyzon merolae* as a model organism

As mentioned, yeast has been an attractive and productive organism for studying splicing, with only around 295 introns located in 280 genes and only nine genes with more than one intron. Nevertheless, the acidothermophile alga *Cyanidioschyzon merolae* was found to have only 39 introns in its genome and a highly reduced spliceosome, which makes it a potentially tractable and simplified organism for investigating intron function and evolution as well as the splicing mechanism.

C. merolae is a small, unicellular organism that inhabits sulphate-rich hot springs (pH three and lower, 42 °C). Of all the photosynthetic eukaryotes whose genomes have been sequenced, *C. merolae* is one of the smallest, with only 16.5 Mbp (Wong *et al*., 2022), containing 4,775 protein-coding genes divided into 20 chromosomes. Widespread intron loss has occurred during the alga's evolution, leaving just 39 introns in 38 genes from a starting pool estimated at 1700 (Wong *et al*., 2022). The small number of introns conserved in *C. merolae* raised the question of whether the full complexity of the canonical splicing machinery has been maintained over the evolutionary process or whether it has lost most of its splicing factors along with its introns (Stark *et al*., 2015). In fact, with only 49 predicted core splicing proteins and only four of the five snRNAs (no U1 has been identified), *C. merolae* appears to have been exposed to intense selective pressure to reduce its spliceosomal complexity, along with its complement of introns (Stark *et al*., 2015). This organism has a dramatically smaller set of splicing machinery than has been found in other organisms. Perhaps due to this reduction, the few remaining introns are spliced inefficiently. As mentioned above, several functions have been discovered for introns besides augmenting proteome diversity by enabling alternative splicing. As in other eukaryotic organisms, any of these functions could be the main reason why *C. merolae* still conserves the remaining 39 introns.

Fortunately for my work, *C. merolae* is a genetically tractable organism. A genetic modification procedure via homologous recombination has been established (Fujiwara *et al*., 2021). Based on this procedure, various genetic techniques such as gene knockout, gene knock-in, stable expression of a transgene without any silencing activity, and inducible/repressible expression of an endogenous gene or transgene have all been developed (Fujiwara *et al*., 2021). To manipulate the *C. merolae* nuclear genome, the uracil synthase (URA) gene and

chloramphenicol acetyltransferase (CAT) gene have been used as selectable markers. When multiple modifications are required on different chromosomal loci in a single strain, they can be achieved by a two-step transformation with URA and CAT markers (Fujiwara *et al*., 2021). This makes *C. merolae* a tractable organism for performing genetic modifications.

1.5. Hypothesis and Thesis Objectives

Since the ancestral red alga is known to have been intron-rich (~1700 introns), this raises the question of why its modern counterpart has conserved only 39 introns and the associated machinery to splice them. Are these introns conserved because they have some important biological function for the cells' survival, or will they eventually disappear? The entire splicing landscape of *C. merolae* may have degraded, so neither the remaining intron-containing genes nor the splicing machinery are needed. This raises the question of whether splicing is essential in this organism. The Rader Lab has demonstrated that splicing occurs in this alga, but, to date, there is no evidence that this process is required for cell viability, either under normal or stress conditions.

I hypothesized that splicing is essential for *C. merolae* to survive. It would be shocking if this were not the case, as splicing is an essential process in all other eukaryotes where it has been studied. The presence of intronic stop codons that produce truncated and non-functional proteins if not removed is a primary reason why splicing is normally essential. Additionally, the Rader lab has discovered sisRNAs (stable intronic sequence RNAs) in *C. merolae* introns (*unpublished*) that may play important roles in various biochemical processes. The presence of the RNase MRP (involved in ribosomal biogenesis) in the intron of CMK142T is also noteworthy. It remains uncertain whether RNase MRP synthesis and processing is splicing-dependent, but if it is, splicing would become an essential process for *C. merolae*. This would explain why, despite its

reduced number of introns and minimal splicing machinery, it still retains the core components of the splicing machinery. In Table 1-1 a list of the genes in *C. merolae* that have been discovered to contain intronic stop codons and their respective function can be seen. Note that many of the genes code for proteins involved in biochemical processes essential for cell survival. If *C. merolae's* intron-containing genes are essential, as their retention suggests, inhibiting core spliceosomal components will lead to deleterious cell effects and cell death. However, if splicing is not essential in *C. merolae*, we will have discovered a novel and exciting phenomenon! By inhibiting essential splicing factors, such as snRNAs and some core spliceosomal proteins (such as Prp8), my aim in this research project was to inhibit the splicing process in *C. merolae* and address the question of whether splicing is essential for these cells to survive. Using RT-qPCR, I analyzed the splicing of 5 intron-containing genes following induction of antisense RNA and protein degradation by degron mechanisms to confirm that my inhibition of splicing was successful.

Table 1-1: Genes that have been discovered to contain intronic stop codons in *C. merolae*. The function of each gene is described.

Gene name (CM)	Gene function
C053C	60S ribosomal protein L35
D067C	Probable prohibitin protein
E034C	Similar to calmodulin
F072C	Glutamate decarboxylase
F136C	Similar to UDP-N-acetylglucosaminepyrophosphorylase
J129C	Histone deacetylase
K245C	ATP phosphoribosyltransferase
K260C	Probable mitochondrial processing peptidase alpha subunit
L049C	Glutamate dehydrogenase
M175C	Similar to divalent cation tolerance protein
O094C	3-isopropylmalate dehydratase small subunit
O257C	Similar to ATP-dependent RNA helicase/Partially identical sequence is found as CMG001C
Q117C	Eukaryotic translation initiation factor eIF-1A
Q163T	Hypothetical transcript
Q270C	Mitochondrial chaperonin hsp60 precursor
Q382C	Similar to U3 snoRNP component Utp15p
R289C	NADH dehydrogenase I (Complex I) iron-sulfur protein 75kDa subunit N-terminal fragment, and the corresponding C-terminus is coded as CMM034C
S262C	60S ribosomal protein L23
S270C	Chaperonin containing TCP1, subunit4 (delta)
S315C	Similar to electron-transfer-flavoprotein, beta polypeptide, mitochondrial precursor
T476C	Probable phosphate transporter Pht2

In chapter two, I describe my attempts to inhibit splicing using antisense RNA. Several antisense RNA techniques have been developed in the past years with the main aim of gene silencing. One of these techniques is RNAi-mediated gene silencing. Due to *C. merolae's* simplicity, the main components of the RNAi mechanism are absent in *C. merolae* cells, making this technique unsuitable. Therefore, I proposed to express the antisense version of spliceosomal genes both transiently from a plasmid and permanently through integration in *C. merolae* cells. It is expected that if the antisense expression is successful, the antisense transcript encounters and base pairs with the endogenous sense transcript of its target, resulting in splicing inhibition and cell death. This is the first time this technique has been employed with spliceosomal genes in *C. merolae*.

In Chapter 3, I explain how I used degron techniques to attempt to inhibit splicing. Degron techniques have been broadly used to achieve protein degradation in several organisms such as zebrafish, humans, *Drosophila,* and bacteria (Willmington & Matouschek, 2016; Banaszynski *et al*., 2006; Caussinus *et al*., 2012; Yamaguchi *et al*., 2019). Therefore, for this dissertation, I proposed targeting core spliceosomal proteins for degradation to inhibit the spliceosome assembly and the splicing reaction. I successfully transformed *C. merolae* cells with genomic constructs containing a degron and attempted to activate protein degradations, therefore splicing inhibition with the small molecule rapamycin. I demonstrated a decrease in splicing and protein degradation in the control strain, but not my experimental ones. The splicing decrease was not, however, sufficient to kill the cells.

2. Chapter 2- Splicing Inhibition by Antisense RNA

2.1. Introduction

Antisense-mediated gene silencing is a post-transcriptional silencing method that uses sequence-specific (antisense) molecules that, through complementary base pairing, suppress the translation of specific target mRNAs (Nielsen & Nielsen, 2013). Gene silencing directed by RNA molecules (RNA interference or RNAi) is one of the best-known pathways for repressing gene expression (Nielsen & Nielsen, 2014). Antisense gene suppression is a powerful tool for analyzing gene function, as the effects generated by silencing a specific target gene can be evaluated in a specific organism. Furthermore, it can be used to silence a particular gene even if it is essential, i.e., when it would not be possible to remove or mutate the gene from the genome or mutate the gene from the genome.

Antisense RNA has been shown to suppress the expression of target genes in *C. merolae* (Ohnuma *et al.,* 2009). However, the *C. merolae* genome does not encode RNAi machinery components such as Dicer. It has been shown that algal species with small nuclear genomes like that of *C. merolae* completely lack Dicer machinery, demonstrating that the RNAi mechanisms have been lost during algal evolution (Casas-Mollano *et al*., 2008). For this reason, gene expression knockdown by antisense RNA has been achieved by transient induction of plasmids encoding the entire antisense sequence of the gene of interest rather than just using antisense RNA oligonucleotides, which activate the RNAi and Dicer machinery (Sumiya, 2014; Ohnuma *et al*., 2009). As *C. merolae* lacks Dicer machinery, it is possible that dsRNAs are not degraded. Instead, the sense/antisense complex could be rendered inert by its inability to fold appropriately, interact with binding partners, or participate in the splicing reaction.

CMP164C and CMS219C are two genes in *C. merolae* that encode similar RNAse III proteins (ribonucleases that cleave dsRNA) (UniProt Protein Database, 2022). Dicer is a type of RNAse III-like protein. The protein encoded by CMS219C seems related to dsRNA processing

and degradation and could be able to process and degrade dsRNA. However, no pathway has been described in *C. merolae* to deal with dsRNA degradation.

I chose U2 and U4 snRNAs, and Cef1 (Figure 2-1 a) and Dib1 (Figure 2-1 b) mRNA as targets for antisense RNA gene knockdown, as they are all essential to pre-mRNA splicing. Within the spliceosome, the snRNAs perform the essential roles of substrate recognition and catalysis (Wilkinson *et al*., 2020). The U2 snRNP recognizes the BP sequence to form the pre-spliceosome or A complex. Cef1 is part of the Nineteen Complex (NTC), which associates with the spliceosome and is essential for the two steps of splicing. The NTC regulates the formation and progression of essential spliceosome conformations required for the two steps of splicing, and it is known as an integral component of spliceosomes from yeast to humans (Hogg *et al*., 2010). Studies have shown that in the absence of the NTC, the U5 and U6 interactions and associations with the active spliceosome are destabilized (Hogg *et al*., 2010). For the spliceosome to be catalytically activated, U4 snRNA must be unwound from U6 snRNA and depart from the spliceosome. Upon U4 snRNP departure, a host of proteins also depart, including Dib1. These departures allow the NTC complex to bind to the spliceosome and perform the needed conformational changes for catalytic activation. Dib1 is essential for preventing premature spliceosomal activation in yeast and *C. elegans.* Dib1 must depart from the spliceosome so that the interaction between the pre-mRNA and U6 and the U5 the loop can occur. These interactions trigger spliceosome activation (Schreib *et al*., 2018).

I tried two different approaches to block the splicing process using antisense RNA. In the first approach, I induced transient gene expression from a plasmid containing the target genes' antisense version (U2 snRNA, U4 snRNA, and Cef1) (Figure 2-2a). In the second approach, I integrated a copy of the antisense version of U2 and Dib1 into a neutral genomic locus in *C. merolae*. For both approaches, targets were under the control of an inducible promoter. The fundamental idea is that our antisense constructs will encounter the endogenously expressed sense version of the same gene and bind to it through base-pair complementarity, blocking any binding sites to other essential splicing factors and inhibiting splicing. As I mentioned, double-stranded RNA is often targeted for degradation, but we do not know whether that is true in *C. merolae*.

a) b)

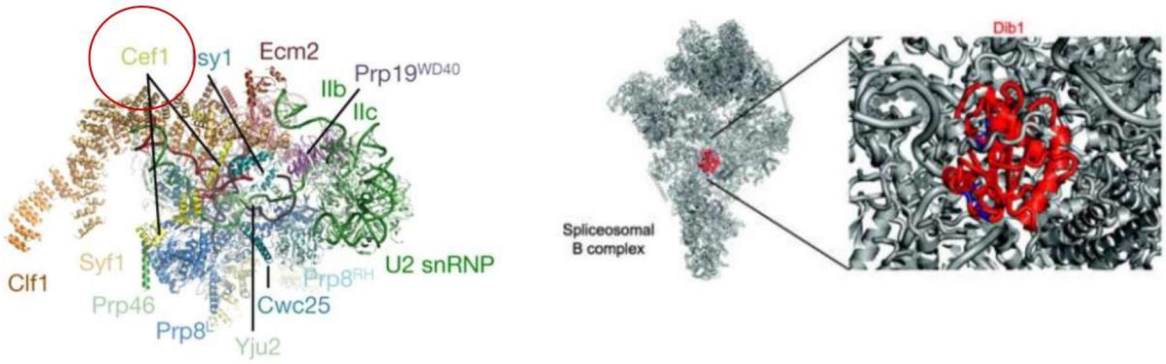


Figure 2-1: a) Cryo-EM structure of the spliceosome immediately after branching, in which Cef1 can be observed as part of the NTC as a core spliceosomal protein (Hogg *et al*., 2010) (Permission to use the figure, cropped from original). b) Cryo-EM structure of Dib1 as part of the spliceosomal B complex (Schreib *et al*., 2018) (Permission to use the figure, cropped from original).

With plasmid transformation for transient gene expression, two markers, CAT and the fluorescent protein mVenus, were included in the vector for transformant selection, whereas for integrated gene expression, only the CAT-selectable marker was included. Selecting for transformed cells relied on antibiotic resistance, with confirmation by cell fluorescence for the transient expression approach and only on antibiotic resistance for the integrated expression approach.

Transient gene expression provides a powerful tool for studying gene function. Transient genes are expressed for short periods after the exogenous nucleic acid, most of the time in the form of a plasmid, has been introduced into a cell. The exogenous genetic material does not fuse with the chromosomal genetic material of the host, resulting in an inevitable loss of the exogenous genetic material after several replication cycles if selection is not maintained. Besides this, inducing transient gene expression under the control of an inducible promoter (which will be activated under certain environmental conditions) is a powerful tool to control at what point and under what conditions the gene will start its expression. Additionally, confirmatory tests were unnecessary for the transient expression approach as the antibiotic resistance selection marker included in the transformed plasmid assured transformed cell survival and death of non-transformed. Overall, this technique offers flexibility and speed for research purposes.

Integrating foreign DNA into an organism's genome can cause genomic destabilization, disrupt cell function, and adversely affect cells if not integrated at the desired locus (Fujiwara *et al.,* 2015). Neutral genomic loci are intergenic regions between coding sequences that do not disrupt endogenous gene expression when exogenous DNA fragments are integrated and are not subject to silencing that would prevent exogenous gene expression. The CMD 184/185 intergenic region is a recognized genomic locus in *C. merolae* that has been extensively used to insert

different types of genes (*e.g.*, antibiotic resistance genes). This locus is short and does not contain promoter activities that would affect gene expression (Miyagishima & Tanaka, 2021; Fujiwara *et al.,* 2015). My target genes were integrated into this neutral locus for the integrated expression approach, taking advantage of *C. merolae's* efficient homologous recombination machinery. Compared to transient gene expression, the advantage of this procedure is that a single copy of the exogenous DNA will be integrated into the targeted genomic locus, assuring just one copy of the antisense version of the target gene and avoiding any overexpression. A 1:1 ratio is expected for endogenous and exogenous gene expression, so there will be one exogenous antisense copy per endogenous sense transcript for the antisense strains. In contrast, when transforming plasmids into cells, the plasmid uptake varies across the population, making it challenging to keep the antisense genes' expression levels uniform. The plasmid uptake affects the number of antisense transcripts, making splicing inhibition more likely in the cells with more plasmid copies. In the Rader Lab, plasmid-transformed cells are not frozen for future use, as it is uncertain if the plasmid will remain intact and unmutated after freezing and thawing, for instance, so when replicating experiments, cells need to be retransformed. With the integration approach, this is not necessary as the foreign DNA is stably integrated into the targeted locus. A disadvantage of this method is that extensive confirmatory tests are required as it is essential to ensure that the integration occurred in the targeted genomic locus and not randomly across the genome.

I needed my transient expression to be inducible for several reasons. First, the antisense versions of our target genes cannot be controlled by a constitutive promoter because we anticipated that antisense expression would cause cells to slow down or die before I could observe and evaluate the effect caused by the expression of the antisense gene. Besides this, studying what happens before and after the induction would be difficult. Second, the fully recovered and selected cells

needed to grow at a standard rate, as for the original strain, and a growth assay needed to be carried out before inducing the antisense gene expression to have good reference data for comparison to the antisense gene expression. Third, RNA samples of the cells must be taken to have comparable data between the RNA composition in the cells before and after induction. This helped us to check for adequate gene expression. Additionally, it was helpful to analyze whether base-pairing occurred between sense and antisense sequences of the target genes, whether dsRNA has been degraded after induction, or if dsRNA was just inert without interacting with other spliceosomal components. Notably, in both cases, splicing was expected to be inhibited. However, this gene must be silent to analyze splicing efficiency before induction. After the antisense target gene expression was induced, splicing efficiency was analyzed, and comparable data was generated.

A nitrogen source-dependent inducible and repressible promoter achieved induced expression (Fujiwara *et al*., 2015; Figure 2-2b). *C. merolae* is usually grown in media containing ammonium, but when ammonium is replaced by nitrate, the cells change their metabolism and adapt. Using nitrate as the primary nitrogen source represents an additional energetic cost for cells as nitrate needs to be reduced to ammonium to be assimilated. Transcriptome analysis in *C. merolae* showed transcriptional induction of essential nitrogen assimilation genes when the cells are switched from ammonium to nitrate media (Fujiwara *et al*., 2014). When this switch occurs, NRT (nitrate/nitrite transporter), NR (nitrate reductase), and NIR (nitrite reductase) genes are stimulated. Using the promoters of these nitrate-assimilation genes, it is possible to regulate gene expression reversibly (Fujiwara *et al*., 2015). These promoters are turned on in nitrate media and are off in the presence of ammonium.

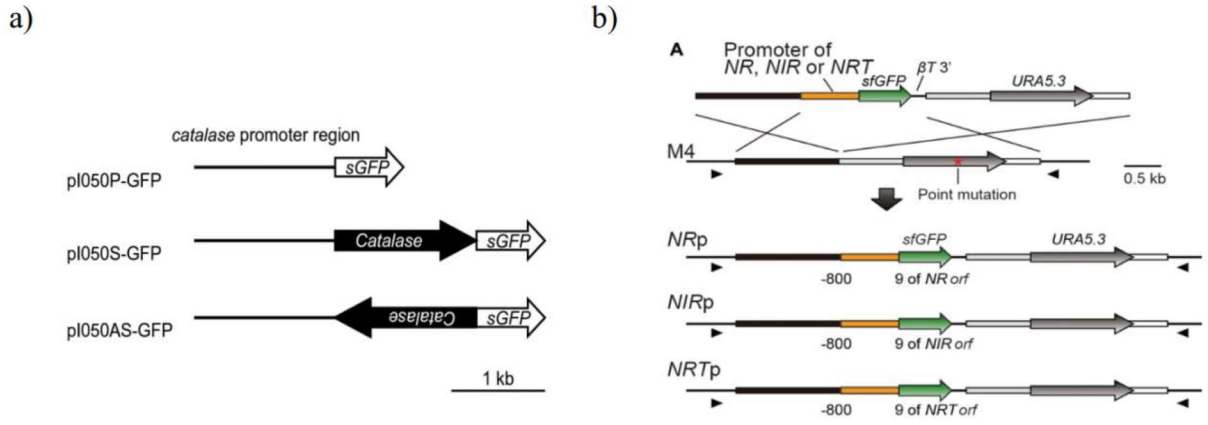
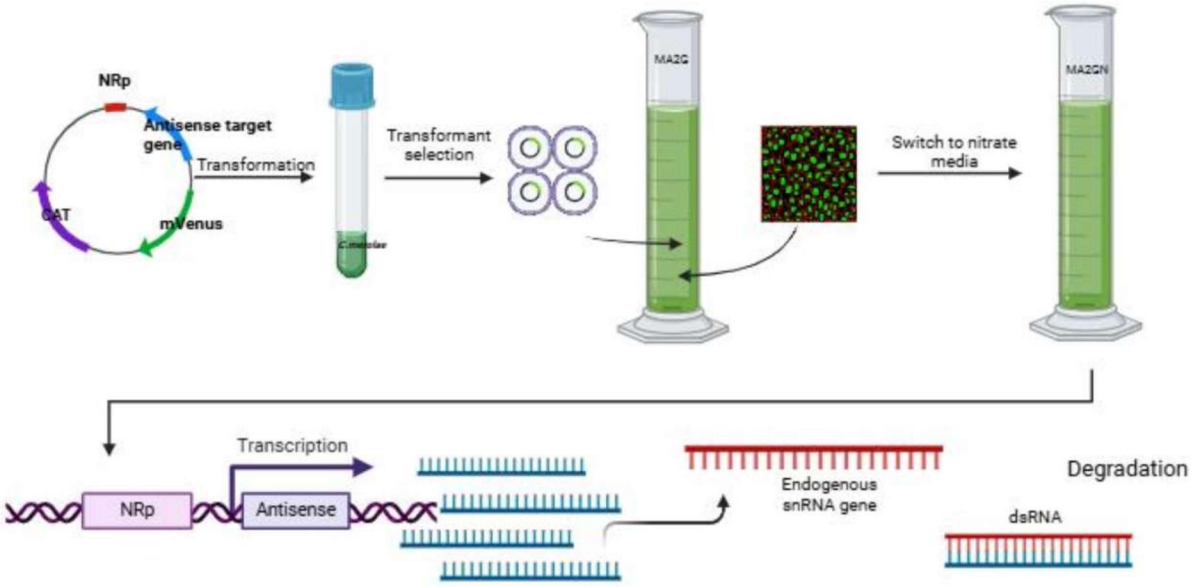


Figure 2-2: a) Plasmids used in Ohnuma *et al*., 2009 for the expression of the antisense RNA of the catalase gene. GFP was fused in frame to the 3' end of the catalase gene as a reporter for transformant selection (CC BY licenced, cropped from original). b) Nitrogen source-dependent inducible and repressible gene expression system DNA constructs used by Fujiwara *et al*., 2015. NRp, NIRp or NRTp + sfGFP constructs were recombined upstream of the URA 5.3 locus in *C. merolae* M4, a strain containing a point mutation in the URA 5.3 gene (CC BY licenced, cropped from original).

For the first antisense approach, I combined the techniques applied by Ohnuma *et al*. in 2009 and by Fujiwara *et al*. in 2015 to create a nitrate-inducible transient gene expression system. The transient vector contained a nitrate inducible promoter driving the expression of the antisense versions of the U2 and U4 snRNAs and the Cef1 mRNA, plus a selectable marker (CAT gene). As a control, vectors containing the sense strand of our target genes were created and transformed along with the antisense constructs to measure the following effects: induced expression, phenotype generated in cells, cell death, and splicing efficiency. According to the Fujiwara study, I chose to use the NRp for the transient gene expression approach because it is the promoter with the weakest activity (Fujiwara *et al*., 2015). In this way, I hoped to avoid the over-expression of the core splicing genes in the control (sense) strains, preventing possible deleterious effects on the cells. As previously mentioned, over-expression of the sense strand may also cause a reduction in

splicing efficiency and negatively affect the cells. Having more copies of a specific snRNA in the cells could lead to an unequal distribution of other splicing factors that interact with that snRNA due to the mismatch between the copies of the snRNA compared to the number of components available for spliceosomal assembly. This can prevent correct spliceosome assembly and consequently inhibit splicing. Besides this, the plasmid copy number per cell is unknown, and using a more potent promoter can cause over-expression in the cells that contain a higher copy number. The backbone plasmid contained the fluorescent protein mVenus (a variation of YFP) under the control of a constitutive promoter to keep track of cell fluorescence during transformant selection. Figure 2-3 shows a detailed schematic representation of how the antisense experiments were designed and are expected to work.

a)



b)

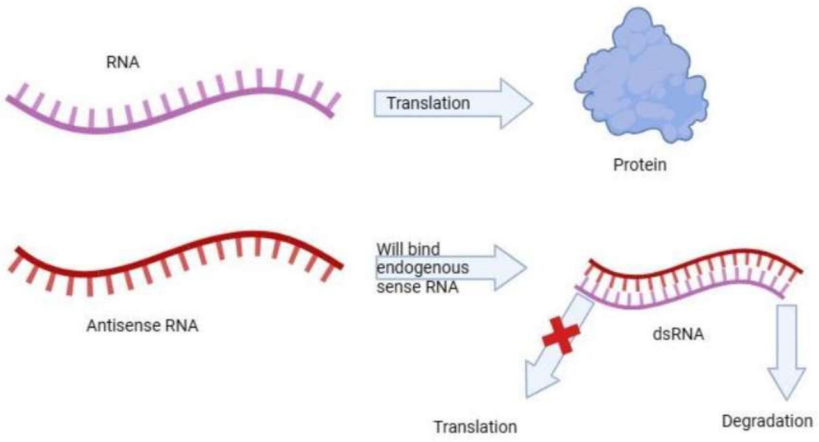


Figure 2-3: a) Experimental setup for the antisense experiments. MA2G= Ammonium media, MA2GN=Nitrate media, NR/NIRp= promoters. b) Representation of what is expected inside *C. merolae* cells and how splicing will be inhibited by antisense mRNA targeting splicing proteins (Cef1).

For the second approach, just the inducible gene expression described by Fujiwara *et al*., 2015 was applied. The NIR promoter was chosen for this approach. As mentioned before, NIR is the strongest promoter of all the nitrate-inducible promoters and was selected to ensure the antisense gene expression without considering the possible over-expression of the control genes. Linear DNA constructs containing the sense and antisense versions of the genes, the CAT selectable marker, and homologous recombination arms for the CMD184/185 neutral loci were designed for transformation into *C. merolae* cells. It was expected that homologous recombination would occur at the targeted locus, resulting in the integration of the constructs. In this approach, copy number is a minor concern as it is unlikely for the integration to occur randomly across the genome. Nevertheless, several screening and confirmatory tests were performed to ensure the desired results.

If the antisense RNAs were expressed, I expected to see a decrease in cell growth and possibly cell death. On the other hand, no decrease in cell growth or cell death was expected for the stains expressing the control sense strands of the genes. Conversely, the absence of a phenotype would suggest that splicing is not essential so long as I could demonstrate the expected molecular effect, namely that the target transcript was degraded or inactivated.

2.2. Materials and Methods

2.2.1. Preparation of *C. merolae* genomic DNA

C. merolae 10D strain (NIES-1332) was provided by the Microbial Culture Collection at the National Institute of Environmental Sciences in Tsukuba, Japan (mcc.nies.go.jp). *C. merolae* genomic DNA was prepared by Martha Stark as described in Stark *et al*. (2015).

2.2.2. Construction of antisense RNA expression vectors

2.2.2.1. Amplification of genomic sequences of the genes of interest by Polymerase Chain Reaction (PCR)

I performed PCRs (polymerase chain reaction(s)) from *C. merolae* genomic DNA for the NR promoter (NRp) and the UBQ 3'UTR terminator from G019C and K296C, respectively. The NRp PCR reaction was performed with the oligos oSDR2191 and oSDR2192 (Appendix 1) using Q5[1] DNA Polymerase (following the manufacturer's guidelines). This primer pair's annealing temperature was 67 °C, and the extension time was 1 minute. The UBQ 3' UTR terminator PCR was performed with oSDR2193 and oSDR2194 (Appendix 1) using Q5 DNA Polymerase. This primer pair's annealing temperature was 68 °C, and the extension time was 30 seconds.

[1] The manufacturer's (New England Biolabs) recommended extension time for Q5 is approximately 30 seconds per kilobase of DNA (kb).

I ran the PCR products on a 1% agarose gel to check for the correct band size. All PCR products along this work were cleaned up with E.Z.N.A Cycle Pure Kit from Omega Bio-Tek to remove primers and dNTPs from the PCR reaction, and their concentrations were measured with the nanodrop spectrophotometer (NanoDropOne^{c} from Thermofisher Scientific).

I performed PCR reactions for U2 and U4 sense and antisense snRNAs (unannotated genes, but location reported in Stark *et al.*, 2015) and Cef1 sense and antisense (R098C) from *C. merolae* genomic DNA as previously described for NRp and UBQ 3' UTR. The primers used and their melting temperatures were as follows:

Table 2-1: Primers used for the amplification of U2, U4 and Cef1 sense and antisense.

Gene	**oSDR number (F/R)**	**Annealing temperature[2] (°C)**
U2 sense	2213/ 2214	64
U2 antisense	2215/ 2216	64
U4 sense	2205/ 2206	66
U4 antisense	2207/ 2208	66
Cef1 sense	2201/ 2202	68
Cef1 antisense	2203/ 2204	68

*I designed all these primers according to the primer design protocol for LIC (Appendix 3). For their description and sequence, refer to Table A1 in Appendix 1.

The extension times for U2, U4, and Cef1 sense and antisense were 15 seconds, 20 seconds, and 2 minutes, respectively.

I ran U2 and U4 PCR products in a 1.5% agarose gel for 40 minutes and Cef1 in a 1% agarose gel for 40 minutes. I then purified the PCR products and measured their concentration.

[2] This annealing temperature corresponds to Q5 DNA Polymerase. The annealing temperatures for the primers vary between the enzymes used for the PCR reaction (i.e.: the annealing temperature for Taq DNA Polymerase for the same primer pair is different from the one for Q5).

2.2.2.2. Insertion of the NR promoter into the transformation vector (pSR1008)

Following the previous procedure, I performed restriction enzyme digestion of the plasmid backbone pSR1008 and the NRp PCR product using XbaI and NcoI (Appendix 2).

Following the digestion reactions, I ran a 2.5 uL sample of the digested backbone plasmid pSR1008 in a 0.7% agarose gel for 2.5 h with an undigested plasmid sample as a control. Following the confirmation of correct digestion, I ran the remainder of the digestion reaction on a 0.7% agarose gel for subsequent gel purification with the E.Z.N.A. Gel Extraction Kit from Omega Bio-Tek (all gel purifications were performed with this kit). The concentration was then measured with the nanodrop. The NRp PCR product was purified to remove the cut pieces produced by the restriction enzymes. Its concentration was measured with the nanodrop. Once the two products were ready, I performed the first cloning reaction following the T4 DNA Ligation and Transformation protocol (Appendix 4) (Figure 2-4).

After incubation, I counted and picked individual colonies and inoculated them in 3 mL liquid LB + 1X ampicillin (stock at 1000X) overnight. I extracted plasmid DNA using the E.Z.N.A. Plasmid DNA Mini Kit I from Omega Bio-Tek (all plasmids from this work were purified as described). The plasmid concentration was measured with the nanodrop.

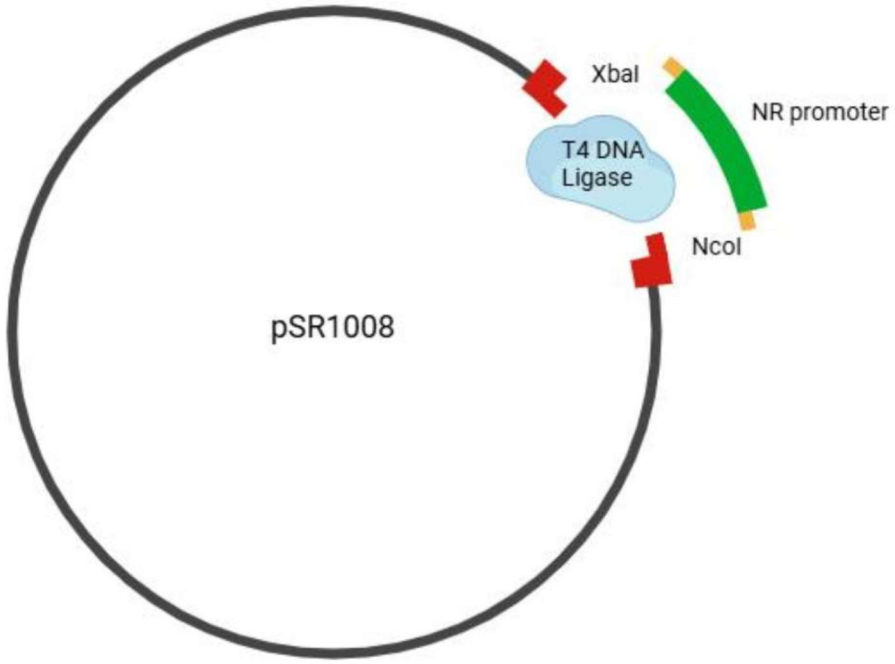


Figure 2-4: Simplified representation of the cloning procedure for the NR promoter into pSR1008.

To check for the correct insertion of the NRp into pSR1008, I performed two restriction enzyme digestions in the plasmids extracted from the previous step with XbaI and NcoI and with EcoRI. I digested 500ng of pSR1008 + NRp and 500ng of pSR1008 with XbaI and NcoI and incubated the digestion for 15 minutes at 37 °C. On the other hand, I digested 200 ng of pSR1008 + NRp and 200 ng of pSR1008 with EcoRI and the digestion reaction was incubated for 15 minutes at 37 °C. I ran the two digestion reactions in a 1% agarose gel for 90 minutes.

2.2.2.3. Insertion of the UBQ 3' UTR terminator into pSR1008 + NRp

The UBQ 3' UTR terminator was inserted into pSR1008 following the standard procedures described in the previous section with the modifications shown in Figure 2-5.

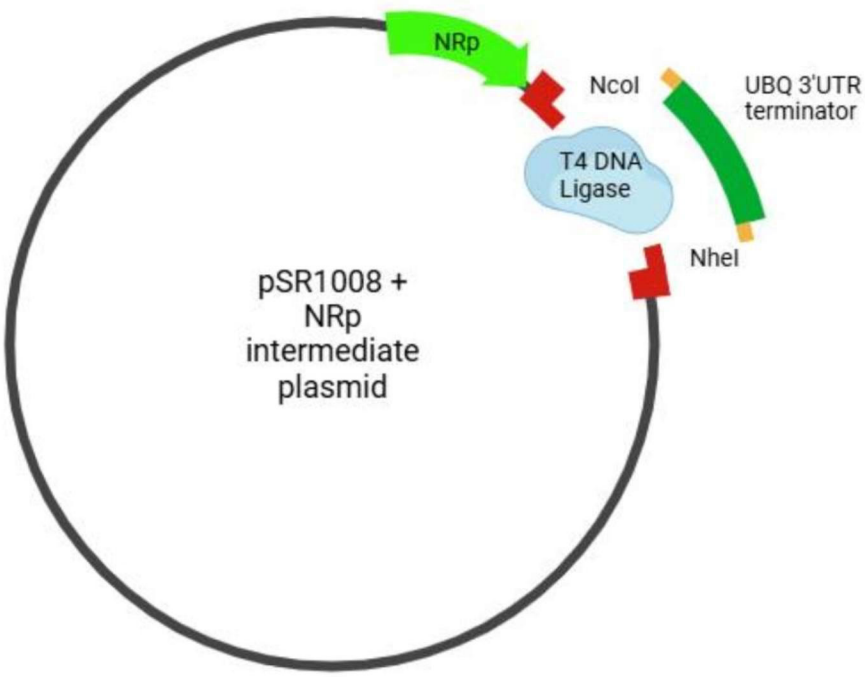


Figure 2-5: Simplified representation of the cloning procedure for the insertion of the UBQ 3' UTR terminator into pSR1008 + NRp intermediate plasmid.

For plasmid confirmation, I digested 500 ng of pSR1025 and 500 ng of pSR1008 + NRp (intermediate plasmid) with NcoI and NdeI. I incubated the reaction for 15 minutes at 37 °C and ran it on a 1% agarose gel for 1 hour. In this step, I prepared a glycerol stock from the remaining liquid bacterial culture corresponding to the correct plasmid for freezing at -80 °C. For the latter, I mixed 80 uL of DMSO (Dimethyl Sulfoxide) and 1 mL of the liquid bacterial culture in a cryotube labelled with the given name for the plasmid (pSR1025).

2.2.2.4. Insertion of the sense and antisense versions of the target genes into pSR1025

I digested 500 ng of pSR1025 with PacI and performed the T4 DNA polymerase treatment on the digested plasmid and the target genes PCR products as described in the LIC Protocol in Appendix 3.

After T4 DNA Polymerase treatment, I set up the LIC reaction following the procedures in Appendix 3. DH5 alpha-competent cells were transformed and plated as described in previous sections. I counted colonies, picked individuals and inoculated them in liquid media (LB) with selection (1X ampicillin). Plasmid DNA was extracted and digested for confirmation.

2.2.2.5. Confirmation of transient gene expression vectors

I digested 500ng of each plasmid using HindIII for pSR1025+U2 sense and antisense and pSR1025+Cef1 sense and antisense, and XbaI and NcoI for pSR1025+U4 sense and antisense following standard procedures. I ran the digestion reactions on a 0.7% agarose gel for 1 hour. Glycerol stocks were prepared from these plasmids.

Table 2-2: Final plasmid names for the sense and antisense experiments engineered vectors.

Plasmid composition	Given name
pSR1025 + U2 sense	pSR1028
pSR1025 + U2 antisense	pSR1029
pSR1025 + U4 sense	pSR1030
pSR1025 + U4 antisense	pSR1031
pSR1025 + Cef1 sense	pSR1034
pSR1025 + Cef1 antisense	pSR1035

2.2.3. Construction of genetic constructs for genomic integration by homologous recombination in the D184/185 neutral locus

As for the transient gene expression approach, the first step was performing the PCRs (polymerase chain reaction(s)) from *C. merolae* genomic DNA for U2 sense and antisense (unannotated) and Dib1 sense and antisense (S018C). I performed the PCR reactions with Q5 DNA Polymerase. The primers used for these PCR reactions and their corresponding annealing temperatures were the following:

Table 2-3: Primers used to amplify sense and antisense U2 snRNA and Dib1 sense and antisense mRNA.

Gene	oSDR	Annealing temperature[3] (primer pair) (°C)
U2 sense	F: 2523 R: 2524	53
U2 antisense	F: 2525 R: 2526	64
Dib1 sense	F: 2542 R: 2543	71
Dib1 antisense	F: 2544 R: 2545	71

*For primer sequences, please refer to Appendix 1.

I ran the U2 snRNA PCR products on a 1.5% agarose gel for 35 minutes and the Dib1 PCR products on a 1% agarose gel for 40 minutes. I then purified them, and their concentration was measured with the nanodrop.

[3] This annealing temperature corresponds to Q5 DNA Polymerase. The annealing temperatures for the primers vary between the enzymes used for the PCR reaction (i.e.: the annealing temperature for Taq DNA Polymerase for the same primer pair is different from the one for Q5) and should be calculated for every specific case. NEB Tm Calculator was used for this project.

The U2 and Dib1 sense and antisense were inserted into the plasmid backbone pSR979 following PCR amplification. I performed restriction enzyme digestion of pSR979 and the U2 and Dib1 PCR products using AflII and NheI restriction enzymes (Appendix 2). The following cloning steps were performed as described in previous sections following standard procedures (Figure 2-6). Table 2-4 shows the numbers of the resulting plasmids.

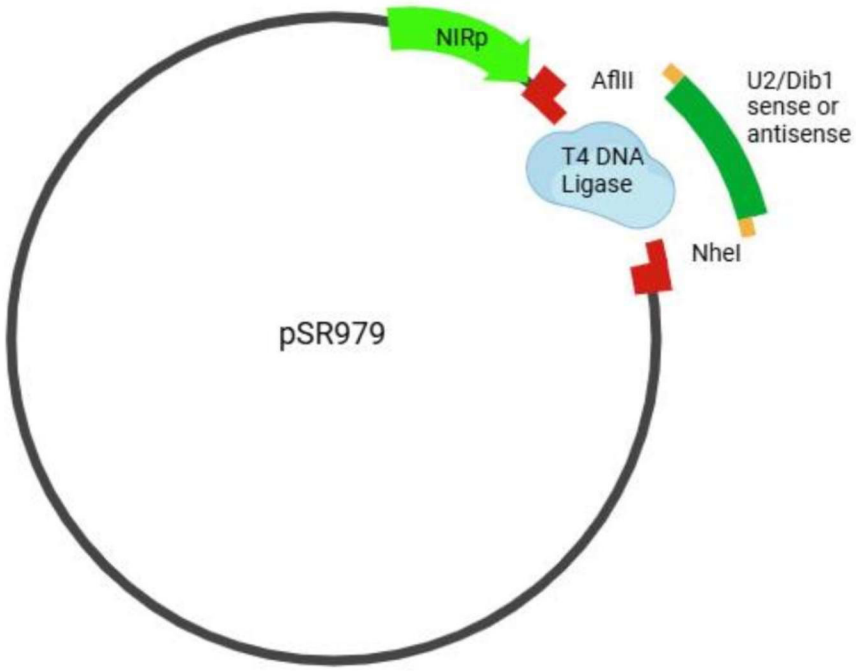


Figure 2-6: Simplified representation of the cloning procedure followed for the insertion of U2 and Dib1 sense and antisense genes into pSR979.

Table 2-4: Plasmid pSR numbers for the resulting engineered U2 and Dib1 sense and antisense plasmids for genomic integration of *C. merolae* cells.

Description	Plasmid pSR number
U2 sense + CAT	1096
U2 antisense + CAT	1097
Dib1 sense + CAT	1099
Dib1 antisense + CAT	1100

To check for the correct insertion of the sense and antisense versions of U2 and Dib1 into pSR979, I performed restriction enzyme digestions for pSR1096/97 and pSR1099/1100, with HindIII and AatII and XbaI and KpnI, respectively. The digestions were performed following standard procedures using pSR979 as a control. I prepared glycerol stocks from the confirmed plasmids as previously described.

2.2.3.1. Amplification of the sequence of interest for genomic integration from engineered vectors by PCR

I performed a PCR reaction to amplify the linear integration construct for transformation and genomic integration from each plasmid containing the target genes (the construct needs to be linear for homologous recombination to occur). I used oSDR2440 as a forward primer targeting the 5' HRA and oSDR2441 as a reverse plasmid targeting the 3' HRA (for primer sequences, refer to Appendix 1). I used PaCeR Polymerase (Gene Biosystems) for this PCR reaction. I prepared the reaction following the manufacturer's guidelines using an annealing temperature of 65 °C and an extension time of 4.5 minutes (1 minute per kb).

I ran 2.5 uL of the PCR reactions on a 0.7% agarose gel to confirm their correctness. I then purified the PCR reactions and measured their concentration with the nanodrop.

2.2.4. Transformation of plasmids and linear constructs into *C. merolae* cells

I transformed the antisense experiment plasmids and linear constructs for transient gene expression and integrated gene expression, respectively, following the PEG-mediated transformation protocol described by Ohnuma *et al*., 2008 with some modifications (Appendix 5).

To prepare *C. merolae* cells for transformation, I grew wild-type cells (WT) in 50 mL MA2G media in a graduated cylinder at 42 °C, 90 umol photons/ m^2s, and 2% CO_2 until an OD_{750} of 0.8 to 1. When the cells were ready to be transformed, I performed the subsequent steps of the transformation protocol.

I carried out the transformations of the U2 sense and antisense, U4 sense and antisense, and Cef1 sense and antisense plasmids independently, one gene at a time. For each gene, I performed four transformations: sense plasmid, antisense plasmid, empty vector control (pSR1025), and control (no DNA, plasmid replaced with water).

I independently transformed U2 sense and antisense and Dib1 sense and antisense linear constructs for genomic integration, one gene at a time. I performed three transformations for each gene: sense gene construct, antisense gene construct, and control (no DNA).

2.2.5. Cell acclimation and transformant selection

For the acclimation procedure, I inoculated PEG-transformed cells in 50 mL of fresh liquid MA2G media in a graduated cylinder at 42 °C, 90 umol photons/m^2s, and 2% CO_2 without

selection (no antibiotic) for 24 hours. After the 24 h period, I either reinoculated the cells in fresh liquid media with selection (250 ug/mL chloramphenicol (Cp)) for plasmid transformation or plated them in solid MA2G + Cp for linear construct transformation (genomic integration).

2.2.5.1. Plasmid transformed cells

After acclimation, the procedures explained in Appendix 5 section Day 2 were followed for transformant selection. As described in Appendix 5, when working with plasmid-transformed cells, the media must be replaced every three days, and fresh selection needs to be added. This ensures proper transformant selection and prevents the cells from losing the plasmid after several replication cycles.

As the plasmids engineered for transient gene expression experiments contained the fluorescent protein mVenus, I performed fluorescence microscopy on day 12 after transformation.

2.2.5.2. Genomic integrated cells

Plating in solid MA2G + Cp was required. I plated the transformations as described in the CAT transformation protocol in Appendix 5, section Day 2.

After plating, colony screening needs to be performed. I performed the colony screening as described in the Colony PCR to Test for Homologous Integrants protocol in Appendix 6. I ran PCR reactions with Taq DNA Polymerase from New England Biolabs following the manufacturer's guidelines. I used two sets of primers to screen the colonies. The primers used were the following:

Table 2-5: Primers used for colony screening by PCR to check for correct genomic integration of my target genes.

Gene	oSDR number	Annealing Temperature (Taq DNA polymerase) (°C)
CMD184	Forward: 1818	59
APCC promoter (CAT Gene)	Reverse: 1816	52
Beta -tubulin terminator	Forward: 1817	59
CMD185	Reverse: 1820	59

*For primer sequences, please refer to Appendix 1.

I set the annealing temperature for oSDR1818 and 1816 at 47 °C and for oSDR1817 and 1820 at 54 °C, with an extension time of 1 minute.

A colony PCR is less efficient than a DNA template PCR. For instance, the product for colony screening should not be bigger than 1kb. This is why the two integration ends must be screened separately. oSDR1818 and 1816 were used to screen the 5' end integration, and oSDR1817 and 1820 were used to screen the 3' end integration. I ran the PCR products on a 1% agarose gel for 40 minutes.

After the colony screening step, I let the correct colonies grow in a 48-well tissue culture plate for a few days. Then, I isolated genomic DNA using the Quick Genomic DNA Isolation protocol (Appendix 7).

I performed confirmatory PCR reactions for the final correct integrant selection. I prepared three reactions: one spanning the 5' end of the integration locus, one spanning the 3' of the integration locus, and one spanning the whole integrated locus (from the 5' to the 3' end). I performed the PCR reactions with Taq DNA Polymerase. The primers used were the following:

Table 2-6: Primers used for confirmatory PCR reactions to check for correct genomic integration of my target genes

Gene	oSDR number	Annealing Temperature (Taq DNA polymerase) (°C)
CMD184	Forward: 1818	59
NOS Terminator	Reverse: 2480	55
CAT	Forward: 1524	61
CMD185	Reverse: 1820	59

*Please refer to Appendix 1 for the primer sequence. These primer sets were used in three combinations to cover the previously mentioned regions.

For the 5' end integration screening, I used oSDR1818 and 2480 with an annealing temperature of 50 °C and an extension time of 2.5 minutes. For the 3' integration screening, I used oSDR1524 and 1820 with an annealing temperature of 54 °C and an extension time of 1 minute. For the 5' to 3' end screening, I used oSDR1818 and 1820 with an annealing temperature of 54 °C and an extension time of 5 minutes. I included a WT genomic DNA control for each primer pair. This allowed the comparison of the integrated locus with the WT locus. I ran the PCR on a 0.7 to 1% agarose gel. I selected the colonies that appeared correct for the three sets of primers for further sequencing and downstream experiments and chose two biological replicates per gene.

For sequencing, I performed a PCR using oSDR1818 and 1820. I cleaned up the PCR product and measured the concentration with the NanoDrop. oSDR1588, 1818, and 2480 (Appendix 1) were used for sequencing the integrated locus. This step was performed as an extra confirmatory test for adequate integration and for discarding any possible mutations in the DNA sequence that might have occurred during the cloning procedure.

2.2.6. Growth assays in MA2G

To evaluate if the transformed plasmids in *C. merolae* cells had any phenotypic effect and to discard the possibility of a leaky nitrate inducible promoter (promoting expression of the

antisense genes even under nitrogen replete (ammonium) conditions), I performed growth assays in MA2G + Cp media for each of the resulting strains (U2 sense, U2 antisense, U4 sense, U4 antisense, Cef1 sense, Cef1 antisense, pSR1025 control, and WT control).

I set up the growth assay for U2 sense and antisense strains with the pSR1025 empty vector control and the WT control in 50 mL graduated cylinders at 42 °C, 2% bubbling CO_2, and 90 umol photons/m^2s. I grew the *C. merolae* strains to an OD of 1 the day prior to the start of the growth assay. For the growth assay, I used a starting OD of 0.05. After setting up the initial cultures and putting them in the incubator under the conditions previously mentioned, I measured the OD after 4, 8, 12, 24, 28, 32, 36, and 48 hours. I then calculated the generation time from the data collected. The generation time was determined by calculating the natural logarithm (LN) of each optical density (OD) measurement at different time points for each strain and each technical replicate. Subsequently, the following formula was applied: LN (2) / SLOPE (x, y), where x represents the natural logarithm of the OD measurements and y represents the corresponding time points at which these measurements were recorded. An average of the three technical replicates per strain doubling times was then calculated for plotting.

I performed the growth assays for U4 and Cef1 sense and antisense and the respective pSR1025 and WT controls (U4 and Cef1 were done independently) as previously described for U2 with the following modifications: I used glass tissue culture tubes and a roller drum instead of graduated cylinders; in this case, CO_2 was not directly injected into the cultures, so the cells depended on the CO_2 present in the incubator's environment; I set up the initial OD to 0.2.

2.2.7. Nitrate induction and growth assays in MA2G nitrate (MA2GN)

2.2.7.1. Transient gene expression experiments

For the induction of the U2, U4, and Cef1 sense and antisense strains, I prepared 60 mL of a fresh culture per strain and let it grow up in MA2G + Cp to OD 1 for the day of the induction. I spun down 58 mL of cells at 2.000xg for 10 minutes and removed the media. I washed the cells 3x in nitrate media to remove any remaining MA2G. After the final wash, I resuspended the cells in 600 uL of MA2GN (nitrate).

I performed the nitrate induction in triplicate (three technical replicates per strain). I used pSR1025 (empty vector), NIRCas9, and WT strains as controls. I prepared 50 mL graduated cylinders with 50 mL MA2GN + Cp media (except for WT, for which no selection was used). I equally split the previously washed 600 uL of cells per strain into three graduated cylinders. As a result, I had three technical triplicates per strain with an initial OD of ~0.4. I then incubated the cells at the same conditions mentioned above. I collected samples for RNA extraction after 4, 8, 12, 20, 24, 30, 42, 48, and 72 hours for U2 sense and antisense induction with controls, and after 4, 10, 24, 28, 34, and 48 hours for U4 and Cef1 sense and antisense induction with controls. I collected the samples from the NIRCas9 strain only at 24 and 48 hours for Western Blotting and protein detection. I used the remaining 2 mL from the starting culture (previous to induction) as uninduced controls for RNA extraction. I measured the OD for each culture at the same time points as the sample collection and developed a growth curve from these data. After 48 hours, I diluted the cultures to an OD of 0.4 and let them grow under the same conditions used across the experiment for seven more days, adding fresh selection (Cp) every three days.

2.2.7.2. Integrated gene expression experiments

I performed the nitrate induction for the integrated gene expression strains, U2 and Dib1 sense and antisense, following the same procedure as for the transient gene expression experiments with the following variations: the experiment was carried out at a smaller scale, using 6-well tissue culture plates and a shaker inside the incubator, I collected samples for RNA extraction only at 24 and 48 hours and measured the OD at the beginning of the experiment and after seven days, and I set up the initial OD for the cultures to 1.5.

2.2.8. RNA isolation and quality evaluation

I isolated RNA for the induced and uninduced samples and controls according to the "RNA Isolation – Cm cold/hot phenol" protocol (Appendix 8), except for the U4 sense and antisense 24-hour samples and their respective controls, in which I isolated the RNA using HiPure™ Total RNA Mini Kit from Gene Biosystems.

I evaluated RNA quality by running a bleach gel with 500 ng RNA per sample following the protocol "RNA Bleach Gel" (Appendix 9).

2.2.9. Northern Blotting

I analyzed sense and antisense RNA induction for transient gene expression and integrated gene expression strains, as well as for the empty vector (pSR1025) and WT controls, by fluorescent northern blot.

I performed Northern blots according to the protocol Fluorescent Northern blots in Appendix 9 with the following modifications: 1) for U2 and U4 snRNA samples from antisense induction of transient gene expression strains, I poured 6% 7 M urea (denaturing)

polyacrylamide gels and loaded 2 ug total RNA; 2) for U2 snRNA samples from antisense induction of integrated gene expression strains, I poured a 6% 7 M urea polyacrylamide gel and loaded 5 ug total RNA per sample; 3) for Cef1 antisense induction 34-hour time point samples, I poured a 6% 7M urea polyacrylamide gel and loaded 20 ug RNA per sample, I increased the gel running time from 45 minutes to 90 minutes, and I increased the transfer time from 30 minutes to 45 minutes; 4) for Cef1 antisense induction 24-hour time point samples I poured a 1.5% formaldehyde (denaturing) agarose gel (refer to section Agarose Northern blots in the protocol) and loaded 20 ug total RNA per sample, I ran the gel for 6.5 hours, and let the capillary transfer overnight; 5) for Dib1 antisense induction 24-hour time point samples, I applied the same conditions as for the Cef1 antisense induction 34-hour time point samples; 6) for Dib1 antisense induction 48-hour time point samples, I applied the same conditions as for the Cef1 24-hour time point samples, except that I loaded 25 ug total RNA per sample; 7) I used a yeast snRNA ladder and U2 or U5 snRNAs as loading controls for Cef1 transient gene expression northern blots and for U2 and Dib1 integrated gene expression northern blots.

I performed Northern blots stepwise, probing first with the sense or antisense probe for the target genes, stripping the blot (if bands appeared), and reprobing with the remaining probe. I included the loading control probe for both probing steps. I loaded four ug of the yeast snRNA ladder into the first lane of the polyacrylamide or agarose denaturing gels. The ladder was independently probed with probes against yeast U1, U2, U4, U5 and U6 snRNAs (see Appendix 1 for probe sequences).

2.2.10. Gene-specific primer reverse transcription and PCR

For the integrated gene expression antisense induction of U2 and Dib1 samples and controls, I selected the RNA of one technical triplicate for each strain and performed a gene-

specific reverse transcription with the BioRad Reliance Select cDNA Synthesis Kit. I performed the reverse transcription with the following gene-specific primers for each strain:

Table 2-7: Gene-specific primers for integrated gene expression antisense induction samples cDNA synthesis.

Gene	oSDR number
U2 sense	2524
U2 antisense	2526
Dib1 sense	2543
Dib1 antisense	2545
CMK260C sense	FUB185
CMK260C antisense	FUB184

*For primer sequence, please refer to Appendix 1.

I set up and performed the reactions according to the manufacturer's protocol. I included a non-reverse transcriptase (NRT) control per sample and one non-template control (NTC) per gene-specific primer reaction set.

Using cDNA as a template, I performed an end-point PCR with specific primer pairs for each gene. I did this to evaluate the RT reaction's efficiency and ensure proper cDNA synthesis. I used the following primer pairs:

Table 2-8: Gene-specific primer pairs used for endpoint PCR with U2 and Dib1 sense and antisense cDNA + controls.

Gene	oSDR number
U2 sense	F:2523 R:2524
U2 antisense	F:2525 R:2526
Dib1 sense	F:2542 R:2543
Dib1 antisense	F: 2544 R: 2545
CMK260C sense	F: FUB184 R:FUB185
CMK260C antisense	F: FUB185 R:FUB184

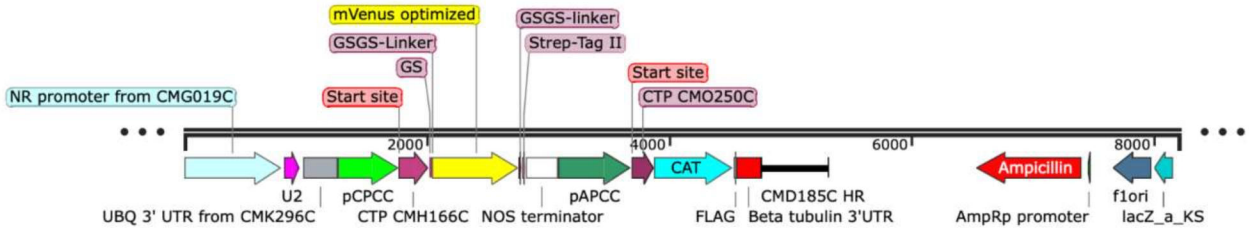
*For primer sequence, please refer to Appendix 1.

It is important to clarify that the reverse transcription reactions were performed with the U2 sense gene-specific primer on the sense, antisense, and control strains RNA and with the U2 antisense gene-specific primer on the sense, antisense and control strains RNA. The same was true for the Dib1 sense and antisense strains and controls.

2.3. Results

2.3.1. Inhibition of splicing factor expression via transient (episomal) antisense expression

a)



b)

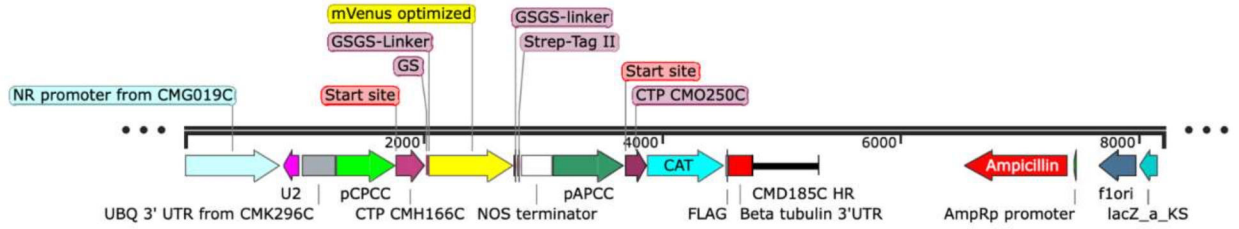


Figure 2-7: Plasmids for the transient inducible gene expression of the antisense and sense U2 snRNA (pink). Note the CAT (sky blue) and mVenus (yellow) markers. a) Vector containing the sense version of U2 snRNA (pSR1028). b) Vector containing the antisense version of U2 snRNA (pSR1029). Note: The plasmids for U4 snRNA and Cef1 are the same except for the target gene.

To attempt to knock down splicing by using antisense gene expression to inhibit the expression of splicing factors, I generated plasmids expressing antisense U2 snRNA, U4 snRNA, and Cef1 under the control of an inducible promoter. Figure 2-7 represents the plasmids' composition after completing all cloning steps.

2.3.1.1. Cloning

I constructed nitrate-inducible, plasmid-borne antisense genes and controls for *C. merolae* transformation and nitrate induction for splicing inhibition attempts. I confirmed successful construction by restriction digest (Figure 2-8).

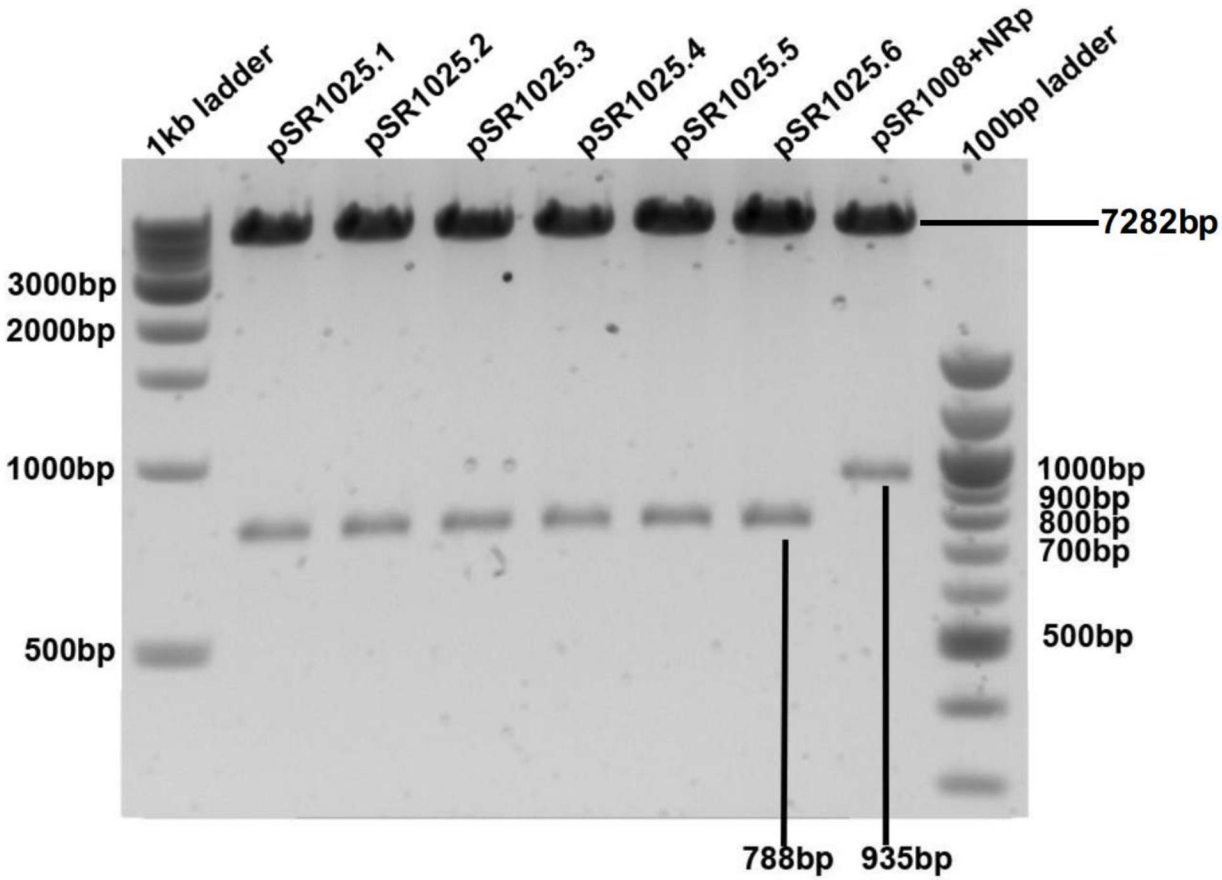
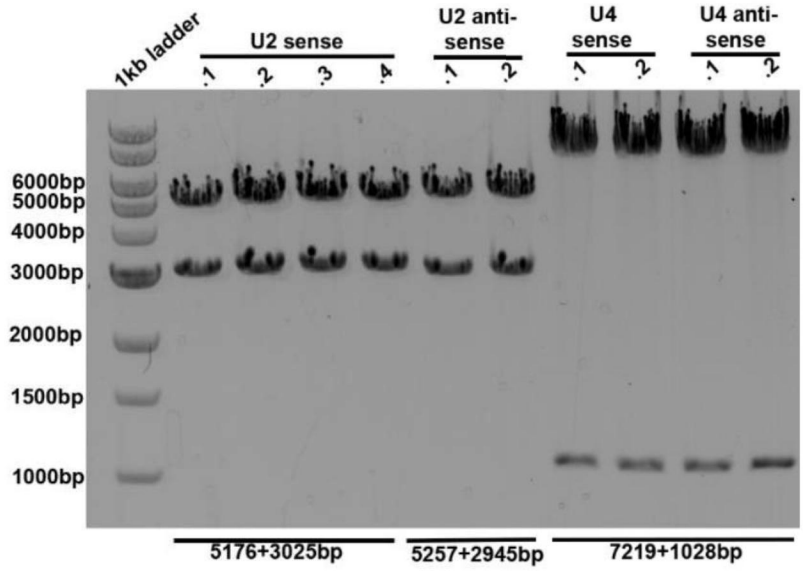


Figure 2-8: Restriction enzyme digestion checks for pSR1025. The digested plasmids with NdeI and NcoI were run on a 0.7% agarose gel stained with EtBr for one hour. The expected band sizes on the gel after restriction enzyme digestion with NdeI and NcoI for the proper UBQ 3' UTR insertion were 7282bp and 788bp (lanes 2-6). The intermediate plasmid pSR1008+NRp was used as a control to compare band sizes to confirm the insertion. Without the insertion, the expected band sizes were 7282bp and 935bp (lane 7).

Figure 2-9 a) and b) show the successful insertion of the target genes into the backbone plasmid after restriction enzyme digestion checks.

a)



b)

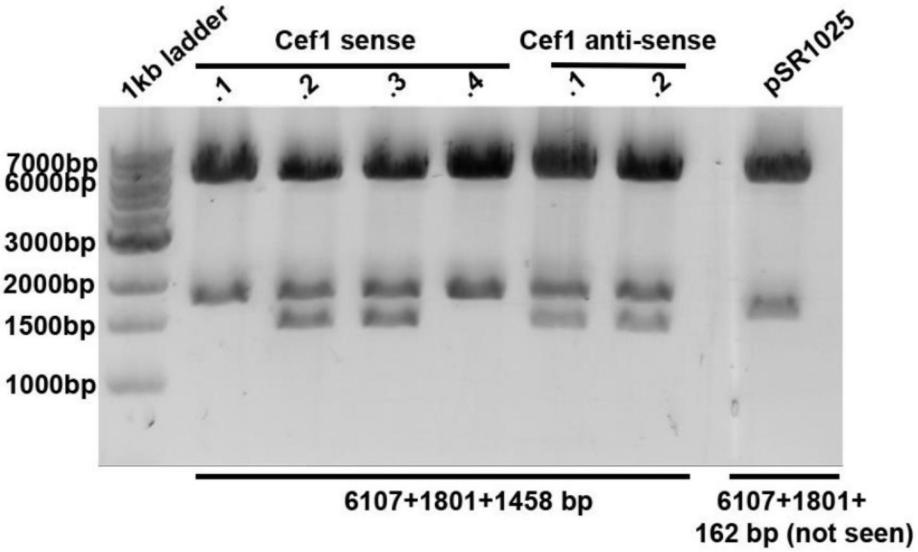


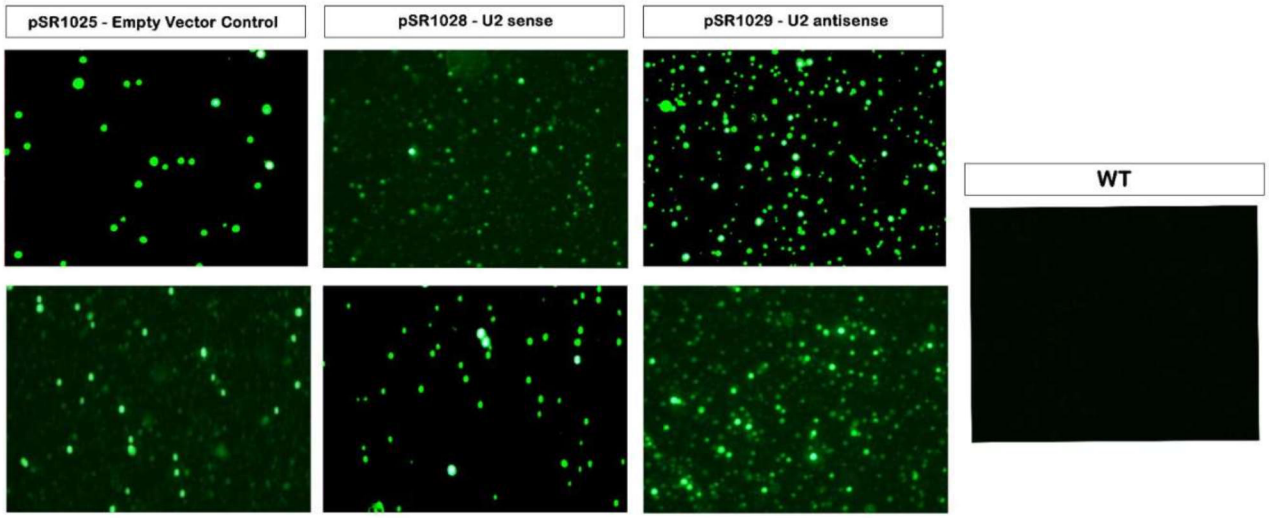
Figure 2-9: a) Restriction enzyme digestion check for pSR1028/1029 (U2 sense (s) and antisense (as), and pSR1030/1031 (U4 s and as). 0.7% agarose gel, 1-hour run. b) Restriction enzyme check for plasmids pSR1034/1035 (Cef1 s and as). 0.7% agarose gel, 40min run. The expected band sizes were the following: for U2 sense snRNAs 5176bp and 3025bp (U2 sense .1-.4, Figure 2-9 a)), for U2 antisense snRNA 5257bp and 2944bp (lanes 6 and 7, Figure 2-9 a), for U4 sense and antisense 7219bp and 1028bp (lanes 8 to 11, Figure 2-9a), and Cef1 sense and antisense 6107bp, 1801bp and 1458bp (lanes 2-7, Figure 2-8 b). pSR1025(no inserts) was included as a control and was digested with HindIII. The expected band sizes were 7348bp and 853bp. pSR1025 was also digested with EcoRI and NcoI. The expected band sizes were 6107bp, 1801bp, and 162bp (not seen).

One correct plasmid for each strain was used for *C. merolae* transformation and nitrate induction antisense experiments.

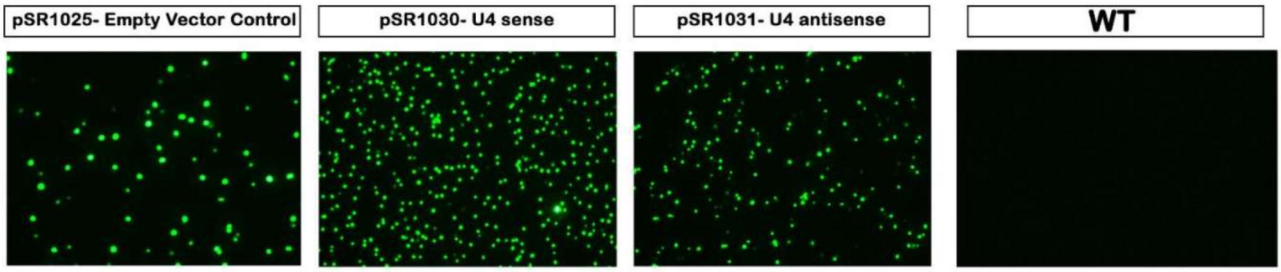
2.3.1.2. Transformation of plasmids into *C. merolae* cells and transformant selection

Antisense nitrate-induction plasmids for U2, U4, and Cef1 were transformed successfully into *C. merolae* cells using the standard methods described in the materials and methods section.

a)



b)



c)

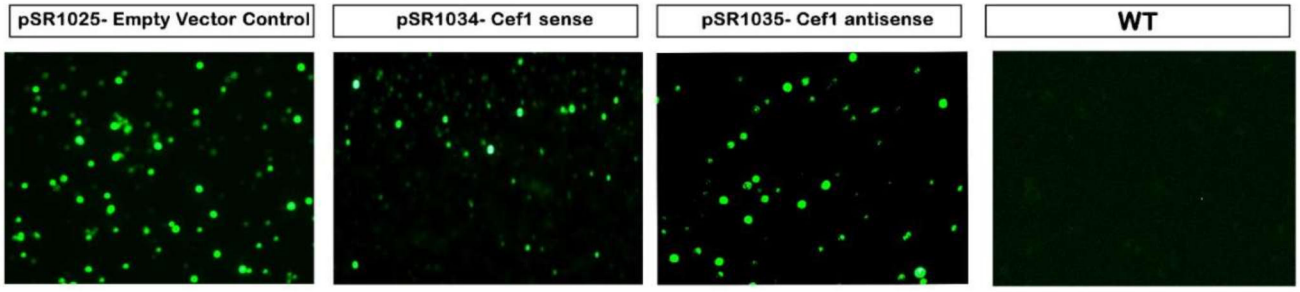


Figure 2-10: Fluorescent micrographs of U2, U4, Cef1, pSR1025 and WT cells after 12 days of transformation with the respective plasmids (FITC fluorescent filter).

Transformed plasmids also contained the fluorescent protein mVenus, another transformant screening strategy. I performed fluorescence microscopy, and cells were proven to be fluorescent for all U2, U4 and Cef1 sense and antisense strains (Figure 2-10 a, b and c, respectively). This confirmed the successful transformation with all the experimental plasmids used. The empty vector control strain (pSR1025) was also fluorescent. A WT negative control was observed under the microscope. The latter showed the background fluorescence in the absence of mVenus.

2.3.1.3. Growth assays in MA2G

I performed a growth assay in MA2G with plasmid-transformed *C. merolae* strains (in triplicate) to rule out any possible toxicity and phenotype caused by the presence of the transformed plasmids that may alter normal growth rate (9-10 hours doubling time for WT) or by a leaky promoter, which would allow transcription of antisense U2, U4, and Cef1 in the absence of induction. This would have complicated the analysis.

I did not expect to observe any growth rate change in the transformed experimental strains compared to the WT strains when the cells were grown in MA2G. This is because the promoter was supposed to be silent when not in nitrate media, and the sense and antisense genes should not be expressed, so they would not have any effect on splicing.

I grew WT cells as a control to compare doubling times between them and the U2 and U4 experimental cells. On the other hand, I grew Z4 cells (a strain containing the Cef1 protein tagged with Z4) for the Cef1 growth assay to compare the doubling times between them and the experimental cells.

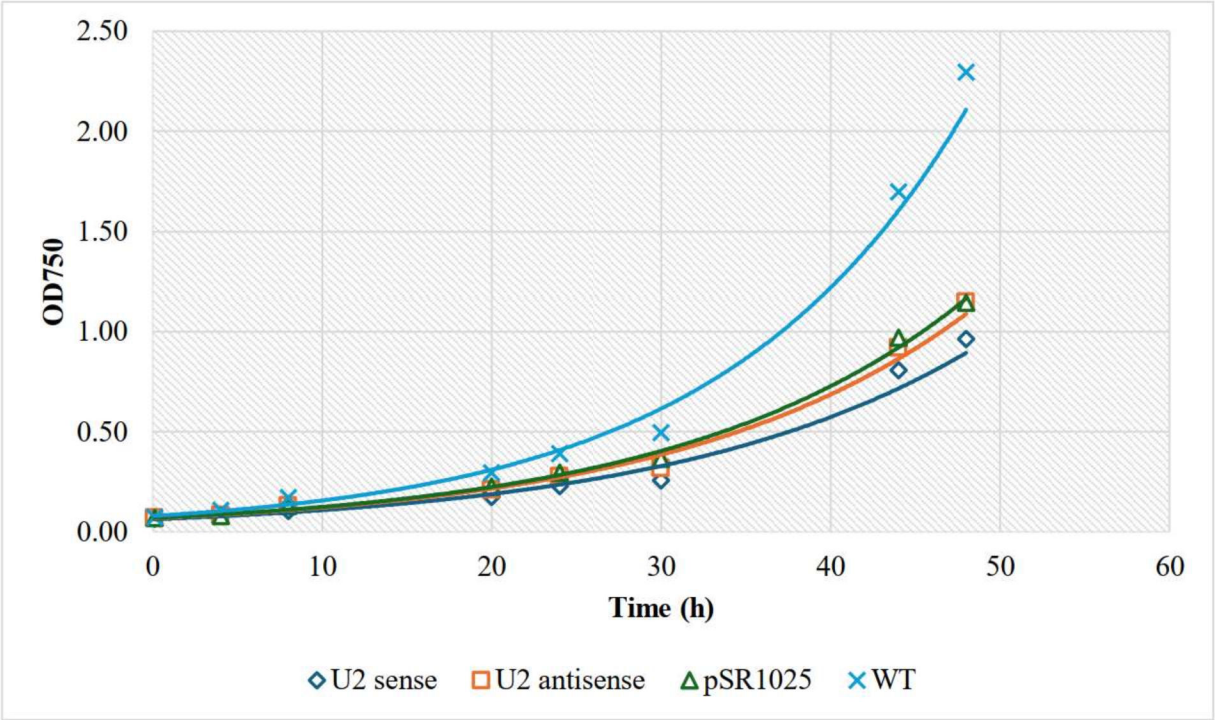


Figure 2-11: Growth curve for U2 sense, U2 antisense, pSR1025, and WT strains in MA2G. The points are the average of the technical triplicates used for the experiment.

U2 sense, U2 antisense, and pSR1025 (empty vector control) strains grew slower than the WT strain (Figure 2-11). I performed the growth assay with chloramphenicol for the experimental strains without selecting WT cells. The doubling times for U2 sense and antisense were very similar, with a slightly faster growth rate for the empty vector control (pSR1025) and an average growth rate for WT cells compared to previous measurements in our lab (Table 2-9, Figure 2-12). Additionally, the calculated p-values do not show a significant difference between the experimental strains and the empty vector growth rate (Table 2-9, Figure 2-12). Since the experimental strains did not exhibit any signs of death or visible phenotype during the assay, the observed growth rate reduction may be linked to antibiotic resistance. Notably, the empty vector control, which lacked any target gene, also showed slower growth compared to the wild-type control. This strongly suggests that the addition of chloramphenicol is contributing to the reduced growth of the cells.

Table 2-9: Doubling times for U2 sense, U2 antisense, pSR1025 and WT calculated from the performed growth assay. Column 3: Calculated the p-value by comparing the U2 sense and U2 antisense doubling time with pSR1025 (empty vector) doubling time.

Strain	Mean doubling time (h)	Standard Deviation (SD)	P-value* relative to pSR1025 (empty vector)
U2 sense (+)	12.5	0.4	0.06
U2 antisense (-)	12.0	0.1	0.13
pSR1025 (empty vector)	11.8	0.1	1.00
WT	10.1	0.1	---

n=3

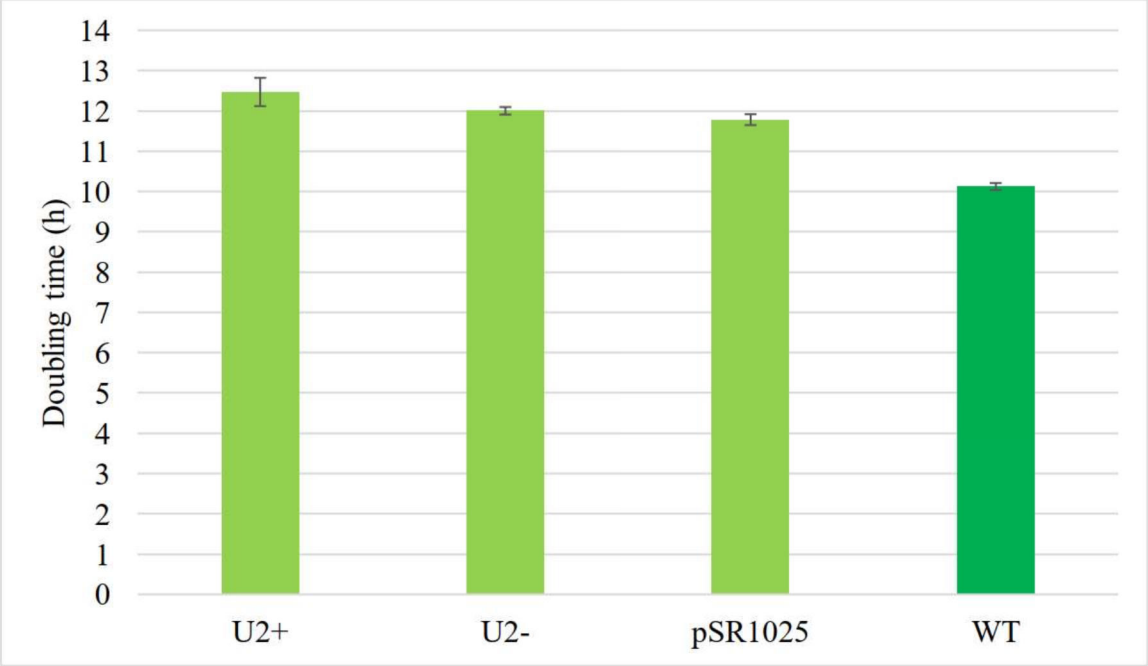


Figure 2-12: Bar graph representing the doubling times for U2 sense, U2 antisense, pSR1025 (empty vector), and WT in MA2G.

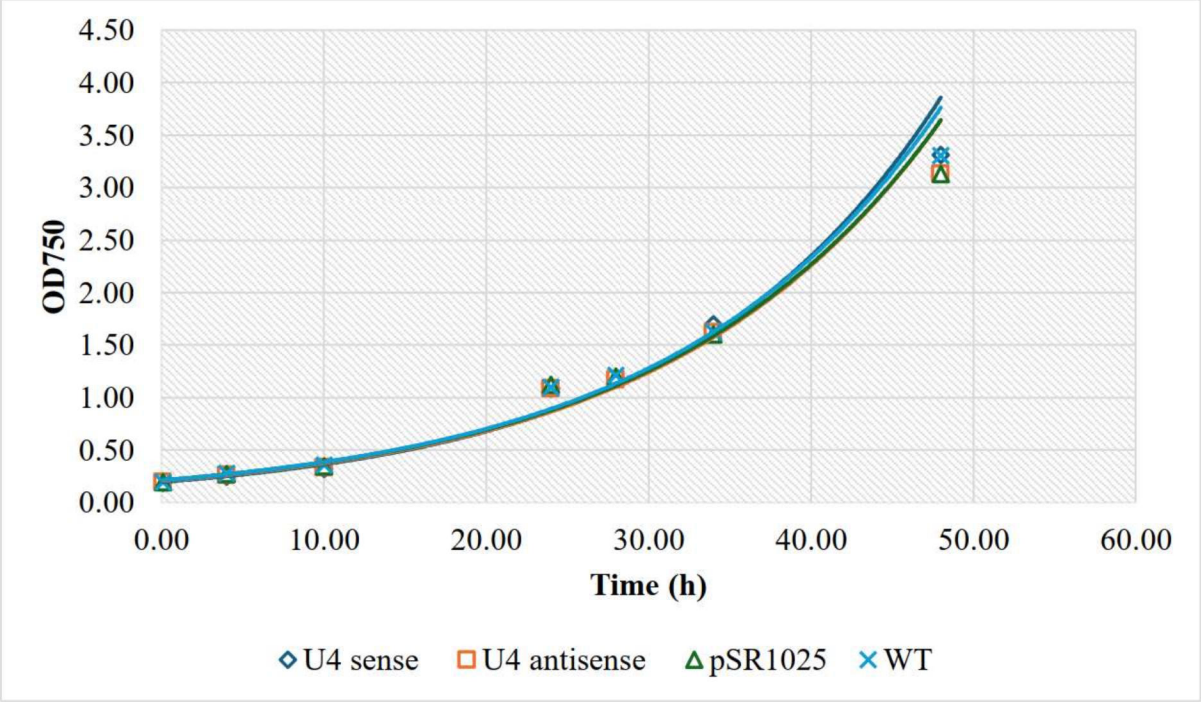


Figure 2-13: Growth curve for U4 sense, U4 antisense, pSR1025 (empty vector), and WT strains in MA2G.

As can be seen in Figure 2-13, the U4 sense, U4 antisense, pSR1025, and WT strains grew at a very similar rate, with the doubling time for all the strains ~11h (Table 2-10, Figure 2-14). The WT strain showed a slightly higher doubling time this time than in the U2 growth assay. All the cultures were grown under the same conditions, except that chloramphenicol was added to U4 sense and antisense and pSR1025 strains. The observation that the doubling time of the experimental cultures closely mirrors that of the wild type (WT) suggests that the transformed plasmids do not have any detrimental effects on the cells. This conclusion is further supported by the calculated p-values in comparison to the empty vector, all of which are greater than 0.05, indicating no significant difference in the growth rates between the WT and the experimental strains.

Table 2-10: Doubling times for U2 sense, U2 antisense, pSR1025 and WT calculated from the performed growth assay. Column 3: Calculated the p-value by comparing the U4 sense and U4 antisense doubling time with the pSR1025 (empty vector) doubling time.

Strain	Mean doubling time (h)	Standard Deviation (SD)	P-value* relative to pSR1025 (empty vector)
U4+	11.2	0.3	0.19
U4-	11.6	0.4	0.76
pSR1025	11.5	0.1	1.00
WT	11.2	0.3	---

n=3

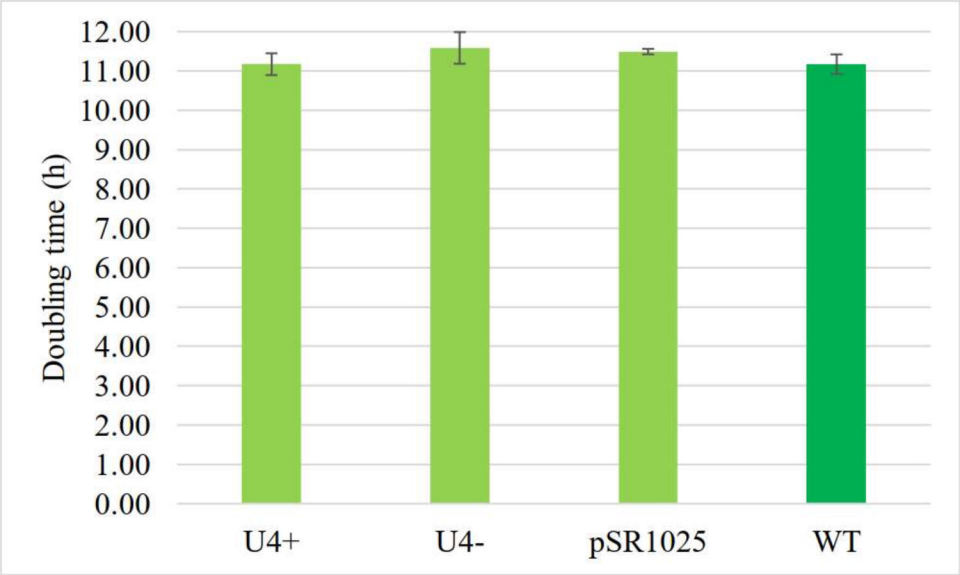


Figure 2-14: Bar graph representing the doubling times for U4 sense, U4 antisense, pSR1025 (empty vector) and WT in MA2G.

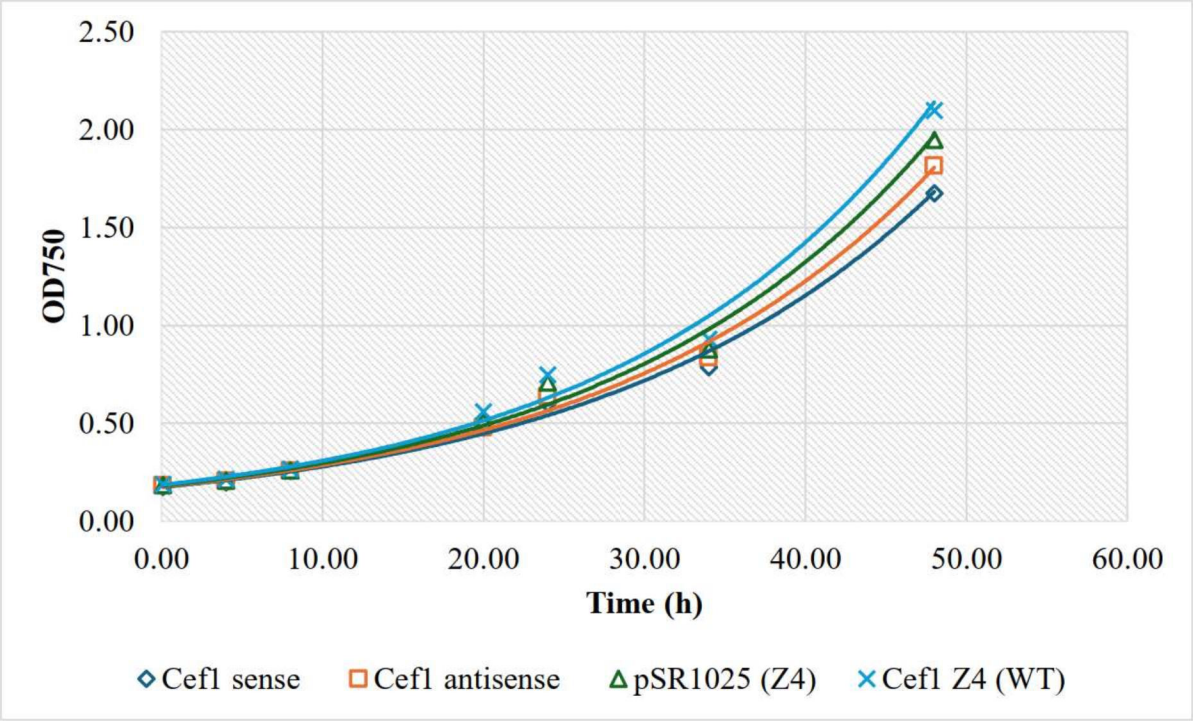


Figure 2-15: Growth curve for Cef1 sense, Cef1 antisense, pSR1025 (empty vector), and WT strains in MA2G.

The Cef1 sense, Cef1 antisense, pSR1025, and Z4 (WT) strains grew at slightly different rates, with Cef1 antisense being the strain with the most similar doubling time to Z4 (Table 2-11, Figure 2-15 and 2-16). As for U2 and U4 growth assays, Cef1 sense and antisense and pSR1025 strains were grown with selection. In this experiment, pSR1025 is transformed into the Cef1 Z4 strain. The p-values calculated by comparing the Cef1 sense, Cef1 antisense doubling times with those of the empty show no significant differences, as described above. There was no significant difference in growth rates between these strains, demonstrating no adverse effect or toxicity due to the plasmids.

Table 2-11: Doubling times for Cef1 sense, Cef1 antisense, pSR1025 and WT calculated from the performed growth assay. Column 3: Calculated the p-value by comparing the Cef1 sense and Cef1 antisense doubling time with pSR1025 (empty vector) doubling time.

Strain	Mean doubling time (h)	Standard Deviation (SD)	P-value* relative to pSR1025 (empty vector)
Cef1+	12.8	0.1	0.31
Cef1-	11.9	0.1	0.04
pSR1025	13.4	0.7	1.00
Z4	11.3	0.2	---

*n=3

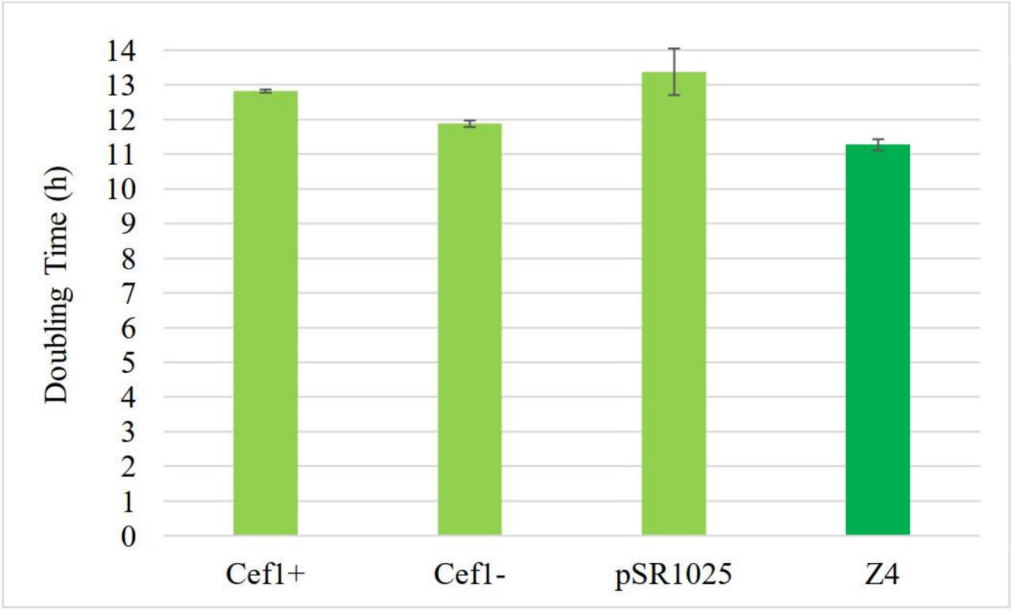


Figure 2-16: Bar graph representing the doubling times for Cef1 sense, Cef1 antisense, pSR1025, and Z4 in MA2G.

Having demonstrated that the plasmids did not cause any change in growth rates, I next addressed whether antisense induction had an impact on cell growth.

2.3.1.4. Nitrate induction and growth assessment

To investigate whether antisense inhibition of splicing factor expression impacted cell growth, I induced the antisense by switching *C. merolae* cells from MA2G to MA2GN (nitrate media). The strains containing plasmids with the antisense versions of U2, U4, or Cef1 were expected to reduce their growth rate throughout the experiment and eventually die once the proteins encoded by the intron-containing genes were sufficiently depleted. I expected the strains containing the plasmids with the sense version of U2, U4, or Cef1 to either grow at the same rate as WT or Z4 cells or slightly slower, the latter, due to the possibility that overexpression of U2 and U4 could be deleterious. As splicing proteins are so limited in *C. merolae,* having more copies pf these genes may reduce spliceosomal assembly. Due to all those extra copies, there will not be enough spliceosomal components to assemble the U2 and U4 snRNPs. In this scenario, splicing might be partially inhibited and growth rates might be partially reduced. The strains containing the pSR1025 (empty vector control) were expected to grow similarly to the WT or Z4 control. An NIR-Cas9 strain was included in the experiments as a control for induction. The doubling time in nitrate media is higher than in ammonium media due to the additional energetic cost of reducing nitrate into ammonium for it to be assimilated by the cells.

U2 sense and antisense nitrate induction showed normal and uniform cell growth. As shown in Figure 2-17, the growth rate of every strain is constant and stable. The U2 antisense strain does not show any growth rate decrease. The doubling time for U2 sense and antisense strains is ~16h, while the doubling time for the pSR1025 and WT strains is slightly lower at ~15h (Table 2-12, Figure 2-18). Remarkably, the p-values calculated by comparing the doubling times from U2 sense and U2 antisense strains with the empty vector showed a non-significant difference between the experimental and control strains (Table 2-12), suggesting no growth

differences between the empty vector control (no sense or antisense expression) and the sense and antisense strains. This suggests a possible lack of induction. Additionally, the p-value calculated by comparing the pSR1025 and WT doubling times indicated a non-significant difference between the growth rates for these two strains (Table 2-12), meaning they are equally growing in nitrate (not shown).

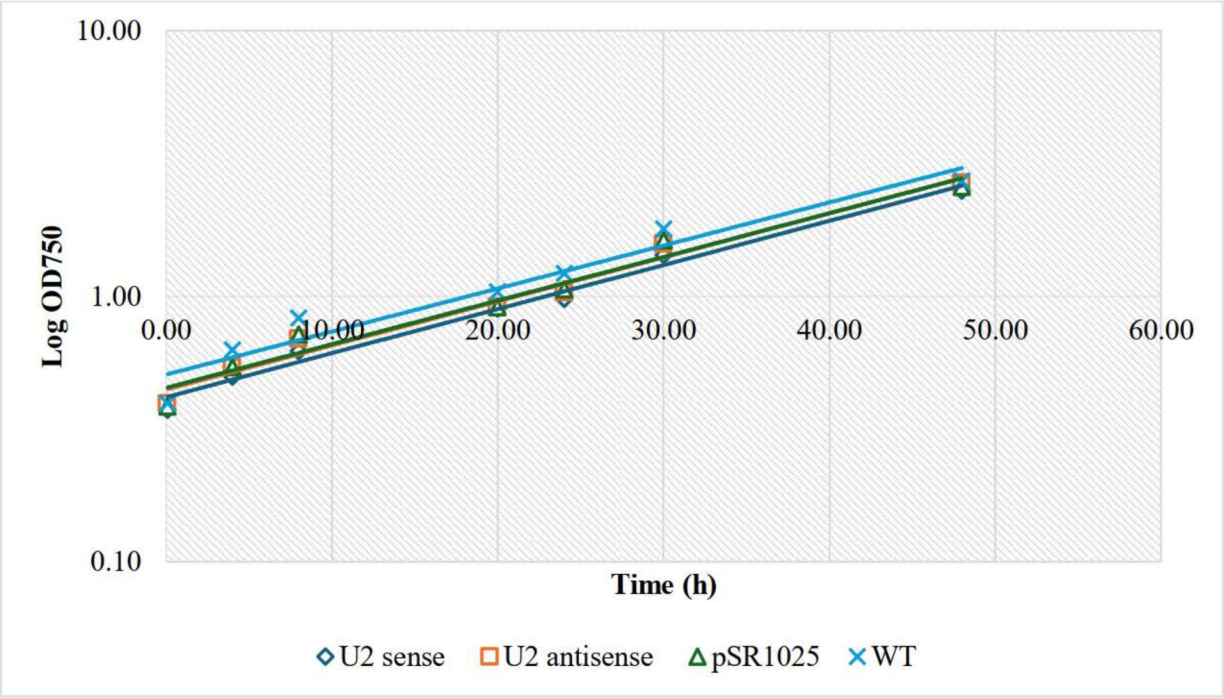


Figure 2-17: Growth curve for U2 sense, U2 antisense, pSR1025, and WT strains in MA2G.

Table 2-12: Doubling times for U2 sense, U2 antisense, pSR1025 and WT calculated from the performed growth assay. Column 3: Calculated the p-value by comparing the U2 sense and U2 antisense doubling time with pSR1025 (empty vector) doubling time in MA2GN.

Strain	Doubling time (h) (Average from triplicates)	Standard Deviation (SD)	P-value* relative to pSR1025 (empty vector)
U2+	16.4	0.2	0.14
U2-	16.7	0.2	0.06
pSR1025	15.8	0.5	1.00
WT	15.7	0.3	---

*n=3

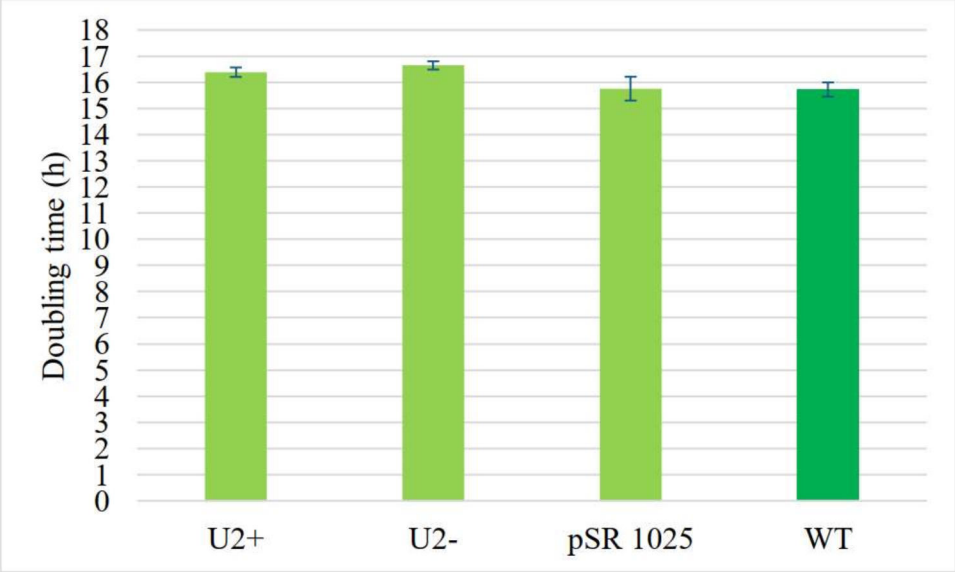


Figure 2-18: Bar graph representing the doubling times for U2 sense, U2 antisense, pSR1025, and WT in MA2GN.

U4 sense and antisense nitrate induction showed normal and uniform cell growth during the experiment for all the strains included. As shown in Figure 2-19, the growth rate of every strain was constant and stable. The U4 antisense strain did not show any growth rate decrease. The doubling time for the U4 sense strain was ~18h, while the doubling time for the U4

antisense and pSR1025 strains was slightly faster at ~17.5h (Table 2-13 a, Figure 2-20). The WT strain was not explicitly included in these experiments as it was included in the U2 sense and antisense induction, and samples were already collected for RNA analysis. Also, the WT strain growth in nitrate was already measured. The statistical significance of this assay was obtained by calculating the p-value between the doubling times of U4 sense and antisense strains and the empty vector control (pSR1025). The calculated p-values indicated no significant difference between the U4 sense and antisense strains and the pSR1025 strain. This suggests that the increased doubling time in nitrate media, compared to that in ammonium media, may simply result from the fact that these strains are growing with nitrate as their primary nitrogen source. The NIR-Cas9 strain was included in this experiment as an induction control (the same prepared media was used for U2, U4 and Cef1 antisense experiments), and its growth rate was also measured. As can be seen in Figure 2-19, the NIR-Cas9 strain grows normal and stable throughout the induction.

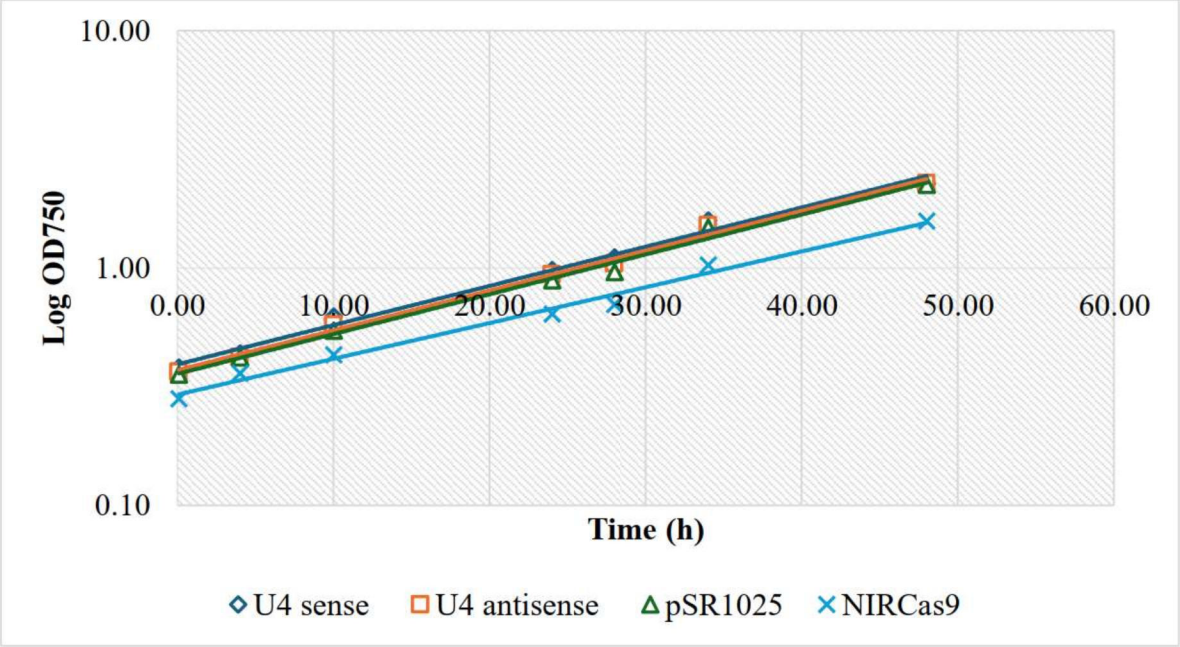


Figure 2-19: Growth curve for U4 sense, U4 antisense, and pSR1025 strains in MA2GN.

Table 2-13: Doubling times for U4 sense, U4 antisense, pSR1025 and NIRCas9 calculated from the performed growth assay. Column 3: Calculated the p-value by comparing the U2 sense and U2 antisense doubling time with pSR1025 (empty vector) doubling time in MA2GN.

Strain	**Doubling time (h) (average of triplicates)**	**Standard deviation (SD)**	**P-value* relative to empty vector control (pSR1025)**
U4+	18.2	0.2	0.36
U4-	17.8	0.1	0.72
pSR1025 (empty vector)	17.8	0.6	1.00
NIRCas9	20	---	---

*n=3

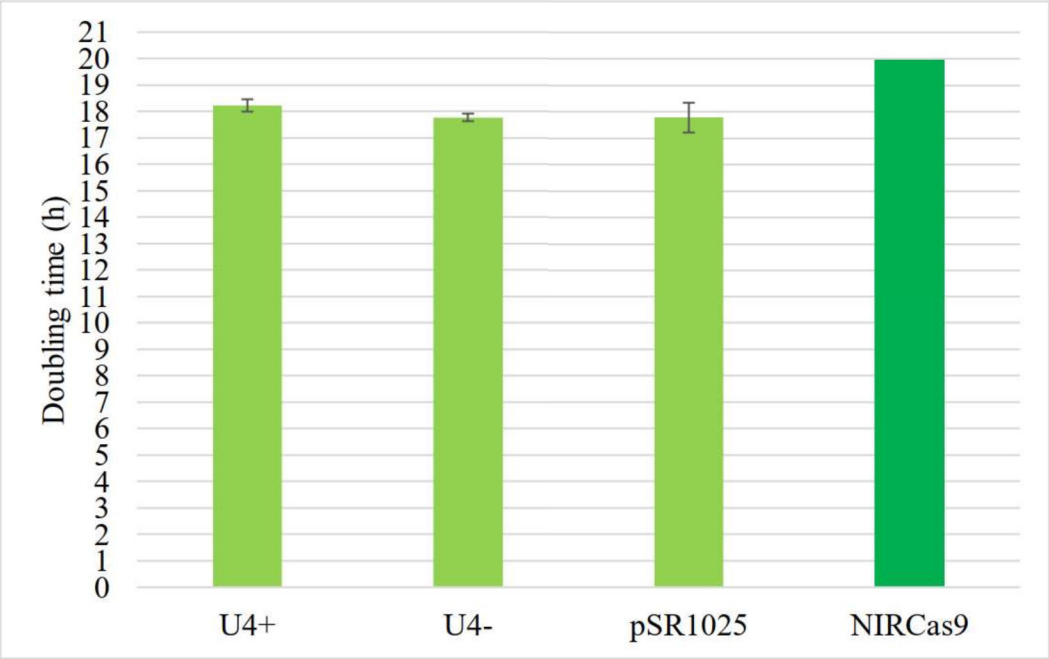


Figure 2-20: Bar graph representing the doubling times for U4 sense, U4 antisense, pSR1025 (empty vector), and NIRCas9 in MA2GN.

Cef1 sense and antisense nitrate induction showed normal and uniform cell growth for all the strains included during the experiment. As shown in Figure 2-21, the growth rate of every strain was constant and stable. The Cef1 antisense strain does not show any growth rate decrease. The doubling time for the Cef1 sense strain was ~18h, while the doubling time for Cef1 antisense and pSR1025 (empty vector) strains was slightly faster with ~16h (table 2-14 a, figure 2-22). The Z4 strain showed a doubling time of ~15 hours, with the fastest growth rate. Overall, no defective growth or cell death could be recorded, especially with the antisense strain, at least in the 48-hour time frame of the experiment duration. The calculated p-values comparing the doubling times for Cef1 sense and Cef1 antisense with pSR1025 (empty vector) showed that there is not a significant difference between the growth rates of the experimental and control strain for Cef1 antisense and that there is a significant difference for Cef1 sense (Table 2-14).

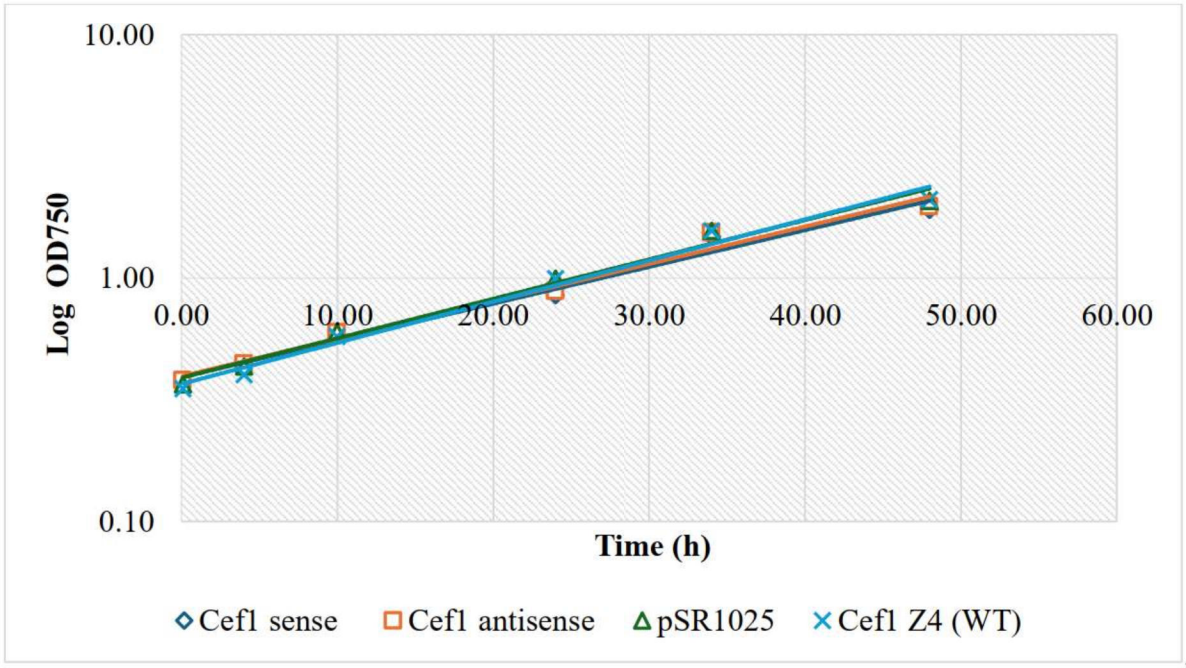


Figure 2-21: Growth curve for Cef1 sense, Cef1 antisense, pSR1025, and WT strains in MA2GN.

Table 2-14: Doubling times for Cef1 sense, Cef1 antisense, pSR1025 and WT calculated from the performed growth assay. Column 3: Calculated the p-value by comparing the Cef1 sense and Cef1 antisense doubling time with pSR1025 (empty vector) doubling time in MA2GN.

Strain	Mean doubling time (h)	Standard Deviation (SD)	P-value* relative to pSR1025 (empty vector control)
Cef1+	17.8	0.1	0.2
Cef1-	15.8	0.7	0.97
pSR1025 (empty vector)	15.8	0.7	1.00
Z4	16.1	0.9	---

*n=3

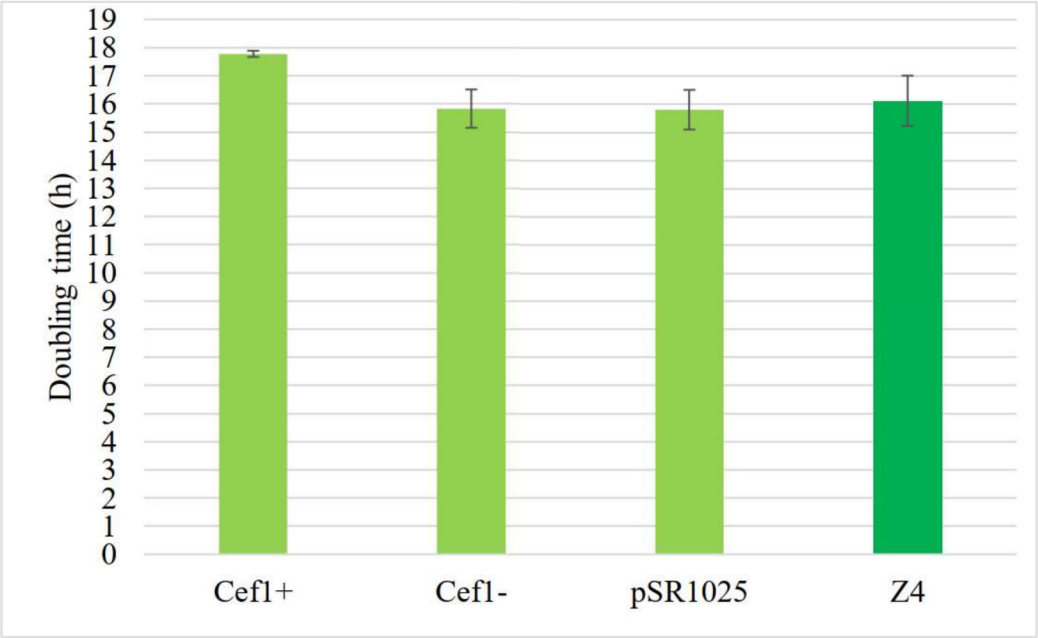


Figure 2-22: Bar graph representing the doubling times for Cef1 sense, Cef1 antisense, pSR1025, and Z4 in MA2GN.

After performing the nitrate induction experiments in a 48-hour time frame and concluding that no growth defects could be seen in any of the strains for any of the three chosen genes, I diluted the cells of all the strains back to OD 0.4. I left them in nitrate media for at least one week to check for long-term effects. None of the U2, U4, or Cef1 strains showed a decrease in growth rate or cell death after a week of monitoring. I measured the OD once daily, and all the strains grew to saturation. I added chloramphenicol (250 ug/mL) to the cultures every three days. The final OD for each strain at day seven after re-dilution to OD 0.4 is shown in the following table (table 2-15):

Table 2-15: OD_{750} of U2, U4, and Cef1 antisense strains 7 days after induced in nitrate. WT and psR1025 (empty vector) for each experiment are included.

Strain	**OD_{750} at day 7 (average of triplicates)**
U2 sense	6.4
U2 antisense	6.4
pSR1025 (U2 experiment)	6.1
WT	17.1
U4 sense	5.7
U4 antisense	5.9
pSR1025 (U4 experiment)	6.0
Cef1 sense	6.1
Cef1 antisense	6.1
pSR1025 (Cef1 experiment)	6.4
Z4	9.0

The OD for all the experimental strains, sense, antisense, and empty vector control for all three genes, oscillated between 5 and 6 on day seven after the induction (Table 2-15). On the other hand, the control strains, WT and Z4, show very high ODs on the same day. First, these strains were not exposed to chloramphenicol, and second, these strains were not transformed with any plasmid, so in a long-term effect, when cultures achieve saturation, these cells have less metabolic costs that will allow them to grow faster. Additionally, it was surprising that the pSR1025 strain had not ended up with an OD like one of the WT or Z4 cells. pSR1025 is an empty vector that does not contain the sense or antisense version of the genes of interest, and that confers antibiotic resistance to the strains. Splicing will remain untouched in these strains. As the OD of pSR1025 for all three experiments is similar to the OD of the sense and antisense strains, we cannot attribute the decrease in growth rate compared to the WT or Z4 to splicing inhibition but to other factors such as the ones previously mentioned.

If it is true that a lack of growth rate reduction or cell death could indicate that splicing is not an essential process for *C. merolae* (assuming that antisense induction was successful),

nothing could be said until actual antisense gene expression was proven by analyzing the RNA for each strain by Northern Blot.

2.3.1.5. Northern Blotting

To test whether the antisense transcripts were successfully expressed upon nitrate induction, I analyzed their expression by northern blot. For the 24 and 72-hour time points, when probing with the U2 sense probe (oSDR2311), a clear band representing U2 snRNA can be seen for all the samples analyzed (131nt) (Figure 2-23). Assuming that *C. merolae* has no double-stranded RNA nucleases, I expected the U2 sense band to be present in all the samples, as endogenous U2 snRNA is present in all the cells. I expected that in the U2 sense strains, a more intense U2 sense band would be seen due to the extra copy of this gene introduced by the transformed plasmid(s) (when transforming plasmids, the copy number is uncertain). In contrast to my expectations, there was no visible difference in band intensities in any strain (Figure 2-23), which I confirmed by quantitation (Figure 2-24). To determine whether antisense U2 was actually induced, I stripped the blot and reprobed it with the U2 antisense probe (oSDR2475). No bands appeared for any of the samples in this case.

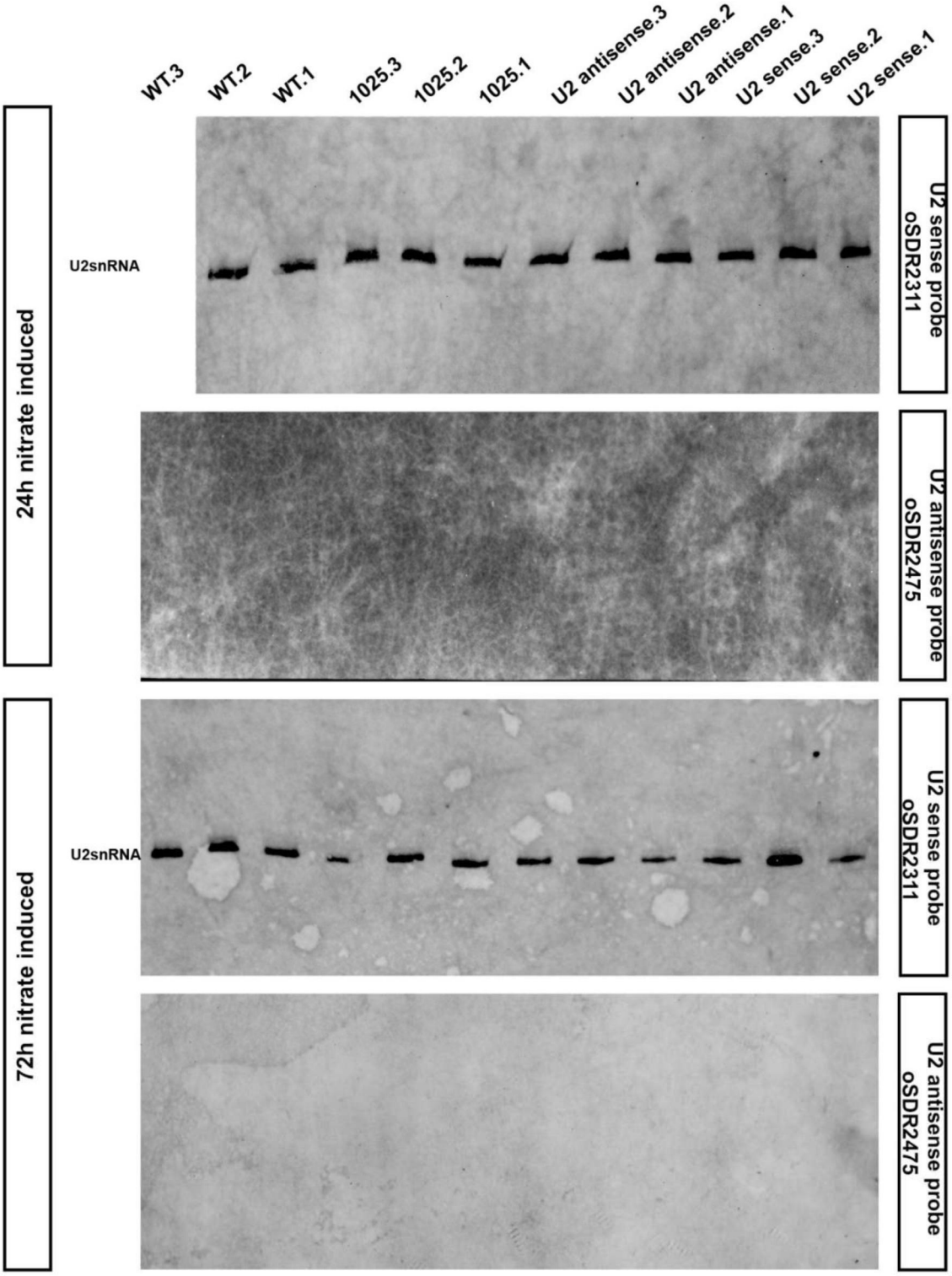


Figure 2-23: Denaturing polyacrylamide northern blot for U2 antisense induction of U2 transient gene expression strains. 24 and 72-hour time point samples were analyzed. 2 ug RNA were loaded per lane, a 6% gel was run for 45 minutes, and the transfer time was set to 30 minutes. For probe sequences, refer to Appendix 1.

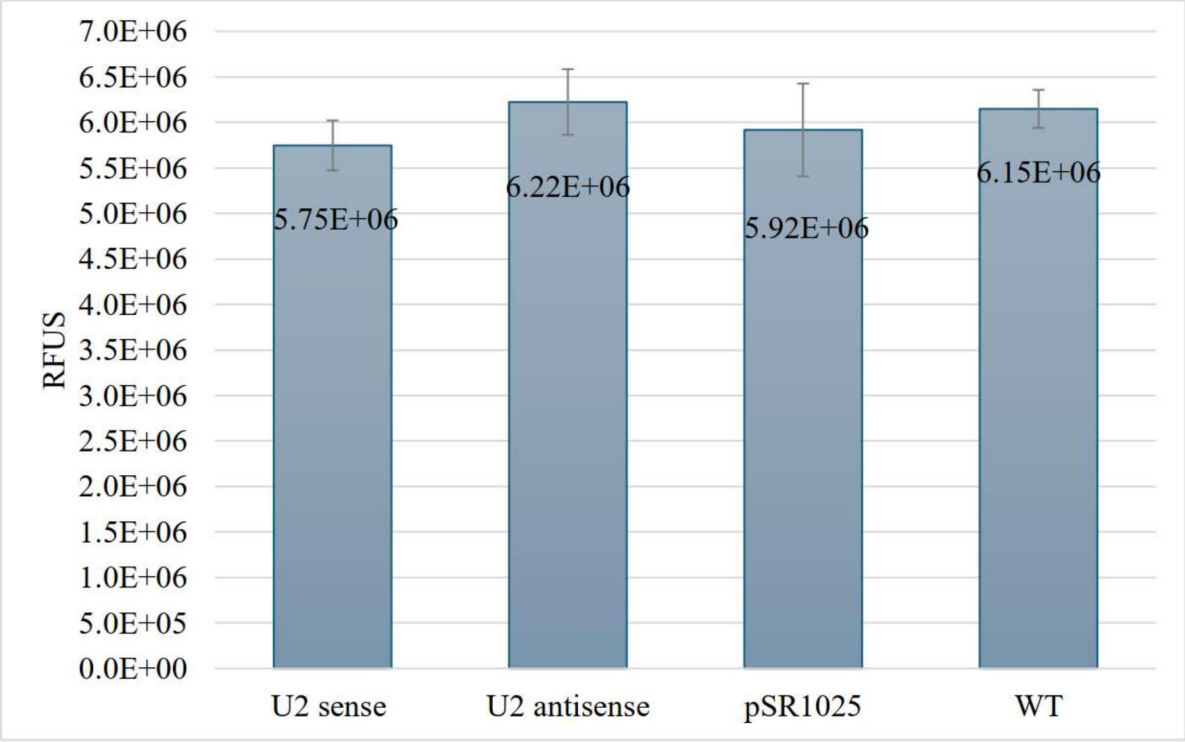


Figure 2-24: Band intensity quantification for U2 sense, U2 antisense, pSR1025 (empty vector) and WT samples from the U2 antisense induction experiment 24-hour samples. Each bar represents the average RFUs from the triplicates analyzed per strain. Band intensity was measured in relative fluorescence units (RFUs).

The same as for U2 was expected for the U4 antisense induction experiment. In this case, I processed and analyzed 24- and 48-hour samples. The results for the U4 induction match those obtained with the U2 induction (Figure 2-25).

The U4 sense band (174nt) is present in all the samples when blots were probed with the U4 sense probe (oSDR2476), and no bands could be detected for any of the samples when blots were probed with the U4 antisense probe (oSDR2477). For the 24-hour U4 nitrate-induced samples, the bands show intensity inconsistencies with even a band absence in sample 1025.1 (Figure 2-25).

More consistent results were observed with the 48-hour time point samples analysis. When probing with the U4 sense probe (oSDR2476), a single band corresponding to U4 snRNA appeared for each sample, but, as for U2, no difference in band intensities between strains was observed (Figure 2-25), which I confirmed by quantitation (Figure 2-26).

In conjunction with the 48-hour time point sample analysis, I analyzed the uninduced U4 sense, U4 antisense and pSR1025 (empty vector). The uninduced samples were not analyzed in triplicate. A single band was observed for the three uninduced samples when probed with the U4 sense probe (Figure 2-25, Lanes 1-3), and their intensity was quantitated. These bands resulted less intense than those from the experimental samples (Figure 2-26)

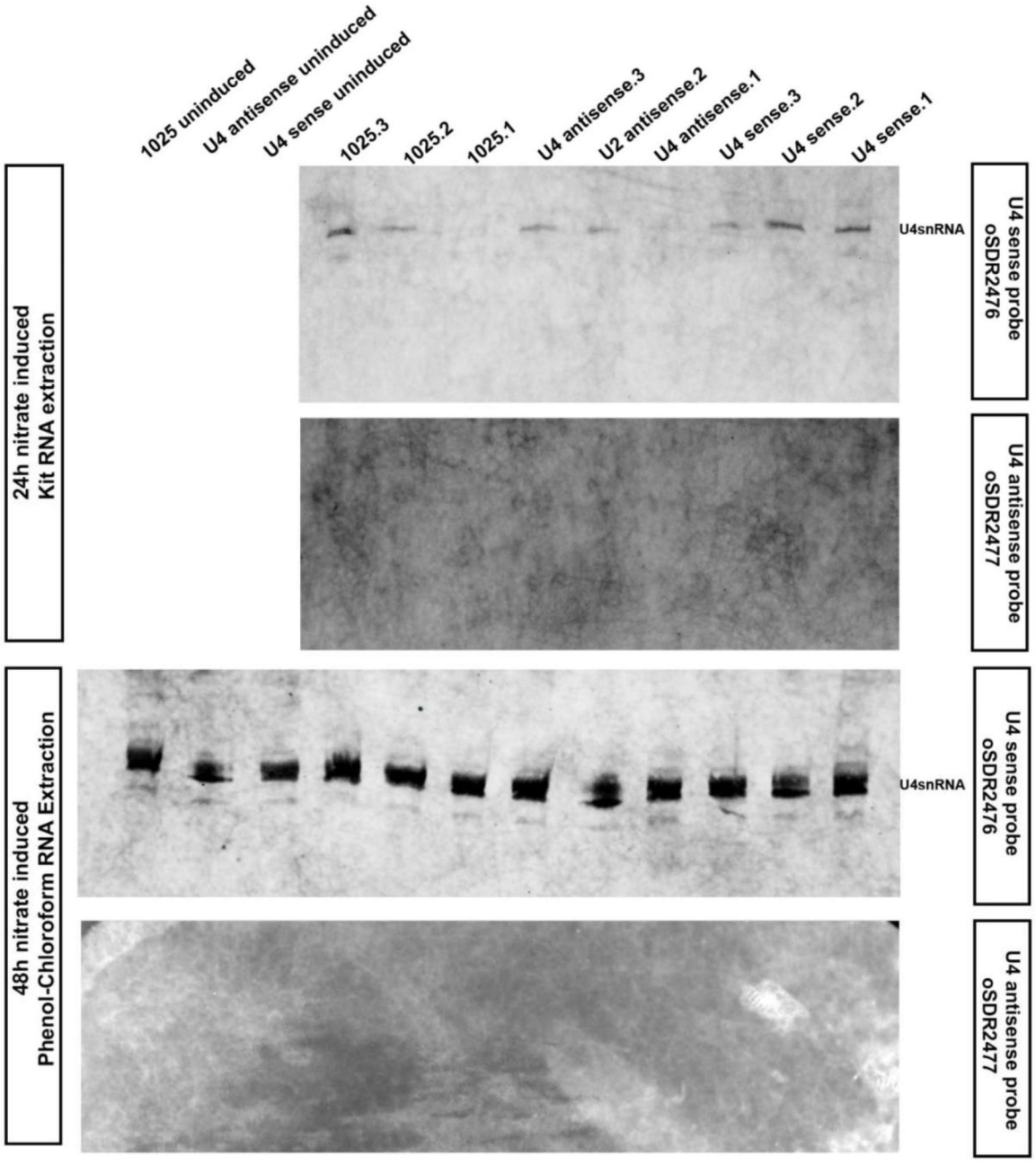


Figure 2-25: Denaturing polyacrylamide northern blot for U4 antisense induction of U4 transient gene expression strains. 24 and 48-hour time points were analyzed. 2 ug RNA were loaded per lane, a 6% gel was run for 45 minutes. For probe sequences, refer to Appendix 1.

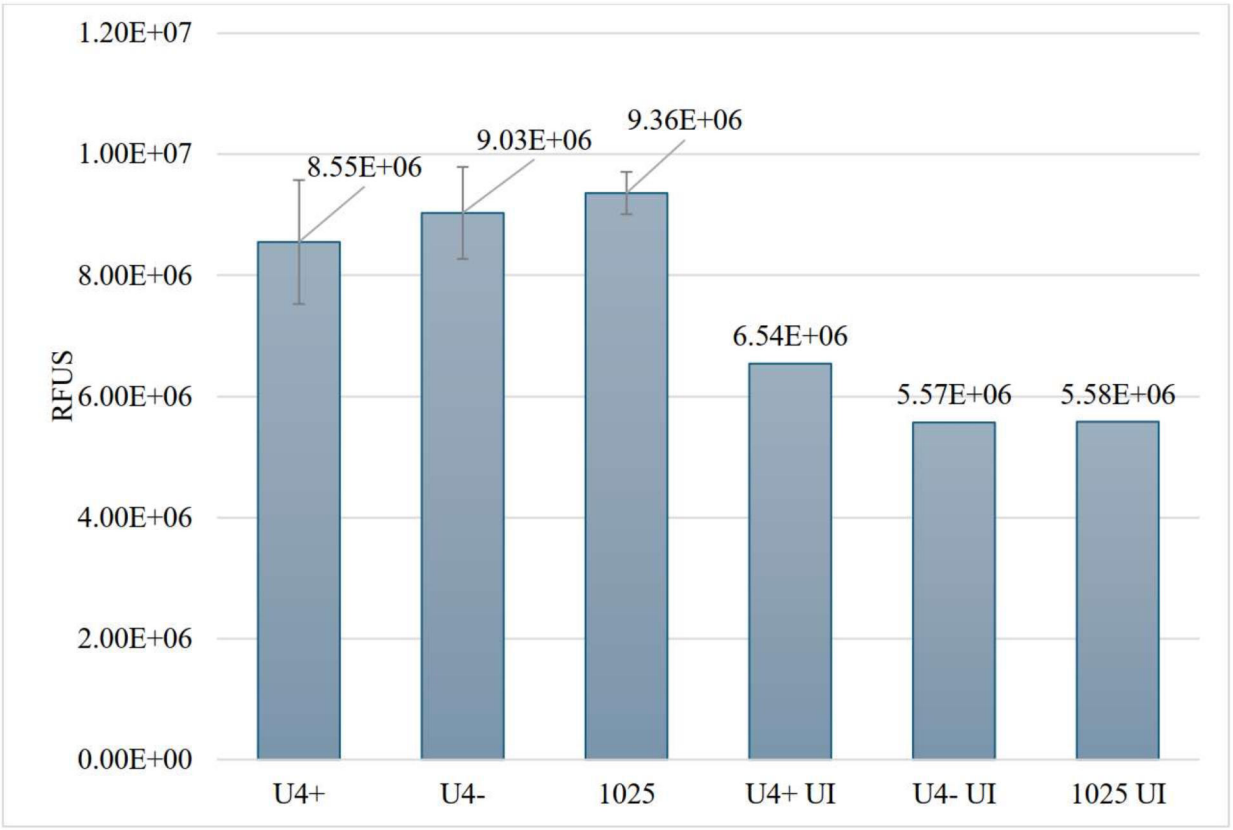


Figure 2-26: Band intensity quantification for U4 sense, U4 antisense, and pSR1025 induced samples from the U4 antisense induction experiment 48-hour samples. U4 sense, U4 antisense, and 1025 uninduced samples were also analyzed. Each bar represents the average RFUs from the triplicates analyzed per strain. Band intensity was measured in relative fluorescence units (RFUs). Note that no error bars exist for the uninduced samples because they were not analyzed in triplicate, and no standard deviation between their values could be calculated.

To confirm whether Cef1 antisense expression was successful upon induction, I attempted two different approaches. First, as the fragment size I was looking for was bigger than 1kb (Cef1 open reading frame is ~2kb), I performed a denaturing agarose northern blot. The blot shows a very faint band for all the analyzed samples, running across the middle of the membrane when probed with the Cef1 sense probe (oSDR2484) (Figure 2-27). The bands were unquantifiable due to their low intensity. It can not be said with absolute certainty that this band corresponds to the Cef1 sense.

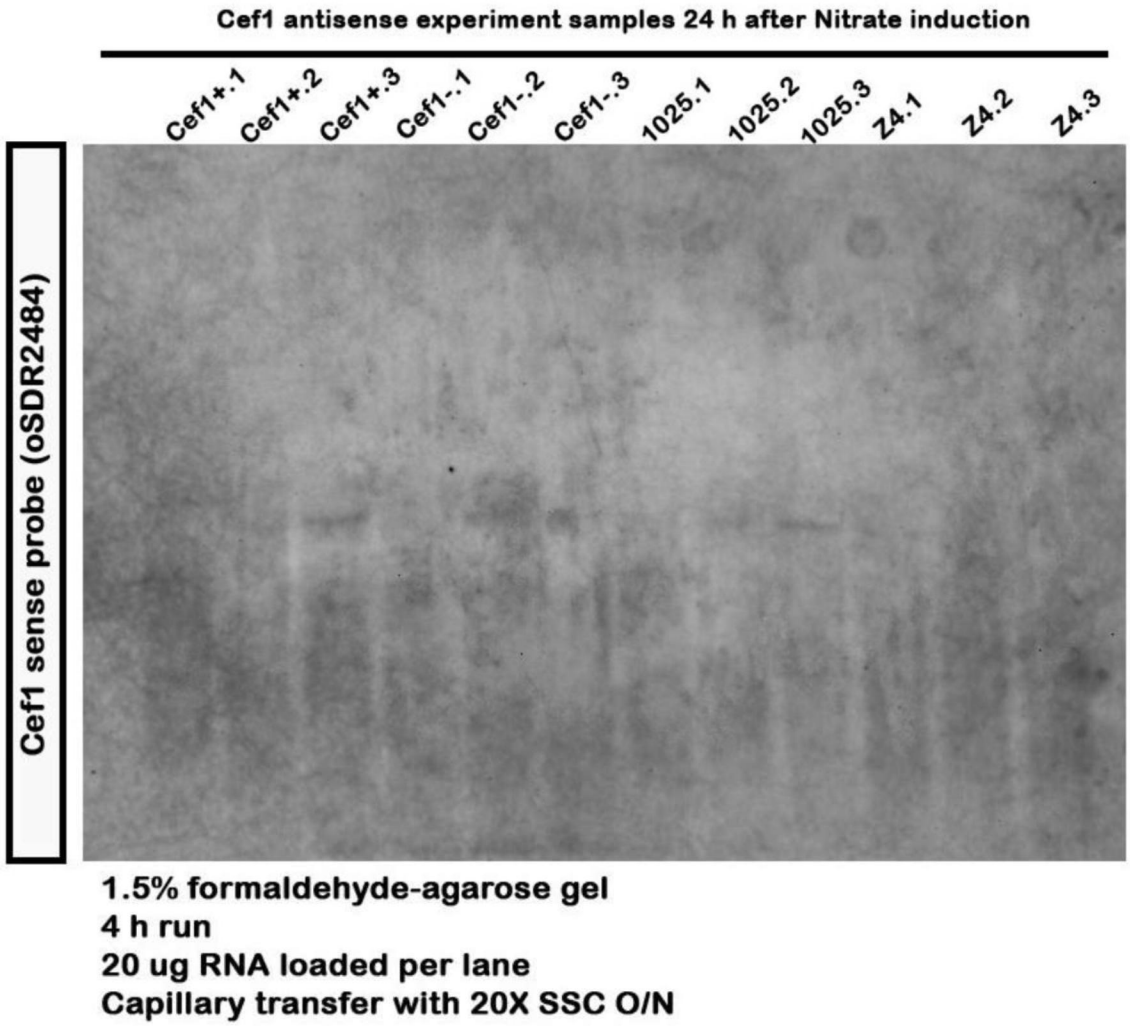


Figure 2-27: Denaturing agarose northern blot to analyze Cef1 antisense nitrate induction 24-hour time point samples. The details of the experiment are included in the figure.

Transfer issues were taken into consideration in the agarose northern blot approach. For this reason, I ran a northern blot using a polyacrylamide gel in the second approach. I performed a 6% denaturing polyacrylamide northern blot with the 34-hour time point samples under the conditions mentioned in the material and methods. U5 was used as a loading control (Figure 2-28).

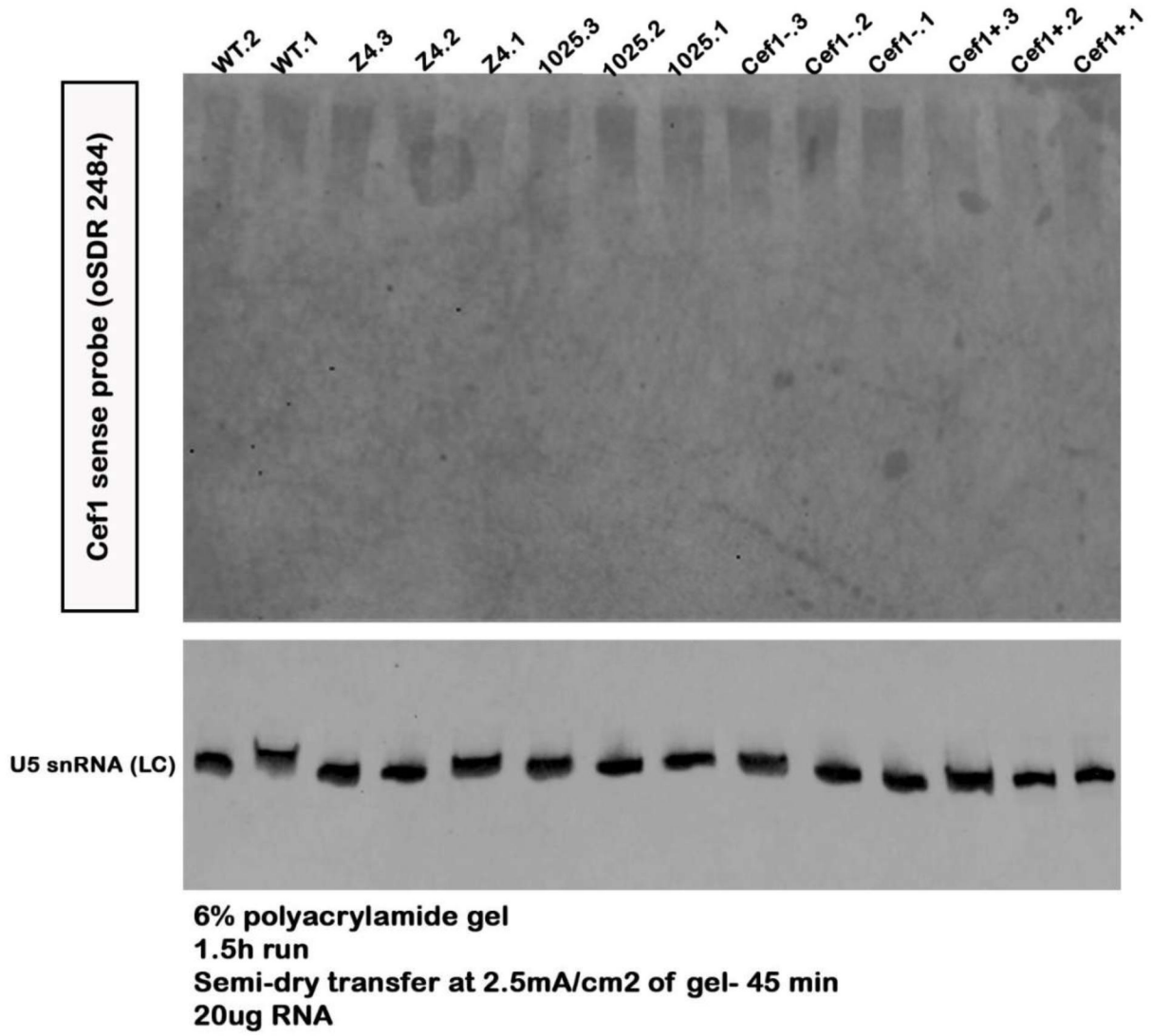


Figure 2-28: Denaturing polyacrylamide northern blot to analyze Cef1 antisense nitrate induction 34-hour time point samples. The details of the experiment are included in the figure.

No specific bands were detected for Cef1 sense in the analyzed samples. The U5 band shows consistency in how much RNA was loaded per lane. It demonstrates that the absence of a Cef1 sense band is not attributable to poor quality or degraded RNA but to other factors. A smear can be observed at the top of the lanes on the blot when probed with a Cef1 sense probe. The smear is faint but could show the presence of the Cef1 sense transcript in the analyzed samples. The smear can be attributed to the different transcript sizes produced during pre-mRNA

processing. The smear was not quantifiable, and once again, nothing could be concluded about the expression of the Cef1 sense or antisense genes. The Cef1 antisense probe was never tried as Cef1 sense detection was unsuccessful.

2.3.1.6. NIR-Cas9 Western Blot

When performing the nitrate inductions of the transient gene expression experiments, I used the NIR-Cas9 strain as a control to ensure induction occurs and rule out any nitrate media issues that would affect the induction effectiveness. According to previous experiments performed in the Rader Lab, the induction peak for the NIR-Cas9 strain is reached at 24 hours. I took samples from this strain 24 and 48 hours after induction and performed Western blots for Cas9. At both, 24 and 48 hours, Cas9 is clearly induced, while the uninduced lane appeared completely clear (Figure 2-29, Lane 2), which indicates the protein was only detectable when cells were switched to nitrate. There is a doublet in the induced lanes, but I assume the top is the correct band based on the expected size of 150 kDa. The other could be a phosphorylated form of the protein. These results show that there was nothing wrong with the media and that, for instance, the previous results in which it was not possible to detect any sense or antisense induction for U2, U4, Cef1, and Dib1 experimental strains, are due to some other unknown factors.

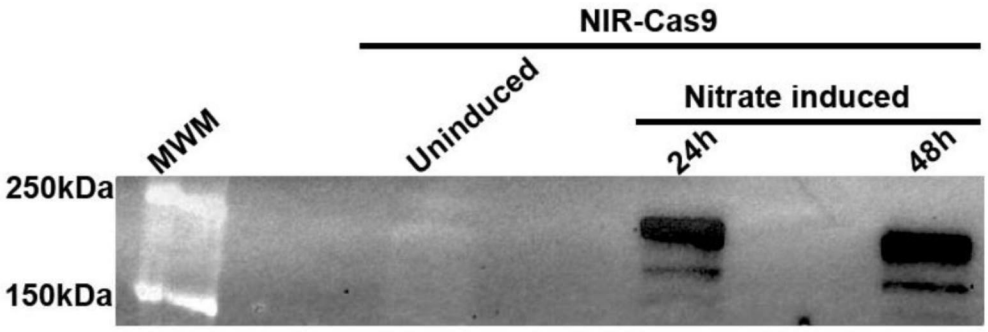
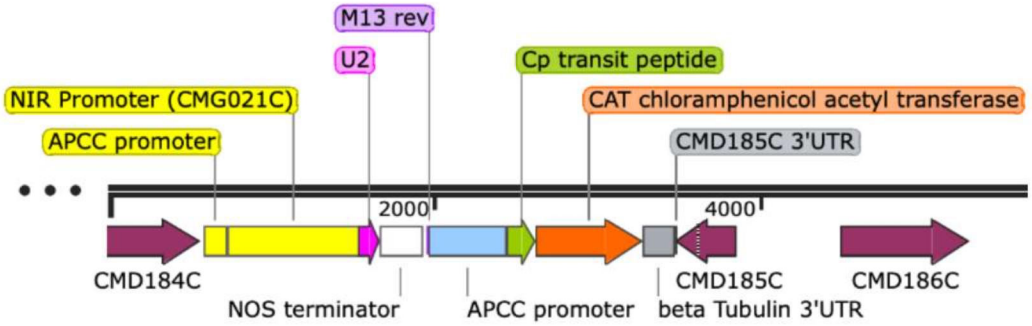


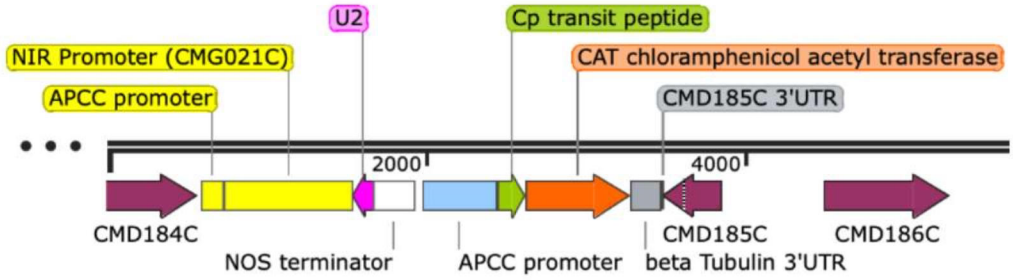
Figure 2-29: NIR-Cas9 western blot for the transient gene expression experiments. Note that the uninduced sample does not show any band. An 8% polyacrylamide gel was run in 5X Tris-Glycine buffer for 1 hour. The transfer time to a nitrocellulose membrane was 1.5 hours at 1.5 mA/cm^2 of gel.

2.3.2. Integrated gene expression experiments

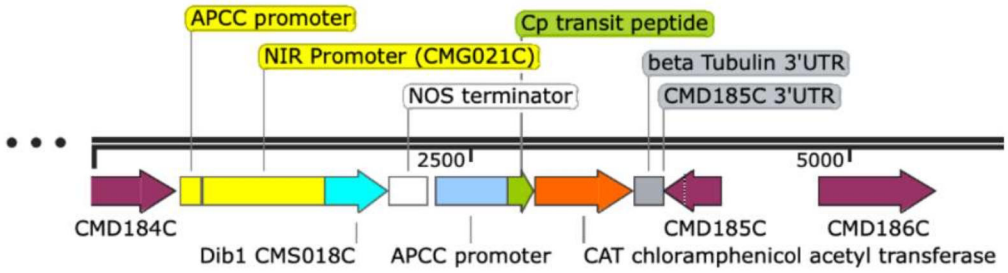
To determine whether higher expression of antisense genes would inhibit splicing, I placed U2 and Dib1 under the control of the NIR promoter, the most strongly induced of the nitrate-inducible promoters. By modifying the pSR979 (NIR-Cas9-NOS terminator) backbone plasmid using restriction enzyme cloning, expression vectors for the integrated gene expression of the antisense U2 and Dib1 were successfully constructed (Figure 2-30).

a)





b)



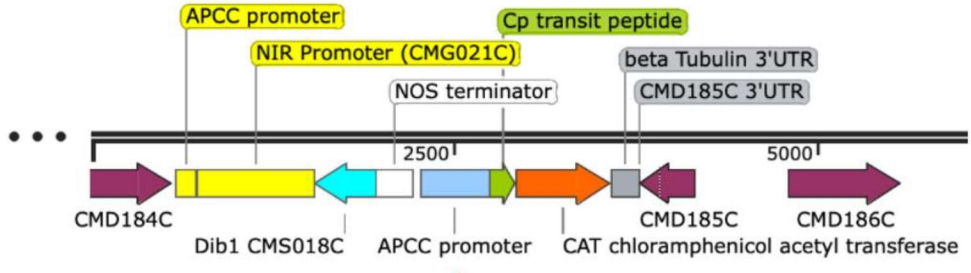


Figure 2-30: Engineered constructs for the genomic integrated gene expression of the antisense and sense U2 snRNA and sense and antisense Dib1 mRNA. Note the NIRp (yellow), the CAT selectable marker (orange), and the HRA for CMD184/185 (purple). a) Linear construct for U2 sense and antisense snRNA (pink) (pSR1096/97). b) Linear construct for Dib1 (sky blue) sense and antisense mRNA (pSR1099/1100).

2.3.2.1.Insertion of U2 and Dib1 sense and antisense into pSR979

I successfully inserted the U2 and Dib1 sense and antisense genes into pSR979. The confirmatory restriction enzyme check showed that the plasmids analyzed were correct for the insertion of the U2 and Dib1 sense and antisense genes (Figure 2-31). The plasmid digestion on the first lane under pSR1100 resulted incorrect, with the lower band being smaller than what was expected. The correct plasmids resulted in pSR1096 and pSR1097 for U2 sense and antisense, respectively, and in pSR1099 and 1100 for Dib1 sense and antisense, respectively (Figure 2-31).

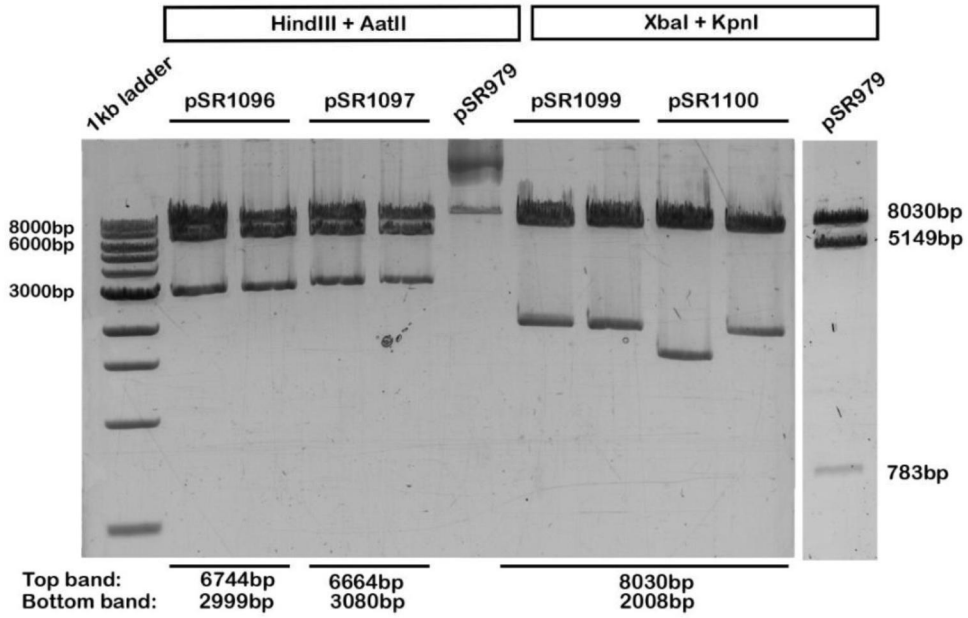


Figure 2-31: Restriction enzyme digest check for pSR1096/1097 (U2 sense and antisense) and pSR1099/1100 (Dib1 sense and antisense). The expected band sizes were the following: for U2 sense 6744bp and 2999bp (lanes labeled pSR1096) and for U2 antisense 6664bp and 3080bp (lanes labeled pSR1097) when digested with HindIII; for Dib1 sense and antisense, the expected band sizes when digested with XbaI and KpnI were 8030bp and 2008bp (lanes labeled pSR1099/1100). Note that the left lane under pSR1100 shows unexpected band sizes. pSR979 (no inserts) was included as a control and digested with HindIII, XbaI, and NcoI. If no insert, the plasmid will linearize when digested with HindIII (lane 6), and the expected band sizes when digested with XbaI and KpnI were 8030bp, 5149bp and 783 bp (lane 11).

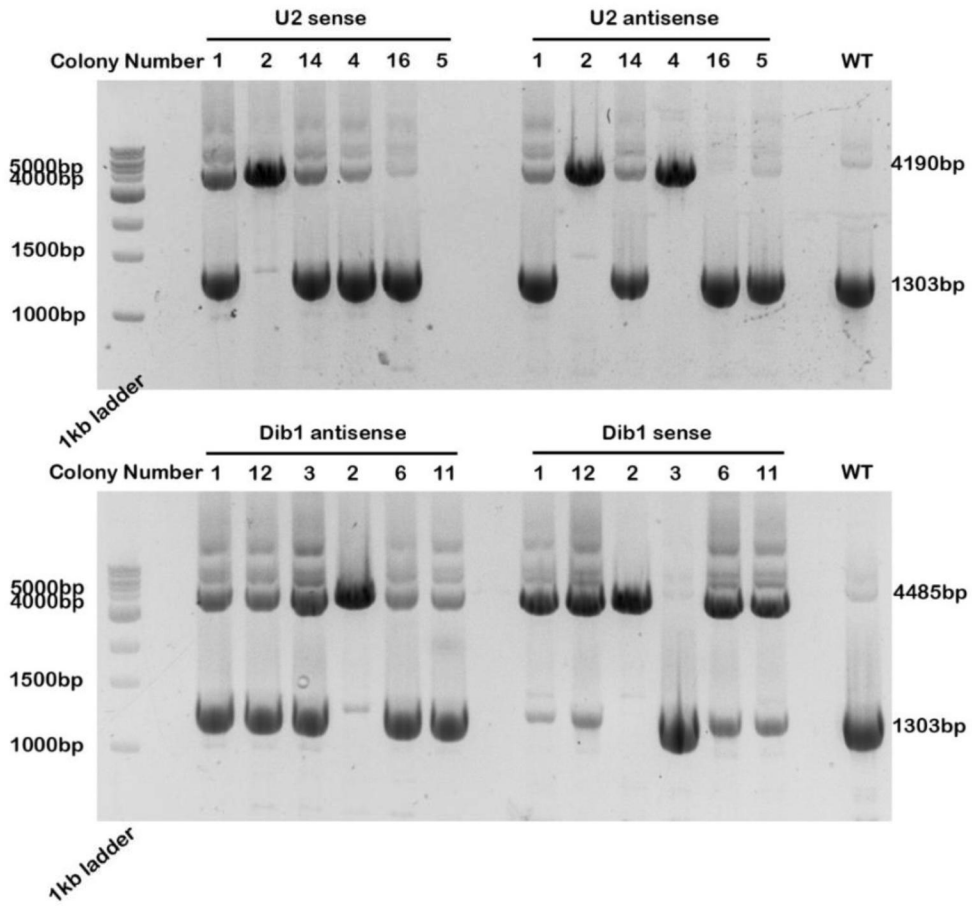
2.3.2.2. Linear construct transformation and transformant selection

Twenty-four hours after transformation, I plated the cells into MA2G media with chloramphenicol and picked individual colonies after ten days. I screened the colonies by colony PCR and chose positive colonies for further screening. I picked six positive colonies per strain and performed PCR with extracted genomic DNA. I used WT genomic DNA as a control to compare the band sizes from the integrated CMD184/185 and WT locus. When performing the PCR with primers flanking the CMD184C (forward) and CMD185C (reverse) genomic locus outside of the homology recombination construct (in our lab referred to as primers H+I (Figure 2-32)), the expected band size for U2 sense and antisense strains was 4190 bp and for Dib1 sense and 4485 bp for Dib1 antisense. On the other hand, the expected band for the WT locus was 1303 bp (Figures 2-33 a and b).

Figure 2-32: Representation of primers H and I location for confirmatory PCR screening.

I confirmed correct integration of my constructs by colony PCR using primers H+I (Figure 2-33a and b).

a)



b)

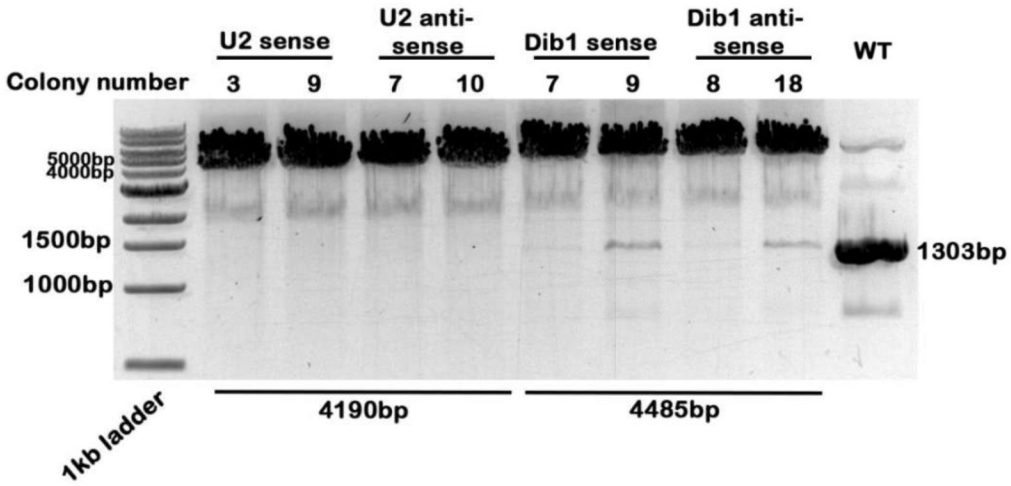


Figure 2-33: a) Confirmatory PCR with primers H+I for U2 sense and antisense and Dib1 sense and antisense strains showing the correct integration of the target genes into the CMD184/185 neutral locus for the first six picked colonies shown positive in the colony PCR screening. 0.7% agarose gel, 1 h run. b) Confirmatory PCR with primers H+I for U2 sense and antisense and Dib1 sense and antisense strains showing the correct integration of the target genes into the CMD184/185 neutral locus for the additional two picked colonies per strain shown positive in the colony PCR screening. 0.7% agarose gel, 1h run.

2.3.2.3. Sequencing

As a second confirmatory test I amplified the region of interest and sequenced it. For this purpose, I performed a PCR with primers H+I from genomic DNA from all the positive colonies from the first and the second confirmatory screenings (3 for U2 sense, 4 for U2 antisense, 3 for Dib1 sense, and 3 for Dib1 antisense) and sent them for sequencing. Sequencing results confirmed the appropriate integration of my target genes into the CMD184/185 neutral locus. I analyzed the results after sequencing and selected two biological replicates per strain based on the smaller number of sequence mismatches from the sequencing results. I used these biological replicates for further experiments. The colony numbers chosen were the following: for U2 sense, colonies 2 and 9; for U2 antisense, colonies 4 and 7; for Dib1 sense, colonies 2 and 7; and for Dib1 antisense, colonies 2 and 8.

Note that from now on, U2 sense 2 and 9 will be referred to as U2 sense A and B respectively, U2 antisense 4 and 7 to as U2 antisense A and B respectively, Dib1 sense 2 and 7 to as Dib1 sense A and B respectively, and Dib1 antisense 2 and 8 to as Dib1 antisense A and B respectively.

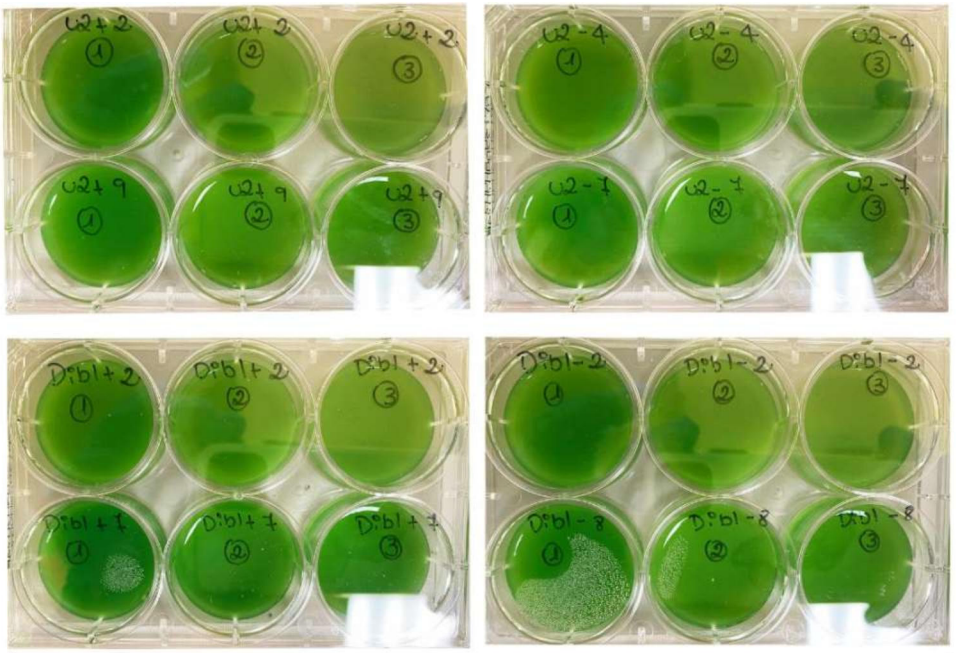
2.3.2.4. Nitrate induction

As for transient expression experiments, I expected that the induction of antisense constructs would inhibit splicing with the integrated strains. When cells were switched to nitrate, the expression of the antisense version of U2 and Dib1 was expected to start, inhibiting splicing and causing a growth rate reduction or cell death. As for the transient gene expression experiments, the sense strains of U2 were possibly expected to show adverse effects on growth for the same reasons explained before (overexpression).

No growth assays were performed for the integrated gene expression nitrate induction experiments. Growth was evaluated considering the initial OD at which the cultures were set up and the final OD after seven days of induction. Samples were only collected 24 and 48 hours after nitrate induction for RNA (U2 and Dib1) and protein (Dib1 and NIRCas9). The initial OD for these experiments was set up to 1.5. The final OD after seven days of nitrate induction can be reviewed in Table 2-16.

Table 2-16: OD_{750} of U2 and Dib1 sense and antisense strains 7 days after nitrate induction.

Sample name	Day 7 OD_{750}
U2+ A.1	13.1
U2+ A.2	15.5
U2+ A.3	10.0
U2+ B.1	12.5
U2+ B.2	14.2
U2+ B.3	15.5
U2- A.1	14.2
U2- A.2	14.6
U2- A.3	15.0
U2- B.1	15.5
U2- B.2	14.2
U2- B.3	15.1
Dib1+ A.1	15.7
Dib1+ A.2	17.7
Dib1+ A.3	16.3
Dib1+ B.1	13.3
Dib1+ B.2	15.1
Dib1+ B.3	15.9
Dib1- A.1	14.2
Dib1- A.2	16.2
Dib1- A.3	14.4
Dib1- B.1	10.4
Dib1- B.2	10.7
Dib1- B.3	8.70
WT.1	9.10
WT.2	10.9
WT.3	11.3
NIR Cas9.1	10.0
NIR Cas9.2	10.3
NIR Cas9.3	11.5



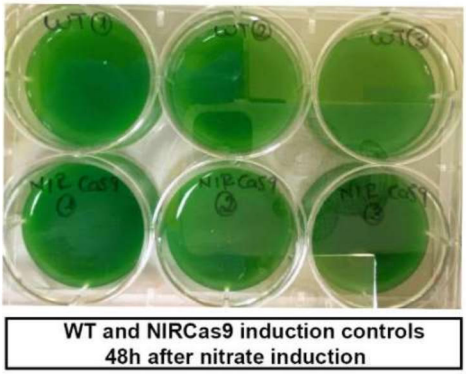


Figure 2-34: U2 and Dib1 sense and antisense, NIRCas9, and WT strains 48 hours after nitrate induction. A 6-well tissue culture plate was used for nitrate induction. Each plate contains the two biological replicates per strain and triplicates for each replicate shown.

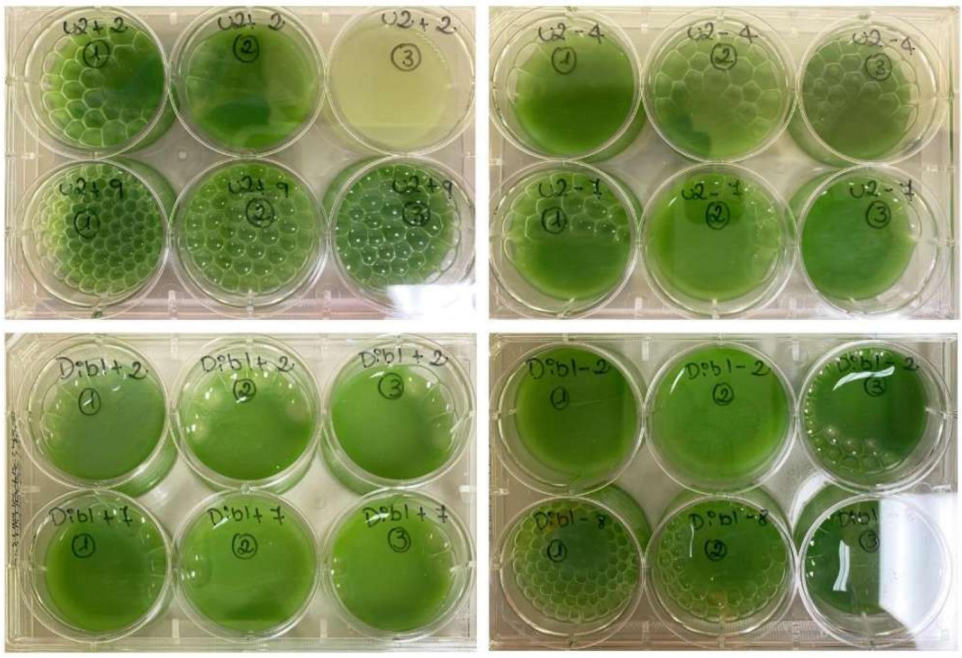


Figure 2-35: U2 and Dib1 sense and antisense strains seven days (d) after nitrate induction.

As for the transient gene expression experiments, I induced the sense and antisense experimental strains for U2 and Dib1 in nitrate by adding chloramphenicol (250 ug/mL), and selection was replenished every three days. The WT control strain was not selected. Most of the experimental strains for both genes grew up to saturation, and no growth rate decrease or cell death could be identified except for one of the U2+ A triplicates, which died after five days (Figure 2-35). The other two U2+ A technical replicates grew normally. The OD for the U2 sense, antisense, and control strains ended up being very similar after day 7 (Table 2-16). The Dib1 antisense strain showed a lower OD than the Dib1 sense and WT control strains at day 7.

2.3.2.5. Northern blotting

To test whether the antisense transcripts were successfully expressed upon nitrate induction in the integrated strains, I analyzed their expression by northern blot. I processed U2 snRNA 24 h samples and performed a northern blot. These experiments used *C. merolae* U5 snRNA (451 nt) as a loading control. In these experiments, I included two biological replicates for U2 sense and two for U2 antisense. When probing with the U2 sense probe (oSDR2311), bands were present in all the samples analyzed for U2 and U5 snRNAs (Figure 2-36). I expected an increase in band intensity for the U2 sense strains. Another possibility was observing a reduction in band intensity for the U2 antisense strains due to dsRNA degradation. Again, as for the U2 snRNA transient gene expression experiments, no difference in band intensity was visually observed for the U2 sense snRNA bands between sense, antisense, and WT strains. I quantified the U2 sense and U5 snRNA bands and normalized the band intensity values for U2 against the band intensity values for U5. The normalization purpose is to discard that any experimental (U2) band intensity variability is due to loading errors. If less RNA was loaded in one lane compared to the other lanes, a reduction in the band intensity for both the experimental and the control bands should be observed. By normalizing the values, the values for this lane can be adjusted to all the other samples, and the variation will be removed.

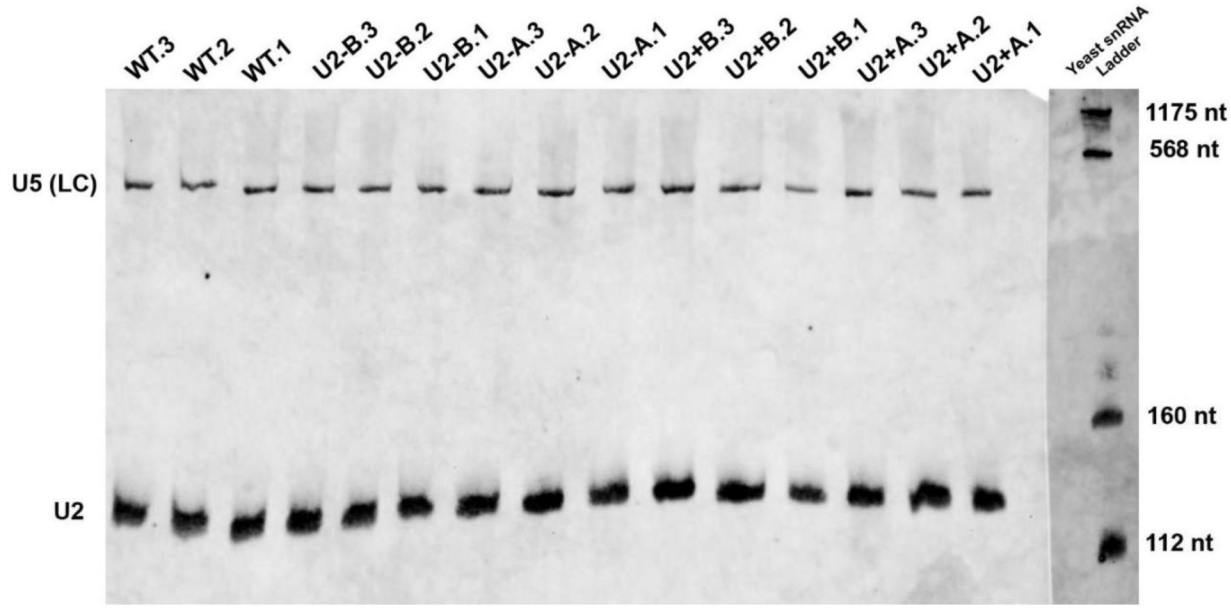


Figure 2-36: Denaturing polyacrylamide northern blot for U2 antisense induction of U2 integrated gene expression strains. 24-hour time point samples were analyzed with a U2 sense probe (oSR2311). 5ug RNA was loaded per lane; a 6% gel was run for 45 minutes. For probe sequences, refer to Appendix 1.

The U2 sense band quantification results, once normalized against the loading control, did not show marked variations between U2 sense, U2 antisense and WT strains (figure 2-37).

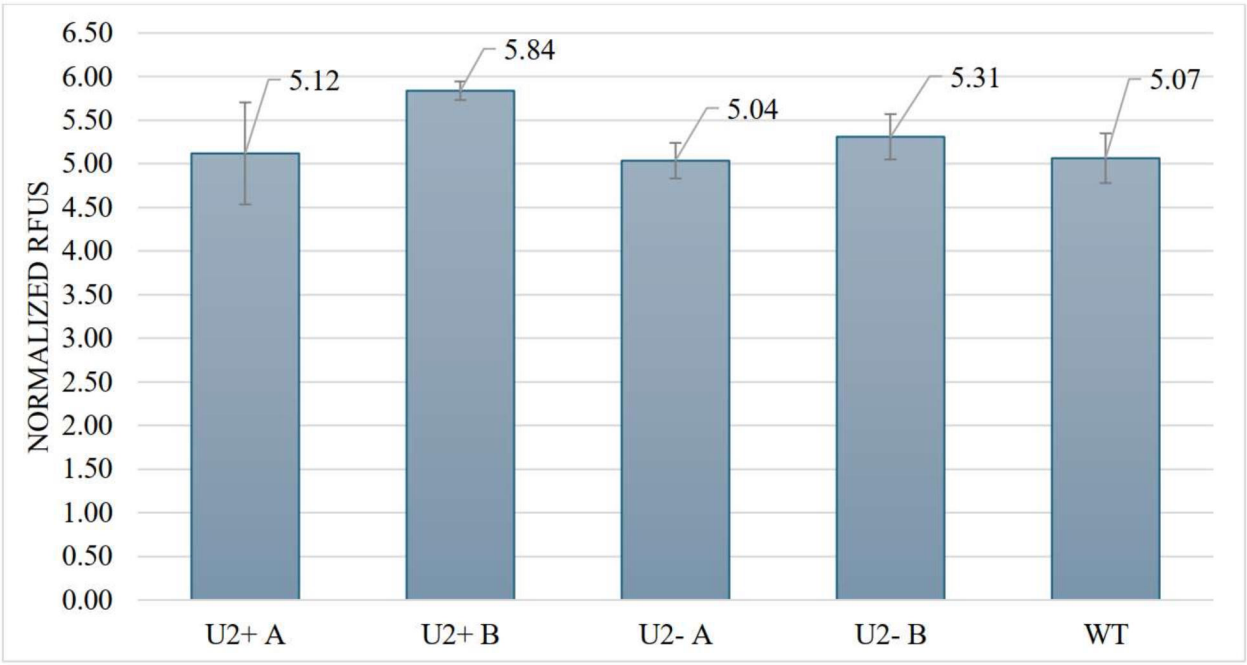


Figure 2-37: Band intensity quantification for U2 sense, U2 antisense, and WT samples from the U2 antisense induction experiment 24-hour samples. Each bar represents the average of the triplicates normalized RFUs analyzed per strain.

I stripped and reprobed the blots with a U2 antisense probe (oSDR2574) as U5 snRNA as a loading control. No bands for U2 antisense appeared for any of the samples (Figure 2-38).

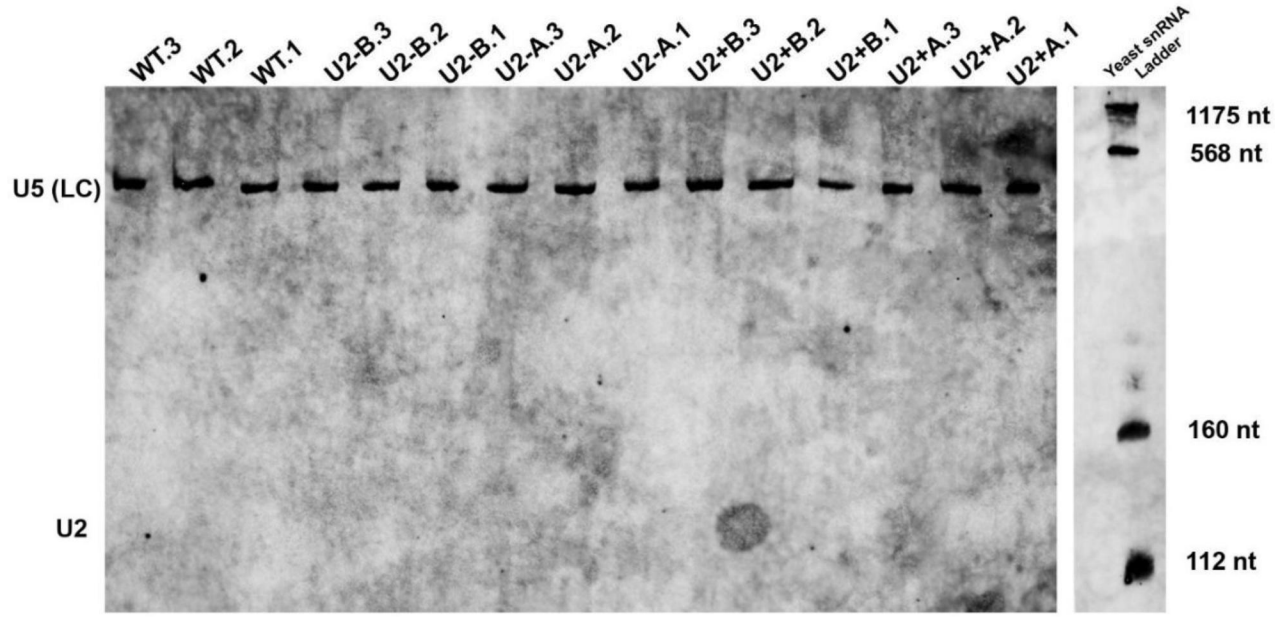


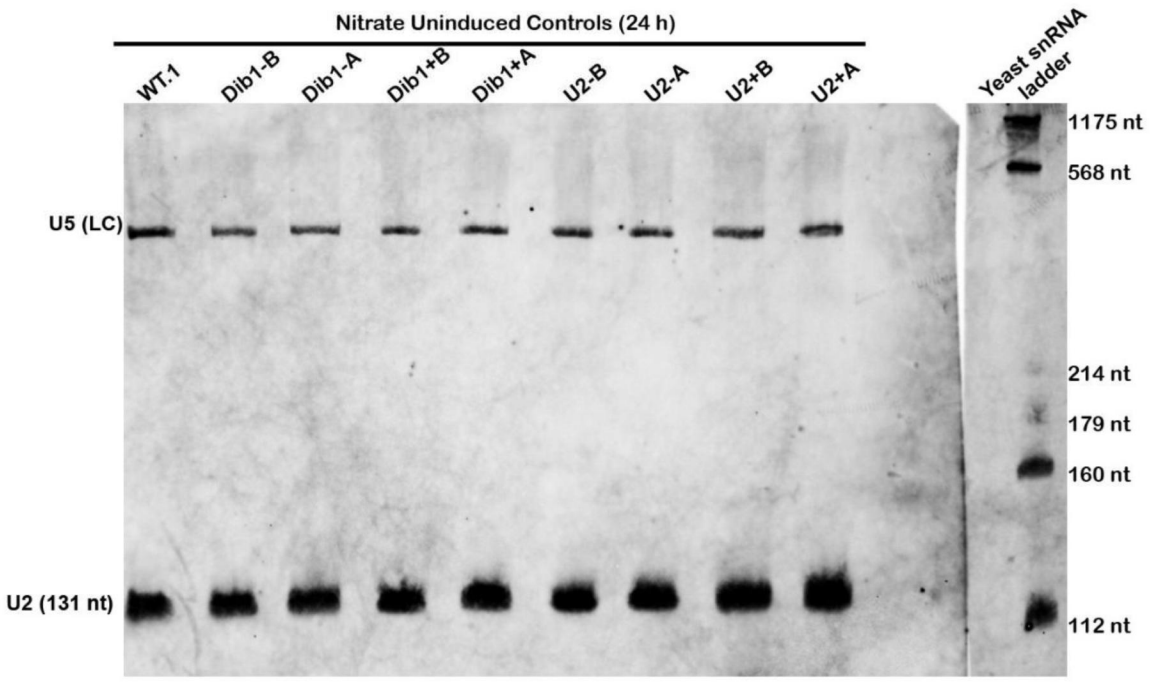
Figure 2-38: Denaturing polyacrylamide northern blot for U2 antisense induction of U2 integrated gene expression strains. 24-hour time point samples were analyzed with U2 antisense (oSDR2574). 5 ug RNA was loaded per lane; a 6% gel was run for 45 minutes. For probe sequences, refer to Appendix 1.

The U2+ A biological replicate shows a lower U2 sense expression than U2+ B (Figure 2-37). Both biological replicates were supposed to behave the same way. However, U2+ A, U2- A and B, and WT show similar U2 expression levels.

To ensure induction of the antisense expression was only driven when cells were induced in nitrate, I analyzed the uninduced samples for each strain as controls. A single band for the U2 snRNA with visibly the same intensity appeared for all the uninduced strains when probing with U2 sense probe (oSDR2311) (Figure 2-39). No bands besides those from the loading control appeared on the blot when probed with the U2 antisense probe (oSDR2574). These results were

the expected, as the cells were never switched to nitrate, and no induction from the sense or antisense U2 genes was supposed to occur (Figure 2-39 a and b). These samples also helped discard any leaky behaviour from the NIR promoter. The uninduced samples were not analyzed in triplicate.

a)



b)

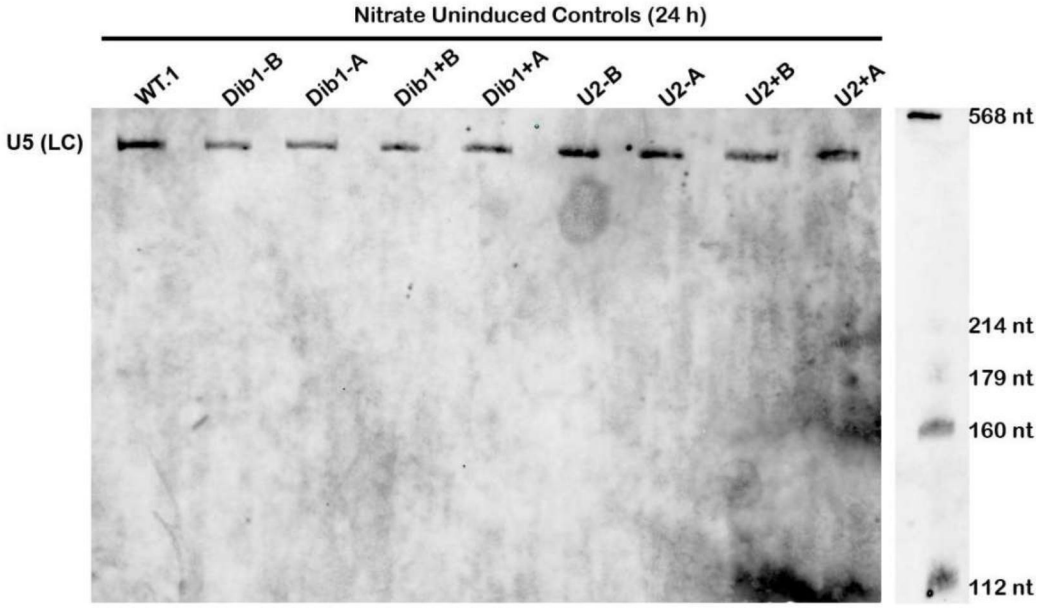


Figure 2-39: a) Denaturing polyacrylamide northern blot for U2 antisense induction of U2 integrated gene expression strains. Uninduced samples for each strain were analyzed with a U2 sense probe (oSR2311). b) Denaturing polyacrylamide northern blot for U2 antisense induction of U2 integrated gene expression strains. Uninduced samples for each strain were analyzed with a U2 antisense probe (oSR2475). 5 ug RNA was loaded per lane; a 6% gel was run for 45 minutes. For probe sequences, refer to Appendix 1.

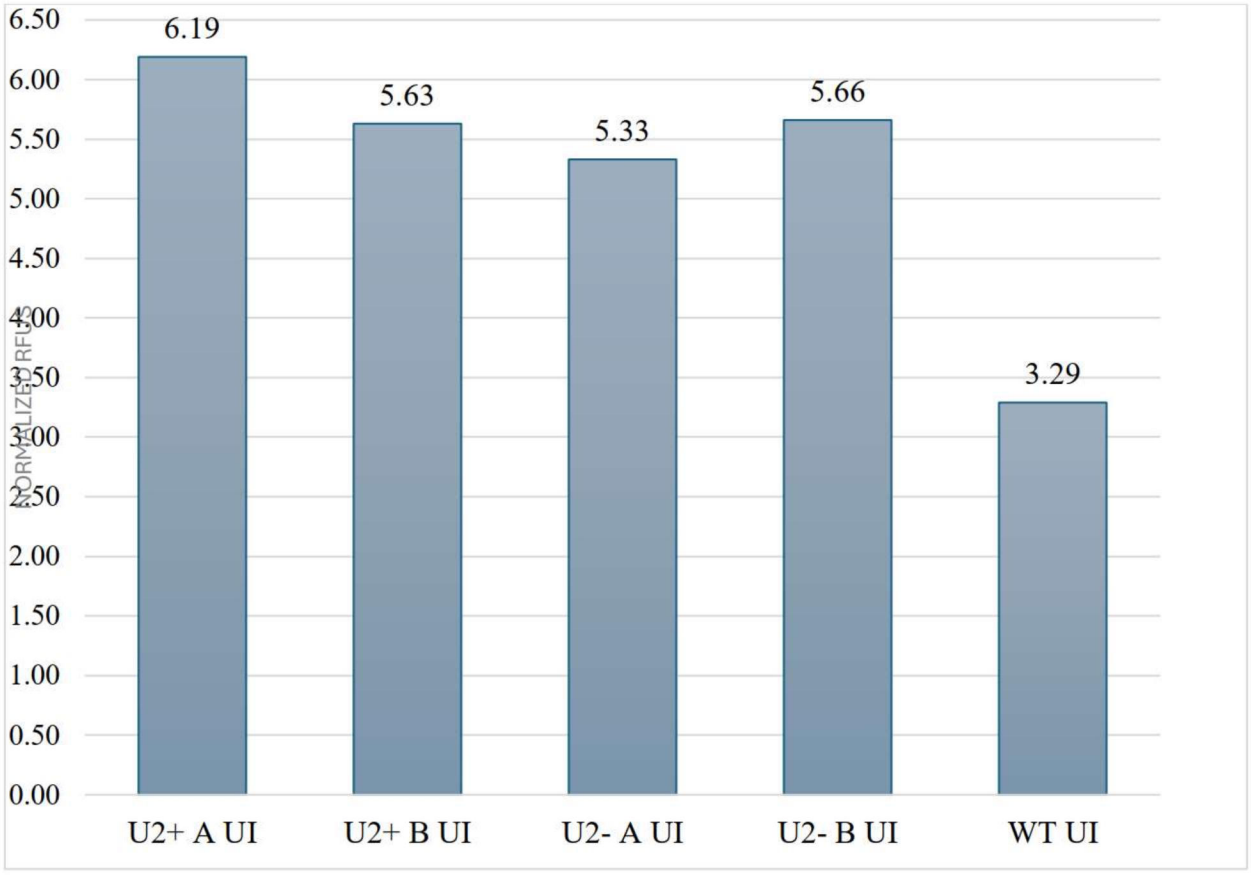


Figure 2-40: Band intensity quantification for U2 sense, U2 antisense, and WT uninduced (UI) samples from the U2 antisense induction experiment. Each bar represents the average of the triplicates normalized RFUs for each strain. Note that no error bars are included in the graph. This is because samples were not analyzed in triplicate, and no SD between their values could be calculated.

To determine whether the band intensities of the uninduced samples were comparable to those of the induced ones, I performed band quantification (Figure 2-40). Note that the RFU normalized values for U2 sense and antisense uninduced strains are very similar to the values obtained for the U2 sense and antisense induced strains when probing with U2 sense. The WT uninduced strain shows a lower U2 sense expression than the U2 sense and antisense integrated strains. As a loading control was used and data was normalized, this should not be attributed to loading errors. However, as mentioned before, the normalized values between induced and

uninduced sense and antisense strains for U2 sense gene expression levels are very similar, and even the WT strain, when in nitrate, shows similar U2 sense levels as the induced and uninduced U2 sense and antisense strains.

To determine whether Dib1 sense and antisense were expressed upon induction, I processed 24h and 48h samples and performed northern blot in two different approaches. For the 24h samples, I performed a denaturing polyacrylamide northern blot and a denaturing agarose northern blot for the 48h samples. For the denaturing polyacrylamide northern blot, I loaded 5 ug RNA per lane (except for the Dib1 uninduced samples in which 2 ug were loaded), and *C. merolae* U2 snRNA (131 nt) was used as a loading control. In these experiments, I included two biological replicates for Dib1 sense and two for Dib1 antisense.

When probing with the Dib1 sense probe (oSDR2597), a smear could be seen at the top of each lane (Figure 2-41). No specific bands appeared for the Dib1 antisense induced, Dib1 sense and antisense uninduced, and WT samples. A specific band appeared for Dib1+ B and Dib1+ A, being the one for the former more intense than the latter. The band was just below the 568 nt band from the yeast snRNA ladder. The size for the Dib1 open reading frame, part of the gene integrated into the CMD184/185 neutral locus, is 474 nt. This band potentially shows the appearance of Dib1 open reading frame transcripts. Interestingly, this band does not appear on Dib1 sense uninduced, Dib1 antisense induced and uninduced, and WT samples. This was unexpected as the Dib1 sense gene is endogenous to all the *C. merolae* strains used for these experiments. Remarkably, the two Dib1 sense biological replicates did not behave similarly, with the Dib1 sense expression higher in Dib1+ B than in Dib1+ A. The U2 snRNA bands, representing the loading control, appeared for all the samples analyzed with visibly the same intensity.

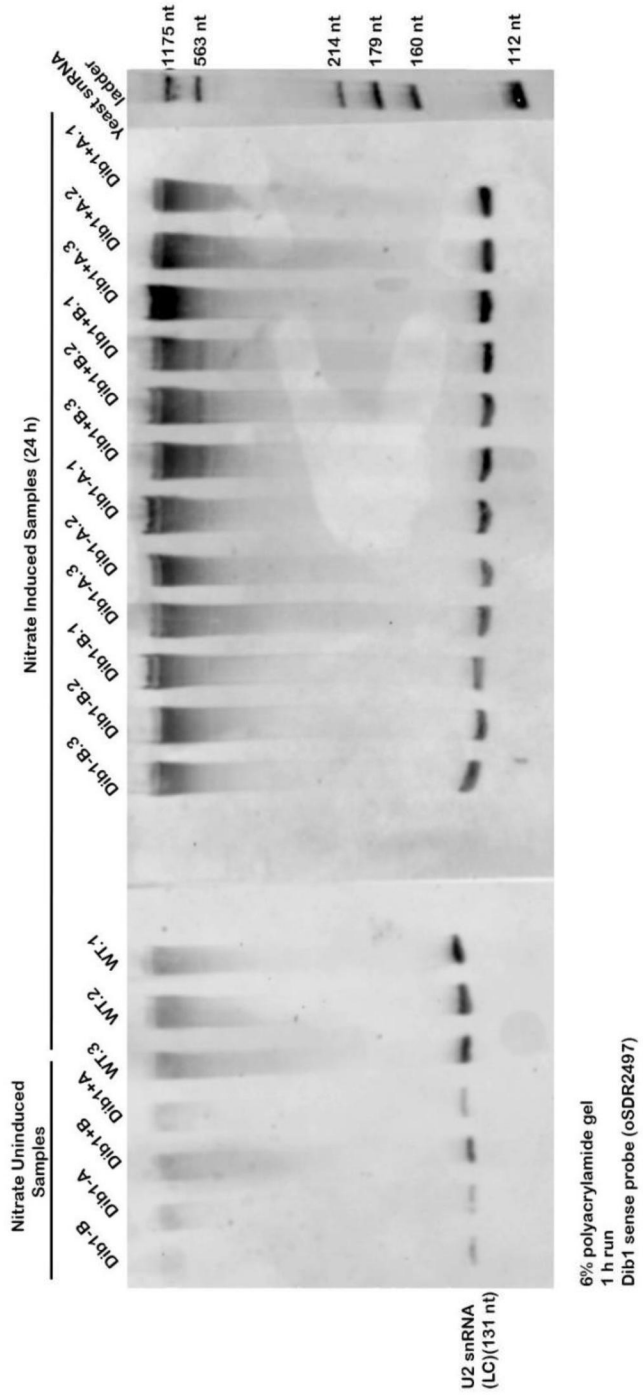


Figure 2-41: Denaturing polyacrylamide northern blot for Dib1 antisense induction of Dib1 integrated gene expression strains. 24-hour time point samples were analyzed with a Dib1 sense probe (oSR2597). 5 ug RNA was loaded per lane; a 6% gel was run for 45 minutes. For probe sequences, refer to Appendix 1.

To determine whether the amount of RNA loaded per lane was the same, I quantified the data by taking the same portion of the smear (including the band for Dib1 sense strains) and normalized it to the loading controls (Figure 2-42). This figure can not give accurate information about what is occurring with the Dib1 sense gene expression on the *C. merolae* strains analyzed, as there is no specific band to show a consistent product for most strains. This graph helps show that the data was appropriately normalized and that regardless the amount loaded, when the experimental band intensity values are normalized against the control, the numbers are all similar to each other.

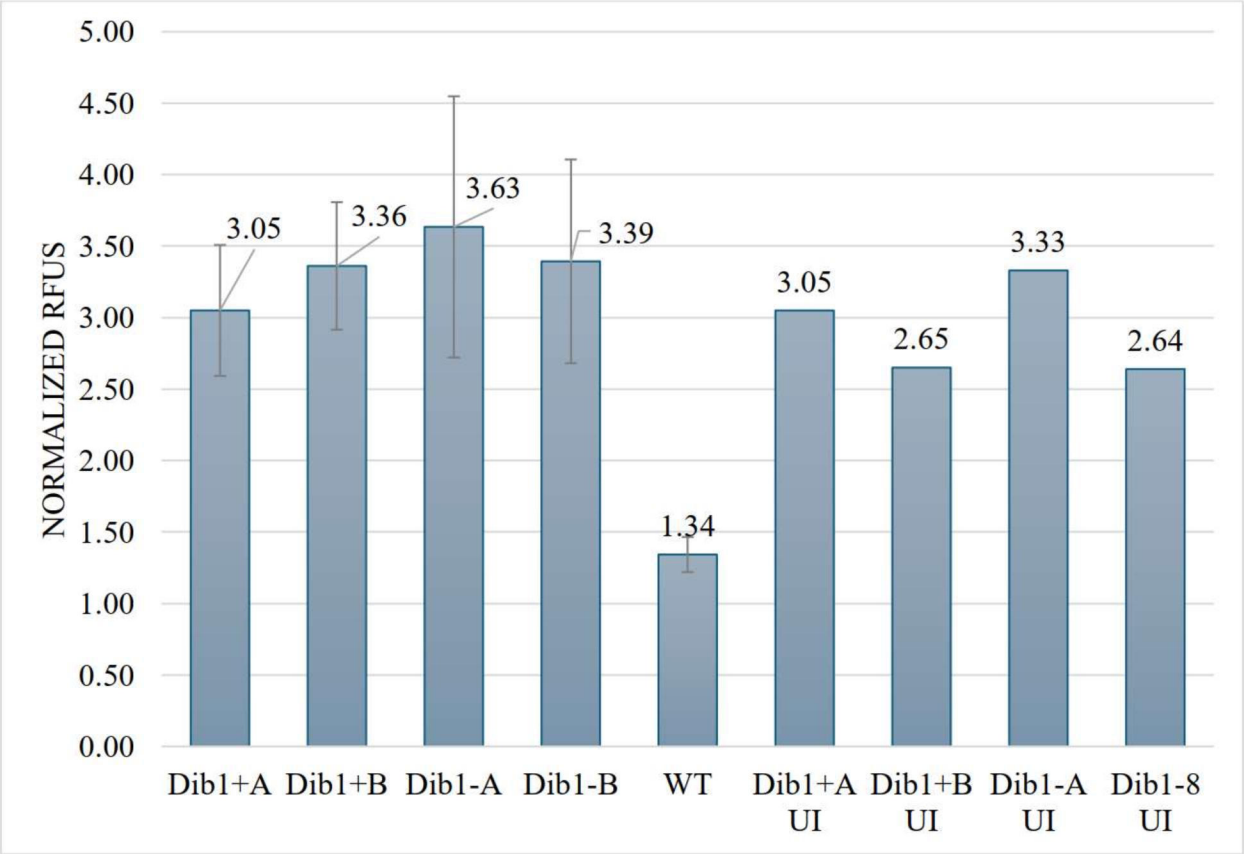


Figure 2-42: Band intensity quantification for Dib1 sense, Dib1 antisense, and WT-induced and uninduced (UI) samples from the Dib1 antisense induction experiment. Each bar represents the average triplicates of the normalized RFUs for each strain. Note that no error bars are included in the graph for the uninduced strains. This is because samples were not analyzed in triplicate, and no SD between their values could be calculated.

To determine if Dib1 antisense was expressed in the antisense strains upon induction, I stripped and reprobed the blot with the Dib1 antisense probe (oSDR2598). As for the Dib1 sense probe blot, a smear appeared at the top of each lane, and no specific bands for any of the samples could be identified (Figure 2-43). The smear for the antisense probe was less intense than the sense probe. The band for U2 snRNA appeared for every sample at approximately the same intensity. This band was previously quantified, showing that the amount of RNA loaded in each lane was approximately the same for all the samples (except for uninduced samples). The scenarios previously described in the U2 antisense nitrate induction from transient gene expression experiments can also be considered.

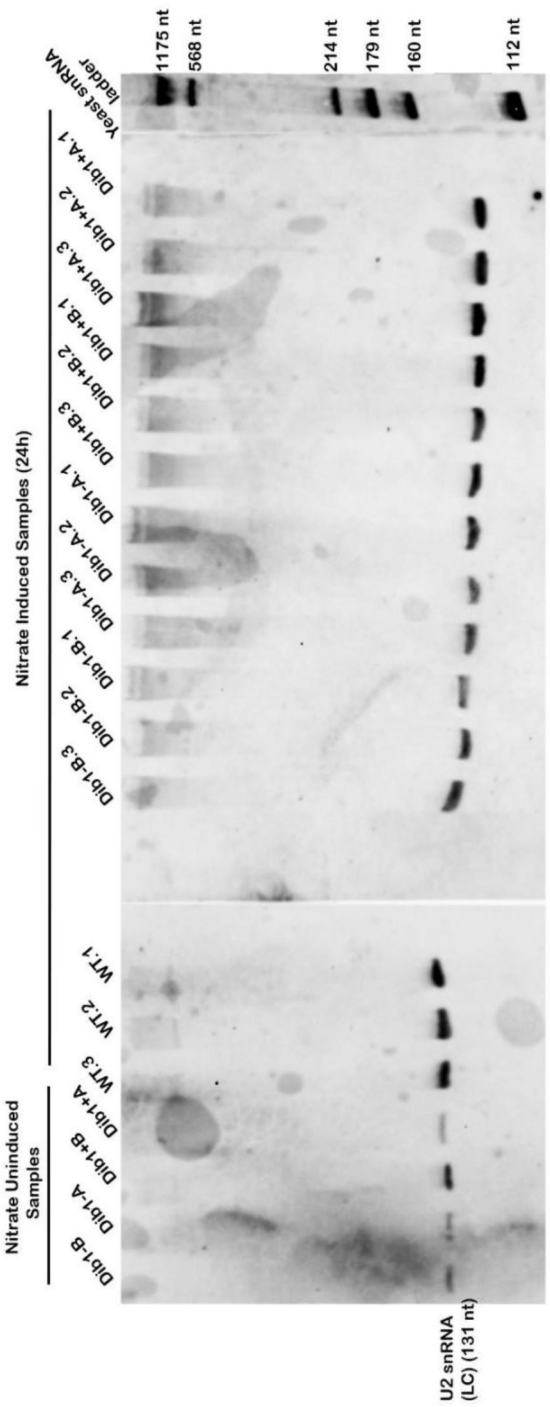


Figure 2-443: Denaturing polyacrylamide northern blot for Dib1 antisense induction of Dib1 integrated gene expression strains. 24-hour time point samples were analyzed with a Dib1 antisense probe (oSR2598). 5 ug RNA was loaded per lane; a 6% gel was run for 45 minutes. For probe sequences, refer to Appendix 1.

To determine whether the Dib1 band absence and smear in the polyacrylamide gel Northern blot was due to the methodology used (polyacrylamide has a high resolution for fragments less than 1 kb – Dib1 ORF is 474 nt but transcripts can include pre-mRNA, mRNA and untranslated regions to up to more than 2 kb) I performed a denaturing agarose northern blot with the Dib1 48-hour samples (Figure 2-44). I loaded 25 ug RNA. Four bands appeared for the Dib1 sense-induced strains when the blot was probed with a Dib1 sense probe. Again, the bands for the Dib1+ B strains were more intense than the ones for the Dib1+ A strain. Interestingly, these bands appear only for these strains and are absent in all the Dib1 antisense-induced strains, Dib1 sense and antisense uninduced strains and WT controls. This result is inconsistent with the result in the polyacrylamide northern blot. However, as agarose has lower resolving power than polyacrylamide, it is possible that the bands do not run true to size. U2 snRNA was used as a loading control, and the corresponding band appeared for every sample analyzed. The bands were not quantified, as there were three additional bands to the one that possibly represents the Dib1 open reading frame. I could not determine what the other bands represent at a molecular level. Additionally, the loading control shows equity in the amount of RNA between samples, confirming that the two Dib1 sense strains are not behaving in the same way.

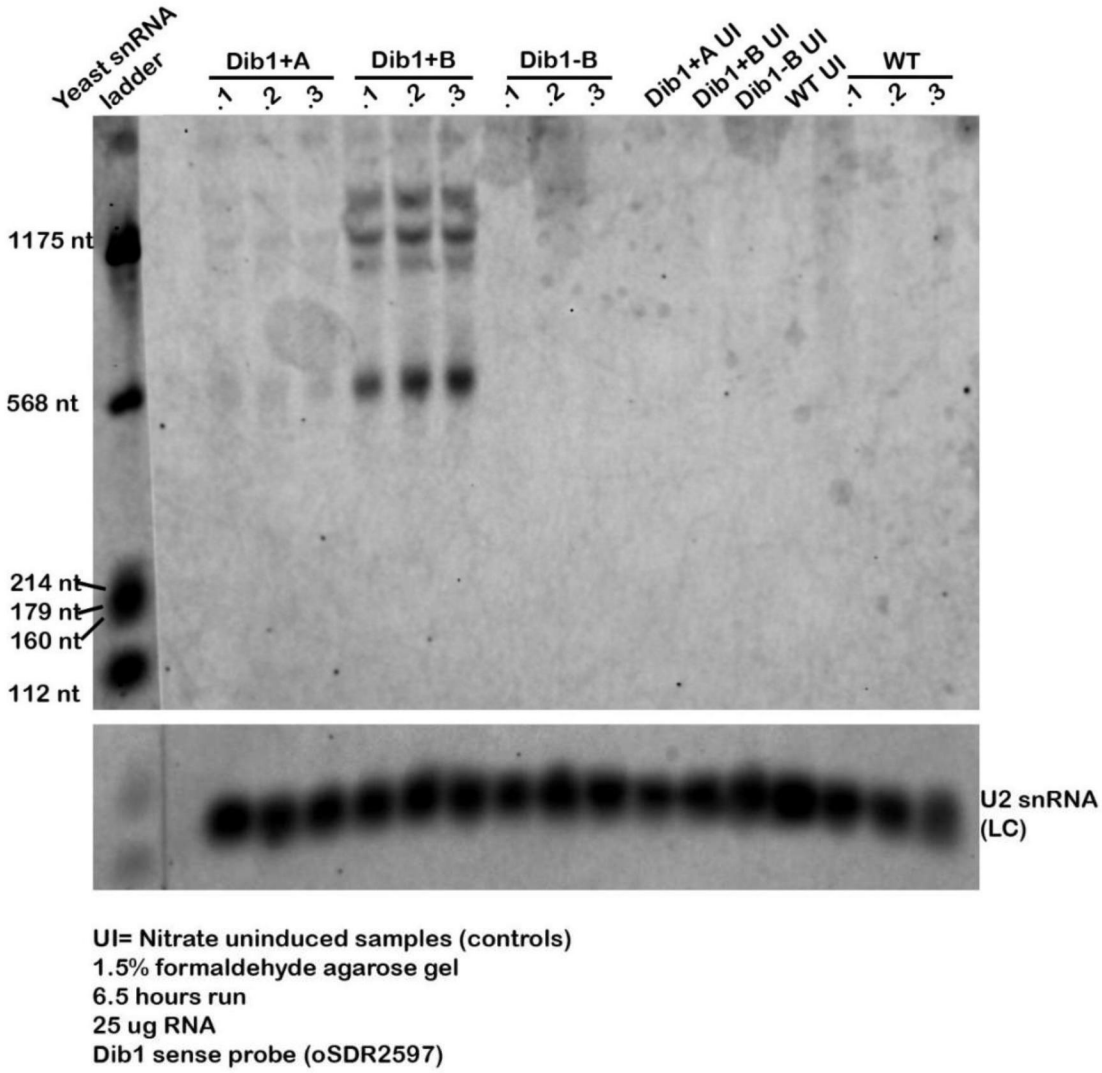


Figure 2-45: Denaturing agarose northern blot for Dib1 antisense induction of Dib1 integrated gene expression strains. 48-hour time point samples were analyzed with a Dib1 sense probe (oSR2597). 25 ug RNA was loaded per lane, and a 1.5% gel was run for 6.5 hours; refer to Appendix 1 for probe sequences.

To determine whether Dib1 antisense expression could be detected with this technique, I stripped and reprobed the blot with a Dib1 antisense probe. No bands appeared for any of the samples (Figure 2-45).

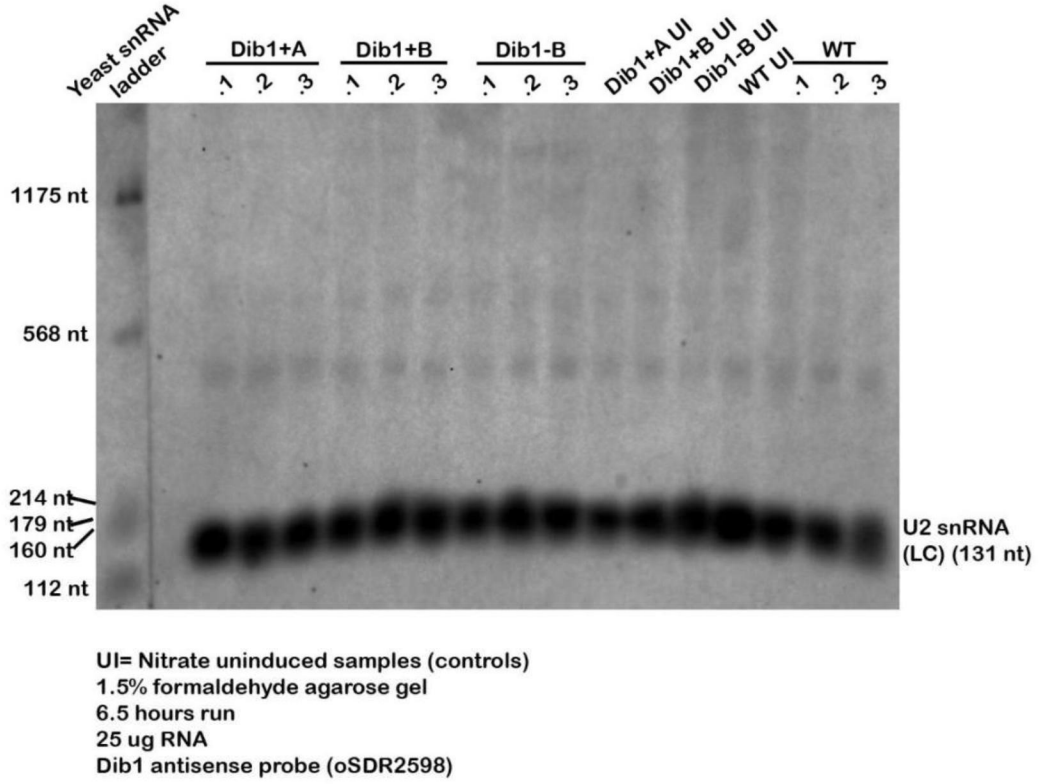


Figure 2-46: Denaturing agarose northern blot for Dib1 antisense induction of Dib1 integrated gene expression strains. 48-hour time point samples were analyzed with a Dib1 antisense probe (oSR2598). 25 ug RNA was loaded per lane; a 1.5% gel was run for 6.5 hours. For probe sequences, refer to Appendix 1.

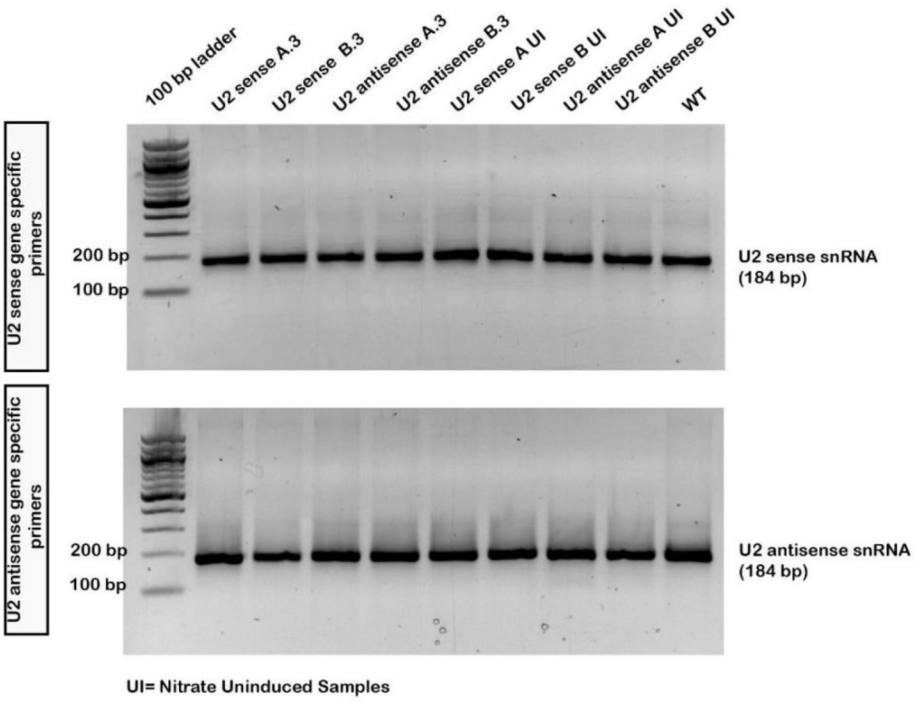
2.3.2.6. Gene-specific primer reverse transcription and PCR for expression evaluation

To discard the possibility that the band absence for the antisense U2 and Dib1 strains was due to northern blotting technical or molecular issues (transfer issues, low RNA abundance which makes it undetectable, etc.), one triplicate of each biological replicate for each nitrate-induced strain from U2 and Dib1 was selected to synthesize cDNA and perform end-point PCR. The uninduced samples for U2 and Dib1 sense and antisense strains were also analyzed. The U2-integrated gene expression nitrate induction RT-PCR results were unexpected. As shown in Figure 2-46a, the end-point PCR for U2 sense and U2 antisense gene-specific primers reverse transcription showed bands for every sample. The expected results were the following:

1. When performing PCR with the reverse transcribed samples with U2 sense primers, bands should appear for every sample, with a higher band intensity for the U2 sense strains.
2. Bands should appear only for the U2 antisense strains when performing PCR with the reverse transcribed samples with the U2 antisense GSP. If bands do not appear for U2 antisense strains in this case, it could be possible, as for the northern blot results, that double-stranded RNA degradation was occurring.

The band intensities for all the samples were the same when performing PCR with the U2 sense reverse-transcribed samples, but, surprinsingly, bands also appeared for every sample when performing PCR with the antisense primers reverse-transcribed samples. This showed no consistent results about U2 sense and antisense gene expression. An NRT control was included for each sample from which cDNA was synthesized to discard the possibility of band appearance due to genomic DNA contamination (Figure 2-46 b).

a)



b)

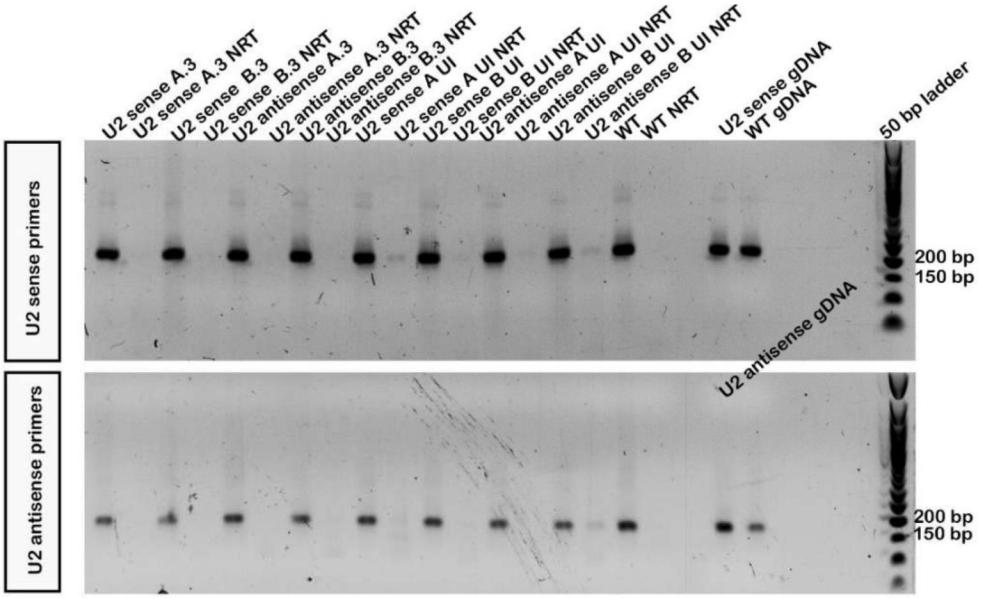
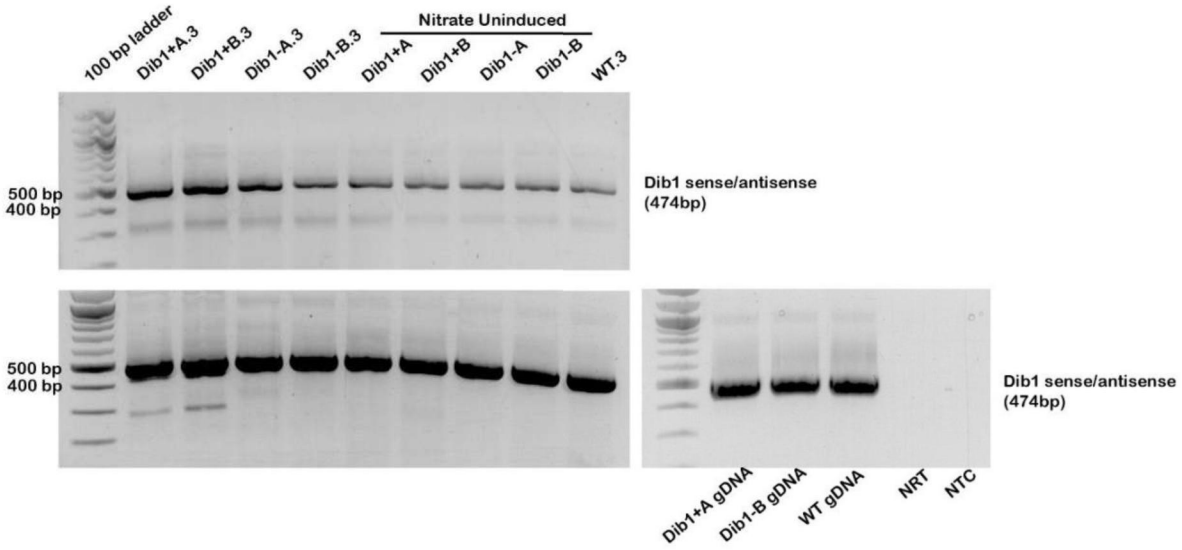


Figure 2-47: An RT-PCR reaction for U2 sense and antisense induced and uninduced stains and controls. Top gel: U2 sense GSP reverse transcribed samples. Bottom gel: U2 antisense GSP reverse transcribed samples. b) RT-PCR reaction for U2 sense and antisense strains and WT controls. Top gel: U2 sense PCR reaction with U2 sense GSP reverse transcribed samples. Bottom gel: U2 antisense PCR reaction with U2 antisense GSP reverse transcribed samples. NRTs are included per sample.

I followed the same procedure for the Dib1 sense and antisense-induced and uninduced strains. When performing the PCR with Dib1 antisense primers from samples that were reverse transcribed with Dib1 antisense primers, bands were only expected for the induced antisense strains. However, bands appeared for every sample, including Dib1-induced and uninduced sense strains, Dib1 uninduced antisense strains, and WT controls (Figure 2-47a). NRT controls were run for each sample, and Dib1 sense, antisense, and WT genomic DNA were used as positive controls (Figure 2-47b). The NRT controls did not show gDNA contamination, therefore, the appearance of non-desirable bands cannot be attributed to this. An NTC control was also included (Figure 2-47 a).

a)



b)

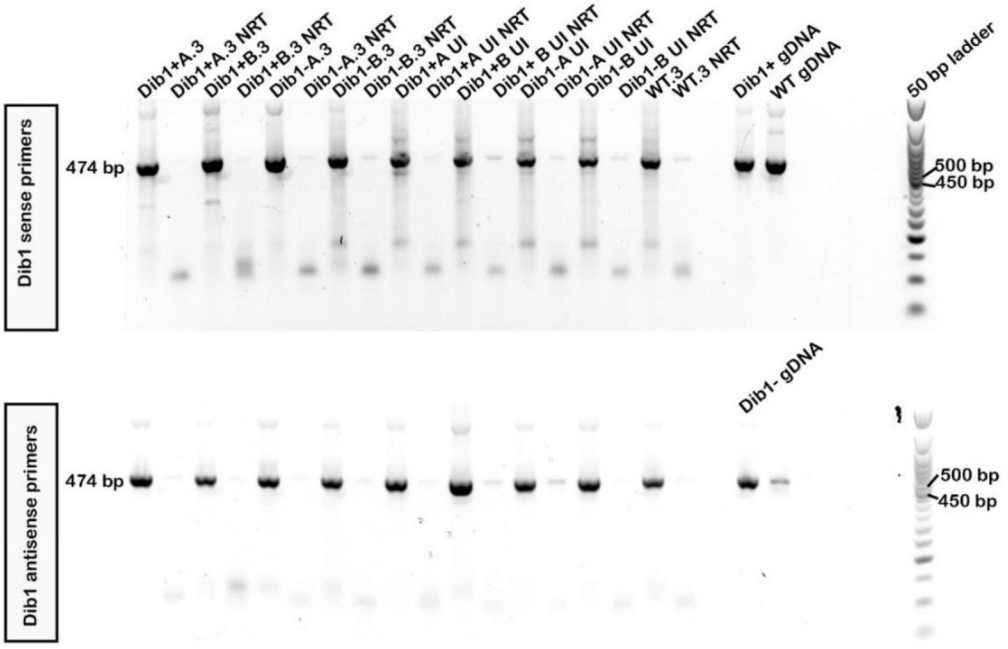


Figure 2-48: a) RT-PCR reaction for Dib1 sense and antisense induced and uninduced stains and controls. Top gel: Dib1 sense primers reverse transcribed samples. Bottom gel: Dib1 antisense GSP reverse transcribed samples. Bottom gel right: Dib1 sense, antisense and WT genomic DNA controls, NRT, and NTC. All 1.5% agarose gels ran for 40 minutes. b) RT-PCR reaction for Dib1 sense and antisense strains and WT controls showing the NRT controls for all the samples. Top gel: Dib1 sense PCR reaction with Dib1 sense primers reverse transcribed samples. Bottom gel: Dib1 antisense PCR reaction with Dib1 antisense primers reverse transcribed samples. NRTs are included per sample. All 1.5% agarose gels ran for 40 minutes.

To confirm the possibility of having gotten the Dib1 antisense bands in the Dib1 sense and WT strains, because of the presence of an overlapping gene, I looked for a *C. merolae* gene that does not have any overlapping sequences and for which some work had been previously performed in our lab. CMK260C was the chosen gene. CMK260C gene-specific primers were used to reverse transcribe RNA from Dib1 sense, antisense and control strains. CMK260C does not have an antisense version in any of the experimental strains. No product was expected to be seen when the reverse transcription was performed with antisense CMK260C gene-specific primers. Surprisingly, I obtained bands for the CMK260C antisense GSP reverse-transcribed samples, just as with the sense CMK260C GSP reverse-transcribed samples (Figure 2-48). NRT controls were not included per sample this time as we had already discarded the presence of genomic DNA contamination for Dib1 strains in the previous experiment. WT genomic DNA was included as a positive control in the PCR reactions.

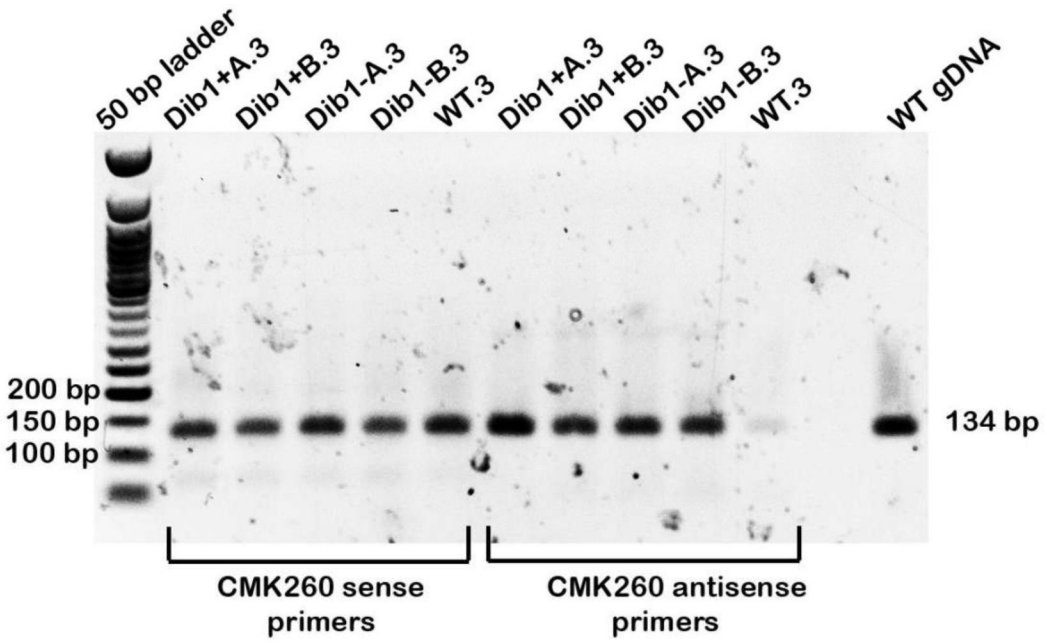


Figure 2-49: RT-PCR reaction for CMK260C in Dib1 sense, antisense strains, and WT controls. Top gel: CMK260C sense PCR reaction with CMK260C sense GSP reverse transcribed samples. Bottom gel: CMK260C antisense PCR reaction with CMK260C antisense GSP reverse transcribed samples. 1.5% agarose gel, 40 min run.

2.3.2.7.NIRCas9 Western Blot

To discard nitrate media issues that would affect the induction effectiveness, I induced a NIR-cas9 strain in nitrate as a control and performed western blots with samples taken 24 hours after induction. The procedure followed was the same as for the western blots performed for transient gene expression experiments, and I used the same antibodies at the same concentrations. Figure 2-49 shows the NIR-Cas9 strain western blot for the integrated gene expression experiments. Note that an uninduced strain and WT samples were included as controls (figure 2-49).

No band appeared for the WT control (Figure 2-49), which was the expected result as this strain does not contain the Cas9 protein integrated gene. However, the uninduced sample shows two thin bands of the same size as the bands representing the Cas9 protein in the induced samples, which, in this case, were run in triplicate.

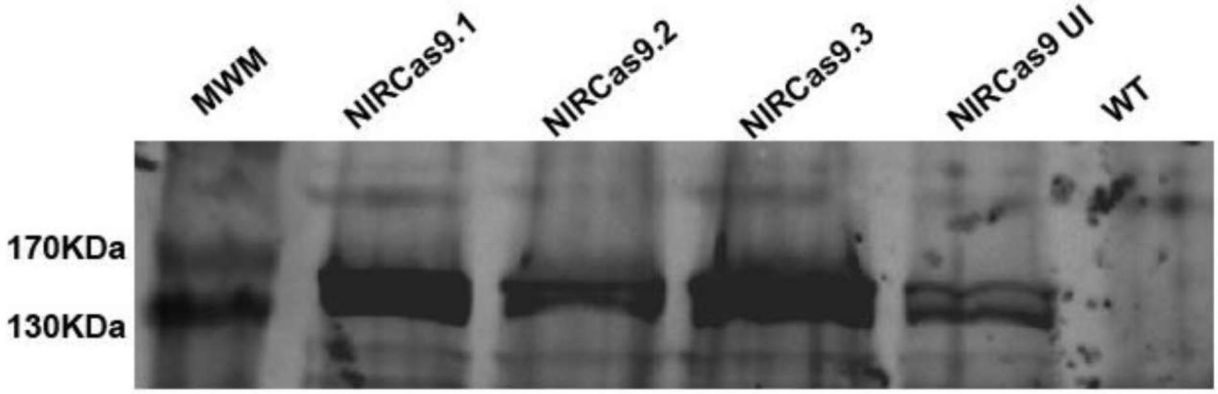


Figure 2-49: NIR-Cas9 western blot for the integrated gene expression experiments 24-hour samples. Note the uninduced sample and the WT control.

2.4.Discussion

Overall, the antisense transient or integrated gene expression of splicing components was unsuccessful and I did not observe splicing inhibition. Antisense RNA oligonucleotides have been extensively used in eukaryotic organisms to trigger gene knock-down of target genes, but, as explained before, this is not possible in *C. merolae.*

Instead, in previous work in *C. merolae* the entire length of the target genes was used for antisense inhibition (Ohnuma *et al*., 2009). In that work, the antisense version of the catalase gene was successfully expressed, resulting in a reduction in catalase activity. The catalase gene was significantly downregulated in the cells transformed with the plasmid that contained the antisense version of the catalase gene. Notably, different and in contrast to my experiments, in Ohnuma *et al*., 2009 the promoter used to express the catalase gene was its own promoter, and its

expression was constitutive and not induced. For this project, constitutive expression was unsuitable as the main aim was to knock down an essential process for cell survival. It is therefore possible that promoter differences somehow led to my lack of success in inducing antisense expression. It is also possible that the intense selective pressure of inhibiting expression of an essential gene rapidly selected for "escape" mutants, for example cells with promoter mutations in the inducible promoter.

When designing the antisense constructs, only the open reading frame for the genes of interest was considered. In the case of the snRNAs, the primary (unprocessed) transcripts are not the 131 nt known for mature U2 or the 174 nt known for U4. These genes are transcribed in very long transcripts, more than 2kb, and then processed to get the final t U2 and U4 snRNAs. It is thought that the cleaved parts of the snRNAs, during their processing, contain critical regulatory elements to control their expression as well as different motifs that serve as binding sites for the complexes involved in the snRNA processing.

Not including the whole length of the transcripts when expressing the antisense version of these genes was probably a significant mistake. In the Rader Lab, a former student working with U5 snRNA demonstrated that overexpression of the gene from a plasmid was possible. This plasmid was designed to include the entire U5 snRNA primary transcript (~4 kb). This suggests that the removed regions during the U5 snRNA processing are, in fact, necessary for the appropriate expression of the gene. For future attempts in splicing inhibition by antisense RNA, a good idea would be designing a plasmid in which the whole snRNA transcript is included and only the final sequence is swapped to the antisense version for transient or integrated gene expression. Swapping the entire original transcript will cause the essential regulatory elements to be in incorrect places and may not be recognized by the snRNA processing machinery.

Another possible reason that the antisense experiments performed for this project with U2 and U4 snRNAs did not work hinges on the secondary structures of snRNAs, which are essential for their interaction with other splicing components. The endogenous U2 and U4 snRNAs acquire these structures as soon as transcribed and processed. The antisense version of these snRNAs should also be able to develop secondary structures. If the sense and antisense versions of the snRNAs are appropriately expressed, but acquire secondary structures, their encountering and interaction will be compromised. It is true that the 5' and 3' ends of the U2 and U4 snRNAs are not base-paired and contain RNA binding sites to interact with other snRNAs from the spliceosome. However, these base-pairing regions are limited and may not be enough to maintain the sense: antisense snRNA duplex with sufficient stability to avoid interaction of the sense snRNA with other spliceosomal components.

Regarding Cef1 and Dib1, the issues that might have been encountered with the antisense snRNA expression do not apply to these genes. Cef1 and Dib1 are protein-coding genes that produce linear transcripts. In *C. merolae* these are intron-less genes and only the open reading frame was considered. It was more surprising that no inhibition was observed for these genes as the experiments were designed similarly to those in Ohnuma *et al*., 2009, except for using the gene's constitutive promoter in the latter.

When Cef1 transient gene expression results were inconclusive, the focus of why the experiments were not working was more on the promoter. The NR promoter is the least strong promoter of the three nitrate-inducible promoters, as was proven by Fujiwara *et al*., 2015. As mentioned previously, there was a concern about gene overexpression and deleterious effects on cells, which is why this promoter was chosen for these experiments. However, as antisense expression was not detected, the thought was that the promoter might be too weak to induce gene

expression. Thus, I switched to the NIR promoter for the integrated gene expression experiments, which is the strongest of the nitrate-inducible promoters (Fujiwara *et al*., 2015). Unfortunately, this promoter did not show any expression either. Why this happened remains unclear as it has been demonstrated that GFP expression can be controlled by the three nitrate inducible promoters, showing successful expression, and expression levels could be monitored by fluorescent microscopy and transcriptomic studies (Fujiwara *et al*., 2015). If the obtained results in Fujiwara *et al*., 2015 have something to do with the fact that GFP is an exogenous protein that was integrated into the *C. merolae* genome and whose expression does not affect any cellular processes, then for this project, the results obtained are not entirely unexpected. This work focuses on splicing genes, which are thought to be essential for *C. merolae* cell survival. It is unknown whether the cells are activating some biochemical mechanism to avoid these foreign sense and antisense genes from being expressed and protect the cells against splicing inhibition, but it could be a possibility. Interestingly, despite the advantages of using inducible systems to control gene expression, it has been shown that heat-shock or nitrate-inducible gene expression systems have a significant disadvantage. They cause physiological defects in cells. For this work, switching the cells into nitrate media to induce the expression of the antisense genes implies a metabolic burden and visible phenotypes such as culture colour and reduced growth rate were observed. The cultures in nitrate turn very starchy and light-yellowish green when cultured up to saturation. The reason why the cells get this colour is unclear but indicates that nitrate is not the preferred nitrogen source for the cells. For expressing a gene that will harm the cells or repressing an essential gene by nitrate-inducible systems, the initial physiology needs to be considered. Trying to inhibit an essential process, such as splicing under non-desirable conditions for the cells, is something to consider for future experimental designs.

An important fact to consider for future attempts, as for the experiments with snRNAs, is to engineer the plasmids considering the entire length of the genes, including the 5' and 3' UTRs. It is known that the 5'UTRs contain essential regulatory elements that control gene expression, such as upstream open reading frames, internal ribosomal entry sites, microRNA binding sites and structural components involved in maintaining mRNA stability (Ryczek *et al.*, 2023). These regulatory elements may play an essential role in the appropriate expression of the target genes, leading to unsuccessful splicing inhibition. The plasmids should be designed with the original 5' and 3' UTRs, and only the open reading frame should be swapped for the antisense version.

It would have been very useful to know with absolute certainty what happens when a sense: antisense RNA duplex forms in *C. merolae* to design and drive the experiments in the most straightforward and effective way. As described in the introductory section of this chapter, some genes that code for enzymes that may be potential double-stranded RNA degraders have been identified in *C. merolae.* However, these have yet to be experimentally confirmed as dsRNA degraders. Two assumptions were then taken for analyzing the results of these experiments: 1) the potential dsRNA degraders in *C. merolae* will degrade dsRNA when the sense: antisense duplexes form, and 2) if any enzyme did not degrade the double-stranded RNA duplexes, the sense and antisense interaction will make the sense version inert and unable to interact with other splicing components in the snRNAs case, and will prevent the mRNA from being translated into protein in the case of Cef1 and Dib1. Both assumptions will lead to splicing inhibition. However, not knowing with certainty what is happening at the molecular level makes the analyses of the results very complex. If dsRNA were being degraded, the antisense version of the genes would not be identified, but this must be consistent with the observation of a reduction in the sense version of the genes in the antisense strains. The antisense RNA for all the genes chosen was

never observed, but a decrease in the sense levels was not observed either, so there was no consistency in the results. Supposing the dsRNA was only inert avoiding splicing components to interact with it (snRNAs) or inhibiting translation to protein (Cef1 and Dib1), the antisense version should have been identified when using denaturing techniques. Unfortunately, this never happened. The only conclusion that could be drawn from these results is that the gene induction of the sense and antisense versions of the genes was unsuccessful. A mixture between both assumptions was observed, but as what was realistically happening inside the cells was unknown, something needs to be consistently concluded from the antisense transient and integrated gene expression experiments.

Northern blot results did not give any clue of nitrate induction. As mentioned in the results section, no bands for U2, U4, or Cef1 antisense were detected for transient gene expression experiments. Bands were detected only for the sense versions of U2 and U4, and nothing was detected for Cef1. Three scenarios could be considered to explain band absence:

- U2, U4, and Cef1 antisense bands were expected to be seen in the antisense strain samples, considering that dsRNA molecules formed when the sense and antisense transcripts encountered inside the cells and did not follow degradation. The denaturing nature of the gel used to run the samples would have taken the double-stranded molecule apart, and the band for the antisense transcripts would have been detected. If this was the case, an increase in the target gene's sense band intensity in the sense strains must be observed for the results to be consistent and so one can conclude that expression is occurring.
- Antisense transcripts were expressed at a very low level, considering that the promoter used for these experiments (NRp) was the weakest from the nitrate inducible promoters

(Fujiwara *et al*., 2015) and, for instance, was undetectable by northern blotting. If this were the case, there would be no increase in band intensity for the sense transcripts of any of the genes in any strain. Also, in this scenario, splicing was not expected to be inhibited as the antisense transcripts' insufficient expression would not effectively block the endogenous sense molecules. This result would be the ideal scenario for the results obtained, but nothing can be told about the occurrence of antisense gene expression. For instance, it can not be concluded that splicing is inhibited, and as the cells did not show growth defects, splicing is not essential for *C. merolae* .

- Antisense transcripts were expressed properly, forming dsRNA molecules with endogenous sense transcripts. *C. merolae* does not possess a known dsRNA degradation machinery, but it has been determined that several genes in the algae code for RNAase III-like enzymes, which in other organisms are in charge of dsRNA degradation (see the introduction for further details). It could be possible that the dsRNA was forming and being degraded; in this case, no bands in the antisense strains would be expected when probing with the antisense probes. These results, though, must be consistent with observing an increase in the band intensity for sense transcripts in the sense strains and a possible decrease in band intensity or band absence for sense molecules in the antisense strains (as sense transcripts are being degraded with antisense transcripts). According to band quantification results (Figure 2-27), none of these could be seen.

The transient gene expression antisense induction experiments were inconclusive. Northern blotting could not detect sense or antisense expression for any of the three genes. Based on these results, nothing can be said regarding antisense gene expression and splicing inhibition.

For U4 antisense experiments, Northern blots reduction in band intensities and band absences, even in control samples as pSR1025 (empty vector), could be attributed to improper transfer and technical issues during the RNA extraction procedure with the kit used. On the other hand, this can not be attributed to adequate antisense gene expression and dsRNA degradation, as the band is present at approximately the same intensity in the other two antisense replicates.

As shown in Figure 2-26, the band intensity for the uninduced samples was significantly lower than for the induced samples. 2 ug RNA was loaded for induced and uninduced samples, but as a loading control was not used, data could not be normalized, and the possibility of loading errors could not be discarded. Remarkably, the band intensities for all three uninduced samples are similar, suggesting that lower amounts of sense U4 snRNA are produced when the cells are not in nitrate. This suggests, for unknown reasons, that the U4 sense is highly expressed when cells are in nitrate media compared to when they are in ammonium. When cells are in nitrate, a whole group of genes involved in the cell's nutrient assimilation pathway are activated, and cellular metabolism is somehow enhanced. It can not be said with absolute certainty that nitrate upregulates gene expression in *C. merolae,* but these results suggest it. Nevertheless, further and more complex studies are needed to support this conclusion, as the uninduced samples were expected to reveal similar amounts of sense U4 as the pSR1025 (empty vector) controls. Also, the induced samples revealed higher amounts of U4 sense snRNA, and the three induced strains showed RFU values of ~$9x10^6$, which means that the U4 sense levels were similar between the U4 sense, U4 antisense, and pSR1025-induced strains. Overall, no U4 accumulation was observed in the U4 sense strains, and no U4 sense reduction was observed in

the U4 antisense strains (regarding scenario number 3, previously explained) when probing with the sense probe.

The Cef1 agarose northern blot showed no band when probed with the Cef1 sense probe. One of the reasons why a band can not be detected on an agarose northern blot may be due to the low abundance of the transcript. Even loading 20 ug of total RNA will not be enough to detect the transcript when it is expressed in very low amounts. Splicing proteins are known to be not abundant in *C. merolae,* so it is possible that their genes are not highly expressed and, for instance, undetectable by northern blot. The Cef1 antisense probe was never tried, as even the sense version of the gene was impossible to detect. Transfer issues were also considered, and for that reason, the polyacrylamide northern blot approach was proposed. This approach was also inconclusive and could not tell anything about appropriate antisense expression and splicing inhibition.

The U2-integrated gene expression nitrate induction results looked more promising than the transient gene expression results. However, as the experimental results were analyzed, they revealed some questionable phenomena. The U2+ B biological replicate showed higher band intensity than the U2+ A. Both biological replicates were supposed to behave equally. There would be a reason to think that the U2+ B biological replicate is showing the expected results due to successful U2 sense induction if we had been able to see either a band that shows antisense expression in at least one of antisense strains when probed with U2 antisense probe, or a reduction in the U2 sense band intensity in at least one of the antisense strains when probed with U2 sense probe. As none of these is accurate, we can not assure that the higher value for U2 sense expression in U2+ B is due to the successful induction of the integrated extra copy of the U2 sense gene.

When analyzing Dib1 antisense induction, a smear appeared at the top of the gel when probing with sense and antisense probes. This could be due to the differences in transcript size found in the cells (poly-A tail nucleotide number variation, transcripts including only the open reading frame and transcripts including 5' or 3' UTRs, etc.). In the Dib1 antisense polyacrylamide Northern blot no band appeared, which could be the expected result if the antisense Dib1 is being expressed, the endogenous sense and antisense molecules are encountering forming an RNA duplex, and the latter is degraded. However, this result must have been consistent with an increase in band intensity for the Dib1 sense strains when probing with the sense probe, showing that the promoter works correctly on every strain, expressing the Dib1 sense additional integrated copy. In this experiment, bands that may represent the Dib1 sense open reading frame appeared for the Dib1 sense strains when probed with the Dib1 sense probe. This made the Dib1 antisense probe results more consistent and promising, and a possible antisense expression was considered. The detection of the Dib1 sense band(s) in the sense strains in polyacrylamide and agarose northern blots may be attributed to the overexpression of the gene, a result of the additional copy present in their genome. As previously noted, splicing proteins are typically low in abundance, and their genes are not highly expressed. Therefore, the expression of Dib1 sense may be insufficient for detection by northern blot in strains lacking this additional copy.

My RT-PCR results were also inconclusive. As described in the results section of this work, when performing reverse transcription with gene-specific primers for the antisense versions of U2 and Dib1, bands appeared also for the sense strains and the WT controls. Antisense U2 snRNA is not present in the integrated sense strains or the WT controls, so the primer used for antisense U2 cDNA synthesis could not bind to a target in these strains. So why

did bands appear in the sense strains and the WT controls when using antisense gene-specific primers? Observing bands for every sample when performing the PCR with the antisense GSP reverse-transcribed samples was surprising, and several ideas were proposed. Snap-back is a phenomenon that can occur when a primer has self-complementarity folding against itself at the 5' or 3' ends, forming a hairpin. This may cause the synthesis of undesirable products in the RT reaction and the appearance of non-expected bands in the subsequent PCR reaction. snRNAs have secondary structures, and it was thought that the U2 snRNA folding may have led to the reverse transcription of U2 sense products with U2 antisense gene-specific primers. Thus, if one end of the antisense U2 snRNA molecule is folded against itself, the reverse transcriptase can recognize this as a primer, being the U2 antisense molecule the template for synthesizing the sense strand. When performing the PCR reaction, the whole molecule will be denatured, but it will contain the U2 sense sequence. This will later be recognized by the antisense PCR primers (the same sequence as the sense primers but switched in positions (sense forward is antisense reverse and vice versa)), and a U2 sense product will be synthesized and observed on the gel.

A product will be seen when cDNA synthesis is performed with primers specific to the sense version of the genes and the PCR reaction with antisense primers. During the denaturation step of the cDNA molecules, sense and antisense (reverse transcription will synthesize the antisense or complementary strand) will separate and be recognized by the antisense primers, triggering amplification. Thinking about all the possibilities involved in these results, when cDNA synthesis is performed with the gene-specific primer for the antisense version of a gene, no reverse transcription should occur in strains with the sense version of the genes or the controls, and nothing should be available to be amplified by PCR. If there is genomic DNA contamination, a product would be seen, but as reported above, no genomic DNA contamination

was observed in any of the samples analyzed, so the possibility of getting a band for this reason was discarded. Snap-back was considered a possibility for explaining this phenomenon in snRNAs, but the results for Dib1 were the same and Dib1 is less likely to form secondary structures as it is a linear transcript. It was then discovered that Dib1 has an overlapping gene in the negative strand, which would act as the antisense Dib1 and, for instance, will be bound by antisense primers, giving a band in the PCR reaction as a result. Additionally, the production of NATs (natural antisense transcripts) is a wide-spread phenomenon across eukaryotic genomes as a gene expression regulation mechanism. NATs regulate gene expression by inhibiting the transcription mechanisms to interact with the sense transcripts, interfering with transcription initiation, or promoting RNA degradation (Wight & Werner, 2015).

RT-PCR with antisense primers for CMK260C in the Dib1 sense and antisense strains and in the WT controls was performed to eliminate the possibility of endogenous antisense transcription muddying the results. CMK260C does not have overlapping genes in the opposite strand at its locus, so performing cDNA synthesis with antisense primers should not yield any product, leading to no bands in the PCR with CMK260C antisense primers. This did not happen, and as for U2 and Dib1, bands appeared for every sample when performing the experiments with antisense RT primers. For this last experiment, as a final attempt, nested primers were used to run the PCR reaction for CMK260. It has been shown that performing the gene-specific primer reverse transcription with a primer different from the one used for the qPCR or PCR reaction is always better. Antisense-specific reverse transcription for CMK260C was performed with FUB184, and a nested primer inside FUB184 was used as the reverse primer for the PCR reaction. The results were the same even with nested primers. After all these attempts, nothing could be concluded for antisense gene expression, and why these results were being obtained

remains unclear. Extensive literature research has been performed, and more must be found to explain this phenomenon coherently.

Neither of the two approaches gave me a clear clue of antisense induction and can be concluded that antisense expression failed with both transient and integrated gene expression approaches.

3. Chapter 3 – Splicing Inhibition by Degron Techniques

3.1. Introduction

As an alternative to splicing inhibition, I explored the degradation of splicing proteins through degron systems. By degrading core spliceosomal proteins, we can prevent the proper assembly of spliceosomes and inhibit the splicing process. If splicing is found to be essential, the degradation of spliceosomal proteins will lead to the death of *C. merolae* cells due to the disruption of this critical process. Conversely, if splicing is not essential, *C. merolae* cells are expected to exhibit no negative phenotypic effects and will continue to grow normally.

Protein function has been traditionally investigated by disrupting the expression of a target gene encoding a protein and analyzing the resulting phenotypic consequences. Loss-of-function experiments are traditionally performed using antisense oligonucleotides (ASOs), RNA of interference (RNAi), zinc finger nucleases (ZFNs), and CRISPR-Cas9 systems (Ludwicki *et al*., 2022). These techniques are broadly used in basic research and are promising for future treatments of genetic disorders. However, challenges and limitations remain in these techniques, including lack of temporal control, unpredictable off-targets, the inability of gene knockdowns to remove proteins already present within cells, leaving some proteins unaffected, and the challenge that deleting an essential gene represents (Ludwicki *et al*., 2022). Here is where degron techniques appear, representing a new inhibition-by-degradation method that takes advantage of the canonical protein degradation pathway mediated by the proteasome that exists within the cells. Degron systems are commonly activated by small molecules that act as a bridging ligand between the degron domain and another protein required to promote protein ubiquitination, which is required for proteasome recognition.

Some of the reasons to control protein concentrations artificially are to elucidate protein function or to study complex biological systems without genetic manipulation. The most common way to artificially manipulate and adjust protein concentrations is by regulating protein synthesis.

However, stable and long-lived proteins remain intact long after their synthesis has stopped, reducing their concentration only as cells divide. This makes modulating protein abundance challenging (Wilmington and Matouschek, 2016). Protein concentrations in the cell are a function of protein degradation and synthesis rates, so another way to control protein abundance is by altering protein degradation pathways (Wilmington and Matouschek, 2016). In most eukaryotic cells, intracellular protein degradation is controlled by the ubiquitin-proteasome system (UPS), which tunes the concentrations of hundreds of proteins. Proteins are targeted to the proteasome by a degradation signal, or degron, with two components: a proteasomal initiation region and a proteasome-binding tag in the form of polyubiquitin chains. When a protein is recruited to a ubiquitin ligase, its ubiquitination occurs. The proteasome receptors recognize the polyubiquitin chains, and degradation initiates in the disordered protein region. The protein is then carried to the proteolytic chamber of the proteasome, where it is hydrolyzed into shorter peptides (Wilmington and Matouschek, 2016).

This system is based on protein ubiquitination, which involves a 3-step process. First, E1 ligases mediate ubiquitin activation; second, E2 ligases mediate ubiquitin conjugation; and finally, E3 ligases mediate ubiquitin ligation (Fujiwara *et al.*, 2024). The E3 ligases are crucial in determining the ubiquitin pathway's high specificity. E3 ligases add ubiquitin molecules to protein residues, marking them for degradation by the proteasome. After ubiquitination, the protein is targeted for degradation. One of the most studied E3 ligases is the SFC (SKP, Cullin, F-box containing) complex, which consists of an F-box protein, S-phase kinase-associated protein 1 (SKP1), Cullin 1 (CUL1), and Ring-box 1 (RBX1) (Fujiwara *et al.*, 2024). Their functions are substrate recognition and binding, acting as an adapter for F-box proteins, providing structural scaffolding, and binding to E2 enzymes, respectively (Caussinus *et al.*, 2011; Fujiwara *et al.*,

2024).

Destabilizing domains (DD) (degrons) fused to a protein of interest make use of this degradation system to modulate the stability of the target protein. The DD interacts with the cellular protein quality control system, leading to proteasome degradation, most likely after ubiquitination (Wilmington and Matouschek, 2016). Some of the most commonly used DD are FKBP12 (FK506 binding protein), FRB (FKBP-rapamycin binding domain from mTOR kinase), and DHFR (dihydrofolate reductase). Usually, temperature or a small molecule or ligand activates the DD and compromises the stability of the entire protein. Chemical inducers of dimerization (CIDs) can be used to control the interaction of two proteins by serving as a bridging ligand. It is this interaction which leads a protein to degradation.

Different systems have been developed using CIDs to control the interaction between two proteins and regulate protein complex formation (Wilmington and Matouschek, 2016). The best established of these systems is the FKBP12 – FRB disordered domains complex formation. These two proteins interact only in the presence of the small molecule rapamycin, where the FKBP-FRB complex forms quickly and tightly, as rapamycin has a high nanomolar affinity for both proteins. FKBP12 is a 12 kDa cytoplasmatic protein, and the FRB is an 11kDa domain derived from the mammalian target of rapamycin (mTOR) (Wilmington and Matouschek, 2016).

For this research, an inducible protein–knockdown system has been developed. The target proteins chosen for degradation were Clf1 and Prp8, both core spliceosomal proteins. Clf1 is part of the NTC complex, whose function was previously described. Prp8 is known as the pivotal component of the spliceosomal catalytic center. It is the largest and most conserved of the spliceosomal components. It participates in multiple interactions with other proteins and RNA components throughout the splicing assembly, catalysis, and disassembly (Dlakic & Mushegian,

2011) (Figure 3-1).

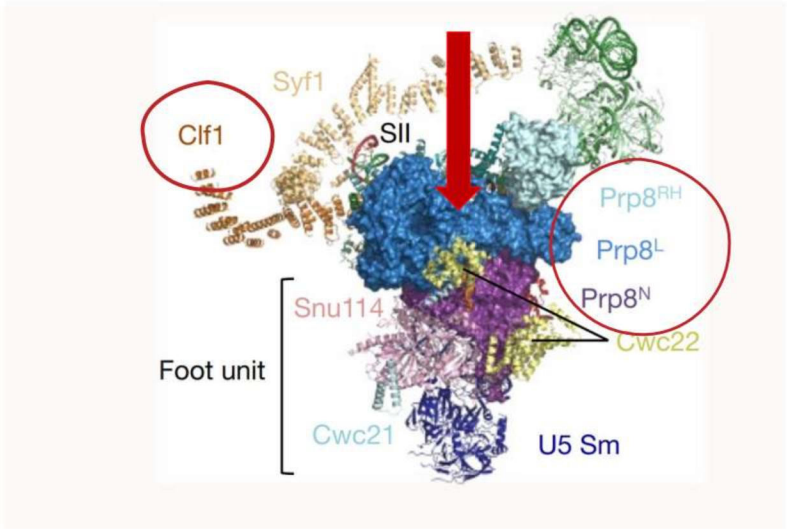


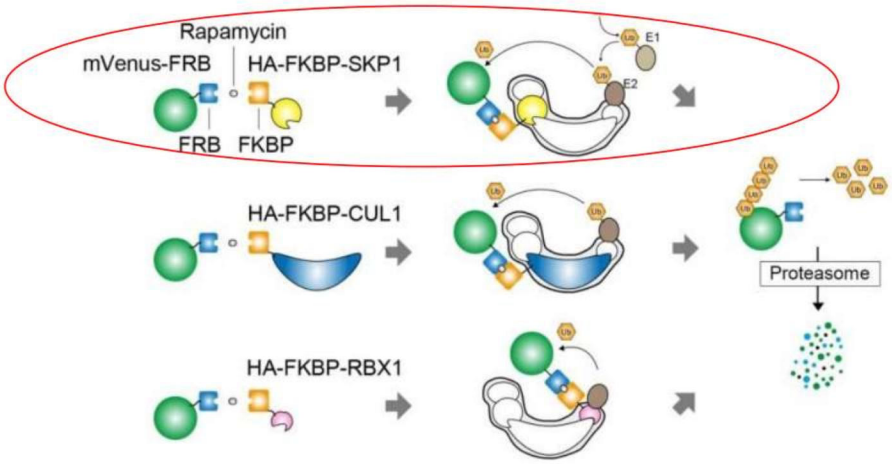
Figure 3-1: Cryo-EM structure showing Prp8 in its essential role of organizing the entire spliceosomal assembly. Note that Clf1 is present in the structure as part of the spliceosome's catalytic centre (Red circle to the left). The red arrow is showing Prp8 at the centre of the spliceosome (Galej *et al.,* 2016). Permission to use the image, cropped from original.

I successfully generated *C. merolae* strains with degron constructs in which the human FRB DD was fused to the C-terminus of Clf1 and Prp8. The degron strains were created from a previously engineered *C. merolae* strain containing a genomic insertion of the FKBP DD fused to the SKP1 subunit of the E3 ligase and tagged with HA (hemagglutinin tag) (HA-FKBP-SKP1) construct (a gift from S. Miyagishima) as an adaptor for degron system (Figure 3-2 b). Including an E3 ligase subunit fused to the FKBP disordered domain will facilitate the incorporation of the target protein into the SCF complex. By fusing the target protein to human FRB and expressing an E3 ligase component fused to the human FKPB in the adaptor, the target protein is expected to be captured by the SCF complex and degraded by the UPS. Thus, in this system, an FRB fusion to a target protein, constitutively reduces the level of the target protein to a certain extent due to the destabilizing effect of FRB (Fujiwara *et al*., 2024) (Figure 3-2 a). If the degradation of the spliceosomal core proteins is successful, the spliceosome will not be able to assemble, and splicing

will be inhibited. If splicing is an essential process in *C. merolae*, degradation of core spliceosomal proteins will result in cell death.

a)

b)



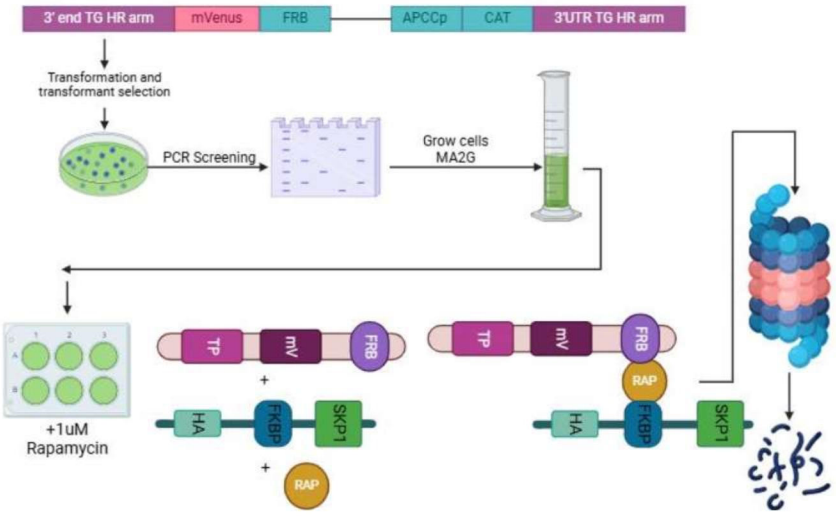


Figure 3-2: a) Diagram of the experimental design of the rapamycin-induced protein-knockdown system (Fujiwara *et al*., 2024, CC BY licenced, cropped from original). The system circled in red is the one used for this project. b) Representation of the procedure for getting the protein-knockdown system to work in *C. merolae*. If successful, splicing protein degradation will be achieved, and subsequent splicing inhibition will occur. TG=Target gene, TP=Target protein.

Samples were taken at 4, 8, and 12 hours after adding rapamycin for protein degradation tracking through Western Blotting and RNA extraction to perform RT-qPCR. RT-qPCR was used to evaluate splicing inhibition in 5 intron-containing genes. A strain containing a mVenus-FRB fusion integrated into the URA 5.3 locus and the previously mentioned adaptor was used as a positive control (gift from S. Miyagishima). This strain is the same one used in Fujiwara *et al*., 2024, which was used in this study to evaluate the efficiency of the protein-knockdown system.

For my project, we tried to mimic the procedures followed by Fujiwara *et al*., 2024, to achieve splicing inhibition by inducing the degradation of core splicing proteins. We expected splicing inhibition and cell death after a certain period when the *C. merolae* cultures were exposed to rapamycin.

3.2. Materials and Methods

3.2.1. Preparation of *C. merolae* genomic DNA

Genomic DNA was prepared as described in section 2.2.1.

3.2.2. Construction of genetic constructs (plasmids) for genomic integration by homologous recombination

I constructed genetic constructs for genomic integration in the target proteins' genomic locus by homologous recombination by modifying the CmC001 (pSR1008) plasmid backbone using restriction enzyme cloning and ligation-independent cloning (LIC). The first cloning step consisted in adding PacI and SwaI sites to the plasmid for LIC of the 5' and 3' homology recombination arms, which will help with the integration and fusion of the destabilizing domain to the C-terminus of the proteins of interest into the *C. merolae* genome. This was accomplished by cloning two duplexes with PacI and SwaI restriction enzyme sites by restriction enzyme cloning. In the second cloning step, the FRB disordered domain had to be inserted into the plasmid just downstream and in frame with the mVenus protein by restriction enzyme cloning. The last step in completing the plasmid was inserting the two homology recombination arms into the previously added PacI and SwaI sites by LIC.

The oligo duplexes were synthesized from the oSDR2195 and 2196 oligos for adding the SwaI site and from oSDR2197 and 2198 oligos for adding the PacI site (for oligo sequence refer to Appendix 1). Fifty picomoles (5 uL of a 10 uM stock) of the two oligos that will form the duplex were mixed in a 1.5mL Eppendorf tube. 0.5 uL of 1 M KCl was added. The tube was then incubated at 65 °C in a heat block for 5 minutes. The mixture was then slowly cooled to room temperature (about 25 minutes).

The oligo duplexes contained restriction enzyme overhangs for ligation. The SwaI duplex contained XbaI and BamHI restriction enzyme overhangs, and the PacI duplex contained SmaI and KpnI restriction enzyme overhangs.

I performed a PCR reaction using pSR1022 as a template to amplify the FRB destabilizing domain with oSDR2199 and 2200 as forward and reverse primers, respectively. The PCR reaction was performed with Q5 DNA Polymerase (NEB), and the primer annealing temperature was 66 °C with an extension time of 10 seconds.

I performed the PCR reactions with Q5 DNA Polymerase from *C. merolae* genomic DNA to amplify Cf1 and Prp8 5' and 3' homology recombination arms (HRA). The primers contained LIC sequences to facilitate their insertion into the backbone plasmids. The primers used with their respective annealing temperatures were the following:

Table 3-1: Primers used to amplify Clf1 and Prp8 5' and 3' HR arms from *C. merolae* genomic DNA.

Gene	**oSDR number**	**PCR products**	**Annealing Temperature (°C)**
Clf1	F: 2233 R: 2234	5' HRA	68
Clf1	F: 2235 R:2236	3' HRA	63
Prp8	F: 2229 R:2230	5' HRA	66
Prp8	F: 2231 R:2232	3' HRA	66

I cleaned up the PCR products as previously described and measured their concentration with the nanodrop.

3.2.2.1. Insertion of SwaI duplex into pSR1008

I digested pSR1008 with XbaI and BamHI following standard procedures. I performed T4 DNA ligation to ligate the digested plasmid backbone with the SwaI duplex following the protocol in Appendix 4. Plasmid DNA was extracted and purified, and the concentration was measured using the nanodrop.

To check for the correct insertion of the SwaI duplex into pSR1008, I performed restriction enzyme digestion with SwaI in the plasmids extracted following standard procedures.

3.2.2.2.Insertion of PacI duplex into pSR1008 + SwaI duplex intermediate plasmid

I followed the same procedure as for the insertion of the SwaI duplex with the following variations:

1. I digested the backbone plasmid with SmaI and KpnI.
2. For the T4 DNA Ligase ligation reaction, I used T4 Blunt DNA Ligase as SmaI is a blunt enzyme (no sticky ends).

pSR1026 was the result of this second cloning step. This plasmid was used to clone FRB and Clf1 and Prp8 homology recombination arms (HRAs).

I confirmed the correct insertion of the PacI duplex by digesting 200 ng of pSR1026 and 200 ng of pSR1008 + SwaI duplex with PacI following standard procedures. A glycerol stock was prepared for this plasmid as previously described.

3.2.2.3. Insertion of the FRB disordered domain into pSR1026

Following standard procedures, I performed restriction enzyme digestion of pSR1026 and the FRB PCR product using EcoRI and AatII (Appendix 2) and T4 DNA Ligation.

The result from this cloning step was pSR1027, which contained the mVenus-FRB construct for modifying target proteins. This plasmid was the backbone for inserting Clf1 and Prp8 HRAs. A glycerol stock was also made from this plasmid.

To confirm the correctness of pSR1027, I digested it with SphI following standard procedures. The correct plasmids were used for the insertion of the Clf1 and Prp8 HRAs by LIC.

3.2.2.4. Insertion of Clf1 and Prp8 homology recombination arms (HRA) into pSR1027

I inserted the Clf1 and Prp8 HRAs following LIC procedures (please refer to Appendix 4). The insertions were performed in a stepwise manner, the 5' arms were inserted first followed by the 3' arms. The 5' arms were inserted by digesting the plasmids with SwaI and the 3' arms by digesting with PacI.

I then confirmed that the insertions of the 5' arms were successful by digesting the Clf1 and Prp8 plasmids with ClaI and MluI, and with StuI, respectively. For the 3' arms the confirmatory digestion was perfomed with ClaI and NcoI and with MluI, for Clf1 and Prp8 plasmids respectively.

The result from these cloning steps was the two final plasmids, pSR1038 and pSR1051 (Figure 3-3), for *C. merolae* transformation and genomic integration of the degron domain into Clf1 and Prp8 genomic loci, respectively. Glycerol stocks were made from these plasmids.

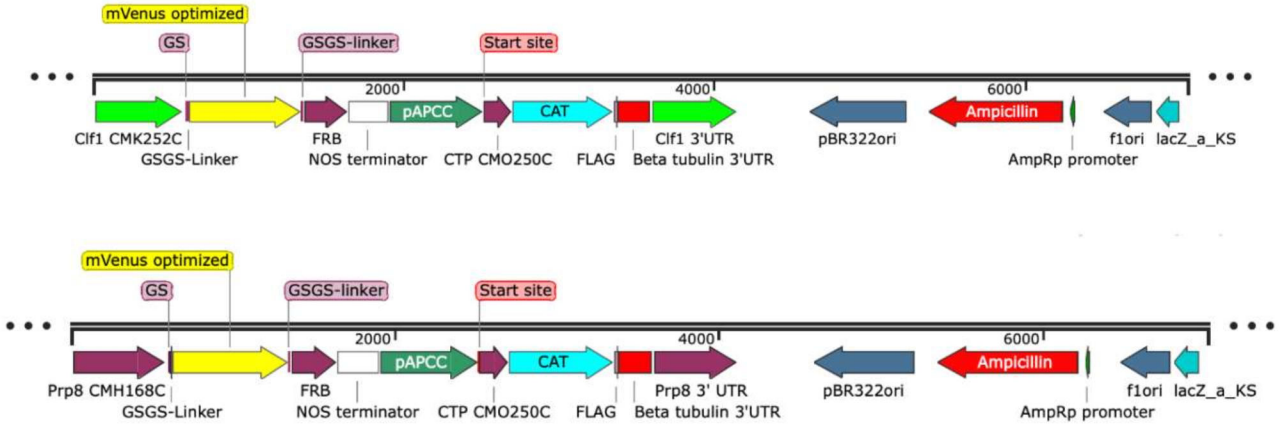


Figure 3-3: pSR1038 and 1051 (degron plasmids). Clf1 (top) HRAs are marked in bright green. Prp8 (bottom) HRAs are marked in purple.

3.2.2.5. Amplification of the sequence of interest for genomic integration from engineered vectors by PCR

I used pSR1038 and pSR1051 as templates for amplifying the linear constructs for genomic integration of the degron domain at the Clf1 and Prp8 genomic loci, respectively. I performed the PCR reactions with Q5 DNA Polymerase with the following primers:

Table 3-2: Primers used for the PCR reactions to amplify the target proteins' linear genomic constructs and transform them into *C. merolae* cells from pSR1038 and pSR1051.

Gene / Template	oSDR number	Annealing Temperature (°C)
Clf1 / pSR1038	F:2233 R: 2236	65
Prp8/ pSR1051	F: 2229 R: 2232	64

*For plasmid sequence, refer to Appendix 1.

The extension time was set up to 1 minute.

I purified the PCR products for transformation, and their concentration was measured with the nanodrop.

3.2.3. *C. merolae* transformation

C. merolae transformation was performed as described in section 2.2.4 for linear DNA constructs with the following modifications: The strain used for the transformation was CSR87, a strain that contains the HA-FKBP-SKPI degron adaptor integrated into the URA 5.3 genomic locus and constitutively expressed, not WT cells.

3.2.3.1. Cell acclimation and transformant selection

C. merolae cells were acclimated after transformation, as described in section 2.2.5. Plating and colony PCR were performed as described in section 2.2.5.2. The primers used for colony PCR were the following:

Table 3-3: Primers used for colony screening by PCR to check for correct genomic integration and fusion of the mVenus-FRB construct into *C. merolae* Clf1 and Prp8 genomic loci.

Gene	oSDR number	Annealing Temperature (°C) (Taq DNA polymerase)
Clf1- 5' HRA	Forward: 2243	59
Clf1– 3' HRA	Reverse: 2244	58
Prp8- 5' HRA	Forward: 2241	58
Prp8– 3' HRA	Reverse: 2242	59
mVenus	Reverse: 2238	58

*Please refer to Appendix 1 for primer sequences. These primers were used in different combinations to screen the 5' and 3' ends of the integration.

I set up two PCR reactions per gene. For Clf1, I used the primer pair 2243 and 2238 to span the 5' end of the integration for the first reaction, and the primer pair 2243 and 2244 span the whole insert from the 5' to the 3' HRA for the second reaction. For Prp8, I used the primer

pair 2241 and 2238, spanning the 5' of the integration for the first reaction, and the primer pair 2241 and 2242, spanning the whole insert from the 5' to the 3' HRA for the second reaction.

I used Taq DNA Polymerase for these reactions. The extension time for each reaction was set up to 20 seconds, looking for a band absence in the second described PCR reaction for both genes. The correct colonies for both PCR reactions were grown in a 48-well plate with selection (Chloramphenicol) for a few days and used for downstream applications.

Genomic DNA was extracted from the integrated colonies for further screening, as described in section 2.2.5.2. I performed confirmatory PCRs with Q5 DNA Polymerase for both genes using the following primer pairs:

Table 3-4: Primers used for confirmatory PCR tests for Clf1 and Prp8 degron strains.

Gene	**oSDR number**	**Annealing Temperature (°C) (Q5 DNA Polymerase NEB)**
Clf1- 5' HRA	Forward: 2243 (Primer C)	67
Clf1– 3' HRA	Reverse: 2244 (Primer D)	67
Prp8- 5' HRA	Forward: 2241 (Primer C)	67
Prp8– 3' HRA	Reverse: 2242 (Primer D)	67
Clf1 – 3' end of the ORF (outside 5' HRA)	Forward: 2301 (Primer H)	66
Clf1- 3' UTR (outside 3' HRA)	Reverse: 2302 (Primer I)	66
Prp8 – 3' end of the ORF (outside 5' HRA)	Forward: 2305 (Primer H)	68
Prp8 - 3' UTR (outside 3' HRA)	Reverse: 2306 (Primer I)	68

I ran two PCR reactions per gene: one with C+D primers and another with H+I primers (Figure 3-4). The extension time for all the PCR reactions was 4.5 minutes. I used WT genomic DNA as a control.

Figure 3-4: Representation of the location of primers C+D and H+I for confirmatory PCR screening.

After PCR determined that the genome of the colonies chosen for the Clf1 and Prp8 degron strains was appropriately integrated, I chose two biological replicates per strain to use in further experiments. These biological replicates were named CSR100 and CSR101 for Clf1 and CSR102 and CSR103 for Prp8.

Next, sequencing was performed as an extra confirmatory test for adequate integration. I performed a PCR with Q5 DNA Polymerase from genomic DNA of the selected biological replicates using oSDR2243 and 2480 for the Clf1 degron strain and oSDR2241 and 2480 for the Prp8 degron strain. The annealing temperature was set to 64 °C, and the extension time was 1.5 minutes for both PCR reactions. The PCR products were cleaned, and the concentrations were measured using the nanodrop. I used oSDR2481 and 2482 (Appendix 1) for sequencing the Clf1 integrated locus and oSDR2481 and 2483 for sequencing the Prp8 integrated locus.

3.2.4. Rapamycin effect on WT controls

I exposed WT cells and the mVenus-FRB (CSR90) positive control strain for the degron system to 1 uM rapamycin and performed a growth assay using non-treated cells as a control. The assessment involved both treated and untreated WT cells, as well as untreated CSR90 cells as controls. I analyzed the mVenus-FRB strain in triplicates, with 3 cultures treated and 3 untreated with rapamycin. The WT strain was analyzed in duplicate, with 2 cultures treated and 2 untreated with rapamycin. The growth assay was carried out to discard any possible effects on cell growth by rapamycin addition. It was performed as previously described in section 2.2.6 for U4 and Cef1.

3.2.5. Cell preparation and exposure to Rapamycin

I set up 50 mL cultures of CSR100, 101, 102, 103, 90 (mVenus-FRB), and 87 (parent strain – contains the adaptor) to grow with selection (250 ug/mL chloramphenicol), except for CSR90 and 87, until an O.D of 1.5.

I exposed each degron strain to rapamycin in triplicate and used triplicates of untreated cells per strain as controls. I split the previously grown culture in 6-well tissue culture plates using 6mL of cells per well. The plates were kept in a shaker so that the cells mixed constantly and avoided settling.

I used rapamycin at a 1 uM concentration and added 100% ethanol at the same volume as the rapamycin to the untreated controls (mock-treated—rapamycin was resuspended in 100% ethanol). Rapamycin was replenished every 4 hours.

I took samples for RNA extraction and protein analysis at 0, 1, and 4 hours for the first experiments and at 4, 8 and 12 hours for the subsequent attempts after rapamycin addition. For

the former experiments, no rapamycin was replenished. For the later experiments, rapamycin was replenished each time samples were taken.

I collected 1-2 OD units of cells for RNA extraction and 0.6 OD units for protein analyses. At the end of the experiment, I prepared a 1:2 dilution from the culture of each treated triplicate and one untreated sample in a 48-well tissue culture plate and measured the OD with a plate reader.

3.2.6. Western blots

I performed Western blots using the protocol in Appendix 11 as a baseline. I modified the protocol depending on the protein to be detected.

For mVenus detection in the mVenus-FRB control strain (CSR90), I ran an 8% SDS-PAGE gel, loading 0.6 OD units of cells resuspended in 20 uL of 2X SDS loading buffer (Laemmli buffer pH 6.8). I transferred in the semi-dry electroblotter into a nitrocellulose membrane at 2 mA/cm^2 of gel for 1 hour. In blocking buffer, I blocked the membrane at room temperature for 1 hour (see Appendix 11 for recipe). I incubated the membrane at 4 °C overnight with a rabbit polyclonal anti-GFP antibody (Invitrogen Cat# A-11122) in a 1:1000 dilution. I washed the membrane as stated in Appendix 11. The blot was incubated with a goat-anti-rabbit-HRP secondary antibody at a 1:5000 dilution at room temperature for 1 hour and then washed, as mentioned in Appendix 11. The detection was performed using SuperSignal™ West Pico PLUS Chemiluminescent Substrate (ThermoFisher Scientific Cat#34577) in a BioRad ChemiDoc MP Imager.

I stripped the blot in high stringency Western Blot Stripping Buffer following the procedure in Appendix 11. This step was performed because the two proteins to be detected,

mVenus – FRB fusion and the HA-FKBP-SKP1 adaptor, were the same size. I incubated the blot with a mouse monoclonal anti-HA antibody (ProteinFind® Anti-HA GeneBiosystems HT-301) at a 1:1000 dilution at 4 °C overnight. I washed the membrane as previously and incubated it with a goat-anti-mouse 800CW fluorescent antibody (IRDye® 800CW Goat anti-Mouse Li-Cor 926-32210) at a 1:10000 dilution at room temperature for 1h. I washed the blot as previously. The detection was performed using the BioRad ChemiDoc MP Imager.

For Clf1 detection in CSR100 and CSR101 I applied the same conditions as for mVenus detection in CSR90, except for the transfer, that was performed for 1h and 15 minutes instead of 1 hour; the primary antibody, which was a mouse monoclonal anti-GFP (ProteinFind® Anti-GFP Mouse Monoclonal Antibody HT801 GeneBiosystems) in a 1: 1000 dilution, and the secondary antibody which was the goat-anti-mouse 800CW in a 1:10000 dilution.

I did not strip the blot and reincubated it with the same anti-HA antibody as mentioned before in a 1:1000 dilution at 4C overnight. The blot was washed as previously and incubated with goat-anti-mouse 800CW in a 1:10000 dilution as a secondary antibody for 1h at room temperature. The blot was washed, and the detection was performed with the BioRad ChemiDoc MP Imager.

For Prp8 detection in CSR102 and CSR103, I ran a 4% SDS-PAGE and attempted two different transfers. I performed a wet transfer using a high molecular weight protein Towbin buffer (see recipe in Appendix 11) at 170 volts for 2 hours. Also, I tried a semi-dry transfer at 2.5 mA/cm2 of gel for 2 h using a high molecular weight Towbin buffer. Prp8-mVenus-FRB is a 315 kDa protein, so I applied high molecular weight protein conditions. The HA-FKBP-SKP1 adaptor was over-transferred with these transfer conditions and could not be detected in the same blot. I performed a second western blot following the same conditions as for the mVenus-FRB

protein with another set of samples to detect the adaptor in these strains. I used the same antibodies for Clf1 and Prp8 detection and followed the same procedure. Also, I stripped the Prp8 blot and reprobed it with the rabbit polyclonal anti-GFP previously used to compare detection quality between the monoclonal and polyclonal antibodies. With this primary antibody, I used the goat-anti-rabbit-HRP previously mentioned and performed the detection the same way.

3.2.7. RNA isolation and quality check

I isolated RNA from CSR100, CSR101, CSR102, CSR103, and CSR87 rapamycin-treated triplicates and from untreated samples from the same strains at the different time points in which sample collection occurred. CSR87 represents the parent strain. I performed RNA isolation as described in Appendix 8 and checked RNA quality through a bleach gel, following the protocol in Appendix 9.

3.2.8. RT-qPCR

I performed RT-qPCR to measure splicing inhibition in 5 intron-containing genes in *C. merolae:* CMK142T, CMS342C, CMQ270C, CMK260C, and CMJ129C. I followed the protocol in Appendix 12.

After RNA was extracted and quality checked, I DNase treated 10ug RNA per sample to remove genomic DNA (gDNA) contamination with the Turbo DNA-free Kit from Invitrogen™ (Cat# AM1907) following the manufacturer's protocol. I measured RNA concentration with the nanodrop. I then performed an end-point PCR with one of the primer pairs to be used in the qPCR reactions to check for any remaining gDNA contamination. The primer pair I used was FUB94/FUB113 (Appendix 1), and the annealing temperature for the reaction was 58 °C with an extension time of 15 seconds. I then ran the PCR reaction on a 1.5% agarose gel for 40 minutes.

If bands appear on the gel for the different DNase-treated samples, the treatment must be performed again, and another PCR needs to be run to ensure no more genomic DNA contamination is present.

I took 1 ug DNase-treated RNA per sample for the reverse transcription step (RT). This step was performed using 2 different methods: BioRad iScript ™ cDNA Synthesis Kit and BioRad Reliance Select cDNA Synthesis Kit. The first method uses oligo-dTs and random hexamers to randomly reverse transcribe the RNA molecules present in an RNA sample with any gene specificity. The second method (mentioned in section 2.2.10) can be used with gene-specific primers that will bind to a specific target, and only the reverse transcription of those target genes will occur. I included a non-RT (non-verse transcriptase) control per sample to discard any gDNA contamination after cDNA synthesis. One NTC (non-template control) per primer and reaction set was included. I performed the gene-specific RT with the following primers for 3 of the 5 intron-containing genes: oSDR1843 for CMK142T, FUB185 for CMK260C, and oSDR2132 for CMJ129C. I used the two methods to ensure the same overall results were obtained and discard any variability due to the methodology used. I set up the reaction for both cDNA synthesis kits as described in the manufacturer's protocol. The efficiency assumption for these reactions was that for 1ug RNA, 1ug of cDNA was synthesized.

I then performed a cDNA quality control end-point PCR with 50 ng cDNA and Taq DNA Polymerase on all the reverse transcribed samples using both methods again using the FUB94/FUB113 primer pair. This procedure aimed to check the RT reaction efficiency and ensure the cDNA was adequately synthesized. I also discarded any unspecific products derived from primer unspecific binding that would reduce the efficiency and accuracy of the qPCR

reactions. The annealing temperature and extension time for these PCR reactions were the same as previously mentioned.

Once the cDNA synthesis was proven successful, I prepared serial dilutions from the stock cDNA: 1:5, 1:25, 1 125, 1:625, 1:3125. A qPCR reaction was performed with these cDNA dilutions in triplicate, including one NRT in technical triplicate per sample and one NTC in technical triplicate per primer pair used. This experiment will produce a standard curve determining the best dilution (cDNA concentration) to amplify with a specific primer set to get the highest efficiency in the qPCR reaction. The dilution with the highest qPCR efficiency for the primer pairs used in this project was 1:125. The qPCR reaction setup is detailed in Appendix 12.

I used two primer pairs per gene, one spanning one of the intron-exon junctions of the pre-mRNA (either exon1 – intron junction or intron – exon2 junction) molecule and another one spanning one of the exons of the pre-mRNA or mRNA molecule (either exon1 or exon2). I used the intron-exon junction primers to measure the pre-mRNA accumulation in the rapamycin-treated samples compared to the untreated controls. Pre-mRNA accumulation will show splicing inhibition. I used the exon primers to measure the relative amount of total RNA, which should not vary between treated and untreated samples. This serves as a control to determine that rapamycin treatment is causing pre-mRNA levels to change and not significantly affecting global gene expression. Also, these primers were used to obtain a ratio of pre-mRNA: total RNA between treated and untreated samples to confirm the previously mentioned accumulation. The primers used were the following:

Table 3-5: Primers used for qPCR reactions to check for splicing inhibition in the degron strains treated with rapamycin.

Gene	Intron -exon junction (F/R)	Exon (F/R)
CMK142T	FUB94/FUB113	FUB93/FUB110
CMS342C	FUB5/FUB6	FUB108/FUB109
CMQ270C	FUB23/FUB24	FUB104/FUB105
CMK260C	FUB182/FUB183	FUB184/FUB185
CMJ129C	FUB97/FUB98	FUB124/FUB125

*For primer sequence, please refer to Appendix 1.

According to the primer validation results, the annealing temperature for all these primer pairs was 58 °C. The thermal cycler reaction conditions are in Appendix 12.

I performed the qPCR for CMK142T, CMS342C, and CMQ270C with iScript™ reverse transcribed samples, and the qPCR for CMK260C and CMJ129C was performed with gene-specific reverse transcribed samples with the Reliance Select cDNA Synthesis Kit. A qPCR for gene-specific reverse-transcribed samples for CMK142T was also performed. I analyzed the results using Bio-Rad CFX Maestro Software and normalized as follows:

1. I independently calculated the variation between the rapamycin-treated samples' Cq values (ΔCq) and the controls from the intron-exon junction primers and exon primers results per strain.
2. To determine the magnitude of pre-mRNA accumulated in the rapamycin-treated samples (intron-exon junction primers), I calculated the fold change (2ΔCq) in pre-mRNA increase between treated samples and controls. I also calculated the results obtained for

the exon primers even though no variation in total RNA was expected between treated and untreated samples.

3. I calculated a ratio between the fold change of the intron-exon junction and exon primers results. This ratio represents the ratio between pre-mRNA and total RNA, which allowed us to determine if there was an actual increase in pre-mRNA levels when samples were treated with Rapamycin.

3.3. Results

3.3.1. Construction plasmids for genomic integration by homologous recombination

After the cloning steps described in the materials and methods section, I successfully created two plasmids containing the Clf1 or Prp8 5' and 3' HRAs and the mVenus-FRB fusion for genomic integration purposes. Figure 3-5 shows the result for the restriction enzyme digestion check for the Clf1 plasmids after all the cloning steps were performed. The plasmids were digested with SmaI and KpnI, MluI and BstxI, and just with MluI, with the expected band sizes as shown in the figure.

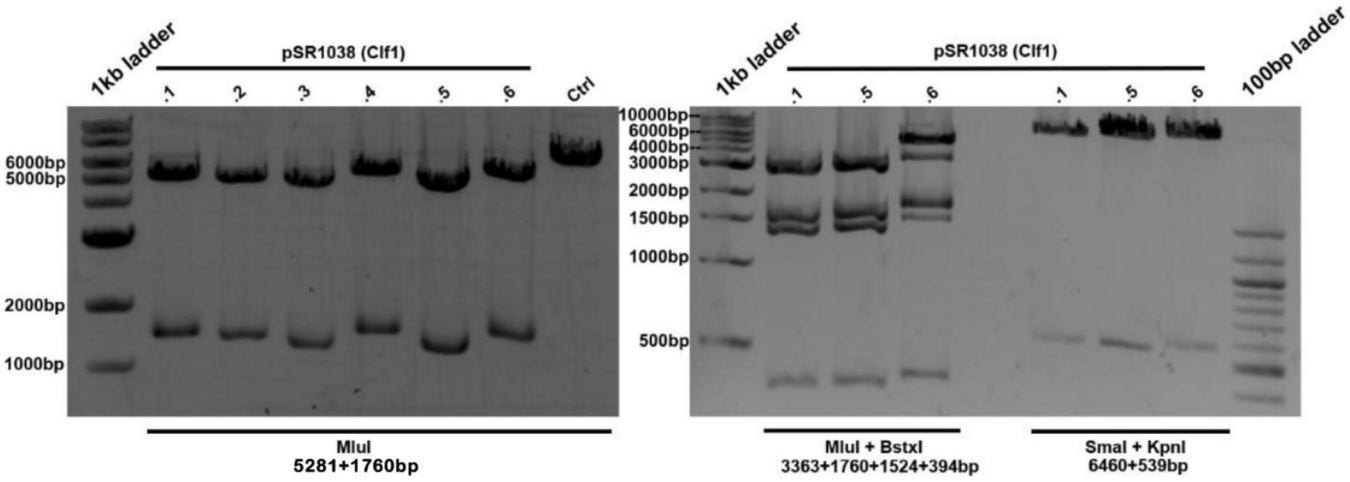


Figure 3-5: Restriction enzyme digest check for pSR1038 with the enzyme(s) listed in the figure. 0.7% agarose gel, 1 h run. The expected band sizes for each restriction enzyme digestion are shown at the bottom of each gel.

Figure 3-6 shows the result for the restriction enzyme digestion check for the Prp8 plasmids after all the cloning steps were performed. The plasmids were digested with ClaI and NcoI. The expected band sizes for this reaction were 3703+3311bp.

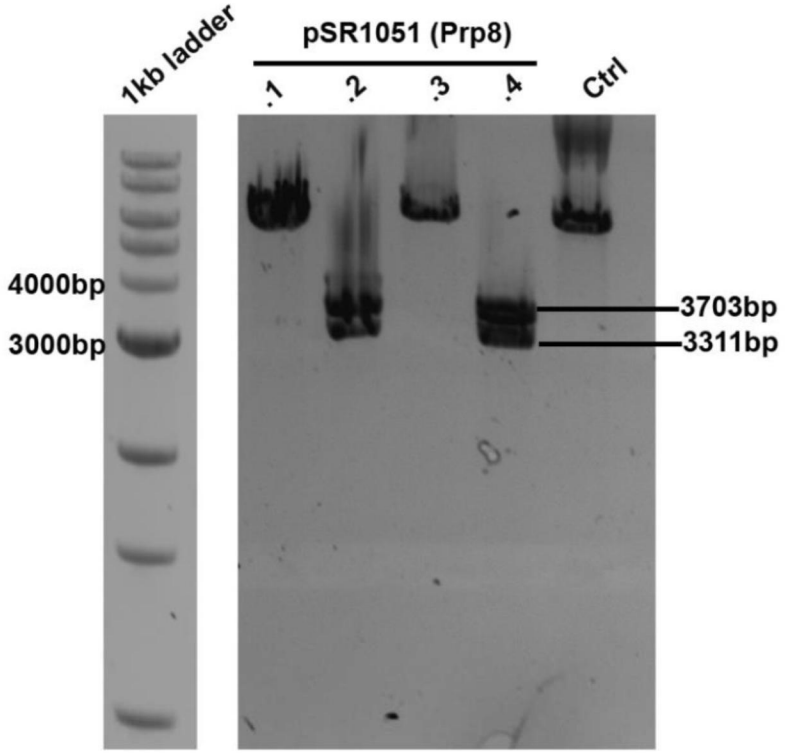


Figure 3-6: Restriction enzyme digest check for pSR1051 with ClaI and NcoI. 0.7% agarose gel, 1.5 h run. Ctrl= Prp8 plasmid with the 5' HRA successfully inserted.

3.3.2. *C. merolae* transformation

3.3.2.1. Genomic DNA isolation and PCR screening

To verify the identity of my strains, I extracted genomic DNA from three positive colonies per strain and analyzed it using PCR. The confirmatory screening showed the correct integration of the mVenus-FRB fusion into the Clf1 and Prp8 genomic loci (Figure 3-7). The expected band sizes for each degron strain and WT control with each primer pair are shown in Figure 3-7. In order to control for phenotypic effects arising from undetected mutations during strain construction, I chose two biological replicates per gene from the resulting strains, which

were named CSR100 and CSR101 for the Clf1 strains and CSR102 and CSR103 for the Prp8 strains.

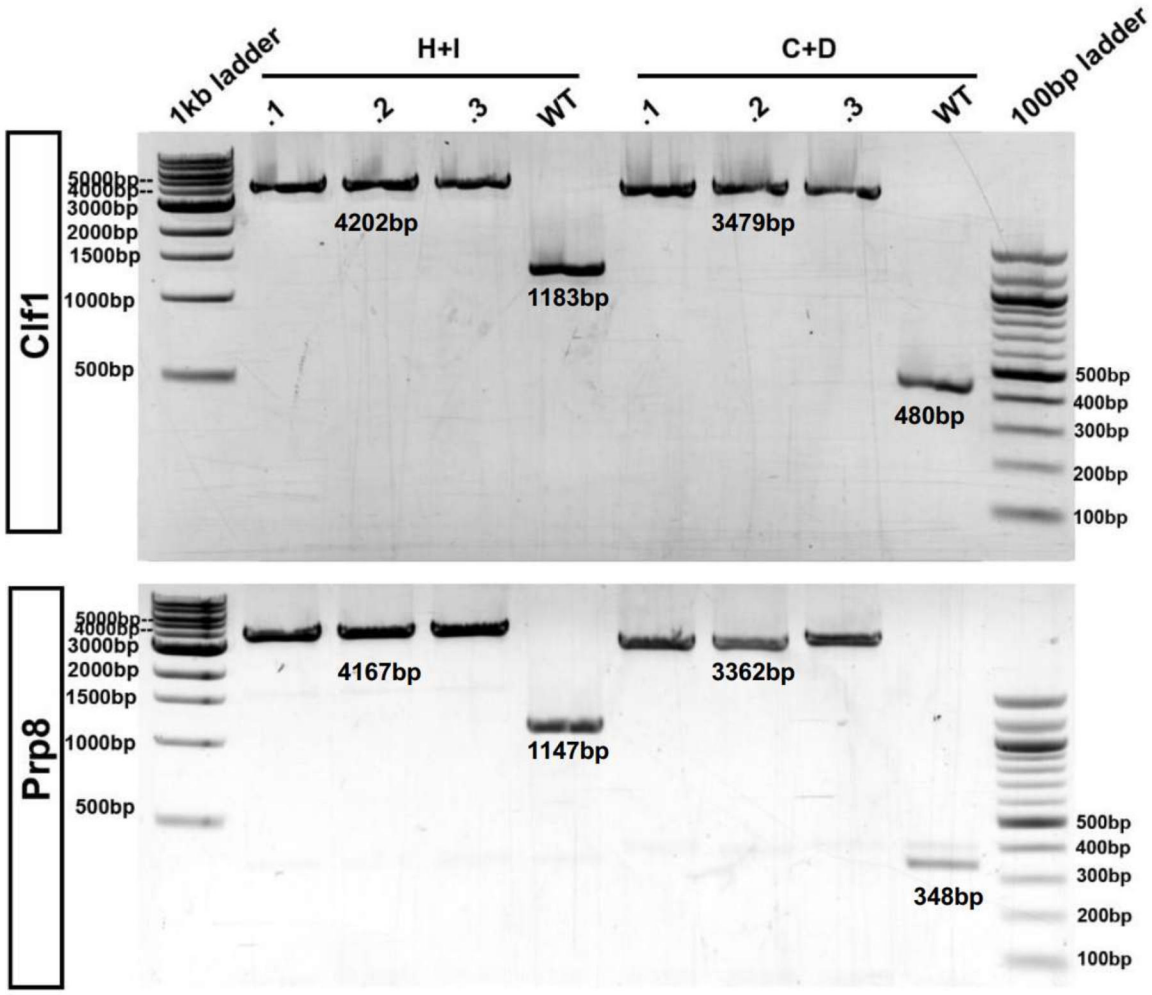


Figure 3-7: Confirmatory PCR showing the appropriate integration of the mVenus-FRB fusion into the Clf1 and Prp8 genomic loci.

The degron strains were designed to incorporate the mVenus fluorescent protein between the target protein and the FRB disordered domain, so I conducted fluorescent microscopy studies to confirm successful transformation. Figure 3-8 depicts the micrographs for the different degron strains using two fluorescent filters, Texas RED and TagYFP. Texas RED captures *C. merolae*

chloroplast autofluorescence, while TagYFP captures mVenus fluorescence. A WT strain was observed under the microscope as a negative control, and the CSR112 (m-Venus-sulfadiazine) strain served as a positive control.

Only the mVenus-sulfadiazine strain exhibited fluorescence under TagYFP. The other degron strains did not show fluorescence under TagYFP, despite robust chloroplast fluorescence. Interestingly, upon reviewing relevant literature, it was noted that this was not unexpected. Fujiwara *et al*. (2024) mentioned that the mVenus-FRB strains they used for their experiments were not brightly fluorescent because the mere presence of the FRB disordered domain attached to one end of mVenus destabilizes the protein, leading to its degradation without the addition of rapamycin. Consequently, the levels of mVenus in those strains are low, making it challenging to detect them by microscopy. It is reasonable to assume that the same phenomenon occurs with the splicing proteins fused to the mVenus-FRB.

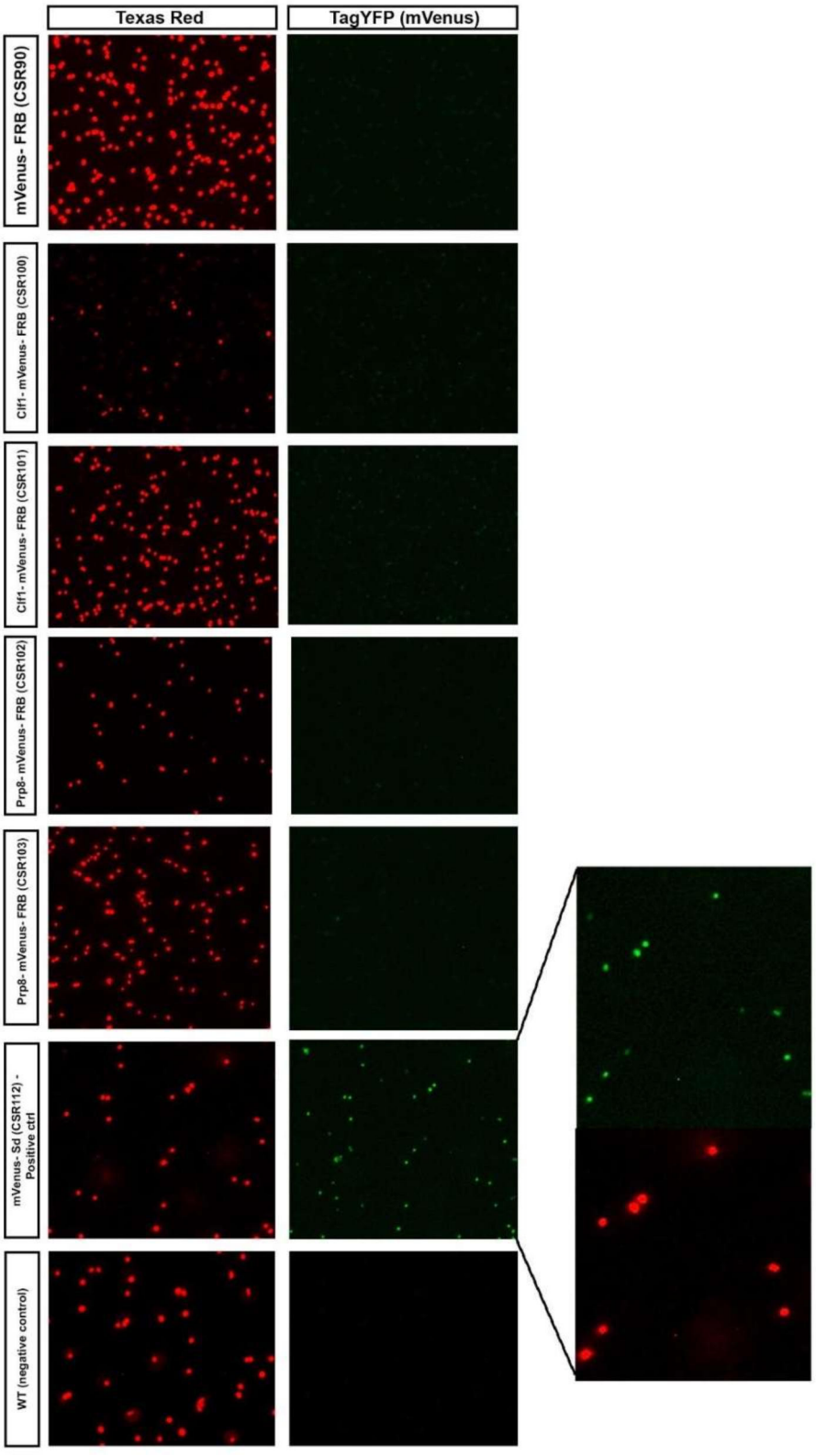


Figure 3-8: Fluorescent micrographs for degron strains and controls. Left column: Texas Red. Right column: TagYFP.

3.3.3. Rapamycin growth assessment

To determine whether rapamycin affects cell growth in the absence of a degron construct, I conducted a growth assessment for the mVenus-FRB control strain (CSR90). Rapamycin had no impact on the growth rate of either strain (Figure 3-9). While the mVenus-FRB strain grew slightly slower than the WT, it was not affected by the presence of rapamycin. The increased doubling time of the mVenus-FRB strain compared to the WT strain can be attributed to the inherent biological characteristics of the strain itself rather than to the addition of rapamycin. These results rule out any potential growth effects caused by rapamycin. Therefore, if any growth changes are observed with the Clf1 and Prp8 degron strains when treated with this chemical, it could be attributed to degradation of the target protein.

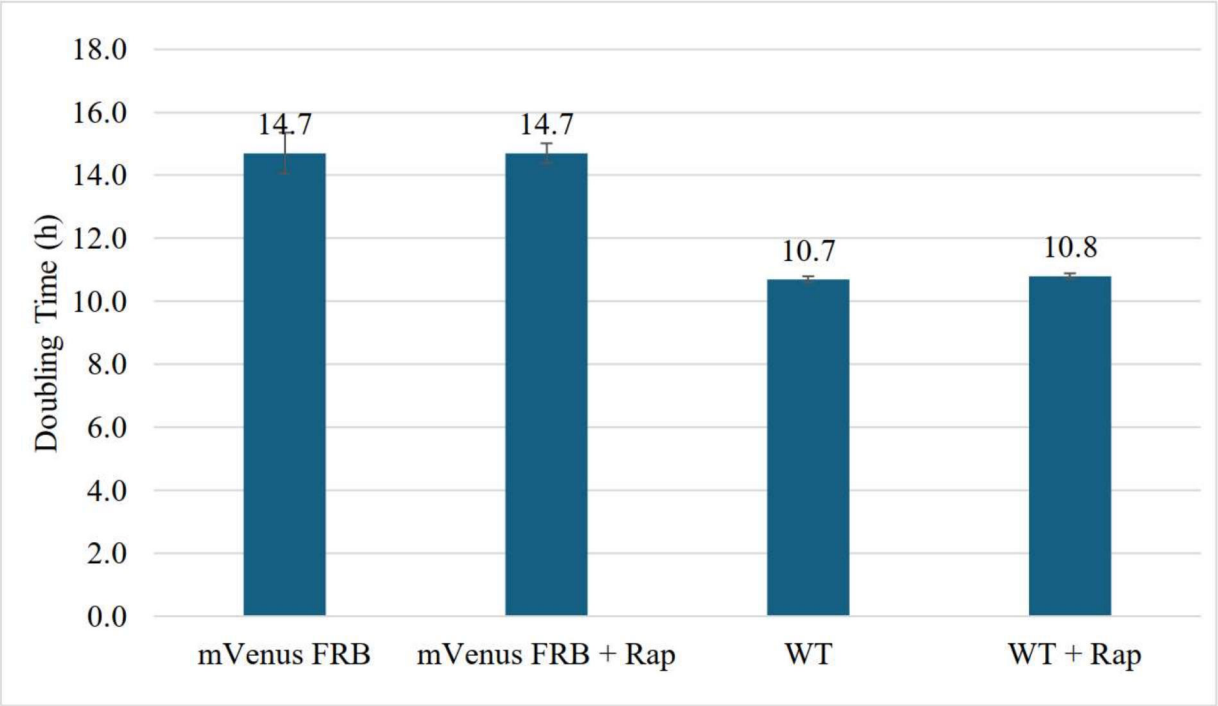


Figure 3-9: Bar graph representing the doubling times for the mVenus-FRB (CSR90) strain and WT strain treated and untreated with rapamycin.

3.3.4. Protein degradation analyses by western blot

To confirm that the degron system was appropriately working in the Clf1 and Prp8 degron strains and that the degradation of the splicing protein was occurring, I performed western blots on the samples taken at different time points for each strain treated with rapamycin and their respective untreated controls. I probed the blots with anti-GFP and anti-HA antibodies to detect Clf1 and Prp8 splicing proteins and the degron adaptor, respectively. In Fujiwara *et al.*, (2024) mVenus becomes almost undetectable by western blot after 4 hours of rapamycin addition. Thus, I performed western blots with the mVenus-FRB control strain (CSR90) treated and untreated with rapamycin to confirm that the system was working correctly in my hands.

As previously explained, I took samples at 0, 1, and 4 hours for the first experiments and at 0, 4, and 8 hours for the second attempt after rapamycin addition. I analyzed these samples by western blot using a rabbit polyclonal anti-GFP antibody and a mouse monoclonal anti-HA antibody. The expected result was to observe the reduction or disappearance of the mVenus-FRB (37kDa) band over time, having a clear band for the 0-hour time point and for the untreated controls and very faint or no bands for the 4 and 8-hour time points treated with rapamycin. The adaptor band (37kDa) was expected to remain constant over time without showing any degradation for the treated samples. The adaptor served as a loading control to demonstrate that the absence of a band is not due to sample loading mistakes but due to protein degradation. I included a WT control in these experiments, expecting not to get any band for this lane with either of the two antibodies. Additionally, I included a positive control (CSR112) corresponding to an mVenus sulfadiazine-resistant strain.

As shown in Figure 3-10, the results were as expected. A band at 37kDa representing the mVenus-FRB fusion protein appeared for the 0-hour time point and untreated controls. Interestingly, the bands for mVenus-FRB were faint compared to those for the adaptor. The mVenus-FRB fusion is integrated into the URA 5.3 locus under the control of the EFTU promoter. On the other hand, the HA-FKBP-SKP1 adaptor is integrated at the same locus just downstream of the mVenus-FRB fusion under the control of the APCC promoter. According to the NCKM (number of counts per kilobase mapped) transcriptomic data, EFTU has a significantly lower expression level than APCC[4]. Due to the low expression level of mVenus-FRB in the CSR90 strain, it was challenging to detect. No bands corresponding to mVenus-FRB

[4] APCC and CPCC are two of the strongest promoters in *C. merolae* (NCKM data and Rader Lab transcriptomic data).

or HA-FKBP-SKP1 appeared for the WT control, and a very bright and thick band appeared for the positive control. There are several bands in the WT control lane may be due to antiserum cross-reactivity with other proteins. It is important to mention that mVenus is under the control of the CPCC* promoter in the CSR112 strain. This is why mVenus is easily detectable in CSR112 and hard to detect in CSR90. The positive control was meant to show that the antibody was working correctly and that the technical skills used for the experiment set-up were adequate.

To detect the adaptor, I stripped the blot with a western blot high stringency buffer. The blot needed to be stripped as the band sizes for mVenus-FRB and HA-FKBP-SKP1 were the same. The bands for the adaptor did not have the same intensity for all the samples analyzed, meaning that not the same amount of protein was loaded per lane. 0.6OD unit samples were taken for each time point, and a whole cell extract was prepared. The OD measurements with the spectrophotometer can be inaccurate, meaning that the number of cells can vary between them even when taking the same OD units for each sample. Thus, the amount of protein loaded per lane is not precisely measured, leading to variabilities in band intensities.

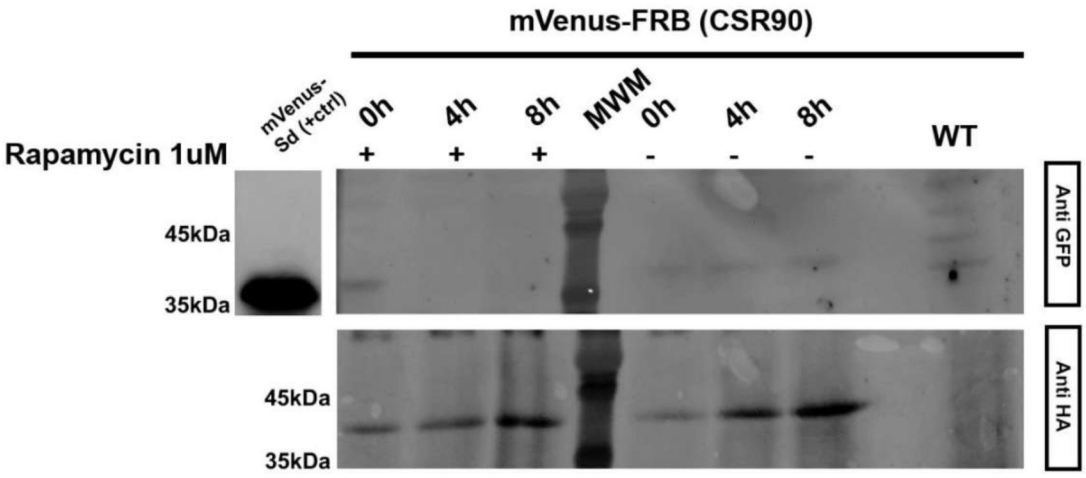


Figure 3-10: Western blot showing the mVenus degradation over time for the CSR90 control strain. Note that a band can be seen only at the 0-hour time point and in the untreated controls. The 4 and 8-hour time points do not have a band. The bands for the adaptor are present in all the samples analyzed.

The results of this experiment demonstrated that the degron system was properly working, as in Fujiwara *et al*., 2024. Subsequently, western blots were performed with the degron strains treated with rapamycin and the respective untreated controls.

Clf1-mVenus-FRB is a 110kDa protein. I did not observe a band in CSR101 samples when the blot was probed with an anti-GFP antibody either for the treated or untreated samples (Figure 3-11). Three different sized bands appeared on the blot, and one (red arrow) corresponds to the size of Clf1-mVenus-FRB. This band does not disappear or turn less intense over time after rapamycin treatment. Notably, the band is still present at the 4h time point, in which, according to Fujiwara *et al*. 2024, 90% of protein degradation can be observed. I cannot say with complete certainty that this band corresponds to the Clf1-mVenus-FRB fusion and that the protein is then not being degraded. The most likely scenario is that this band corresponds to something not specific to these experiments, highlighting. the importance of a WT control sample.

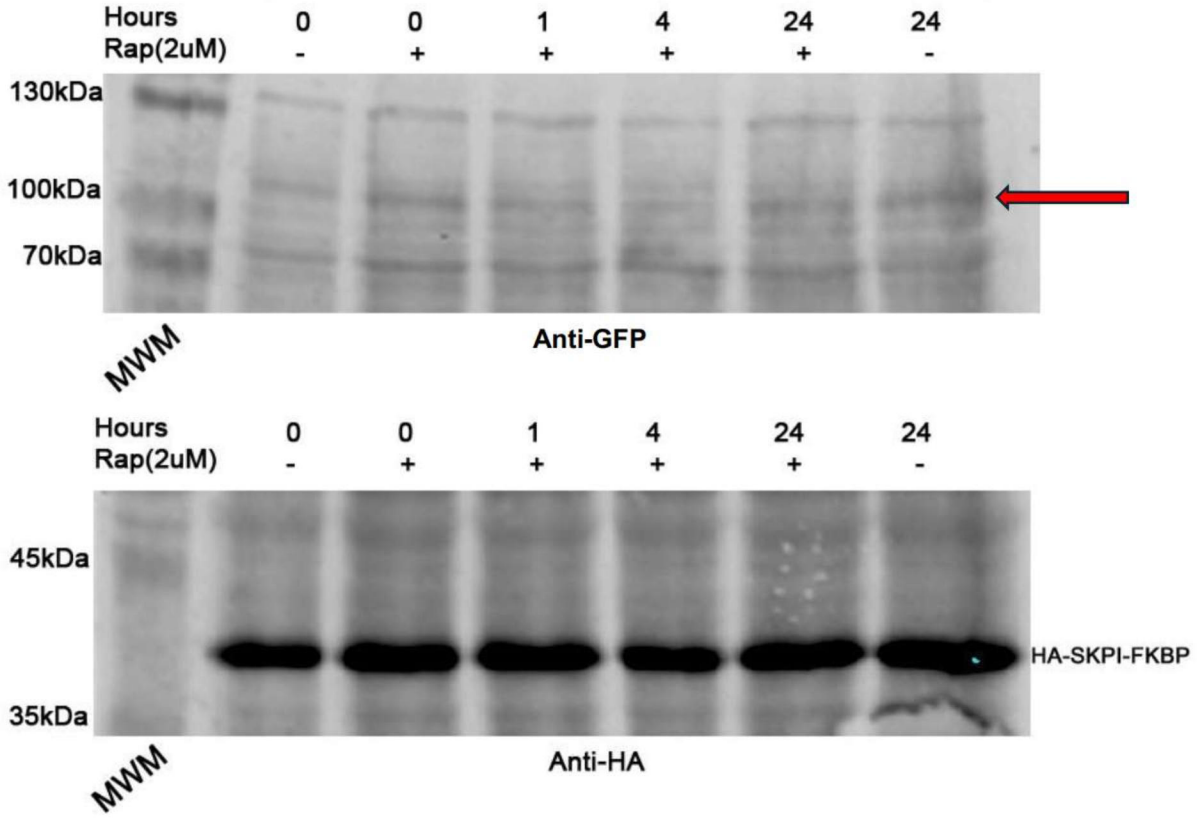


Figure 3-11: Western blot for CSR101 rapamycin treated and untreated samples showing the HA-FKBP-SKPI adaptor (band at ~37kDa). No specific band appeared for the Clf1-mVenus-FRB fusion protein. The red arrow is pointing to the possible Clf1 band that appeared in the blot.

The band for the adaptor was bright and clear and remained constant over time for all the samples analyzed, regardless of whether they were treated or not with rapamycin. This suggested that the strains were properly engineered and that the degron system was present in the cells. The latter affirmation was supported by the qPCR results. So, the reason why Clf1 was undetectable and degradation could not be observed remains unclear.

Prp8-mVenus-FRB is a ~315kDa protein. Despite all the attempts to detect Prp8 by western blot, I never detected it. Several modifications to the western blot protocol in Appendix 11 were

performed according to literature reviews and protocols taken from other laboratories (Stanek Lab in the Czech Republic and Furber Lab from my institution).

To detect Prp8 and detect the adaptor, I used different sample sets. I was not able to analyze all the samples on a single blot as the adaptor is a 37kDa protein and Prp8+mVenus+FRB a 315kDa protein. For Prp8 to be transferred to the membrane, transfer times of more than 2 hours and a high amperage (2.5 mA/cm^2 of gel) were needed. The adaptor was over-transferred under these conditions and, for instance, undetectable when the blot was probed with anti-HA. I performed a western blot with the 8-hour time point protein samples to prove the presence of the adaptor following the same procedure as for its detection in the mVenus-FRB (CSR90) strain. As can be seen in Figure 3-12 the band for the adaptor is present in all the samples treated and untreated with rapamycin from the CSR102 and 103 strains (Prp8 degron strains). The parent strain (CSR87) was included in this experiment, and the adaptor can also be seen for the treated and untreated samples. A WT control was included in the blot, and no band for the adaptor appeared, which was the expected result.

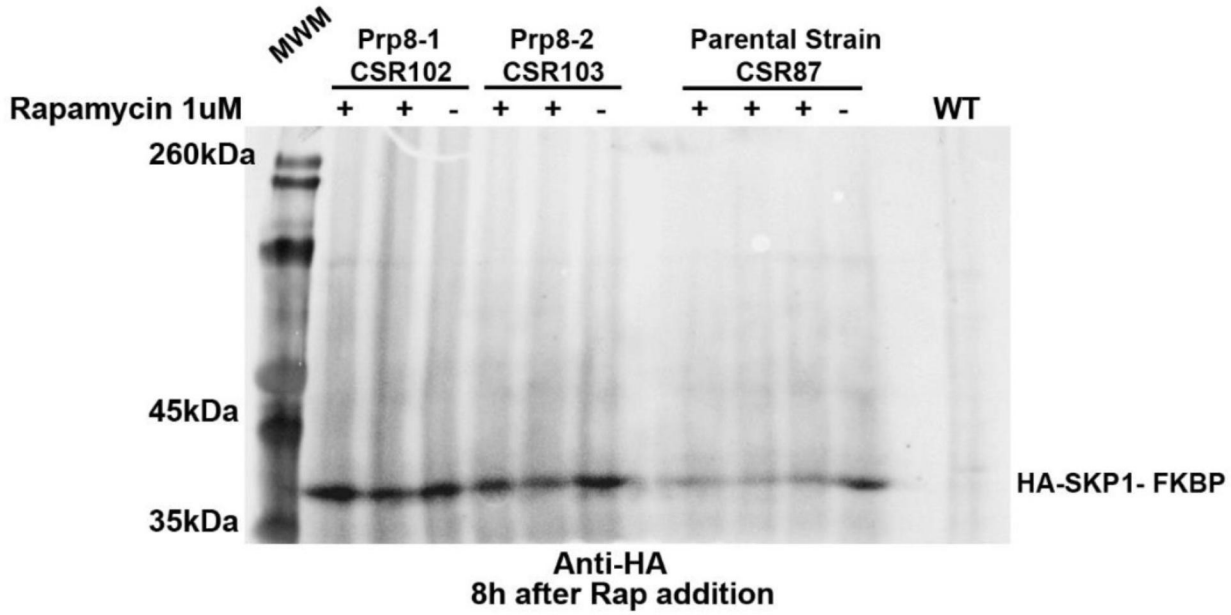


Figure 3-12: Western blot was performed in the CSR102 and CSR103 rapamycin-treated and untreated strains to detect the HA-FKBP-SKP1 adaptor. A WT control was included in the blot.

These results show that the Prp8 strains were appropriately engineered and that the degron system was likely working appropriately in the Prp8 degron strains. As for Clf1, this affirmation is supported by the qPCR results. The non-detection of the Prp8-mVenus-FRB fusion protein is likely due to some other factors, such as the ones previously mentioned, but the real reasons remain unclear.

3.3.5. Effect of degron induction on cell growth

To determine if splicing is essential for cell viability, I monitored cell growth after degron induction. I grew cultures to an initial OD of 1.5, added rapamycin, and let them grow for 12 hours with rapamycin replenishment every 4 hours. Figure 3-13 shows how the cultures looked at the end of the experiment. I chose one biological replicate per strain to measure the OD by

making a 1:2 dilution of the cultures 24 hours after the first rapamycin addition. The results are shown in Table 3-6.

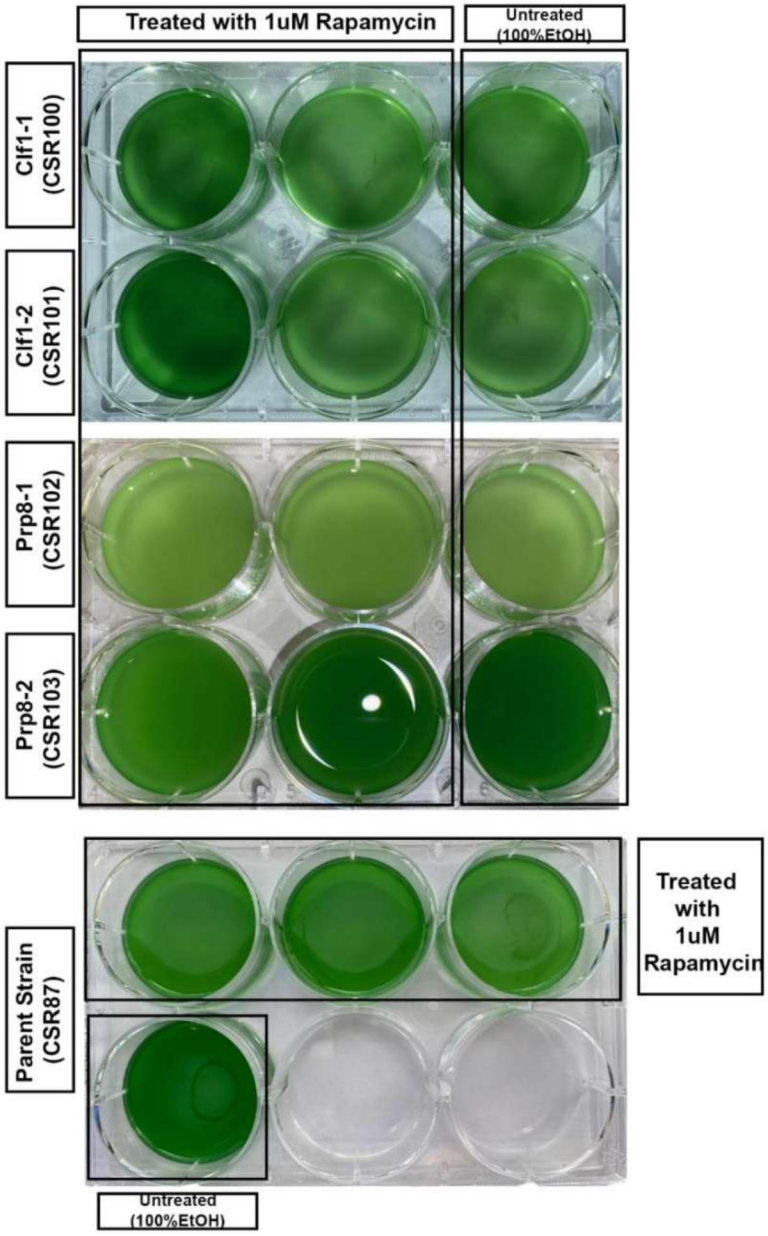


Figure 3-13: Degron strains cultures 12h after rapamycin addition. 6-well tissue culture plates were used. Duplicates for the Clf1 and Prp8 treated strains and triplicates for the treated parental strain are shown.

Table 3-6: OD_{750} measurements for CSR101 and CSR103 degron strains 24h after rapamycin addition.

	Triplicates 1 uM Rap			
Strain	**1**	**2**	**3**	**Control**
CSR101 (Clf1)	1.1	1.1	---	1.2
CSR103 (Prp8)	1.1	1.0	---	1.2
CSR87 (PS)	1.1	1.1	1.0	1.2

*Numbers represent duplicates or triplicates and control=untreated with rapamycin. PS= Parent strain.

The optical density (OD) measurements for the control samples are slightly higher than those for the treated samples across all analyzed strains. However, the difference is minimal, suggesting that the degron induction has a limited effect on cell growth over 12 hours. The OD for the parent strain control is equivalent to the controls for the experimental degron strains, indicating that they behave similarly to their parent strain when not treated with rapamycin or otherwise induced.

3.3.6. RNA isolation, RNA quality control, and reverse transcription for qPCR

To assess whether splicing was inhibited following degron induction, I extracted RNA from both the induced Clf1 and Prp8 strains and control samples (first experiment's 4 and 8-hour time point for Clf1 and Prp8, and second experiment's Prp8 and parent strain 12-hour time point) . I assessed the quality of the RNA (Prp8 12-hour time point and parent strain) and subsequently reverse-transcribed it (all RNA extracted) to perform qPCR on a set of five intron-containing genes. The RNA isolation was carried out as outlined in previous sections, and its quality was evaluated using a bleach gel (Fig. 3-14), which indicated good RNA quality, acceptable for RT-qPCR: the doublets corresponding to 28S and 18S rRNA were distinctly separated, with no

observable smear. Additionally, the intensity of the top band for 28S rRNA was roughly double that of the 18S rRNA doublet as expected. The bands for 28S rRNA were located at approximately 1.5 kb, while those for 18S rRNA were at around 0.8 kb, consistent with anticipated results. This gel demonstrates that high-quality RNA was obtained for subsequent applications.

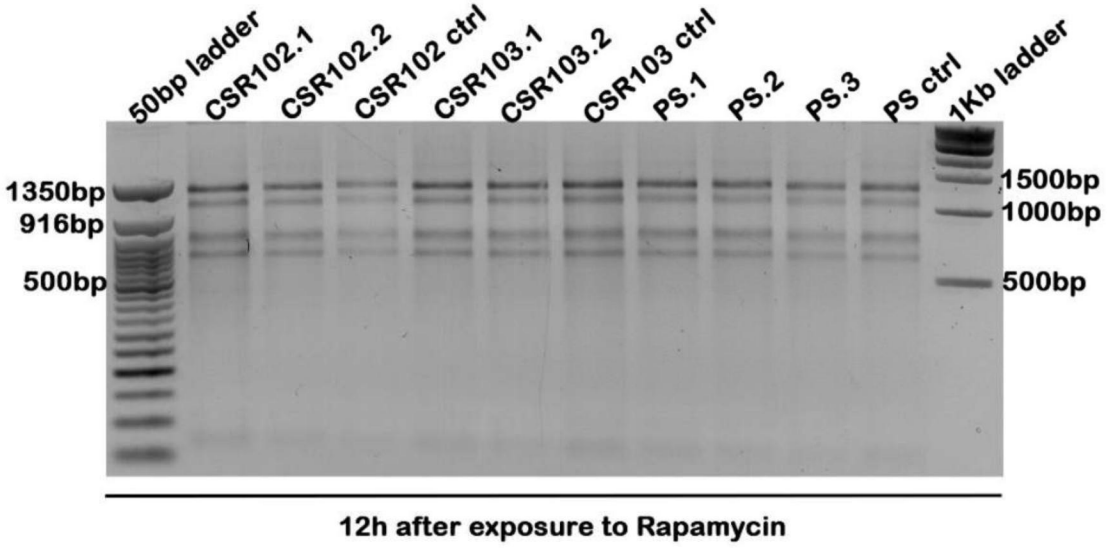


Figure 3-14: RNA bleach gel for the Prp8 degron strains (CSR102 and 103) and parental strain rapamycin-treated and untreated samples after 12 hours of the first rapamycin addition. PS=Parent Strain

When I performed gene-specific reverse transcription, an NRT control was included for each sample, and a cDNA quality check was performed using FUB93/114 primers (intron-exon junction primers for CMK142T). This was done to rule out any possible genomic DNA contamination that could affect the results of downstream applications such as qPCR. Figure 3-15 shows the cDNA quality check for the degron samples corresponding to CSR102 (Prp8-1) and CSR103 (Prp8-2) degron strains and the CSR87 parent strain for the 12-hour time point after

rapamycin addition with their respective NRT controls. The NTC was clean for every sample, and the NRT controls were largely blank, showing that gDNA contamination was not a concern.

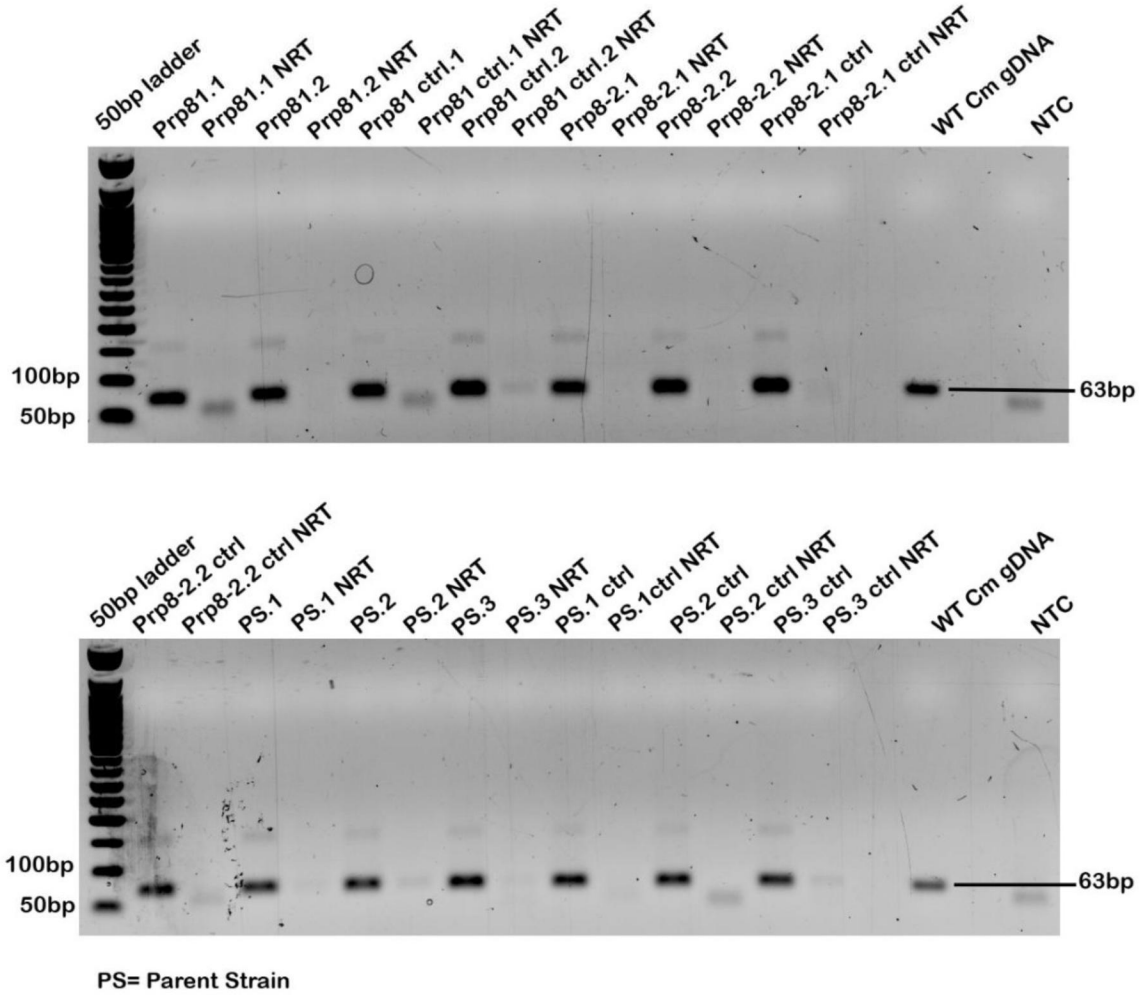


Figure 3-15: cDNA quality check and genomic contamination discard (NRT controls) for the 12h after rapamycin addition time point for the Prp8 degron and parental strains. Treated and untreated controls are included. The expected band size was 63 bp.

3.3.7. Measuring splicing inhibition with RT-qPCR

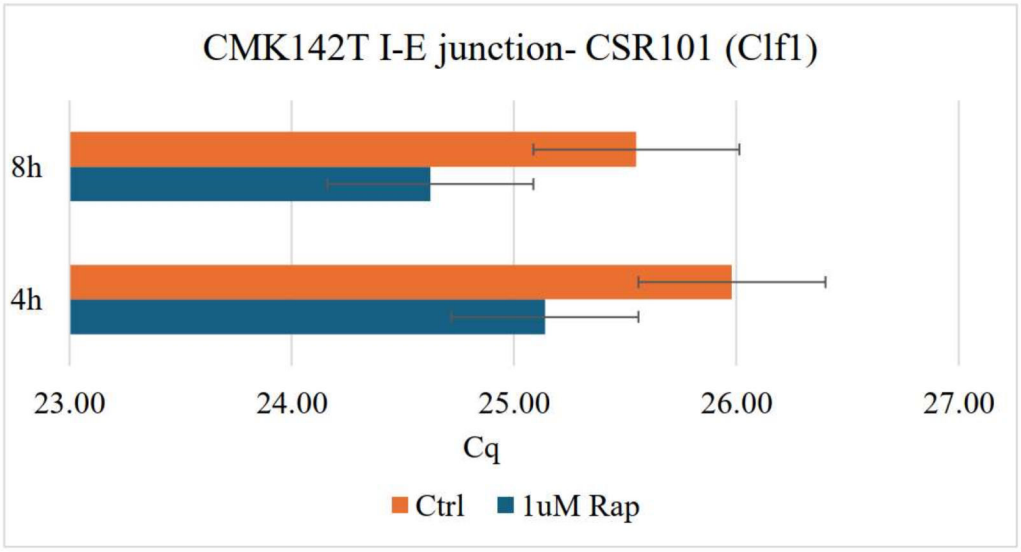
I performed RT-qPCR on RNA isolated from CSR101 (Clf1 degron strain), CSR102/103 (Prp8 degron strains), and CSR87 (parent strain), both in rapamycin-treated and untreated

conditions, to assess splicing inhibition following degron induction in a set of five intron-containing genes. Two separate experiments were conducted. The first experiment focused on samples from the CSR101 (Clf1) and 103 (Prp8) strains, which were treated and untreated with rapamycin and collected at 4 and 8 hours post-rapamycin addition. I used intron-exon junction and exon primers for each gene. The intron-exon junction primers were intended to evaluate pre-mRNA accumulation resulting from splicing inhibition in the rapamycin-treated samples compared to untreated ones. Conversely, the exon primers aimed to provide an estimate of total RNA for each gene, allowing for the calculation of a ratio between pre-mRNA and total RNA results to confirm if the accumulation reflects a change in splicing or a change in RNA abundance. I analyzed CMK142T, CMS342C, and CMQ270C in this first experiment. CMK142T was selected because our laboratory has determined that its intron contains RNase MRP, which is crucial for ribosome biogenesis. We hypothesize that its synthesis and processing depend on splicing. Splicing inhibition could prevent the synthesis of this essential enzyme, negatively impacting ribosome biogenesis and ultimately harming the cells. CMQ270C was chosen for analysis as it is the most spliced gene in *C. merolae*; thus, if splicing is inhibited, the most significant impact would likely manifest in this gene. Lastly, CMS342C was selected due to its demonstrated accumulation under heat-stress conditions. Additionally, it has recently been revealed that this intron encodes a sisRNA, whose function remains unknown. If splicing is inhibited and the intron is indeed crucial for *C. meroale's* survival, the growth of the cells will be compromised.

When analyzing CMK142T with intron-exon junction primers in Clf1 and Prp8 strains, the Cq values for rapamycin-treated samples are systematically lower than the untreated controls for both strains and at both time points (Figure 3-16). This suggests that there was more pre-

mRNA in the treated samples, in which splicing is being inhibited by addition of rapamycin and consequent protein degradation, and less or no pre-mRNA accumulation in the untreated samples, in which the addition of rapamycin, induced the degron system. This accumulation is more evident for CSR103 (Prp8) and the eight-hour time point shows higher pre-mRNA accumulation for both strains. The same was true for CMS324C and CMQ270C for both strains and time points between each treated/untreated pair.

a)



b)

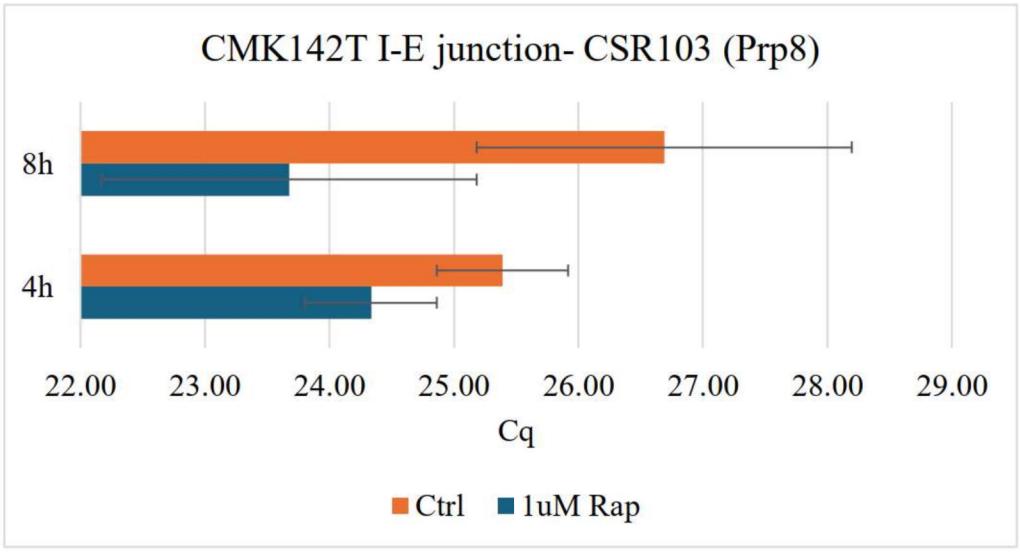


Figure 3-16: a) Bar graphs for CSR101 (Clf1) intron-exon junction primers results for CMK142T rapamycin-treated and untreated samples at 4 and 8 hours after rapamycin addition. b) Bar graphs for CSR103 (Prp8) intron-exon junction primers results for CMK142T rapamycin-treated and untreated samples at 4 and 8 hours after rapamycin addition.

To determine whether the differences were statistically significant, I calculated p-values. Table 3-7 shows the p-values calculated for the results in Figure 3-16. As can be seen, there is a

statistical significance between the treated and non-treated strains for both strains and both time points. There was a significant difference between the Cq values from the treated and untreated samples for both degron strains at 4 and 8-hour time points.

Table 3-7: P-values calculated for the results shown in Figure 3-16.

Strain	Time Point	Standard Deviation (SD)	p-value
CSR101	4h	0.46	$3.7x10^{-4}$
CSR101	8h	0.48	0.015
CSR103	4h	0.52	0.001
CSR103	8h	1.5	$4.8x10^{-7}$

n=3

To determine whether the accumulation of pre-mRNA was due to splicing inhibition rather than a change in RNA abundance resulting from rapamycin exposure, I calculated the pre-mRNA/total RNA ratio based on the analyses conducted using I-E junction and exon primers for each strain, time point, and gene (refer to Tables 3-8, 3-9, and 3-10). This ratio was obtained by dividing the calculated fold-change ($2^{\Delta Cq}$) between the treated samples and the controls from the intro-exon junction primers (representing pre-mRNA) by the values from the exon primers (total RNA). The methodology for these calculations is elaborated on in the materials and methods section.

A ratio higher than 1 was expected if real pre-mRNA accumulation was occurring. Conversely, if pre-mRNA levels are increasing because transcription is increasing, the ratio would be 1 or less. As seen in tables 3-8, 3-9, and 3-10, the ratio was higher than one for both strains and both time points for the three genes analyzed, except for the CSR101 4h time point regarding CMS342C. CMS342C was the gene that showed the most inconsistent results. For CMS270C and CMK142T, the ratio for the 4h time point in both strains is lower than the ratio

for the 8h time point, which is what was expected if splicing inhibition increases over time. However, for CMS342C, the CSR103 4 hour time point ratio was higher than the one for the 8h time point, suggesting a higher pre-mRNA accumulation at 4h than at 8h after rapamycin addition, which was not the expected result. Additionally, the ratio for CSR101 4h time point was lower than one, suggesting that there is not a pre-mRNA accumulation but rather an increase in total RNA in samples treated with rapamycin versus the controls, which was not the expected result either.

Table 3-8: Pre-mRNA/ total RNA ratio for CSR101 and CSR103 strains at 4h and 8h after rapamycin addition for CMQ270C.

Gene	Strain	Time point	Ratio pre-mRNA/ Total RNA
CMQ270C	CSR101	4h	1.27
CMQ270C	CSR101	8h	1.96
CMQ270C	CSR103	4h	1.19
CMQ270C	CSR103	8h	2.09

Table 3-9: Pre-mRNA/ total RNA ratio for CSR101 and CSR103 strains at 4h and 8h after rapamycin addition for CMS342 C.

Gene	Strain	Time point	Ratio pre-mRNA/ Total RNA
CMS342C	CSR101	4h	0.24
CMS342C	CSR101	8h	1.39
CMS342C	CSR103	4h	3.69
CMS342C	CSR103	8h	2.67

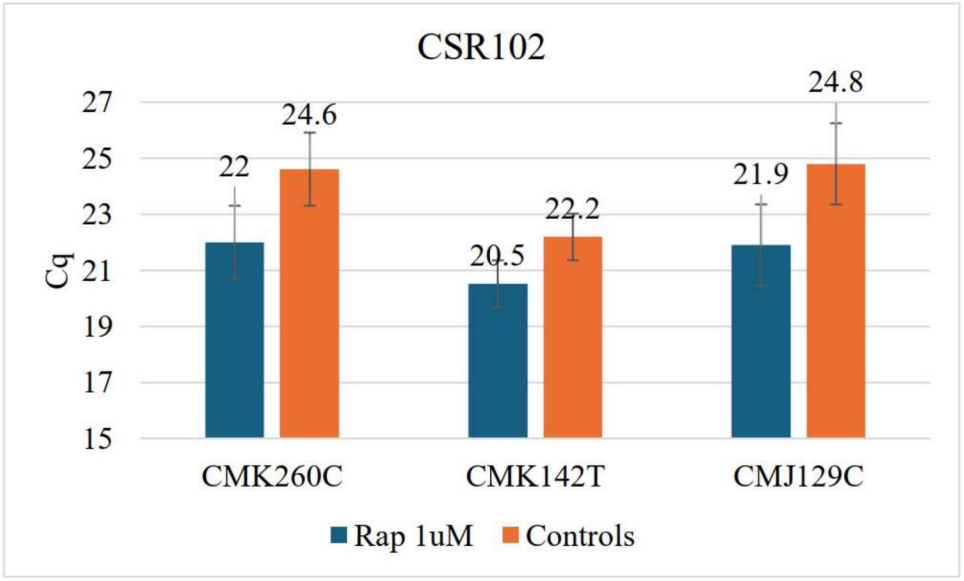
Table 3-10: Pre-mRNA/ total RNA ratio for CSR101 and CSR103 strains at 4h and 8h after rapamycin addition for CMK142T.

Gene	Strain	Time point	Ratio pre-mRNA/ Total RNA
CMK142T	CSR101	4h	1.33
CMK142T	CSR101	8h	1.78
CMK142T	CSR103	4h	3.04
CMK142T	CSR103	8h	4.70

I concluded that activation of the degron led to splicing inhibition, so to validate these results and to confirm that this was not an artifact of rapamycin treatment, I conducted a second experiment. With this experiment, I sought to validate the results from the first experiment with a more rigorous analysis that included more replicates, the parental strain as a control, and only used the tagged Prp8 strains (CSR102 and CSR103) that had shown stronger effects. Analyzing the parental strain (CSR87) when exposed to rapamycin was an essential part of confirming that the degron system was effectively working in the degron strains and discarding the possibility of pre-mRNA being accumulated due to some effect caused by the exposure to rapamycin and not because of splicing inhibition caused by the degron system. No pre-mRNA accumulation should be identified in the rapamycin-treated samples, as the parental strain is only engineered with the adaptor. The amount of pre-mRNA should remain the same for the treated and untreated samples across the duration of the experiment.

For this second experiment, I included duplicates of treated and untreated samples for each degron strain and triplicates of treated and untreated samples for the parent strain. I chose two new genes to be analyzed: CMK260C and CMJ129C. I analyzed CMK142T as the third gene for these experiments. The results from these experiments were expected to be consistent with the ones of the first experiments, and a pre-mRNA accumulation was expected to be observed. Pre-mRNA accumulation was observed for the three genes in both CSR102 and CSR103 Prp8 degron strains 12h upon Rapamycin addition (Figure 3-17 a and b). The calculated p-values show a significant difference in the pre-mRNA amount between treated and untreated samples (Tables 3-11 and 3-12) .

a)



b)

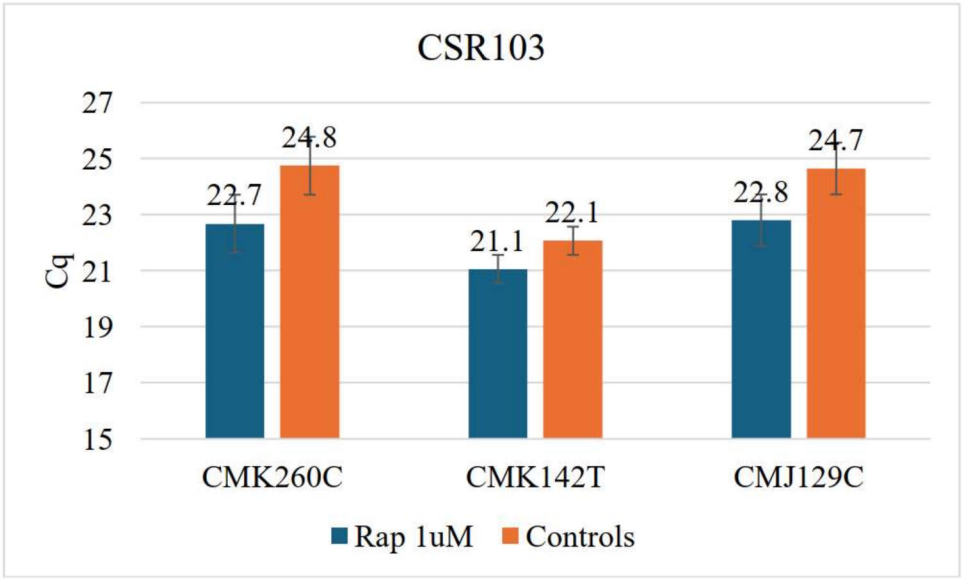


Figure 3-17: a) Bar graph showing the difference in pre-mRNA between rapamycin-treated and untreated samples in the Prp8-1 degron strain (CSR102) for CMK260C, CMK142T, and CMJ129. b) Bar graph showing the difference in pre-mRNA between rapamycin-treated and untreated samples in the Prp8-2 degron strain (CSR103) for CMK260C, CMK142T, and CMJ129.

Table 3-11: Calculated p- value by comparing the Cq values obtained for rapamycin-treated and untreated samples from the following genes (CSR102):

Gene	SD	p-value
CMK260C	1.30	$1.35x10^{-7}$
CMK142T	0.84	$2.5x10^{-4}$
CMJ129C	1.45	$4.3x10^{-7}$

n=3

Table 3-12: Calculated p- value by comparing the Cq values obtained for rapamycin treated and untreated samples from the following genes (CSR103):

Gene	SD	p-value
CMK260C	1.03	$1.87x10^{-5}$
CMK142T	0.50	$5.86x10^{-5}$
CMJ129C	0.93	$1.35x10^{-6}$

n=3

I performed a slightly different analysis for these experiments in which, as previously described, a ratio of pre-mRNA/total RNA was calculated per strain. However, then, using the *C. merolae* transcriptomic data, the standard unspliced percentage of cells growing in rich media at 42 °C (*C. merolae* ideal growing conditions) was pulled out and compared with the pre-mRNA accumulation calculated ratio to get the percentage unspliced after rapamycin addition. This was to address concerns that the measured increase in pre-mRNA was inconsistent with the amount of pre-mRNA in the cell. For example, if a gene is only spliced 50% under standard conditions, pre-mRNA could not logically increase more than two-fold without a corresponding change in expression.

For the three genes, the percentage unspliced after rapamycin addition is greater than the percentage unspliced under normal growth conditions, suggesting a ~30% splicing inhibition for the three genes (Table 3-13). CMJ129C has a low splicing level, with a standard splicing percentage of only around 30%. Thus, the unspliced fraction lays around 70% under normal

growth conditions. After 12h of rapamycin addition, the unspliced fraction goes up to 100%, meaning that for this gene, splicing was completely inhibited (Figure 3-18). Similarly, splicing of CMK260C decreases by ~20% after rapamycin treatment, while that of CMK142T decreases by ~30%. This provides a fairly consistent estimate of ~30% splicing inhibition over all three genes tested

Table 3-13: Splicing analysis shows the percentage unspliced under normal growth conditions (green) and the percentage unspliced after the addition of rapamycin (red). The cells marked in yellow show the average pre-mRNA/total RNA ratio calculated for both Prp8 strains—RM=Rich media.

Strain	CMK260C	CMK142T	CMJ129C
Prp8-1 (CSR102)	2.15	1.63	1.58
Prp8-2 (CSR103)	1.46	1.50	1.24
Average Prp8	1.81	1.57	1.41
Parent (CSR87)	0.92	1.10	0.99
%spliced (RM, 42 °C - transcriptomics)	0.71	0.43	0.27
%Unspliced (RM, 42 °C)	0.29	0.57	0.73
%Unspliced 12h after rapamycin addition	0.52	0.89	1.03

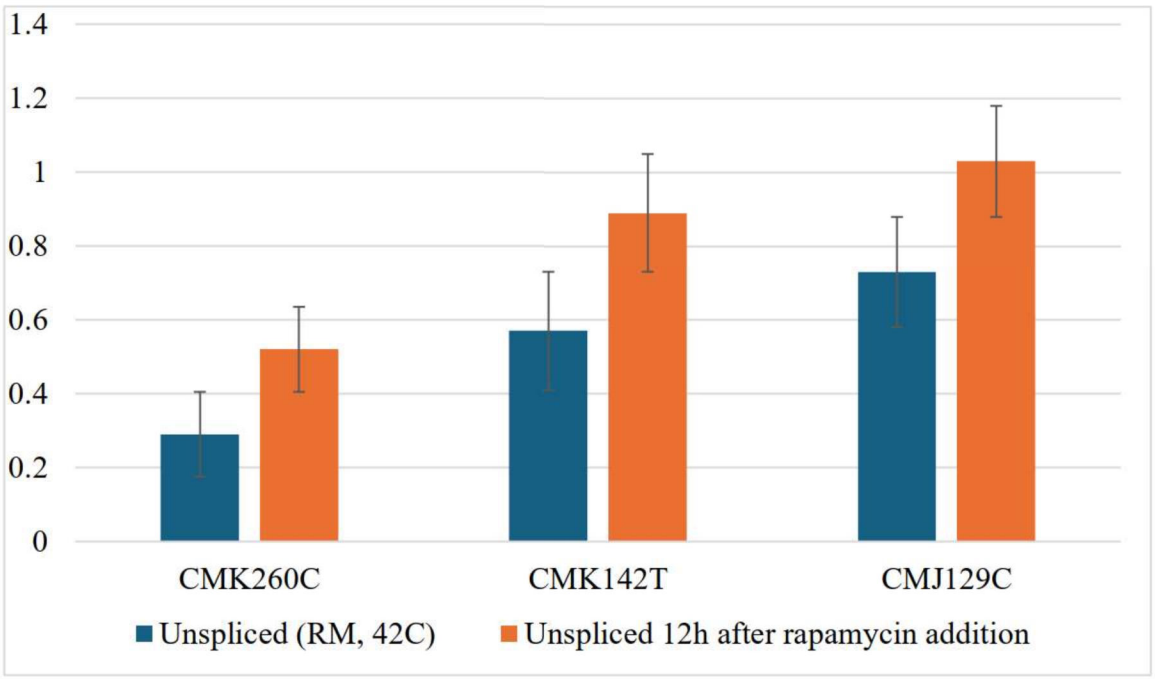


Figure 3-18: Bar graph showing the percentage spliced at normal or ideal growth conditions vs the percentage spliced after 12h of rapamycin addition for CMK260C, CMK124T, and CMJ129C.

Regarding the pre-mRNA/total RNA ratios, these are greater than one for the three analyzed genes, meaning, as for the first experiments, that the increase in the pre-mRNA levels shown by the intron-exon junction primers is larger than any increase in total RNA shown by the exon primers in the treated versus the control strains. Combined with the calculated unspliced fractions after 12 h rapamycin addition, these results confirm an actual pre-mRNA accumulation and splicing inhibition for the Prp8 degron strains.

As mentioned before, analyzing the parent strain was a vital part of consistently confirming that the pre-mRNA accumulation observed in both experiments was caused by the correct functioning of the degron system and as part of a rapamycin addition effect. In Table 3-13 the row named parent shows the ratio pre-mRNA/total RNA calculated for the tree genes

analyzed. The values for two of the three genes are below one, and CMK142T is just 0.1 units above 1. This means that pre-mRNA variation is equivalent to the variation in total RNA for the treated and untreated samples after 12 hours of rapamycin addition and that pre-mRNA accumulation did not occur. This was the expected result for this strain. As can be seen in Figure 3-19, the variation in pre-mRNA between the treated samples and the controls for the parental strain is very small, especially for CMK142T, in which the Cq value for the treated samples and the controls is around ~22 for both. The calculated p-values between the treated and untreated controls showed a non-significant difference (Table 3-14) in pre-mRNA amounts between the treated and untreated samples after 12 hours upon rapamycin addition.

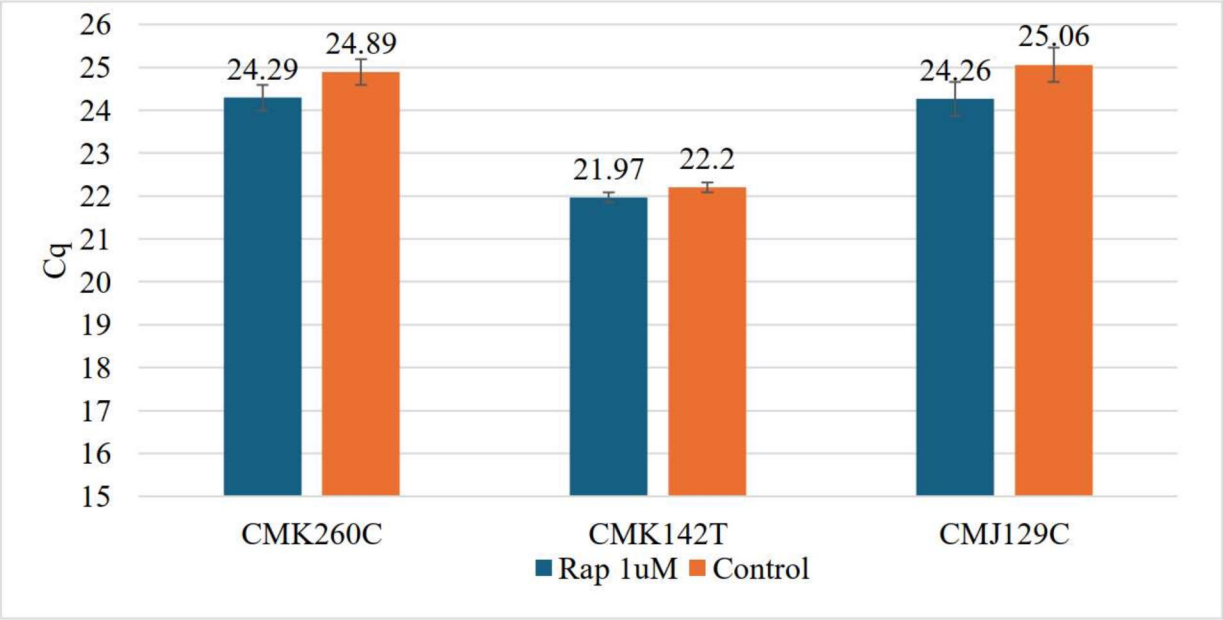


Figure 3-19: Bar graph showing the difference in pre-mRNA between rapamycin-treated and untreated samples in the parental strain (CSR87) for CMK260C, CMK142T, and CMJ129.

Table 3-14: Calculated p- value by comparing the Cq values obtained for rapamcycin treated and untreated samples from the following genes:

Gene	SD	p-value
CMK260C	0.3	0.18
CMK142T	0.12	0.34
CMJ129C	0.4	0.02

n=3

In conclusion, the degron system in Clf1 and Prp8 strains seemed to work appropriately based on the results obtained by the preliminary and confirmatory experiments. From the five intron-containing genes analyzed, the 5 showed pre-mRNA accumulation after 4, 8 and 12h of rapamycin addition. The results for the samples in which cDNA synthesis was performed with iScript™ and those in which cDNA was synthesized with gene-specific primers were consistent. They showed the same effect on pre-mRNA accumulation for the rapamycin-treated samples versus the controls. These analyses showed that splicing was partially or totally inhibited for the different analyzed genes; for instance, it can be assumed that the activated degron system degrades Clf1 and Prp8. However, these results need to be supported by western blot results, in which protein degradation can be seen over time.

3.4. Discussion

In this chapter, I sought to address whether splicing is essential in *C. merolae* by means of a degron system meant to specifically degrade target proteins. Fujiwara *et al*., 2024 recently published the degron system used for this dissertation, which promised to facilitate the study of essential genes and mechanisms in *C. merolae.*

RT-qPCR emerged as the technique that most effectively supported the functionality of the degron system. I can excitingly say that for the first time, splicing inhibition using degron systems was achieved in *C. merolae.* However, it is noteworthy that complete splicing inhibition was only

achieved for one of the five intron-containing genes analyzed, probably because it is poorly spliced to begin with. The results from both the first and second experiments were consistent, demonstrating pre-mRNA accumulation in samples treated with rapamycin at various time points in both Clf1 and Prp8 degron strains. The calculated p-values comparing the Cq values between treated and untreated samples for Clf1 and Prp8 degron strains at different time points upon rapamycin addition, showed statiscal significance. The consistency between experiment 1 and 2, also indicates that the different methodologies employed for cDNA synthesis in each experiment—using iScript™ for the first and gene-specific primers with the BioRad Reliance cDNA Synthesis Kit for the second—did not affect the observed pre-mRNA accumulation.

During the analysis of CMS342C in the initial experiment, several inconsistencies emerged. Notably, this gene was the only one to exhibit pre-mRNA/total RNA ratios lower than 1 at certain time points analyzed. This indicates that the variation in total RNA observed with exon primers between treated and untreated samples for both strains was greater than the pre-mRNA accumulation. While this observation is perplexing, it may relate to the nature of the protein encoded by the gene when exposed to rapamycin. CMS342C encodes for a V1-ATPase, an enzyme responsible for generating an electrochemical proton gradient across cell membranes, which is essential for energy production in various cellular processes (Vasanthakumar & Rubinstein, 2020). It also plays a role in acidifying intracellular compartments and facilitating the uptake of environmental substances. Is this gene upregulated when cells are exposed to rapamycin? It is established that rapamycin inhibits the TOR kinase, which is a key player in the TOR nutrient assimilation pathway. When TOR kinase activity is suppressed, cells may experience a starvation-like state, necessitating the function of enzymes such as V1-ATPase to generate energy and activate other metabolic pathways for nutrient assimilation and uptake from the environment.

According to the RT-qPCR results, the increase in pre-mRNA for the treated samples was 4 – 6 fold higher compared to the untreated samples. Assuming that from total RNA, just a fraction of it is pre-mRNA, the logical maximum fold increase that could be observed in the treated samples is a 2-fold change. So, why are we observing an increase in pre-mRNA that is suggesting an increase in total RNA when treating the degron strains with rapamycin? Rapamycin is a nutrient-sensing pathway inhibitor, and even though *C. merolae's* endogenous FKBP is not sensitive to rapamycin, the chemical can cause some other effects on the algae's gene expression. It is unknown if *C. meroale's* genes are upregulated or downregulated or both depending on the gene, upon the addition of rapamycin but either could happen as a response to a change in the ideal growing conditions (stressor, environmental change). The fact that rapamycin is activating a protein degradation system in the degron strains as a means of splicing inhibition, can make the cells activate several molecular defence mechanisms causing up or downregulation of different genes. The qPCR results for the treated degron strains suggest an increase in gene expression upon the addition of rapamycin as the fold-change shows an increase in total RNA in treated versus untreated samples. Further analyses need to be performed regarding gene expression in rich media added with rapamycin to elucidate which genes are upregulated upon its addition and the possible reasons for it. For now, even though the fold-change in pre-mRNA increase was more than what was expected, it can be said with accuracy that pre-mRNA accumulation occurred upon rapamycin addition and that splicing was partially inhibited.

This protein knock-down system offers several advantages but has several limitations. First, rapamycin was shown to be active in *C. merolae* cultures for a period as short as 4 hours, so repetitive doses need to be administered to maintain the target protein levels low enough across the duration of the experiment. Second, although growth assays and transcriptomic analyses

(Fujiwara *et al*., 2024) do not show substantially pronounced side effects on the cells upon rapamycin addition, caution may be taken when analyzing cellular processes. Regarding this point, in the present project, it was demonstrated that the controls (untreated with rapamycin), grew slightly faster than the rapamycin-treated samples, and the final OD of the former was higher than the one of the latter at the end of the experiments, although the difference was not great. Overall, it could be concluded that rapamycin did not substantially affect the cells. Additionally, these concerns were solved when the parent strain was analyzed at an RNA level, confirming that the effects seen in splicing for the Clf1 and Prp8 degron strains were not due to rapamycin addition but due to the activity of the inducible protein knock-down system. Third, the fusion of the FRB disordered domain to the N or C-terminus of a protein increases the levels of protein degradation constitutively, without the addition of rapamycin.

The first limitation affected the experiments in this project mainly because a limited amount of rapamycin was available at this stage, and the cells could not be maintained for more than 12 hours in rapamycin. This time frame was enough to observe changes in splicing patterns for the analyzed genes. However, it was insufficient to achieve 100% splicing inhibition and check for cell phenotypes, deleterious effects or death. Thus, whether splicing is essential or not in *C. meroale* remains unclear. Rapamycin addition overnight would have been complex without an adequate delivery system to add the appropriate rapamycin amount to the cultures every 4 hours. A peristaltic pump would have been a valuable piece of equipment to try. To overcome this limitation, it would have been helpful to screen for a rapamycin derivative that sustains its effect in acidic media for more than 4 hours. One of these rapamycin analogs is known as RapaLog or MaRap, which has been used for protein knock-down systems in human cells (Willmington & Matouschek, 2016). The use of these analogs will also solve the second

limitation. Rapamycin is a TOR (mTOR) inhibitor, and rapamycin analogs such as MaRap exhibit a substantially lower affinity for the WT FRB domain of TOR but a higher affinity for FRB mutants. These modifications must be used when working with organisms with rapamycin-sensitive TOR pathways. In Willmington and Matouschek, 2016, a mutated version of the human FRB domain and RapaLog activated the protein knock-down system. This study was performed in human cells, so adding rapamycin would have inhibited the mTOR pathway, causing adverse physiological effects on cells. The potential use of rapamycin analogs was not considered for these experiments, not only because they were not available at the moment the experiments were performed but also because they were unnecessary. The rapamycin-based system can be applied to *C. merolae* with no issues, as it has been shown that the endogenous FKBP is insensitive to rapamycin, in contrast to FKBP in other organisms. As a result, rapamycin does not affect *C. merolae's* TOR pathway. However, it does affect and inhibit TOR (mTOR) pathways in other organisms, such as humans or superior plants (*A. thaliana)* (Imamura *et al*., 2013). *C. merolae* is, therefore, the first photosynthetic organism in which a rapamycin-induced system can be used.

The third limitation was supported by the results obtained in the experiments shown for this project. The mVenus-FRB control strain and the Clf1 and Prp8 degron strains (despite containing the mVenus fluorescent protein) were never fluorescent. As shown in the results for the mVenus-FRB strain in Fujiwara *et al.*, 2024, which is the same mVenus-FRB strain used for this project, the strain was barely fluorescent. A strain containing mVenus under the control of the EFTU promoter was used as a control without the disordered domain fused to it. This strain was shown to be fluorescent, and its fluorescence level was set at 100%. The mVenus-FRB strain, in which mVenus-FRB was under the control of the same promoter without exposure to rapamycin,

showed a fluorescence level of 16% compared to the control. The fluorescence of mVenus was reduced by 84% just by having the FRB domain attached to it. This means the protein levels dramatically reduced when the FRB domain was added to mVenus. This explains why the Clf1 and Prp8 strains used for this work showed no fluorescence. Additionally, as mentioned before, according to our lab's NCKM data, Clf1 and Prp8 are very lowly expressed. If the increased degradation level is added to the low expression of these two splicing proteins, it is not unexpected that cells were not fluorescent. This might explain why the proteins were also undetectable by western blotting. Fujiwara *et al*., 2024 tried adding a Stabilon tag to counteract the effect of the FRB-disordered domain in enhancing protein degradation, and the experiment was successful. The reduction of mVenus levels due to the presence of the FRB domain was partially mitigated when adding the Stabilon-Tag. The maltose binding protein has also been used as a stabilizing domain for human degron systems (Willmington & Matouschek, 2016). To counteract the effect of the disordered domain, the FRB domain can be fused to the adaptor. This is a critical aspect for future degron experiments on splicing proteins. To counteract the effect of protein degradation due to the fusion of a disordered domain, the disordered domain can be included in the adaptor instead of in the target protein. So for this work, the HA-SKP1 adaptor will contain the FRB domain being HA-FRB-SKP1, and the FKBP will be fused to the splicing protein (e.g., Prp8-mVenus-FKBP). The degron system will function equally efficiently with this modification as the rapamycin will act as the bridging ligand between FRB and FKPB, leading proteins to degradation. This will reduce constitutive protein degradation of splicing proteins and may facilitate their detection by western blotting.

This project used only one adaptor, in which the SKP1 E3 ligase subunit was included to promote protein ubiquitination, which will mark the protein for degradation by the proteasome.

Other E3 ligase subunits have been used as ubiquitination machinery for rapamycin-induced degron systems. The RBX and CUL1 subunits were used by Fujiwara *et al*., 2024. This study's other two adaptor components remained the same for the three variations. The efficiency of the different adaptors was tested, determining that the adaptor that most efficiently promoted protein degradation was the one containing CUL1 (HA-FKBP-CUL1). This is another thing to try for future inducible degron systems for splicing proteins in *C. merolae.* The adaptor used for this dissertation, HA-FKBP-SKP1, was demonstrated to be efficient in promoting rapid protein degradation. In a 4-hour time frame, mVenus-FRB/GFP was 90% degraded (Fujiwara *et al*., 2024; Willmington & Matouschek, 2016). Adaptors used for degron systems in human cells contain proteasome binding tags instead of E3 ligase subunits. In this case, the ubiquitination step is omitted, and the proteasome automatically recognizes the protein for degradation. Using adaptors as the UbL domain of the human Rad23b protein will make this system more direct and will remove one step in the process of inducing protein degradation (Willmington & Matouschek, 2016). Adding a fluorescent protein to the adaptor, which is not supposed to be degraded, will help keep track of it throughout the experiments. Notably, adding a stabilizing domain would also be a good idea if the FRB domain is included in the adaptor and not fused to the target protein (Willmington & Matouschek, 2016).

Other degron techniques are not based on rapamycin. These techniques do not fuse the human FRB-disordered domain to target protein or include an adaptor containing the FKBP so that, upon rapamycin addition, protein degradation will be triggered. These systems take advantage of the WD40 domains in the F-box proteins, which are also an E3 ligase subunit. F-box proteins are the ones in charge of determining substrate specificity. F-box proteins are bound to SKP1 through the F-box domain and perform substrate recognition through the WD40 interaction motifs.

A degron system published by Caussinus *et al*. in 2012 replaces the WD40 interaction motifs of E3 ligases F-box proteins with a nanobody (nano-antibody) against GFP or derivatives. The system does not need any small molecule to be activated, as the nanobody will automatically recognize the target protein, and the E3 ligase will ubiquitinate the protein, leading it to degradation.

The Fujiwara *et al*., 2024 study does not target splicing proteins but proteins related to transcription (transcription factors) and chloroplast division. Due to the low expression level and abundance of the splicing proteins, it is unknown to what extent this might have affected the efficiency of the system. If a rapamycin inducible degron system is used in the future with splicing proteins in *C. merolae*, a helpful recommendation would be not to fuse the splicing protein to mVenus and FRB but only to FRB domain and use either a commercial antibody against the protein itself for its detection or tag the protein with another tag as Flag, His, or HA, for which commercial antibodies are available and has been demonstrated that are efficient for protein pull-dows or detection. In Fujiwara *et al*., 2024 for the experiments performed with the E2F transcription factor, the constructs were designed without fusing mVenus to E2F but just fusing E2F to a 4xFLAG tag and the FRB domain. Degradation was successfully detected.

The degron-disordered domain can be fused either to the C-terminus or to the N-terminus of the target proteins. In the study performed by Fujiwara *et al*., 2024 the FRB was fused to the chloroplast division protein DRP5B at the N-terminus and the E2F transcription factor at the C-terminus. Both proteins were successfully degraded, and nothing is mentioned about the efficiency of fusing the protein to the N or C-terminus of the target protein. Fusing the FRB or FKBP domain to the N-terminus of the protein would be worth trying for future experiments, and those results should be compared with those of the protein with the FRB fused to the C-terminus.

Also, targeting smaller and not core but essential splicing proteins such as Prp22 would be something to try in the future.

Western blots presented in the study performed by Fujiwara *et al*., 2024, were successful for every protein analyzed. However, the process followed for western blotting was not explained in detail in this paper's materials and methods section. Despite all the attempts to get nice-looking western blots and detect splicing proteins, nothing was detected.

I attempted several experiments to try to detect Clf1-mVenus-FRB by modifying the western blot protocol (transfer time, transfer buffer, antibody concentration, antibody type) with different samples from the CSR100 and CSR101 strains treated with 1 and 2uM rapamycin. None of these attempts showed the Clf1-mVenus-FRB band or its degradation over time. Splicing proteins are not abundant due to their low expression level. The NCKM data showed that Clf1 is expressed at a level of 1126 counts per kilobase mapped, which is a very low expression level compared with one of the most highly expressed intron-containing genes in *C. merolae*, CMC053C, which is ~96000 counts per kilobase mapped. Higher amounts of protein must be loaded on the SDS-PAGE gel to detect very low abundant proteins. The most suitable way to concentrate protein is to perform an immunoprecipitation or ammonium sulphate precipitation with a high starting cell concentration or prepare protein extract instead of whole cell extract. This will allow the loading of higher total protein amounts.

The rapamycin exposure degron experiments were designed to be performed in low volumes and at ODs below the saturation point (OD 3), where cells are actively and exponentially growing. The volumes needed to be low due to rapamycin limitations. Rapamycin needed to be added at a 1 uM concentration to 16 cultures and replenished every 4 hours. The rapamycin vial needed to be enough for performing the complete experiment, so maximum volumes of 6mL

were used per sample. Considering an initial OD of 1.5 and the fact that samples needed to be taken at each time point (4 time points) for RNA and protein, the number of cells was limited, and a high starting cell concentration was not possible for performing any of the mentioned experiments to concentrate the protein and get more chances of detecting it. Whole-cell extract using the same OD units per sample taken was the most suitable track to follow but certainly not the best one. In one attempt in which untreated cells from the two Clf1 and Prp8 strains were grown up to OD10, protein extract was prepared from 100OD units (10mL of cells at OD10), protein concentration was measured, and 20 ug of protein was loaded on the gel. This western blot failed to detect Clf1, but the adaptor was detected. A rabbit polyclonal anti-GFP antibody was tried in some other westerns to detect Clf1 instead of the mouse-monoclonal antibody used for the experiment in Figure 3-20. The Clf1 detection was also unsuccessful. It is important to mention that for the experiment in Figure 3-19, the rabbit polyclonal anti-GFP antibody was used as an initial attempt to detect mVenus-FRB with the mouse-monoclonal anti-GFP, which was unsuccessful. This might have implied that the mouse monoclonal antibody was not correctly working. However, once the experiment turned unsuccessful, as with the rabbit polyclonal antibody, it was concluded that Clf1 detection was unsuccessful for reasons that were not clear.

The same as for Clf1 is valid for Prp8 regarding using different antibodies and preparing protein extract to load 20 ug of protein per lane on the gel. As mentioned, Prp8 was never detected. Prp8 is the most conserved protein among eukaryotes and is considered the core of the spliceosome in eukaryotic organisms (Moore *et al*., 1999). Although Prp8 is highly expressed in humans, in *C. merolae,* it is lowly abundant, and according to the NCKM data, its expression level is as low as 255 counts per kilobase mapped. This number is ~4 times lower than the one for Clf1, making it even less likely to detect Prp8. Additionally, Prp8 is a giant protein, which

makes transferring more challenging. Several studies in the literature have shown that human Prp8 has been detected by western blot using a standard western blot protocol (Moore *et al.*, 1999). Also, anti-GFP was used for Prp8 detection in *C. merolae* instead of an anti-Prp8 antibody, as used for human Prp8 detection in Moore *et al.*, 1999.

As previously mentioned, qPCR results indicated partial splicing inhibition for most of the genes analyzed. Was a 12-hour time frame sufficient for splicing analyses? Cells require prolonged exposure to rapamycin to observe the long-term effects of the treatment and to determine if cell death follows 100% splicing inhibition. According to CMJ129C qPCR results, complete splicing inhibition for the remaining genes could be observed if qPCR analyses are conducted beyond the 12-hour time frame. Once this is established, growth phenotypes can be evaluated. I would also recommend targeting more than one splicing protein for degradation: perhaps if two or three splicing proteins were simultaneously tagged with FRB there would be more complete inhibition of splicing.

As an overall conclusion, the appropriate function of the degron system for this project was supported by the qPCR results, in which splicing inhibition was shown by pre-mRNA accumulation when the degron strains were treated with rapamycin. Thus, the interaction between the Clf1 and Prp8 fusion proteins with the adaptor upon adding rapamycin was successful, and proteins were degraded. For reasons previously discussed in this chapter's results and discussion sections, the splicing protein degradation was not observed by western blot. However, the adaptor was present in all the strains and samples analyzed. The low expression level of splicing proteins, the constitutive degradation caused by the FRB domain fusion, the promoter used for the expression of mVenus-FRB in the control strain versus the promoter used for the adaptor, etc, were factors that potentially influenced the results from these experiments.

As a reminder, mVenus-FRB in the control strains is under the control of the EFTU promoter, whose expression level is low compared to the APCC promoter, which is driving the expression of the adaptor. That is likely why the adaptor was more easily detected than the mVenus-FRB. However, nothing can be concluded about whether splicing is essential for *C. merolae* as, because of the limitations with rapamycin previously mentioned, the experiment could not be carried out for extended periods. An effect on splicing was indeed observed, but a deleterious cell phenotype due to this effect was not observed. According to our calculations based on the qPCR results, the splicing of CMJ129C was utterly inhibited after 12 hours of rapamycin addition, and the cells were growing normally. Does this mean that splicing is not essential? All the mentioned considerations are worth trying for future experiments, and further, more conclusive experiments need to be performed to answer this question.

Although the degron system requires further investigation and additional experimental design for future experiments, in this project it is demonstrated that the degron system degrades splicing proteins and partially inhibits splicing in a 12-hour time frame.

4. Chapter 4 – Concluding Remarks

The hypothesis that splicing is essential for *C. merolae* cannot be definitively accepted or rejected based on the results of this project. Antisense RNA experiments aimed at inhibiting splicing did not yield conclusive results, as the expression of antisense genes was undetectable and no discernible effects on splicing or cell growth were observed. Using degron techniques to inhibit splicing for the first time in *C. merolae* showed some promise, with q-PCR results indicating splicing inhibition upon the addition of rapamycin for the splicing protein degron strains. However, the inability to detect protein degradation over time using Western Blot and the absence of growth defects during rapamycin treatment complicated the interpretation of these results. Although splicing inhibition of 5 intron-containing genes was observed, further comprehensive experiments are necessary to firmly establish whether splicing is truly essential for *C. merolae.*

5. References

1. Agafonov, D. E., Deckert, J., Wolf, E., Odenwälder, P., Bessonov, S., Will, C. L., Urlaub, H., & Lührmann, R. (2011). Semiquantitative proteomic analysis of the human spliceosome via a novel two-dimensional gel electrophoresis method. *Molecular and Cellular Biology*, *31*(13), 2667–2682. https://doi.org/10.1128/MCB.05266-11
2. Banerjee, S., Galarza-Muñoz, G., & Garcia-Blanco, M. A. (2023). Role of rna alternative splicing in t cell function and disease. *Genes*, *14*(10), 1896. https://doi.org/10.3390/genes14101896
3. Caussinus, E., Kanca, O., & Affolter, M. (2012). Fluorescent fusion protein knockout mediated by anti-GFP nanobody. *Nature Structural & Molecular Biology*, *19*(1), 117–121. https://doi.org/10.1038/nsmb.2180
4. Csuros, M., Rogozin, I. B., & Koonin, E. V. (2011). A detailed history of intron-rich eukaryotic ancestors inferred from a global survey of 100 complete genomes. *PLoS Computational Biology*, *7*(9), e1002150. https://doi.org/10.1371/journal.pcbi.1002150
5. Fabrizio, P., Dannenberg, J., Dube, P., Kastner, B., Stark, H., Urlaub, H., & Lührmann, R. (2009). The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Molecular Cell*, *36*(4), 593–608. https://doi.org/10.1016/j.molcel.2009.09.040
6. Fredericks, A., Cygan, K., Brown, B., & Fairbrother, W. (2015). RNA-binding proteins: Splicing factors and disease. *Biomolecules*, *5*(2), 893–909. https://doi.org/10.3390/biom5020893
7. Fujiwara, T., Hirooka, S., & Miyagishima, S. (2021). A cotransformation system of the

unicellular red alga *Cyanidioschyzon merolae* with blasticidin S deaminase and chloramphenicol acetyltransferase selectable markers. *BMC Plant Biology*, *21*(1), 573. https://doi.org/10.1186/s12870-021-03365-z

8. Fujiwara, T., Kanesaki, Y., Hirooka, S., Era, A., Sumiya, N., Yoshikawa, H., Tanaka, K., & Miyagishima, S.-Y. (2015). A nitrogen source-dependent inducible and repressible gene expression system in the red alga *Cyanidioschyzon merolae*. *Frontiers in Plant Science*, *6*. https://doi.org/10.3389/fpls.2015.00657
9. Fujiwara, T., Hirooka, S., Yamashita, S., Yagisawa, F., & Miyagishima, S. (2024). Development of a rapamycin-inducible protein-knockdown system in the unicellular red alga *Cyanidioschyzon merolae*. *Plant Physiology*, kiae316. https://doi.org/10.1093/plphys/kiae316
10. Galej, W. P., Wilkinson, M. E., Fica, S. M., Oubridge, C., Newman, A. J., & Nagai, K. (2016). Cryo-EM structure of the spliceosome immediately after branching. *Nature*, *537*(7619), 197–201. https://doi.org/10.1038/nature19316
11. Gildea, M. A., Dwyer, Z. W., & Pleiss, J. A. (2020). Multiplexed primer extension sequencing: A targeted RNA-seq method that enables high-precision quantitation of mRNA splicing isoforms and rare pre-mRNA splicing intermediates. *Methods*, *176*, 34–45. https://doi.org/10.1016/j.ymeth.2019.05.013
12. Gonorazky, H. D., Naumenko, S., Ramani, A. K., Nelakuditi, V., Mashouri, P., Wang, P., Kao, D., Ohri, K., Viththiyapaskaran, S., Tarnopolsky, M. A., Mathews, K. D., Moore, S. A., Osorio, A. N., Villanova, D., Kemaladewi, D. U., Cohn, R. D., Brudno, M., & Dowling, J. J. (2019). Expanding the boundaries of RNA sequencing as a diagnostic tool for rare Mendelian disease. *The American Journal of Human Genetics*, *104*(3), 466–483.

https://doi.org/10.1016/j.ajhg.2019.01.012

13. Hegele, A., Kamburov, A., Grossmann, A., Sourlis, C., Wowro, S., Weimann, M., Will, C. L., Pena, V., Lührmann, R., & Stelzl, U. (2012). Dynamic protein-protein interaction wiring of the human spliceosome. *Molecular Cell*, *45*(4), 567–580. https://doi.org/10.1016/j.molcel.2011.12.034
14. Hogg, R., McGrail, J. C., & O'Keefe, R. T. (2010). The function of the NineTeen Complex (NTC) in regulating spliceosome conformations and fidelity during pre-mRNA splicing. *Biochemical Society Transactions*, *38*(4), 1110–1115. https://doi.org/10.1042/BST0381110
15. Hudson, A. J., Stark, M. R., Fast, N. M., Russell, A. G., & Rader, S. D. (2015). Splicing diversity revealed by reduced spliceosomes in *C. merolae* and other organisms. *RNA Biology*, *12*(11), 1–8. https://doi.org/10.1080/15476286.2015.1094602
16. Ikpa, P. T., Bijvelds, M. J. C., & de Jonge, H. R. (2014). Cystic fibrosis: Toward personalized therapies. *The International Journal of Biochemistry & Cell Biology*, *52*, 192–200. https://doi.org/10.1016/j.biocel.2014.02.008
17. Imamura, S., Kawase, Y., Kobayashi , I., Sone, T., Era, A., Ohta, H., Tanaka , K., Miyagishima, S., & Shimojima, M. (2015). Target of rapamycin (TOR) plays a critical role in triacylclycerol accumulation in microalgae. *Plant Molecular Biology* , *89*, 309–318. https://doi.org/DOI 10.1007/s11103-015-0370-6
18. International Human Genome Sequencing Consortium, Whitehead Institute for Biomedical Research, Center for Genome Research:, Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., …

Morgan, M. J. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860–921. https://doi.org/10.1038/35057062

19. Irimia, M., & Roy, S. W. (2014). Origin of spliceosomal introns and alternative splicing. *Cold Spring Harbor Perspectives in Biology*, *6*(6), a016071–a016071. https://doi.org/10.1101/cshperspect.a016071
20. Koonin, E. V., Csuros, M., & Rogozin, I. B. (2013). Whence genes in pieces: Reconstruction of the exon-intron gene structures of the last eukaryotic common ancestor and other ancestral eukaryotes: Whence genes in pieces. *Wiley Interdisciplinary Reviews: RNA*, *4*(1), 93–105. https://doi.org/10.1002/wrna.1143
21. Lambowitz, A. M., & Belfort, M. (2015). Mobile bacterial group II introns at the crux of eukaryotic evolution. *Microbiology Spectrum*, *3*(1), 3.1.04. https://doi.org/10.1128/microbiolspec.MDNA3-0050-2014
22. Lipinski, K. A., Senn, K. A., Zeps, N. J., & Hoskins, A. A. (2023). Biochemical and genetic evidence supports Fyv6 as a second-step splicing factor in *Saccharomyces cerevisiae*. *RNA*, *29*(11), 1792–1802. https://doi.org/10.1261/rna.079607.123
23. Lord, J., & Baralle, D. (2021). Splicing in the diagnosis of rare disease: Advances and challenges. *Frontiers in Genetics*, *12*, 689892. https://doi.org/10.3389/fgene.2021.689892
24. Lukačišin, M., Espinosa-Cantú, A., & Bollenbach, T. (2022). Intron-mediated induction of phenotypic heterogeneity. *Nature*, *605*(7908), 113–118. https://doi.org/10.1038/s41586-022-04633-0
25. Majumdar, R., Rajasekaran, K., & Cary, J. W. (2017). RNA interference (RNAi) as a potential tool for control of mycotoxin contamination in crop plants: Concepts and considerations. *Frontiers in Plant Science*, *8*. https://doi.org/10.3389/fpls.2017.0020

26. Matera, A. G., & Wang, Z. (2014). Erratum: A day in the life of the spliceosome. *Nature Reviews Molecular Cell Biology*, *15*(4), 294–294. https://doi.org/10.1038/nrm3778

27. Matsuzaki, M., Misumi, O., Shin-i, T., Maruyama, S., Takahara, M., Miyagishima, S., Mori, T., Nishida, K., Yagisawa, F., Nishida, K., Yoshida, Y., Nishimura, Y., Nakao, S., Kobayashi, T., Momoyama, Y., Higashiyama, T., Minoda, A., Sano, M., Nomoto, H., … Kuroiwa, T. (2004). Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature*, *428*(6983), 653–657. https://doi.org/10.1038/nature02398

28. Misumi, O., Matsuzaki, M., Nozaki, H., Miyagishima, S., Mori, T., Nishida, K., Yagisawa, F., Yoshida, Y., Kuroiwa, H., & Kuroiwa, T. (2005). *Cyanidioschyzon merolae* genome. A tool for facilitating comparable studies on organelle biogenesis in photosynthetic eukaryotes. *Plant Physiology*, *137*(2), 567–585. https://doi.org/10.1104/pp.104.053991

29. Nielsen, T., & Nielsen, J. (2013). Antisense gene silencing: Therapy for neurodegenerative disorders? *Genes*, *4*(3), 457–484. https://doi.org/10.3390/genes4030457

30. Ohnuma, M., Misumi, O., Fujiwara, T., Watanabe, S., Tanaka, K., & Kuroiwa, T. (2009). Transient gene suppression in a red alga, *Cyanidioschyzon merolae* 10D. *Protoplasma*, *236*(1–4), 107–112. https://doi.org/10.1007/s00709-009-0056-5

31. Parenteau, J., & Abou Elela, S. (2019). Introns: Good day junk is bad day treasure. *Trends in Genetics*, *35*(12), 923–934. https://doi.org/10.1016/j.tig.2019.09.010

32. Parenteau, J., Durand, M., Véronneau, S., Lacombe, A.-A., Morin, G., Guérin, V., Cecez, B., Gervais-Bird, J., Koh, C.-S., Brunelle, D., Wellinger, R. J., Chabot, B., & Abou Elela,

S. (2008). Deletion of many yeast introns reveals a minority of genes that require splicing for function. *Molecular Biology of the Cell*, *19*(5), 1932–1941. https://doi.org/10.1091/mbc.e07-12-1254

33. Ryczek, N., Łyś, A., & Makałowska, I. (2023). The functional meaning of 5′utr in protein-coding genes. *International Journal of Molecular Sciences*, *24*(3), 2976. https://doi.org/10.3390/ijms24032976
34. Sarharkar, M., Chow, V., & Kangueane, P. (2004). *Distributions of exons and introns in the human genome*. *4*(4), 387–393.
35. Schärfen, L., Zigackova, D., Reimer, K. A., Stark, M. R., Slat, V. A., Francoeur, N. J., Wells, M. L., Zhou, L., Blackshear, P. J., Neugebauer, K. M., & Rader, S. D. (2022). Identification of alternative polyadenylation in *Cyanidioschyzon merolae* through long-read sequencing of mRNA. *Frontiers in Genetics*, *12*, 818697. https://doi.org/10.3389/fgene.2021.818697
36. Schreib, C. C., Bowman, E. K., Hernandez, C. A., Wilson, A. L., Potts, C. H. S., & Maeder, C. (2018). Functional and biochemical characterization of Dib1's role in pre-messenger RNA splicing. *Journal of Molecular Biology*, *430*(11), 1640–1651. https://doi.org/10.1016/j.jmb.2018.04.027
37. Stark, M. R., Dunn, E. A., Dunn, W. S. C., Grisdale, C. J., Daniele, A. R., Halstead, M. R. G., Fast, N. M., & Rader, S. D. (2015). Dramatically reduced spliceosome in *Cyanidioschyzon merolae*. *Proceedings of the National Academy of Sciences*, *112*(11). https://doi.org/10.1073/pnas.1416879112
38. Sumiya, N., Fujiwara, T., Kobayashi, Y., Misumi, O., & Miyagishima, S. (2014). Development of a heat-shock inducible gene expression system in the red alga

Cyanidioschyzon merolae. *PLoS ONE*, *9*(10), e111261. https://doi.org/10.1371/journal.pone.0111261

39. Turunen, J. J., Niemelä, E. H., Verma, B., & Frilander, M. J. (2013). The significant other: Splicing by the minor spliceosome: Splicing by the minor spliceosome. *Wiley Interdisciplinary Reviews: RNA*, *4*(1), 61–76. https://doi.org/10.1002/wrna.1141
40. Vasanthakumar, T., & Rubinstein, J. L. (2020). Structure and roles of V-type ATPases. *Trends in Biochemical Sciences*, *45*(4), 295–307. https://doi.org/10.1016/j.tibs.2019.12.007
41. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., … Zhu, X. (2001). The sequence of the human genome. *Science*, *291*(5507), 1304–1351. https://doi.org/10.1126/science.1058040
42. Vosseberg, J., & Snel, B. (2017). Domestication of self-splicing introns during eukaryogenesis: The rise of the complex spliceosomal machinery. *Biology Direct*, *12*(1), 30. https://doi.org/10.1186/s13062-017-0201-6
43. Ward, A. J., & Cooper, T. A. (2010). The pathobiology of splicing. *The Journal of Pathology*, *220*(2), 152–163. https://doi.org/10.1002/path.2649
44. Wilkinson, M. E., Charenton, C., & Nagai, K. (2020). RNA splicing by the spliceosome. *Annual Review of Biochemistry*, *89*(1), 359–388. https://doi.org/10.1146/annurev-biochem-091719-064225
45. Wilmington, S. R., & Matouschek, A. (2016). An inducible system for rapid degradation of specific cellular proteins using proteasome adaptors. *PLOS ONE*, *11*(4), e0152679.

https://doi.org/10.1371/journal.pone.0152679

46. Wong, D. K., Grisdale, C. J., Slat, V. A., Rader, S. D., & Fast, N. M. (2022). The evolution of pre-mRNA splicing and its machinery revealed by reduced extremophilic red algae. *Journal of Eukaryotic Microbiology*. https://doi.org/10.1111/jeu.12927
47. Xu, H., Fair, B. J., Dwyer, Z. W., Gildea, M., & Pleiss, J. A. (2019). Detection of splice isoforms and rare intermediates using multiplexed primer extension sequencing. *Nature Methods*, *16*(1), 55–58. https://doi.org/10.1038/s41592-018-0258-x

6. Appendixes

6.1. Appendix 1 – Oligonucleotide table

Table 6-1: Oligonucleotides used throughout this research project. The name, purpose, target gene, and sequence are shown in each column.

Name	Purpose	Product/ target	Sequence (5' to 3')
oSDR1524	Colony PCR	CAT	TGTTTATGAACTCTATTCAGGAATTGTCAGATAGGCC
oSDR1588	PCR	NIR Promoter	cctatcGGTACCCCGCTATCAATATCCGACGATATGCA
oSDR1816	Sequencing	pSR886 - APCC promoter	CTGGCACTTTTCCATCTC
oSDR1817	Sequencing	Beta Tubulin 3'UTR	GGGAGAAGTTGCTGCGGTAT
oSDR1818	colony PCR	CMD184C	GATGTTGGCGACCTCCTGAA
oSDR1820	colony PCR	CMD185C	GCAGCAGAGTCGCTATCCAA
oSDR1843	RT-PCR	CMK142T	GTGGATAACATGCCTCATCGC
oSDR2132	Reverse Transcription	CMJ129C	GCCGTCGCCATGATGTACGT
oSDR2135	Reverse Transcription	CMJ129C	GTGCCTTGACGACCTGCT
oSDR2191	PCR	NR promoter CMG019	gatcgcagat TCTAGA ATGCACCATCATGCGTG
oSDR2192	PCR	NR promoter CMG019	gcagat CCATGG GTGGATTACATGTTAAGTGGATTAATTAACATAC
oSDR2193	PCR	UBQ 3' UTR CMK296	gcagat CCATGG GGTTTTCGGTGGTATGTGATTTAC
oSDR2194	PCR	UBQ 3' UTR CMK296	gcagat GCTAGC TCAGCCACGACAGAGAC
oSDR2195	Duplex	XbaI-SwaI-BamHI	ctagaGcttatctcaatATTTAAATtggtgattgttCGg
oSDR2196	Duplex	XbaI-SwaI-BamHI	gatccCGaacaatcaccaATTTAAATattgagataagCt
oSDR2197	Duplex	SmaI-PacI-KpnI	gggCAGTTGAAGTATGTTAATTAATCCACTTAACATGggtac
oSDR2198	Duplex	SmaI-PacI-KpnI	cCATGTTAAGTGGATTAATTAACATACTTCAACTGccc
oSDR2199	PCR	EcoRI-FRB 5'	gcagat GAATTC GAGATGTGGCATGAGGG
oSDR2200	PCR	AatII-FRB 3'	gcagat GACGTC TTAAAGTTGTTTGCTAATACGTCGG
oSDR2201	PCR	Cef1 Sense 5'	AGTTGAAGTATGTTAC atgacccgccgcttc
oSDR2202	PCR	Cef1 Sense 3'	ATGTTAAGTGGATTAC tcacgttcgctcccg
oSDR2203	PCR	Cef1 Antisense 5'	AGTTGAAGTATGTTAC TCACGTTCGCTCCCG
oSDR2204	PCR	Cef1 Antisense 3'	ATGTTAAGTGGATTAC ATGACCCGCCGCTTC
oSDR2205	PCR	U4 Sense 5'	AGTTGAAGTATGTTAC ATACTTGCGCAGTGTCGG
oSDR2206	PCR	U4 Sense 3'	ATGTTAAGTGGATTAC CTTTCCAAAAATTTCCACCAAGC
oSDR2207	PCR	U4 Antisense 5'	AGTTGAAGTATGTTAC CTTTCCAAAAATTTCCACCAAGC
oSDR2208	PCR	U4 Antisense 3'	ATGTTAAGTGGATTAC ATACTTGCGCAGTGTCGG
oSDR2209	PCR	U6 sense 5'	AGTTGAAGTATGTTAC GGTGCGCCTTTATCGG
oSDR2210	PCR	U6 sense 3'	ATGTTAAGTGGATTAC AAAAAGGTATACCTCGAGACGATT
oSDR2211	PCR	U6 antisense 5'	AGTTGAAGTATGTTAC AAAAAGGTATACCTCGAGACGATT

oSDR2212	PCR	U6 antisense 3'	ATGTTAAGTGGATTAC GGTGCGCCTTTATCGG
oSDR2213	PCR	U2 sense 5'	AGTTGAAGTATGTTAC CTCATGGTGTATCGAGAGC
oSDR2214	PCR	U2 sense 3'	ATGTTAAGTGGATTAC AACGAAAAATTAGTTAAGAGATGCAG
oSDR2215	PCR	U2 antisense 5'	AGTTGAAGTATGTTAC AACGAAAAATTAGTTAAGAGATGCAG
oSDR2216	PCR	U2 antisense 3'	ATGTTAAGTGGATTAC CTCATGGTGTATCGAGAGC
oSDR2229	PCR	Prp8	CTT ATC TCA ATA TTT G AACGCGGATACAGTGGTTTG
oSDR2230	PCR	Prp8	AAC AAT CAC CAA TTT G AGTTCCCTCTTCGATCGGTG
oSDR2231	PCR	Prp8	AGT TGA AGT ATG TTA C ACTAGCACGGGATGATGC
oSDR2232	PCR	Prp8	ATG TTA AGT GGA TTA C AGCAGCGTCAAACTCGATAC
oSDR2233	PCR	Clf1	CTT ATC TCA ATA TTT G TCTGGAGCGATGTAGGGAG
oSDR2234	PCR	Clf1	AAC AAT CAC CAA TTT G GTCAGACGGATCGCGGATAT
oSDR2235	PCR	Clf1	AGT TGA AGT ATG TTA C AAACTTGTAGATTTGGGGCA
oSDR2236	PCR	Clf1	ATG TTA AGT GGA TTA C AATTGCCGCCTGTACAAAAC
oSDR2237	Colony PCR	Hsh155 HR arm	GGAGTGCTGAATGCCTACCT
oSDR2238	Colony PCR	mVenus	CCGACACCGAGAACTTGTG
oSDR2239	Colony PCR	B Tubulin	CACGTTATCGACCAGCTCTC
oSDR2240	Colony PCR	Hsh155 HR arm	TCAGAAGAAGTTGGAAGCAGT
oSDR2241	Colony PCR	Prp8 HR arm	CGATCCTCAATCCGTGCATG
oSDR2242	Colony PCR	Prp8 HR arm	CCCATACTCCAGCACGACAT
oSDR2243	Colony PCR	Clf1 HR arm	GCTTTTACTGGCGAAGGAGG
oSDR2244	Colony PCR	Clf1 HR arm	ACGAATACAAGACCCAGGCA
oSDR2301	PCR	Clf1	GCCGCTCGAGATGCTATTTT
oSDR2302	PCR	Clf1	TAGATCTGGTGGGGCATTGG
oSDR2303	PCR	Hsh155	CATAGCGACCTTACAGAGCC
oSDR2304	PCR	Hsh155	GCGCTGGAAGAACCAAAAAG
oSDR2305	PCR	Prp8	CTTTTGCTCGCGGGTATGAC
oSDR2306	PCR	Prp8	TTCAAGCTCGAGATCGGTGG
oSDR2311	Northern Blot	U2	/5Biosg/GATGCAGGCTCCCTGGAATATAAAATATCCC
oSDR2475	Northern Blot	U2 antisense	/5Biosg/GGGATATTTTATATTCCAGGGAGCCTGCATC
oSDR2476	Northern Blot	U4	/5Biosg/AAATTGTTTGTGTTCAGCATACCGTT
oSDR2477	Northern Blot	U4 antisense	/5Biosg/AACGGTATGCTGAACACAAACAATTT
oSDR2480	PCR	NOS Terminator	CCCATCTCATAAATAACGTCATGC
oSDR2481	Sequencing	NOS Terminator	ATAATCATCGCAAGACCGGC
oSDR2482	Sequencing	3'end Clf1 ORF	GGCGTTACGGAGAACACAC
oSDR2483	Sequencing	3'end Prp8 ORF	CGCAAAAGATGGTCGGCTG
oSDR2484	Northern Blot	Cef1 Sense	/5Biosg/TCT CCA AAA ACC TTT CAA TGA GCT TCC GCG
oSDR2485	Northern Blot	Cef1 Antisense	/5Biosg/CGC GGA AGC TCA TTG AAA GGT TTT TGG AGA
oSDR2523	PCR		GCAGAT GCTAGC CTCATGGTGTATCGAG

oSDR2524	PCR		GCAGTA CTTAAG AAACTTTATTGCCAAATGTTTGAACGATCAACGAAAAATTAGTTAAGAG
oSDR2525	PCR		GCAGAT GCTAGC AACGAAAAATTAGTTAAGAGATGCAGG
oSDR2526	PCR		GCAGTA CTTAAG AAACTTTATTGCCAAATGTTTGAACGATCCCTCATGGTGTATCGAGAGC
oSDR2542	PCR	Dib1 sense	GCAGAT GCTAGC ATGGACAGTGCACCGTTGGT
oSDR2543	PCR	Dib1 sense	GCAGTA CTTAAG AAACTTTATTGCCAAATGTTTGAACGATCCTAGAGTCGGAACGGCGCCA
oSDR2544	PCR	Dib1 antisense	GCAGAT GCTAGC CTAGAGTCGGAACGGCGCCA
oSDR2545	PCR	Dib1 antisense	GCAGTA CTTAAG AAACTTTATTGCCAAATGTTTGAACGATCATGGACAGTGCACCGTTGGT
oSDRFUB5	qPCR	CMS342 C	TCAATGACAGAAACTAACCTGTGA
oSDRFUB6	qPCR	CMS342 C	TGTCGATAGTGTCGCAGTGG
oSDRFUB94	qPCR	CMK142T	GCGATCCTGAATCTGGTCAA
oSDRFUB 113	qPCR	CMK142T	ACCTGCTTCAGTTCCTTGGAC
oSDRFUB 23	qPCR	CMQ270C	AAGCAAACTAACCGTGCGTG
oSDRFUB 24	qPCR	CMQ270C	TCGCCAATCTTCTCATCACCA
oSDRFUB 184	qPCR	CMK260C	GATTCGGGTCTGTTTGGGAT
oSDRFUB 185	qPCR	CMK260C	GTTTGATTCTTGGCTCGCAC
oSDRFUB 97	qPCR	CMJ129C	AACCGTTTCATCAGTGCGAA
oSDRFUB 98	qPCR	CMJ129C	GGAAACCGCACAGAAGCAG
oSDRFUB 108	qPCR	CMS342 C	GCCTAGCGGAGATGCCAG
oSDRFUB 109	qPCR	CMS342 C	CGCTCATAGAAGGAAGCCAGA
oSDRFUB 93	qPCR	CMK142T	TGTCCGTGGACGTATTCAC
oSDRFUB 110	qPCR	CMK142T	AAGCAACATTAGCTTATTGAGTGTG
oSDRFUB 104	qPCR	CMQ270C	GGATCTCAAACGTGGCATTGA
oSDRFUB 105	qPCR	CMQ270C	ATTTCATTCTTGCCGCCGAT
oSDRFUB 182	qPCR	CMK260C	GCAAGGACACGCAATTACAA
oSDRFUB 183	qPCR	CMK260C	GTTGCCACCAGCGAAAATAA
oSDRFUB 124	qPCR	CMJ129C	TTGAGGATTCCCCCTGTTTTG
oSDRFUB 125	qPCR	CMJ129C	TGATAGCCACGTCGCAGAAA
oSDRFUB 263	qPCR	CMK260C	CGCGAGTGCGGAGTCT
oSDRFUB 264	qPCR	CMK260C	GCCCCATCTTGGCAATCTCT
oSDRFUB 266	qPCR	CMK260C	ACCACGGTGTTACTCTCGAGATTC

6.2. Appendix 2 – Restriction Enzymes

Table 6-2: Restriction enzymes used for cloning purposes. The recommended manufacturer's reaction buffer and the optimum activity temperature are shown (all the enzymes are from New England Biolabs).

Enzyme	Buffer	Temperature (°C)	Time Saver?*	Heat inactivated?
XbaI	CutSmart (NEBuffer 4)	37	Yes	Yes
NcoI	CutSmart (NEBuffer 4)	37	Yes	Yes
NheI	CutSmart (NEBuffer 4)	37	Yes	Yes
NdeI	CutSmart (NEBuffer 4)	37	Yes	Yes
PacI	CutSmart (NEBuffer 4)	37	No	Yes
SwaI	NEBuffer 3.1	25	No	Yes
AflII	CutSmart (NEBuffer 4)	37	Yes	Yes
BamHI	CutSmart (NEBuffer 4)	37	Yes	Yes
SmaI	CutSmart (NEBuffer 4)	25	Yes	Yes
KpnI	CutSmart (NEBuffer 4)	37	Yes	Yes
EcoRI	CutSmart (NEBuffer 4)	37	Yes	Yes
AatII	CutSmart (NEBuffer 4)	37	Yes	Yes
ClaI	CutSmart (NEBuffer 4)	37	Yes	Yes
MluI	NEBuffer 3.1	37	Yes	Yes
StuI	CutSmart (NEBuffer 4)	37	Yes	Yes
SphI	CutSmart (NEBuffer 4)	37	Yes	Yes

*Enzymes time-saver qualified will digest unit assay substrate in 5-15 minutes under the reaction conditions indicated by the manufacturer. These enzymes can be safely incubated overnight.

6.3. Appendix 3- LIC Protocol

LIC of PCR Product into PacI/SwaI

Primer design

To make the PCR primers for cloning into the PacI and SwaI sites (…= your gene sequence):

SwaI forward: 5'- CTT ATC TCA ATA TTT **G**...

SwaI reverse: 5'- AAC AAT CAC CAA TTT **G**…

PacI forward: 5'- AGT TGA AGT ATG TTA **C**...

PacI reverse: 5'- ATG TTA AGT GGA TTA **C**...

Design primers to have an annealing temperature of around 65C (use the NEB Tm calculator with Q5 polymerase). Only include your gene specific sequence in the Tm calculation.

*homology arms for integration into Cm need to be a minimum of 500bp

This is what the SwaI LIC site looks like in the plasmid after digestion with SwaI

```
                        |
GCTTATCTCAATATTT  AAATTGGTGATTGTTC
CGAATAGAGTTATAAA TTTAACCACTAACAAG
                        |
```

This is what the site looks like following treatment with T4 DNA pol in the presence of dGTP. The ends of your PCR primers are homologous to these sticky ends, but contain an extra **G**, which is necessary for the T4 DNA pol treatment of the PCR product (which you do in the presence of dCTP).

```
G                       AAATTGGTGATTGTTC
CGAATAGAGTTATAAA                      G
```

This is what the PacI LIC site looks like in the plasmid after digestion with PacI

```
                              |
CAGTTGAAGTATGTTAAT TAATCCACTTAACATG
GTCAACTTCATACAAT TAATTAGGTGAATTGTAC
                              |
```

This is what the site looks like following treatment with T4 DNA polymerase in the presence of dCTP. The ends of your PCR primers are homologous to these sticky ends, but contain an extra **C**, which is necessary for the T4 DNA pol treatment of the PCR product (which you do in the presence of dGTP).

```
C                          TAATCCACTTAACATG
GTCAACTTCATACAAT                        C
```

For knocking out genes the frame is not important, but if you are tagging a gene or making a mutation within a gene then you need to make sure that you add an extra base to your primer before the unique sequence to keep the reading frame intact.

LIC vector prep

Digest 500ng vector (~100-250fmol) with 0.5ul PacI(5U) or SwaI(5U) in 10ul with 1x cutsmart buffer for 3h at 37C (Pac) or in 10uL with 1x NEB buffer 3.1 for 3h at 25C (Swa), followed by 20min at 65C to kill the enzyme. Treat half the digest with T4 DNA polymerase (NEB). This half of the digest is enough for 10 LIC reactions.

5ul digest

0.5ul 10x cutsmart buffer

0.5ul 100 mM DTT

0.5ul 50 mM dCTP(Pac); dGTP(Swa)

0.4ul (1U) T4 DNA pol (NEB)

3.1 uL water

30 min 22 °C, then heat inactivate T4 20 min 75 °C

LIC PCR prep

Get rid of dNTPs from PCR rxn with PCR purification kit. Quantify PCR product.

250fmol PCR product

1ul 10x cutsmart buffer

0.5ul 100mM DTT

0.5ul 50mM dGTP(Pac); dCTP(Swa)

0.4ul (1U) T4 DNA pol NEB

water to 10ul

30min 22C, then heat inactivate T4 20min 75C

LIC Reaction

Combine 1ul vector (25 ng;5-10 fmol) + 1ul PCR product (25 fmol) and incubate 5min at RT. Control = 1ul vector + 1ul water.

*Adjust as necessary to keep molar ratio 1:3 - 1:5, and total mass of DNA <75 ng.

Transform *E. coli* using normal protocol (3 uL rxn/50 uL DH5alpha). Plate on selective plates. You should have a handful of colonies on your control and many times more on your LIC reaction plate. Pick a couple of colonies, grow up, miniprep, and test by RE digest to make sure they contain your insert.

6.4. Appendix 4 – DNA Ligation with T4 DNA Ligase protocol / Transformation of DH5alpha competent cells (Ligation Reaction) (Stark, 2011; updated December 2023).

DNA Ligation with T4 DNA Ligase (NEB)

1. Set up the following reaction in a microcentrifuge tube on ice. (T4 DNA ligase should be added last. Table shows a ligation using a molar ratio of 1:3 vector to insert. Ratios between 1:2 and 1:6 may be used. The final DNA concentration should not exceed 10ng/uL.) Sambrook says to achieve maximum ligation efficiency, set up reactions in as small a volume as possible (5-10ul).

Component	20ul Reaction
10x T4 DNA Ligase Buffer*	2ul
Vector DNA (3kb)	50ng (0.025 pmol)
Insert DNA (1kb)	50ng (0.076 pmol)
Water	To 20ul
T4 DNA Ligase	1ul

*The Ligase Buffer should be thawed and resuspended at room temperature. Repeated freeze/thaw degrades ATP so make 10-20ul aliquots of fresh buffer. If buffer is old you can add more ATP (1mM final) and DTT (10mM final).

2. Gently mix the reaction by pipetting. Don't vortex.

3. For sticky ends, incubate 16C overnight or room temp for 10 min.

4. For blunt ends or single base overhangs, incubate 16C overnight or room temperature for 2 hours (alternatively, high concentration T4 DNA Ligase can be used in a 10-minute ligation).

5. Chill on ice and transform 1-5ul of the reaction into 50ul competent cells.

For 1:3 ratio

Insert mass in ng = 3 x (insert length in bp/vector length in bp) x vector mass in ng

I have routinely used 100 ng vector successfully for a 3-4 kb plasmid. If your backbone is larger, use less mass. Set up the ligation rxn in 10 uL. Mix linearized plasmid backbone (~30 fmole, 50-100ng) with digested insert (~100 fmole; 1:3 molar ratio) plus 1ul DNA ligase and 1ul 10x buffer containing ATP, and water up to 10ul. Keep the total mass of vector + insert under 150 ng, making sure that you keep your molar ratio at 1:3. Set up a control reaction exactly the same way but replace the insert with H2O. Incubate on your bench for 10-15 min. Transform half the reaction.

MRS 23/12

Transformation of DH5alpha (Ligation reaction)

1. Thaw competent cells on ice 5-10min. Spin down 1 sec in bench top centrifuge. Gently mix cells with pipet tip (do not pipet up and down) and aliquot 45-50 uL of cells for each transformation into pre-chilled microcentrifuge tubes. Aliquots are 100 or 200 uL – use the larger aliquot if you have more than 2 ligation reactions to transform. Make sure you put exactly the same amount of cells in each of your tubes so that you can compare the transformation efficiency of your experimental vs. control ligations. Unused cells can be refrozen one time – make 10 uL aliquots and write 10 uL on the cap. KEEP CELLS ON ICE AT ALL TIMES. Cells should be competent to $>10^6$ colonies/ug DNA.

2. Add 5ul DNA to 50ul competent cells (not more than ~75 ng, less is better)*. Mix gently, do not pipet up and down.

3. Incubate on ice 30 min.

4. Heat shock 45 sec at 42 °C.

5. Immediately place tubes on ice for 2 min.

6. Add 500 uL LB if the plasmid has resistance other than Ampicillin. If the plasmid is Amp resistant, add 150 uL LB and plate the entire 200 uL on an LB Carbenicillin plate using 4-5 glass beads.

7. If you've added 500 uL LB, incubate 45min at 37 °C in shaker, in beaker with tubes on their sides. (Not necessary with Amp resistant plasmids, although efficiency will be reduced by about half.)

8. Spin down 10sec, remove all but about 200 uL LB, resuspend, and plate out entire transformation on LB agar + appropriate antibiotic.

9. Incubate overnight upside down at 37C, 14-18 h. If you incubate too long, or if your antibiotic plates are too old, you will get lots of tiny satellite colonies coming up all around your real, antibiotic-resistant colonies.

*Sambrook says add up to 25 ng DNA per 50 uL of competent cells in a volume not exceeding 5% of that of the competent cells. (I have added 5-7.5 uL containing ~50ng DNA.)

MRS 23/12

6.5. Appendix 5 - *C. merolae* transformation protocol (Ohnuma *et al.*, 2008; modified by Martha Stark and updated August, 2023)

CAT Transformation updated 2023.08.24 MRS

Transform WT cells with 1 pmol PCR product containing at least

500 bp homology to your favourite gene- [CAT gene]-500 bp homology to your favourite gene. For transient transformation use 5 ug of plasmid.

2-3 days before transformation (15 min)

Dilute WT cells so that you have an actively dividing culture that will have an OD750 < 3.0 one day before transformation. The doubling time of WT cells grown optimally (42 C, 2% CO2, 90uE light) is ~9h. If you are starting from a saturated culture, the lag time is ~5h.

1 day before transformation (2.5 h)

Dilute cells to OD750 ~0.15 in 50 mL. Aim to transform cells at OD750 0.8-1.0

Pour 0.77XMA2G + 250 ug/mL chloramphenicol (Cp) plates (1/experimental transformation). Leave out O/N to dry, wipe off condensation and store upside down in plastic sleeve at 4C. Always pour fresh Cp plates.

Grow up an extra 50-100 mL of WT cells to OD 1.0 to make MA2GC (conditioned media). Calculate how much you will need based on how many transformations you are doing. You will need this for the final resuspension of the cornstarch as well as for dilution of your cells for plating. You do not need this for transient transformation.

Day of transformation (Day 1) (1 h)

Make fresh PEG4000 (60% w/v in MA-I = 0.9 g PEG + 750 uL MA-I in 2 mL tube; dissolve at 42 °C with occasional inversion– this takes a little while so start it first)

Spin down 40 mL cells OD750 0.8-1.0, 2000xg 10 min, wash with 1 mL warm MA-I buffer and transfer to eppendorf tube, spin 2000xg 1 min, completely remove liquid and resuspend cells in ~100 uL warm MA-I. Total final volume should be 200 uL (=200x concentrated). This is enough for 8 transformations. (40mL OD 1 = 20 x 10e7 cells/200 uL; use 25 uL cells/transformation -> ~2.5 x 10e7 cells/transformation). If you have fewer/more cells, resuspend in a correspondingly lower/higher volume.

While the cells are spinning, mix at RT in eppendorf tube 1 pmol linear DNA (PCR product with 0.5-1.5 Kb homology arms) or 5 ug circular plasmid (for transient) + H2O to 84 uL + 10 uL 10x MA-I + 6 uL 10 mg/mL sonicated salmon sperm DNA (boiled 5 min and then put on ice). Also set up a control with no DNA.

Add 25 uL cells to DNA mixture, then, one sample at a time, add 125 uL PEG, mix quickly by flicking your wrist so that the PEG and the cells/DNA are completely mixed. The cells will form clumps upon contact with the PEG. Immediately add 1 mL warm MA2G, and then quickly pour into 49 mL warm MA2G in graduated cylinder. Use some media from the cylinder to rinse remaining cells from the tube. Grow 24 h with light and 2% bubbling CO2 at 42 C.

Spin down culture for MA2GC max speed 15 min. Sterile filter the resulting media through a 0.2 um syringe filter into a sterile bottle.

Day after transformation (Day 2)

Wash the required amount of cornstarch 3x with MA2G and then resuspend as a 20% v/v slurry in MA2GC. You'll need 2.5 mL 20%/per plate (= 1000 uL 50%/plate). Dry plates open in the 37 °C incubator for ~20-30 min before spotting to make sure that the plates are dry enough to quickly absorb the liquid from the cornstarch. Place cornstarch slurry in a sterile trough and spot 16 uL aliquots onto plates using a multichannel pipette. Avoid poking holes in the Gelzan or blowing bubbles in your spots. Resuspend the slurry frequently, as the cornstarch settles out rapidly. Spot 144 spots/square dish. Allow spots to completely dry with the lid off. You will need 1 plate/transformation. Do not spot your transformation onto wet cornstarch!

24 h after transformation spin down cells 2000xg 10 min, resuspend in 300 uL MA2GC and spot 10 uL of cells onto cornstarch spots following the dilution guidelines below. Make all dilutions in the same media. Spot 10 uL of some chloramphenicol resistant cells on 9 of the spots around the plate to act as nurse cells. Allow spots to completely dry on your bench with the lid off. Single wrap plates in grafting tape and incubate right side up with light and CO2 for 10-14 days. Colonies should appear in 7 days. You want to have spots with just a few colonies on them to test by colony PCR.

Make your 1:9 dilution in the 50 mL conical that you spun your cells down in. Shake out the media with several quick wrist flicks in order to remove as much as possible. Resuspend cells in 300 uL media and then add another ~2300 uL so that the total volume is 2700 uL. This is now the 1:9 dilution. Pipet 120 uL of your 1:9 dilution/well across the top row of a 96 well plate (ok to use flat-bottomed plate). Pipet 133 uL MA2GC/well into the next 3 rows using blue 50-300 uL multichannel (in drawer just to the left of the fume hood).

Make serial dilutions (use the same set of tips for the media and for all of the dilutions):

67 uL from Row 1 -> Row 2 1:27

67 uL from Row 2 -> Row 3 1:81

67 uL from Row 3 -> Row 4 1:243

Spot on your 12 x 12 cornstarch grid using the gray multichannel pipettor. Start with the 1:243 dilution and use the same set of 12 tips for plating all the cells. Spot 24 x 1:9, 36 x 1:27, 36 x 1:81, 36 x 1:243. If you have used 0.5 pmol to transform, spot more of the less dilute cells and less of the more dilute cells (e.g. 36 x 1:9 and 24 x 1:243, or forget the 1:243 dilution all together). Singly spot your nurse cells on top of 9 of the transformation spots, as you will have covered all of the spots. Spot 4 x 1:9 control transformation on 1 plate.

You can grow any remaining cells in MA2G + Cp liquid in a well of a 6-well plate or in a tube in the roller drum as a backup.

*If you have transformed a plasmid you must keep the cells under constant selection in order to maintain the plasmid. 24 h after transformation spin down the cells 2000xg 10 min. (If you want to do microscopy on cells maybe add Tween-20 to a final concentration of 0.01%? Or try growing on a shaker, which seems to help disrupt clumps?) Resuspend cells in 1 mL MA2G + 150 ug/mL chloramphenicol and then dilute into 14 mL of the same media in a 25 mL graduated cylinder. Change the media every 3 days by spinning cells down, removing media and adding fresh media + Cp. After 3 days in selection raise Cp concentration to 250 ug/mL. After day 6 the control and experimental cultures should both look sickly yellow-green. By day 9 the experimental culture should be dark green, while the control culture should remain yellow-green. This also works well in a 6-well tissue culture dish with about 6-7 mL media. The cells settle to the bottom so that you don't have to spin them down to change the media - just pipette off the old media and add fresh, resuspending the cells in the new media.

Alternatively, instead of spinning cells down to add fresh Cp every 3 days, just add 150ug/mL fresh Cp to the culture every 3 days (e.g. if you count all of the Cp in the culture as active, then day 1 of selection = 150 ug/mL Cp, day 4 = 300 ug/mL, day 7 = 450 ug/mL - cells should be dark green by day 10, so at this point, I dilute into fresh 250 ug/mL Cp media) *** This is the easiest option and I have done it successfully**

The chloramphenicol stock is 34 mg/mL in 100% EtOH. (This = 227x for 150 ug/mL final, 136x for 250 ug/mL, or 85x for 400 ug/mL.)

0.77XMA2G + 250ug/mL chloramphenicol (Cp) plates

Autoclave 1g Gelzan (Cedarlane G024-1KG- this is made by Caisson Labs 71010-52-1; plant agar is cheap and also works – DuchefaBiochemie P1001) + 167 mL H20 in 500 mL Erlenmeyer flask with stir bar for 20 min. (Be careful not to splash gelzan up the sides of the flask as it doesn't melt very well.)

Cool on stir plate about 15-20 min.

While autoclaving make up 50mL sterile 3.3x MA2 and warm in the 42 °C water bath. pH will be ~3, so don't need to add any H2SO4 – if you use a different brand of gelling agent, you may need to adjust pH. See Kobayashi 2010.

16.7 mL Sol I

1.67 mL Sol II

0.167 mL Sol III

0.667 mL Sol IV

2.17 mL 37.5% (5 M) glycerol

28.6 mL H20

Once you can hold your hand on the glelzan flask, add the warm 50mL sterile 3.3x MA2G and stir briefly to mix. Add 1.6 mL chloramphenicol to 250ug/mL (stock is 34mg/mL in EtOH). Stir briefly to mix. Pour thick plates – 217mL makes 3 square plates. Mark with two blue stripes. Leave on bench top O/N, then wipe off any condensation and store upside down at 4C in a plastic plate sleeve. If you don't need 3 plates pour 1 plate before adding Cp, and then add 1 mL Cp to the remaining ~145 mL before pouring the two Cp plates. Or just pour 1 Cp plate using the liquid gelzan and pre-made concentrated media according to the directions on the bottle.

PCR screening for HR

Pick 24 colonies into 16 uL MA2G in a 96-well plate. Screen for HR by PCR with Taq. Add 200 uL MA2G + Cp to the remainder of your resuspended colonies. Remove one of the tips from the multichannel pipettor and pipet 100 uL water in between the wells surrounding your cell suspension (adds extra humidity to the plate so cultures evaporate less). Wrap plate in grafting tape and grow in incubator for ~3 days until you have a small clump of green cells at the bottom. Transfer the positive cultures to 2 mL MA2G + Cp in a 24-well plate and grow for ~4 days. Once your cultures are green isolate genomic DNA from 0.25 mL (using the Quick Edward's Genomic DNA prep) and retest by PCR. Grow the remaining 1.75 mL in 12 mL MA2G + Cp. Once the PCR tests confirm the strain, give it a CSR number based on the master glycerol stock document.

Make 9 glycerol stocks (1 mL OD750 2.5-3.0 + 80 uL DMSO/1.5 mL tube) and place in the correct stock box in the -80 freezer. Label a tissue culture flask and make a room temperature stock.

To test for multiple integration events, isolate genomic DNA using the Omega EZNA Plant DS DNA kit and then digest the genomic DNA with appropriate restriction enzymes and perform a Southern or test by quantitative PCR (see Fujiwara 2013 PLOS).

*MA2G = MA2 + 50mM glycerol

*MA2GC = conditioned MA2G (grow WT cells in MA2G to an OD750 1.0, spin 3000xg 15min to remove cells, filter through 0.2 um filter) - use to dilute cornstarch and cells for plating after transformation.

6.6. Appendix 6 – Colony PCR to test for homologous integrants protocol (updated by Martha Stark in August 2022)

Colony PCR to Test for Homologous Integrants (updated 22/08 by Martha Stark)

1. Make sure you have the appropriate plates available for re-spotting your positive colonies after screening.
2. Use NEB TmCalculator to determine the annealing temperature of your primers; G+H and I+J. Smaller product sizes will be more successful. Max size is 2000 bp, but this might not always work. Primers H and I should be in the targeted locus, outside the fragment that you transformed with. G and J should be in sequences unique to your transformation fragment that are not normally found at the genomic locus site of integration (i.e. selectable marker, it's promoter/terminator). If your product sizes allow, consider multiplexing your reactions – all 4 primers in one reaction.
3. Make master mixes (20uL x number of colonies + 5%) with all components except the colony suspension (1 uL colony suspension/20 uL PCR rxn). Use Taq polymerase for screening, 0.1 uL/20 uL rxn
4. Pick 24 colonies into 16 uL appropriate media in U-bottom 96 well sterile plate. Only pick colonies that you are sure are not touching any others. Stir and pipet up and down to disperse the cells.
5. Divide each master mix evenly across 12 wells of 96 well plate
6. Use gray multichannel pipettor to aliquot 19 uL/well of 96 well PCR plate. Use only one set of tips for each mix.
7. Use white multichannel pipettor to aliquot 1 uL/well of colony resuspension. Before pipetting, use tips to stir suspension. Visually check that you have liquid in each tip. You want to avoid sucking up any green clumps of cells.
8. Dispense into PCR plate containing the master mix and give a little stir to mix. Visually check your tips to make sure you have dispensed the samples.
9. Cover the PCR plate with a sheet of sticky film. Roll to seal. Briefly take the plate off ice and set on bench top. Using the back edge of the roller scrape the plate all over in both directions to make sure that it is well sealed, and then put back on ice.
10. Program the PCR machine, choosing the colony PCR program. Adjust the annealing temperature and the extension time according to the length of the expected product (1min/kb). If the products will be longer than 1000 bp, extend slightly longer than 1min/kb.
11. Pour a big agarose gel with two of the special clear 25-well combs.
12. Use one set of tips to add dye and load each row of your plate. Set gray multichannel to 17 uL. Put a small amount of 10x loading dye into the reagent reservoir. Suck up about 2 uL 10x loading dye into tips and add to first row of PCR plate. Pipet up and down to mix and then suck up 17 uL and load onto gel (will go into lanes 1, 3, 5, etc.; second row from PCR plate will go into lanes 2, 4, 6, etc. – remember this when you are labeling your gel image). Keep the same set of 12 tips on the pipettor for loading all your samples. Pipet up and down in the buffer at the bottom of the gel box in between samples to rinse the tips. Save the leftover loading dye in your own tube to use again – do not add it back to the 15 mL stock tube.

13. Load only 1 uL DNA ladder or it will be too bright!
14. Expect 30% homologous recombination frequency (~7/24; may be much lower if you are making an intron deletion). You need to have a minimum of two positives, and ideally, 3-4 (in case not all of them turn out to be correct), so that you end up with 2 biological replicates. Screen more colonies if you don't get enough in the first round.
15. Spot the remainder of the colony suspensions which are positive by colony PCR onto cornstarch spots on an appropriate plate (~14 uL). Allow to grow up into a solid green spot (usually 4-7 days). If you do not have time to run your PCR reactions on the same day you must spot all of your colony suspensions onto plates.
16. Pick the colony disk into 1 mL appropriate media, leaving a few cells on the plate in case you need to come back to them. If your disks are dark green the cells will all stick together, making this easy to do with a blue tip. Resuspend.
17. Transfer 750 uL into 5-10 mL media and grow to an OD750 2.5-3 to make glycerol stocks of your strains if you confirm that they are correctly integrated.
18. Isolate genomic DNA (quick Edward's prep) from the remaining 250 uL and retest by PCR (use Q5 if your products are >3 kb) using additional primers (especially I+H outside primers, and if you have made a deletion a primer that binds in the deleted region – should give no product) to confirm integration. Use 1 uL of a 1:10 dilution of your gDNA for PCR. Use 0.5 uL WT genomic DNA, diluted 1:400 (stock is already diluted 1:40) as a control for each of your primer pairs.

    *If you are testing a deletion strain with primers that bind in the gene you deleted (looking for the absence of product) and you are getting faint product, try using Taq polymerase, and you may need to set the annealing temperature several degrees above the recommended temperature. This is only something that I recommend doing if all of your other PCR tests are convincingly correct. Cm WT genomic DNA contamination seems to abound in the lab, making it difficult to show that your gene is deleted when using primers that can give a product from WT DNA.

<u>20 uL PCR rxn</u>

2 uL 10X Taq buffer

0.4 uL 10 mM dNTPs

0.4 uL 10 uM forward primer

0.4 uL 10 uM reverse primer

0.1 uL Taq polymerase NEB

1 uL colony resuspension

15.7 uL H2O

<u>PCR conditions</u>

1x 95 °C 5 min

35x 95 °C 30 sec; 51-59 °C 20 sec; 68 °C 20 sec-2 min (1 min/Kb)

1x 68 °C 5 min

Primer pairs

C+D 200-300bp no insert; 200-300 bp+marker size with insert

*G+H only get product if targeted insertion

*I+J only get product if targeted insertion

*H+I length of homology arms no insert; length of homology arms + marker with insert; WT locus should give a band of a different size

Tip use for screening 96 colonies

12 yellow tips for aliquoting 16 uL media in 96 well plate from reagent reservoir

96 yellow or white tips for picking colonies

1 yellow tip for aliquoting master mix across 1 row of 96 well plate

12 yellow tips for aliquoting 19 uL master mix into PCR plate

96 white tips for adding 1 uL colony resuspension to PCR plate

12 yellow tips for adding 2 uL loading dye + loading samples on gel.

6.7. Appendiz 7 – Quick *C. merolae* (Cm) genomic DNA isolation protocol

Quick Cm genomic DNA Isolation (Hu and Lagarias 2020 bioRxiv)

1. Spin down 1-2 ODU cells. Much less is also fine. As long as you have even a tiny green pellet you will get enough gDNA for PCR.
2. Resuspend in 200 uL Edwards Buffer*, vortex 5 sec. Use 100 uL if pellet is small.
3. Add 200 uL Isopropanol, mix by inversion. Use 100 uL if pellet is small.
4. Centrifuge 13,000 rpm, 5 min
5. Decant supernatant, invert tube on paper towel to air dry pellet < 2 min
6. Add 100 uL water (or 50 uL if pellet is small), resuspend with pipetman. A lot of the DNA won't resuspend right away, but enough will get into solution for PCR. If you allow this to sit on your bench O/N before freezing more of it will go into solution.
7. Centrifuge 13,000 rpm 1 min to sediment insoluble material
8. Dilute this 1:10, and use 1 uL for PCR (always include a WT control, also dilute WT g. DNA 1:10)
9. Use Taq for products < 3 kb, Q5 for larger products and extend >30 sec/kb
10. If you are seeing any WT bands in your samples that you were convinced were correctly deleted, try tweaking your PCR conditions (dilute the g DNA 1:100, increase the annealing temp, add GC enhancer, try Taq instead of Q5). Ultimately, you will need to confirm deletion by Southern.

*Always check your Edward's buffer b/f using to make sure that the SDS has not precipitated! If it has, place in the 42 °C H2O bath until it goes back into solution.

Edwards Buffer

200mM Tris-HCl, pH7.5

250mM NaCl

25mM EDTA

0.5% SDS

6.8. Appendix 8 – RNA isolation cold/hot phenol (Stark, 2023)

RNA Isolation – Cm cold/hot phenol (2023.07.06)

1. Spin down up to 2-10ml cells OD750 ~1.0 (1 OD unit = 0.5 x 10^7 cells).
2. Resuspend in 400 ul cold phenol lysis buffer (*see note below)
3. Sonicate setting 2, 5-10sec. Be careful of foaming. SDS will lyse the cells, sonication is to shear the DNA.
4. Acid phenol:chloroform extract 2x with equal volume – for hot phenol first extraction, heat at 65C for 5 min; chloroform extract 1x. *** *See note below for use of phase lock tubes.** If not using phase lock tubes, remove 300ul first time, 250ul second time. Use p200 yellow tip and stay well away from the white interphase. Be very conservative in how much aqueous phase you remove from phenol extractions: LESS IS MORE
5. Precipitate with 1ml cold EtOH
6. Spin max speed 30 min at 4 C
7. Wash with 70% EtOH
8. Briefly air dry the pellet and resuspend in 25-50 uL H_20. Dilute 1:10 before measuring the concentration by Nanodrop. Measure 3x and take an average the concentrations. Use the corrected value if the Nanodrop offers one. (If it is giving you a corrected value it is because you have phenol contamination in your sample – phenol absorbs at 230 nm – which means that you need to be more careful in your extractions). 2 ml cells OD750 1.0 should give ~30 ug relatively clean RNA. The $A_{260/280}$ and the $A_{260/230}$ should be 2.0 and 2.0-2.2, respectively.
9. Check 500 ng of the RNA on a 1.5% agarose/1% bleach gel to ensure that the RNA is not degraded (see separate protocol). You can use your remaining 1:10 diluted RNA for this. You should see two bands corresponding to the 28S rRNA (upper) and two bands corresponding to the 18S rRNA (lower). The top band of 28S rRNA should be twice as abundant as the top band of 18S.
10. Aliquot the RNA so as not to freeze/thaw the whole stock every time you want to use some.
11. If you need your RNA to be DNase-free, treat with the Turbo DNase kit. You do not need to do this for Northern analysis. I have successfully done this by adding the whole amount of DNase at once for 30 min, but it may be more effective to add half the amount for 30 min, and then the second half for another 30 min. See the manufacturer's protocol.

**Note: use of phase lock tubes (QuantBio 2302830 heavy 2mL, sold by VWR, or you can make your own in 2mL tubes by squirting in some Dow Corning High Vacuum Grease) makes this much easier and the yield is much greater. Do two sequential phenol extractions in one 2 mL phase lock tube, then one more phenol extraction in a regular Eppendorf to make sure all of the green has gone into the organic phase. Chloroform extract and precipitate in a single tube. 10 ODU/400 uL buffer may not work in phase lock (lock floats to top -homemade or stays at bottom – QuantBio). Try resuspending in 1 mL buffer and divide between two phase lock tubes. Or, if you have a lot of cells, you should consider doing the extractions in a 50 mL conical.

1. Transfer to phase lock tube that has been prespun 20 seconds at 15,000 x g

2. Spin 5min at RT, 15,000 x g

*Cold/hot phenol lysis buffer

200mM Tris-HCl, pH 7.5

500mM NaCl

10mM EDTA

1% SDS

6.9. Appendix 9 – RNA Bleach Gel (Stark, 2023)

Bleach gel (2023.06.16)

1. Add 1.0% w/v agarose to 1x TBE buffer
2. Add 1.0% v/v bleach and incubate at RT for 5 min with occasional swirling
3. Heat the suspension to melt the agarose
4. Allow to cool before adding EtBr to a final concentration of 0.5 ug/mL
5. Pour solution into gel tray and allow to solidify 30-40 min
6. Load 500 ng RNA mixed with 10x DNA loading buffer to 1x. Also load a DNA ladder.
7. Electrophorese in 1x TBE at 120V for 45 min. Run in buffer designated for bleach gels (This is kept separate from our DNA TBE running buffer so that we are not running our DNA gels in bleachy buffer). Take note of the temperature of your gel buffer. If the gel gets too hot you will compromise the integrity of your RNA and could short out the power supply. This buffer can be reused 5 times.
8. Image. Compare intensities of 28S and 18S rRNA bands. Quantify using image tool if necessary.

Note: The bleach serves to inhibit RNases and somewhat unfolds RNA secondary structure due to sodium hypochlorite. This is not a true denaturing gel so rRNAs will not run true to size (28S ~4.8 Kb/runs ~1.5 Kb, 18S ~1.8 Kb/runs ~0.8 Kb). Not sure why, but these run as doublets in Cm.

Note: Good quality RNA should have a 28s/18s ratio of 2 where the 28s is two-fold higher in intensity.

6.10. Appendix 10 – Fluorescent Northern blots (Stark, 2023)

RNA is detected indirectly with a biotinylated oligo (IDT) which is then bound by streptavidin conjugated to a near-infrared dye. In our hands a 5' biotinylated oligo bound to IRDye 800CW streptavidin (LiCor) detected at least 0.05 fmol of RNA. A 3' biotinylated oligo also detected 0.05 fmol of RNA, but the intensity of the signal can be 2-3x greater. The 5' bio oligos are cheaper and should be sufficient for most applications. (^{32}P-labeled probes have a lower detection limit of 0.005–0.01 fmol.) IRDye 680RD streptavidin is another fluorophore that could be used. Order 25 nmol scale 5' bio oligo (100 nmol for 3' bio oligo) and resuspend to 100 uM in H2O. /5biosg/ is the IDT prefix for 5' biotin; /3bio/ is appended at the end of the nt sequence for 3' biotin.

1. Oligo Design

Characteristic	Recommendation
Length	26–45 nucleotides
Tm value	78–90°C
GC content	47–62%

Check the Tm using OligoCalc, setting the salt concentration to 500 mM and using the salt-adjusted Tm. (http://biotools.nubic.northwestern.edu/OligoCalc.html). Nucleotide BLAST your primer against Cm to make sure that it is specific for your gene of interest.

2. RNA Isolation

Isolate RNA from Cm by the cold phenol method using phase-lock gel tubes (see separate protocol) or using a total RNA kit. These column-based kits have a size cut-off of about 200 nt. They can be modified to decrease the cut-off to about 50 nt (GenebioSystems kit: step 2 – replace RNA Binding Buffer with 100% EtOH; step 6 – replace RW1 Buffer with RWC Buffer = 1 mL RTL Buffer + 1.5 mL 100% EtOH; I haven't confirmed this experimentally). 2 OD units of cells should yield 15-30 ug of total RNA. Aliquot the RNA into several tubes so that it does not undergo many cycles of freeze/thaw. If you know that you will/might need to load a large amount (20 ug) of RNA then start with more cells. Check your RNA on a bleach gel to make sure that it is intact before using it.

3. Denaturing Gel
   a) Pour a 6-15% denaturing polyacrylamide gel for RNAs < 1000 nt. (For larger RNAs use a denaturing (formaldehyde) agarose gel – can be transferred on semi-dry blotter***).
   b) Pre-run the gel 15 minutes at 400 V in 1x TBE. Using a syringe with a bent needle blow out the urea from the wells and the trapped air from the bottom of the gel before starting.
   c) Mix equal volumes of RNA (1-10 ug) with 2x formamide loading buffer. It is best to keep the volume under 20 uL.
   d) Denature samples at 65C for 3 min immediately before loading on gel. Quick spin and place on ice.
   e) Again, blow out the urea from the wells of the gel with the syringe.
   f) Load the samples using an elongated gel-loading tip. Use 1 tip/gel, rinsing the tip in the TBE buffer in between samples.
   g) Run the gel at 400 V for 45 -90 min, depending on the size of your RNA.

% acrylamide	Xylene cyanol (nts)	Bromophenol blue (nts)
5	130	35
6	106	26
8	75	19
10	55	12
12	28	8

4. Membrane Transfer with Semi-Dry Electroblotter
   a) Cut 6 pieces of Whatman paper and 1 piece of Hybond+ nylon membrane (GE Healthcare, RPN303B) slightly larger than your gel (14.25 x 12.5 cm for full gel). If you know that your RNA is in the top half of the gel, for example, then only transfer this portion. If you have not loaded all of the lanes then adjust the width of the membrane and Whatman, so as not to waste the membrane.
   b) Label the top left back of the membrane with the date and an experiment identifier using a pencil. Cut off the top right corner.

c) Place 2 pieces of Whatman paper pre-wetted in 1x TBE on the semi-dry blotter, ensuring no bubbles are trapped underneath. Use roller if necessary (broken-off 10mL pipet). Add a little extra buffer.
d) Transfer gel from glass plate to dry Whatman by pressing Whatman down on gel and then peeling up. If the gel is 9% or greater it doesn't stick to the Whatman easily so you may need to turn the glass plate over and press the Whatman and gel together using the edge of the bench, peeling the gel/Whatman off the plate as you slide it along the bench edge.
e) Put the gel on the Whatman paper down next, gel side up. Lower slowly, allowing the Whatman to become wet as you lower it.
f) If there are bubbles under the gel, wet the gel with a little buffer, wet the roller, and gently roll or slide the bubbles out the edge of the gel.
g) Using forceps, pre-wet the membrane and place it on the gel, making sure there are no bubbles. Only touch the membrane with gloved fingers on two corners.
h) Cover with 3 more pre-wetted Whatman papers, making sure there are no bubbles.
i) Wipe up any puddles of buffer surrounding your gel stack.
j) Gently put the top of the blotter on and attach the leads. Do not use the screws.
k) Transfer 30-45 min at 2.5 mA/ cm^2.

5. Crosslink RNA to Membrane
   a) Using forceps, transfer the membrane, RNA side up, onto a piece of Whatman paper.
   b) Immediately cross-link the RNA to the membrane using the auto cross-link setting on the Stratalinker.

All of the following steps are done in hybridization tubes in the hybridization oven, with rotation.

6. Pre-hybridization
   a) Preheat ULTRAhyb™–Oligo Buffer (Invitrogen, AM8663) to 42-68C until completely resolubilized (42 °C H2O bath).
   b) Set hyb oven to 42 °C.
   c) Using forceps, place the blot(s) in a hybridization bottle, RNA facing in.
   d) Pour 5-10 ml ULTRAhyb on the blot, depending on whether your blot is half-size or full-size (or 2 half-size blots in one tube). It is fine for your blot(s) to overlap – just make sure that you allow buffer to completely wet the blot that is underneath.
   e) Incubate at 42 °C for 30 minutes.

   Note 1: ULTRAhyb appearance upon its arrival and during storage can vary from white, pink or yellow. The variation in buffer color does not affect the product performance or stability. The buffer remains stable with repeated heating to 68 °C, thus the entire contents of the bottle can be prewarmed to 68 °C before removing an appropriate amount for hybridization.

   Note 2: If you are probing 2 blots with the same probe, place them in one tube, minimizing the overlap. Add the prehyb and lift the overlapping blot slightly with forceps to allow the blot underneath to become saturated with buffer. It is much easier to put dry membranes in the tube now, rather than combining them later at the blocking step when they are wet.

Note 3: If you will probe your blot with a probe that is already in hyb buffer, pour off the prehyb into a 15 mL conical and save at 4 °C to reuse 1-2 more times when you have pre-made probe/hyb buffer.

Note 4: If you see nonspecific bands on your blot you can try increasing the hybridization temperature.

7. Hybridization
    a) Add 5 pmol/ml of biotinylated oligonucleotide directly into the pre-hybridization buffer. Do not add oligo onto the blot.
    b) Hybridize 2–24 hr at 42 °C. A 2-6 h hybridization will give about 40% of the maximum signal. 14 h is sufficient.
    c) Transfer the hyb solution into a 15 mL conical tube and store at 4C for future use.

    Note 4: Hyb solution can be reused multiple times (at least 4, maybe more). To use a probe for the second time prehyb the blot as usual, but save the prehyb at 4 °C to use again on another blot (can be used 3 times). Warm your used probe/hyb in the 42 °C H2O bath to solubilize before adding to your blot.

8. Washing (2x SSC, 0.5% SDS, stored at 37 °C)
    a) Quick rinse your blot with 5-10 mL of wash buffer to remove the majority of the unhybridized probe.
    b) Wash the blot with about 20-25 ml buffer for 30 min at 42 °C.
    c) Discard the washing buffer.

9. Blocking
    a) Add 5 mL blocking buffer (Licor, 927-70001) to the blot. If you have multiple smaller blots that were hybridized with different probes you can combine them into one tube.
    b) Incubate for 1 h at RT (leave the oven door open).

All of the following steps are carried out in the hyb oven IN THE DARK. Put the DARK ROOM sign on the door so that no one turns on the light!

10. IRDye 800CW Streptavidin Binding

    a) In the dark, add 0.5 uL Streptavidin-IRDye 800CW conjugate (Licor, 926-32230) into the blocking buffer (=1:10,000; if you have multiple tubes you can make a 1:100 dilution in blocking buffer and then add 50 uL/tube).

        Note: To prevent background, before adding dye, add 100 uL 10% Tween-20 (0.2% final concentration) and 50 uL 10% SDS (0.1% final concentration)

    b) Incubate 30 min at RT in the dark.
    c) Discard the buffer.

    Note 5: Divide the IRDye 800 stock into 25 uL aliquots in 1.5 mL Eppendorf tubes and store at 4C in the dark. Do not expose the dye to light!

11. Washing (1X PBS, 0.1% Tween-20 = PBST)
    a) Quick rinse your blot with 5-10 mL of wash buffer to remove the majority of the unbound dye.
    b) Wash with 20-25 mL wash buffer 3 x 5 min at RT

c) Wash 1 x 5 min with 1X PBS at RT (Removing the detergent will improve fluorescence).

12. Detection
    a) Place the blot on a piece of Whatman paper wet with water on an acetate sheet, cover with saran wrap, and put into the Bio-Rad Imager. If the blot dries, you will be unable to strip it!
    b) Choose the IR-Dye 800 CW setting and capture the image with auto exposure.
    c) You can manually choose the exposure time if you want to take a longer exposure.

13. Analysis
    Use Bio-Rad Image lab software 6.1 (BioRad; free version available upon request at https://www.bio-rad.com/) for measuring the levels of relative fluorescence units (RFU).

14. Stripping (0.2% SDS)
    a) If you may want to reprobe your blot with a different probe, strip it before the blot dries out.
    b) Heat about 50 mL buffer in the microwave to almost boiling.
    c) Incubate with blot in hybridization bottles for 10 min at RT, with rotation.
    d) Repeat.
    e) Rinse with about 50 mL 2x SSC.
    f) Rinse blot with H2O.
    g) Re-expose the blot on the Imager for at least as long as your original exposure time to make sure that you have removed all of the probe.
    h) Wrap the blot in saran wrap and store at RT (e.g. in your drawer) if you will not reprobe the same day.
    i) Stripping will remove a small amount of the RNA from the blot, but it is no problem to strip once, or even twice.

6% / 7M Urea Denaturing Polyacrylamide Gel (15 mL)

2.25 mL 40% acrylamide 19:1
6.3 g urea
750 µL 20x TBE
7.3 mL H_2O
150 µL 10% APS* (*Make fresh every month)
15 µL TEMED

Change the amount of acrylamide and water accordingly for 4.5%, 9%, 12%, 15% gel.

Unpolymerized acrylamide is a neurotoxin. Wear gloves and a lab coat.

1. Weigh urea into a 50 mL beaker
2. Add acrylamide using a 10 mL disposable pipette
3. Add dH2O with the same pipette. Keep a separate bottle of dH2O for this purpose only. Put the pipette in its cover and save for pouring the gel.
4. Add 20x TBE and a small stir bar. Put the beaker on the hotplate/Stirrer set to 100 °C, and mix until the urea is completely dissolved. Do not leave on the hotplate once dissolved.

5. Add 10% APS and stir with the pipet tip.
6. Add TEMED and stir with the pipet tip.
7. Pour the gel using the 10 mL disposable pipette.
8. Leave the remaining gel in the beaker to solidify. Once solidified, remove the stir bar and scoop out the solidified acrylamide into the trash.

***FOR DENATURING AGAROSE NORTHERN BLOT

Buffers:

<u>10X MOPS</u>

- 0.2 M MOPS (It is not the FREE ACID!)
- 50 mM Sodium Acetate
- 10 mM EDTA (disodium salt) – EDTA is hard to dissolve. It is better if you use the 0.5M EDTA pH 8.0 stock solution)
- pH 7.0 (if using MOPS sodium salt (not free acid) adjust pH with GLACIAL ACETIC ACID)

Weigh out chemicals and dissolve in the appropriate volume of milliQ H_2O. Adjust pH to 7.0. Measure final volume in graduated cylinder and adjust to desired final volume. Does not need to be autoclaved.

Note: Store at RT in the dark (cover with foil)!

<u>20X SSC (3 M NaCl, 0.3 M Na Citrate) - 1L</u>

- 175.3 g NaCl (FW 58.4)
- 88.2 g Na Citrate (citric acid trisodium salt - $2H_2O$, FW 294.1)

Weigh out chemicals and dissolve in ~0.75L milliQ H_2O. Adjust pH to 7.0 with a few drops of 1M HCl. Measure final volume in graduated cylinder and adjust to 1L. Does not need to be autoclaved.

<u>Running Buffer</u>

1X MOPS + 7% formaldehyde (1L)

- 100 mL 10X MOPS
- 20 mL 37% formaldehyde (stock)

<u>50X NAQ Agarose Northern Blot Transfer Buffer</u>

- 0.2M MOPS (It is not the FREE ACID!)
- 50 mM Sodium Acetate
- 5 mM EDTA (disodium salt) – EDTA is hard to dissolve. It is better if you use the 0.5 M EDTA pH 8.0 stock solution)
- pH 7.0 (if using MOPS sodium salt (not free acid) adjust pH with GLACIAL ACETIC ACID)

Weigh out chemicals and dissolve in the appropriate volume of milliQ H_2O. Adjust pH to 7.0. Measure final volume in graduated cylinder and adjust to desired final volume. Does not need to be autoclaved.

Note: Store at RT in the dark (cover with foil)!

1% Agarose Formaldehyde Gel

Note: Adjust the amount of agarose according to the desired percentage. Use the big tray and pour a 100mL gel.

- Weigh out 1 g agarose in 87 mL milliQ water.
- Weigh before and after melting the agarose in the microwave – you will lose water due to evaporation when boiling.
- Add water to bring to initial weight after melting.
- Cool down until able to touch.
- Add 10 mL 10X MOPS.
- Add 3 mL 37% formaldehyde.

IMPORTANT – Prepare and pour in fumehood!

1. Running the gel
    a. Once your gel has solidified, pour running buffer in the electrophoresis chamber and let the gel sit 15-30 minutes to equilibrate it.
    b. Prepare your samples in the meantime – sample preparation is the same as for denaturing polyacrylamide gel northern blots.
    c. Load your samples with your P20 and yellow tips.
    d. Run the gel at 125 V for 2.5 to 3h (depends on the size of your RNA) IN FUMEHOOD. Or 50 V for 6-8 h.

Note: In 1% agarose gels xylene cyanol migrates at a rate of a 4-5kb DNA fragment and bromophenol blue migrates at a rate of a 300 – 400 bp DNA fragment.

2. Transfer in Semi-dry Electroblotter
    a. While the gel is still on the tray after running, cut the portion in which your RNA of interest is.
    b. Rock your gel in a plastic box with transfer buffer 2x 10min for equilibration.

Note 1: This is important to wash out part of the formaldehyde and especially when your running and transfer buffers are different. E.g.: when your running buffer is 1X MOPS and your transfer buffer is 1X TBE.

Note 2: If you are using 50X NAQ transfer buffer it is ok to equilibrate your gel 1x for 10min.

    c. Cut 6 pieces of Whatmann Paper and 1 piece of Hybond Nylon membrane to the size of you gel.
    d. Set up the sandwich as stated above, pre-wetting the Whatman and the membrane with transfer buffer.

Note 3: Do not use a piece of dry Whatman to press on your gel to peel it off. Pick it up with your hands and place it upside down on top of the first 3 pre-wetted Whatman papers.

e. Transfer at 1.5 mA/cm2 of gel for 45 minutes (depends on the size of your RNA). This needs to be optimized!

3. Capillary transfer:
    a. Cut 3 pieces of Whatman 13.5 x 23 cm (wick).
    b. Cut 3-4 more pieces of Whatman to the size of your blot.
    c. Cut 1 piece of HyBond-N^+ membrane to the exact same size of your Whatman paper.
    d. Cut about a 3–4-inch stack of paper towels to same exact size as your gel, Whatman and membrane. This depends on the gel percentage and transfer time.
    e. Use an electrophoresis chamber as the solid support.
    f. Pre-wet the Whatman and membrane in 20X SSC (transfer buffer).
    g. Build your sandwich as follows: wicks, gel, membrane, 3x Whatman, stack of paper towels. No Whatman, membrane or paper towels should be overhanging the gel.
    h. Pour transfer buffer on both sides of the electrophoresis chamber so that there is enough buffer for completing the transfer but so that it does not reach the flat surface in which the wicks are sitting.
    i. Place a glass plate and a 500 g weigh on top of the paper towel stack.
    j. Wrap the assembly with plastic film and let transfer overnight.

Pre-hyb, hybridization, washing, blocking, IRDye binding, detection and stripping are performed as described above.

6.11. Appendix 11 - Western blots

Reagents

1) Nitrocellulose 0.45 um or PVDF Immuno-Blot 0.2 um
2) Wash Solution: 0.1% Tween 20 in 1X PBS (1X PBS-T).
- For 500mL of a 10X stock of PBS: 40 g NaCl + 1 g KCl + 7.2 g Na_2HPO_4 + 1.2 g KH_2PO_4 + milliQ dH_2O. pH to 7.4 with HCl. Bring volume up to 500mL. AUTOCLAVE
- For 50 mL of a 10% stock of Tween20: 5 mL of 100% Tween20 into a 50 mL conical (pour until you hit the 5mL mark), then top up to the 50mL mark with milliQ dH_2O. Mix completely.
- For 500mL of 1X PBS-T: 455mL milliQ dH_2O + 5mL of 10% Tween20 + 50mL of 10X PBS
3) Blocking Solution: 5% Milk in 1X PBS-T
- 2.5 g of milk powder per 50 mL of 1X PBS-T
4) 1° Antibody Solution: 1° diluted in 1X PBS-T –appropriate dilution determined by dot blot or manufacturer's guidelines: aFLAG 1:2500, aStrepII 1:1000
- (e.g. For 3 mL of 1:1500, use 2 µL 1° ab + 2998 µL 1X PBS-T)
5) Pre-Bleed Solution: Pre-Bleed Serum diluted in 1X PBS-T
- (dilute same as primary)
6) 2° Antibody Solution: 2° diluted in 1X PBS-T - appropriate dilution determined according to manufacturer's guidelines: Goat a Rabbit-HRP 1:5000, Goat a Rat-HRP 1:5000, Goat a Mouse-HRP 1:20,000
- (e.g. For 3mL of 1:5000, use 0.6 µL 2° ab + 3000 µL 1X PBS-T)
7) Final Wash Solution: 1X PBS
8) Luminol Reagent: 1:1 ratio of the two solutions, 0.125 mL/cm^2 of Nitrocellulose
9) Towbin Buffer (25 mM tris, 192 mM Glycine, 20% Methanol, pH 8.3)
- For 1L: 3.03 g Tris base + 14.4 g Glycine + 200 mL Methanol + 500 mL milliQ dH_2O
- Allow solids to dissolve and top up to 1 L with dH2O
- Check pH – should be between 8.1 - 8.5
- For high molecular weight proteins reduce the methanol to 10% and add SDS at 0.01%.
10) 6 Whatman papers and 1 nitrocellulose membrane cut to match the size of your resolving gel that contains samples (about 6 x 9 cm if you are transferring a full gel)
11) Prestained marker
12) Ponceau stain (0.1% w/v in 5% acetic acid)
13) Stripping buffer (For 100 mL)

Note: Make fresh every time or make a higher volume without the BME and add BME before each use.

Component	Volume (mL)
0.5 M Tris-HCl, pH 6.8	12.5
10% SDS	20
Beta-mercaptoethanol (BME)	0.8
Water	67.5

Protocol

1. Run an SDS-PAGE gel
   - *Resolving and stacking gel recipe below
   - Include a prestained marker and a positive control (i.e. the purified protein that was used to inject the rat/rabbit – you should know how much to load based on your dot blot – if you want to see it with Ponceau load >100ng)
   - Carefully cut off the stacking layer of the gel prior to transfer
2. While the gel is running
   - Cut 6 pieces of Whatman paper the same size as the SDS-PAGE
   - Wearing gloves, cut a single piece of nitrocellulose or PVDF. Keep it between the package sheets until the transfer. Use only forceps to directly touch the membrane. Use PVDF Immuno-Blot 0.2um membrane when transferring very small proteins (10-30kDa).
   - Using a pencil, write the date in the top left corner. This will be the back side of the membrane. Cut off the top right corner.
   - If using PVDF, wet the membrane briefly in 100% MeOH, then equilibrate 10min in Towbin Buffer. Also, equilibrate your gel in Towbin Buffer for 10min to remove some of the SDS, which interferes with protein binding the membrane. (SDS helps proteins come out of the gel, so for larger proteins we don't pre-soak the gels).
3. When the gel is finished running pour a small amount of Towbin buffer into the glass container and remove the lid of the Owl Semi-dry Transfer Apparatus.
4. Using forceps, dip a piece of Whatman paper in the Towbin buffer, and lay down in the center of the Owl Transfer apparatus, making sure there are no bubbles underneath it. Repeat 2 more times so that there are 3 wet Whatman papers on top of each other, as per Fig. 1.
5. Remove the glass plate covering the SDS-PAGE gel. Use one of the grey plastic spacers to cut/remove the stacking layer.

6. Carefully lay the SDS-PAGE gel onto the Whatman paper so that it is perfectly aligned. Make sure there are no air bubbles trapped underneath. If necessary, wet the plastic roller and use to roll out any bubbles.
7. Using forceps, carefully remove the nitrocellulose from the package paper, and dip it in the Towbin buffer to wet. Place it on top of the gel by aligning the bottom of the nitrocellulose with the bottom of the gel and then lowering it down towards the top of the gel. Have the date facing up and at the top of the gel. Roll out any trapped bubbles.
8. Wet and place each of the 3 remaining Whatman papers on top of the nitrocellulose, avoiding air bubbles.
9. If there are 2 gel sandwiches, place them side by side and centered, so that the total area is minimized
10. Using paper towel, carefully wipe up all of the excess pooled liquid
11. Place the lid down carefully, making sure it's aligned first (do NOT move or shift afterwards)
12. Connect the positive and negative leads to the Owl and the power source
13. Turn on the power source, and set the power and time
    - 0.8 – 2.2 mA per cm^2 of gel
    - 30 min to 2 hour, depending upon protein size and gel percentage
    - smaller proteins transfer faster than larger proteins – time and current have to be determined experimentally. A good place to start is 0.8mA/cm2 for 1h. For small proteins (<20kDa) have had success with 1.5mA/cm2 for 30-40min. Make sure that the Volts don't exceed 15 during your transfer.
14. Press the "On/Off" button and hold for several seconds. If the power supply shows an error, try placing a water bottle on top of the Owl to help complete the circuit, and press the On button again
15. When the transfer is complete, use forceps to remove the Whatman paper until you reach the Nitrocellulose.
16. If you peel up the edge of the nitrocellulose and see that your marker has not completely transferred, you can reassemble the top of the sandwich and transfer longer, assuming that you have not disturbed the alignment of the nitrocellulose on the gel.

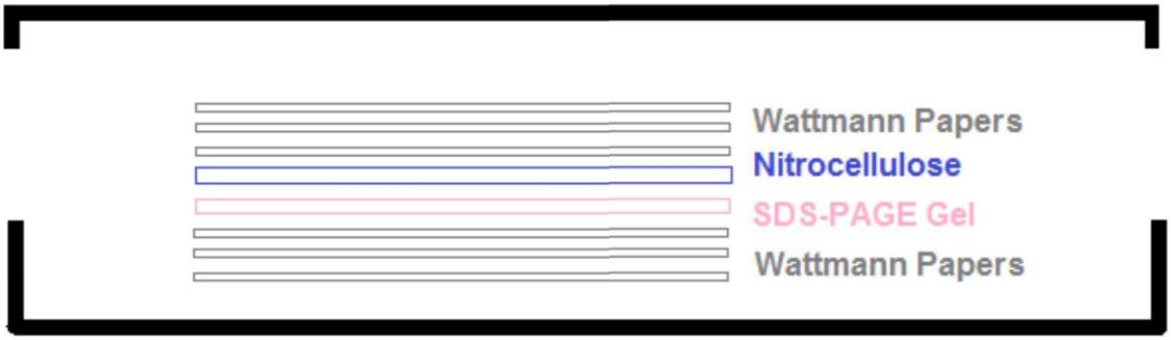


Figure 1. Set-up for Transfer of protein sample from SDS-PAGE to Nitrocellulose for Western Blot.

17. Place nitrocellulose, protein side up, in small plastic box
18. Add 10-15mL Ponceau stain and rock 1-5min, then rinse repeatedly with H2O until background disappears and red protein bands stand out. This step is optional and only works if you have enough protein on the blot (100ng/band minimum detection) **Don't need to do this for most Westerns**
19. Take a picture on the gel doc using reflective white light. Do this so that you have a good picture of the prestained marker (it will fade during the following steps). You can alternatively use your phone, or mark with a pencil next to each of the marker bands
20. Add approx. 25-50 mL of blocking buffer (5% milk in 1X PBS-T), and rock at room temperature (RT) for 1 hour or on rocker in cold room overnight
21. Briefly rinse the nitrocellulose using 2 changes of PBST (~25mL)
22. Place the Nitrocellulose in a sealable bag and seal the edges so that they are close to the membrane. Leave one edge unsealed.
23. Add 2-4 mL of the 1° antibody solution, depending on the size of blot. Seal the bag, making sure to remove all bubbles. Incubate on rocker at RT for 1 hour, or on rocker in cold room overnight
24. Remove Nitrocellulose from sealable bag and put back into plastic container. Save the 1° antibody solution for later Western blots (add NaN3 to 0.01%).
25. Wash 3 x 5min PBST
26. If your secondary antibody will be used at a low dilution, place the Nitrocellulose in a sealable bag, and seal 3 of the edges. Otherwise, place in small box (e.g. if 1:10000 or greater)
27. Add 2-4 mL of the 2° antibody solution. Seal the bag. Incubate on rocker at RT for 1 hour, or in cold room overnight
28. Wash as before
29. Wash 1x5min with PBS (no Tween)

30. Take a piece of acetate sheet (in red box on top shelf, middle bench), and clean both sides very well with 70% EtOH, DO NOT TOUCH WITH GLOVES
31. Remove nitrocellulose from the wash and drain excess buffer by blotting edge of membrane on paper towel. Lay the Nitrocellulose protein-side up on acetate sheet
32. Take out Luminol Reagents.
    - The reagents (A and B) are added to each other 1:1
    - Mix A+B so that you have 0.125 mL/cm^2 of membrane. This is actually more than you need. 1.5ml for half a gel and 3mL for a whole gel is enough.
    - As soon as the reagents are mixed, use the 1000 µL pipette and immediately pipet evenly over entire sample. As soon as the mixture has been applied to the Nitrocellulose, time 1 minute.
    - Return Luminol to fridge.
33. After the 1-minute detection, pick up the Nitrocellulose using forceps and blot excess liquid off of it by touching the edge of membrane to a paper towel. Place the Nitrocellulose PROTEIN-SIDE DOWN on a dry part of the acetate and clean up all excess/pooled liquid.
34. Detect adjusting the exposure in Bio-Rad ChemiDoc MP imager.
35. Strip if needed.
    a. For Stripping:
        i. Rinse the membrane in water.
        ii. Warm up stripping buffer to 50 °C.
        iii. Incubate membrane protein-side up in stripping buffer with gentle agitation IN FUMEHOOD for 30 – 45 minutes.
        iv. Wash membrane 6X (5 min) with agitation with PBS-T.
        v. Re-block membrane before incubating with primary antibody.

6.12. Appendix 12 – RT-qPCR Protocol (Geertz and Stark, 2023)

RT-qPCR Protocol (23/08/09)

Initial Considerations

- Avoid pipetting less than 5 µL

1. Design primers

Using NCBI primer blast, design primers that are between 70 – 200 bp. The lower the amplicon length the higher the efficiency of the reaction will be (as it does not need to extend as long).

a. Pick an annealing temperature between 58 – 61 °C. Lower temperatures result in non-specific binding while higher temperatures will decrease the amount of product formed.
b. Ensure the GC content is between 50 – 60 %.
c. Avoid any primer pairs that include runs of greater than 4 As or Ts and runs of more than 3 Gs or Cs. Avoid dinucleotide repeats such as ATATATAT. This reduces the amount of non-specific binding. However, primer pairs with GC clamps (G or C at the 3' end of primer) are favourable since they bind tightly.
d. Check for primer dimers using Oligoanalyzer on IDT. Check for hairpin structures and self-dimers. The melt temperatures for hairpin structures should be roughly 20 °C lower than the primer annealing temperature (ex. If the annealing temperature is 58, the hairpin structure melt temperatures should be no higher than 38 °C). The self-dimer ΔG should be no larger than – 6 kj/mol. The hetero dimer can also determine if primer pairs will bind to each other. The ΔG should be also no larger than – 6 kj/mol.

2. Checking for primer specificity

Run the primers on PCR to ensure that there are no primer dimers or any non-specific binding. The extension time for PCR will depend on the amplicon length.

a. For a primer pair that has a predicted melt temperature at 58 °C, extending a 150 nt amplicon, using NEB Taq polymerase, a master mix should look like:

Components	Volume (µL)
Buffer (10X)	2
dNTPs (10 mM)	0.4
Forward primer (10 µM)	0.4
Reverse primer (10 µM)	0.4
gDNA (10 ng)	1
Taq polymerase	0.1
Water	15.7
Total	20

b. For the same master mix a PCR reaction should be prepared as:

Step Number	Temperature (°C)	Time (sec)
1	95	30
2	95	20
3	58	30
4	68	15
5	Return to Step 2 (34 X)	
6	68	30
7	12	Infinity

The product can then be visualized on a 2 % agarose gel at 120 V for 60 minutes. A single band should be observed under the gel doc.

3. Assess the RNA quality using a bleach gel

A 1 % bleach, 1.5 % agarose gel with 1X TBE or TAE should be used (for example, a 50 mL gel would include 49.5 mL of TBE, 0.5 mL of bleach and 0.75 g of agarose). Bleach will denature the RNA to fully extend the strands. This step should be performed after RNA extraction.

a. Use 250 – 500 ng of RNA for each well and add 10X loading dye to each RNA sample (Use 10 µL of RNA and 1 µL of loading dye). Include a 100 bp and 1000 bp ladder.
b. Run the gel for 60 minutes on 120 v. The 28s rRNA should be twice as intense as the 18s rRNA. Smearing of the 28s rRNA will indicate poor RNA quality (no upper band – just a smear) Two clear bands should be seen. Quantify the RNA using the Nanodrop.

4. Removing gDNA contamination

To remove any gDNA in the RNA sample, RNA samples should be treated with the Turbo DNase kit. This ensures that the only DNA you will eventually quantify through qPCR represents only the RNA that you extracted, not any gDNA that may have been present.

a. Use 10 µg of RNA with only half of the DNase and incubate for 30 minutes and then add the second half of DNase and incubate for an additional 30 minutes (ex. 0.5 µL of DNase each time). This follows the rigorous DNase treatment with ONLY 1 µL of DNase.
b. Once RNA has been DNase treated, perform a PCR reaction with at least one primer pair and run with 5 and 40 ng of RNA to ensure no gDNA contamination remains. Use gDNA as a positive control (40 ng). If any band appears for the RNA samples, DNase treatment must be performed again. Quantify the RNA samples using the Nanodrop and ensure that the A260/A280 ratio is no lower than 1.80 (should be at 2.00 – indicating pure RNA) and that the A260/A230 ratio is also no lower than 1.80 (indicating no inhibitors such as Phenol in the sample).

5. Reverse transcription of RNA

Reverse transcription converts the RNA to cDNA. The BioRad iScript RT Supermix is an efficient enzyme. This reverse transcriptase (RT) uses a blend of both oligo dTs (mRNA specific) and random hexamers (targets total RNA). However, if the RNA sample contains secondary structures, it may be more beneficial to use an RT that works at a higher temperature such as the BioRad Reliance Select RT. This ensures that all the RNA is denatured before being reverse transcribed.

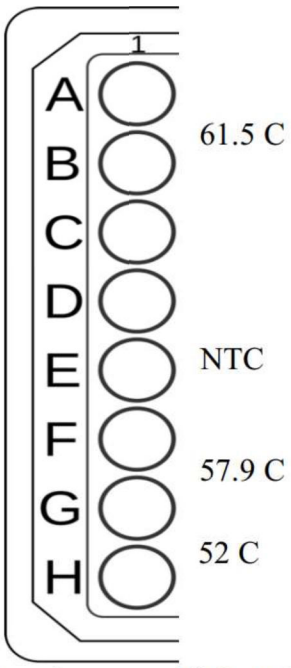
a. To make sure that cDNA will be synthesized correctly, use 1 µg of RNA and perform one reaction on the thermal cycler.
b. Once reverse transcribed, a single PCR reaction should be performed to ensure that cDNA product was formed with at least one primer pair (use gDNA as a control). If you are using 1 µg of RNA and have created a 20 µL sample, assume a 1:1 reverse transcription from RNA to cDNA. As a result, a 50 ng/µL sample should be created. If the product has been confirmed, reverse transcribe enough cDNA to last all remaining optimization steps and pool the samples together.
c. When testing experimental cDNA, a NRT control (No reverse transcriptase control) should be included according to the MIQE guidelines. For this, perform the exact same reaction with a new RNA sample, however, replace the enzyme mix with water. Create enough NRT sample to complete the experimental tests.

6. Thermal gradient

To determine the optimal annealing temperature of the primers, a thermal gradient must be set up. This ensures that primers are binding optimally to the cDNA.

a. Test each primer across 7 different temperatures on the thermal cycler (T100A or T100B), ranging 10 °C. Use a range so that the predicted annealing temperature is in the middle or closer to the higher end of the range. For example, if the predicted melt temperature is 58 °C, choose a range from 52 °C to 62 °C. Ensure that at least one well has the predicted annealing temperature. Temperatures range vertically so put the samples in from top to bottom.
b. Add 5 ng of cDNA to each well and include a NTC well (replace the cDNA with water). Include the NTC well close to the predicted annealing temperature well so that it is a more accurate representation of what might happen during experimental conditions.

For example, if the predicted melt temperature is 58 °C, a possible plate layout for a single primer on an 8 well plate may look like:



Set up the PCR reaction the same as the primer specificity PCR reaction, except include the thermal gradient instead of a fixed annealing temperature.

c. Run the products on a gel using 120 V for 60 minutes. Choose the temperature that results in the most intense single band. If there are multiple products, this is indicative of either primer dimers or non-specific binding. Primer dimers will appear at roughly 30-50 bp. If primer dimers are visible but very faint, it is ok to continue testing them on the standard curve. However, if possible, try another primer pair before testing on the standard curve. If multiple primer pairs were run together, choose the temperature that produces the most intense single band for all primer pairs without non-specific binding or primer dimers.

7. Standard curve

The standard curve is used to test the efficiency of the amplicon doubling. A 100 % efficiency would indicate that the amplicon is perfectly doubling every cycle. A lower efficiency would indicate the presence of inhibitors, secondary structures, or degraded cDNA, while a higher efficiency can indicate primer dimers but can also be a result of high Cq values (low cDNA amounts).

a. To perform a standard curve, at least 5 concentrations of cDNA are needed. Create a 1:5 serial dilution of the cDNA so that for the standard curve, 1/5, 1/25, 1/125, 1/625 and 1/3125 cDNA samples are available. If the starting assumed concentration of cDNA is 50 ng/µL, then a 10 – 0.016 ng/µL samples will be used for the standard curve.
b. For each primer pair, technical triplicates should be performed for each dilution factor. Include a NTC in technical triplicate for each primer pair.
c. A typical reaction for qPCR might look like:

Components	Vol (µL)
LUNA Master Mix (2X)	10
Forward primer (10 µM)	0.5
Reverse primer (10 µM)	0.5
cDNA	8
Water	1
Total	20

A NTC reaction for qPCR will look like:

Components	Vol (µL)
LUNA Master Mix (2X)	10
Forward primer (10 µM)	0.5
Reverse primer (10 µM)	0.5
Water	9
Total	20

A 12:8 ratio of master mix to cDNA should be used to reduce variation between technical triplicates and to ensure that the same pipet is used.

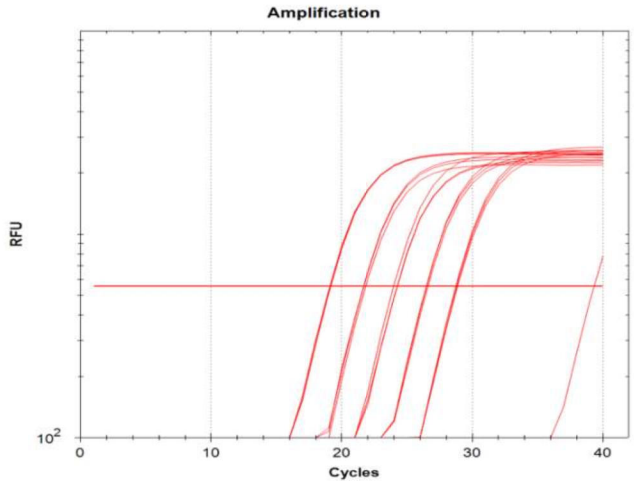
A typical reaction for a 150-nucleotide amplicon may look like:

Step Number	Temperature (°C)	Time (sec)
1	95	45
2	95	15
3	58	30
4	Return to Step 2 (39 X)	
5	68	5
6	95	0.5

d. When loading the plate, make sure to run the dilution series vertically so that higher concentrations are at the top of the plate and the lowest are at the bottom. Run the results on CFX96 and upload the results to CFX Maestro. Once setting up the plate on the software, a standard curve will be generated. The efficiency should fall between 90 – 110 % according to the MIQE guidelines while the R^2 value should be greater than 0.985. The Cq standard deviation between technical triplicates should ideally be less than 0.2, however, some labs consider less than 0.4 acceptable. Removing the highest or lowest concentration well can increase the efficiency of the reaction.

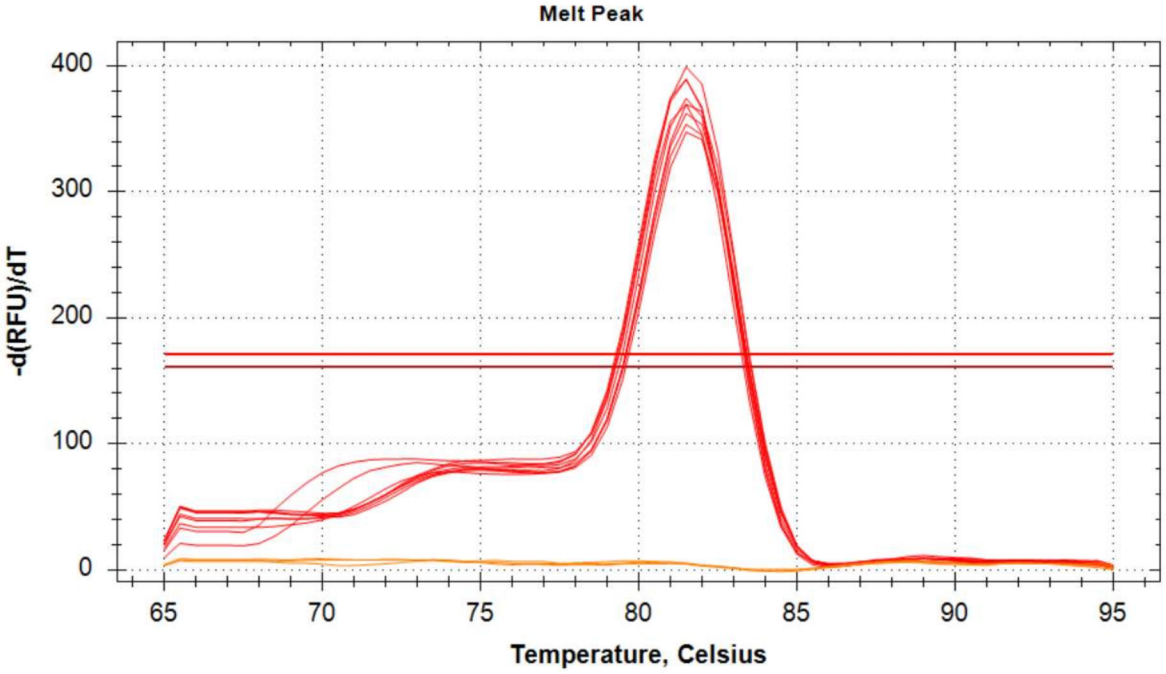
e. Choose the concentration that falls in the middle of the standard curve for further experimentation. As a result, if the experimental conditions change the cDNA amounts considerably, they will still fall within the standard curve Cq values. Cq values that are larger than 32 and smaller than 14 should be disregarded. Large Cq values indicate very low quantities of cDNA and can be misinterpreted for background noise.
f. NTC should not produce any signal. However, some primer dimers may produce a small signal. If a NTC produces a signal, Cq values should be larger than 10 Cq values away from the experimental Cq values. For example, if the cDNA concentration at the middle of the standard curve has a Cq of 25, the NTC Cq value should be 35 or greater.

A standard curve will look like:



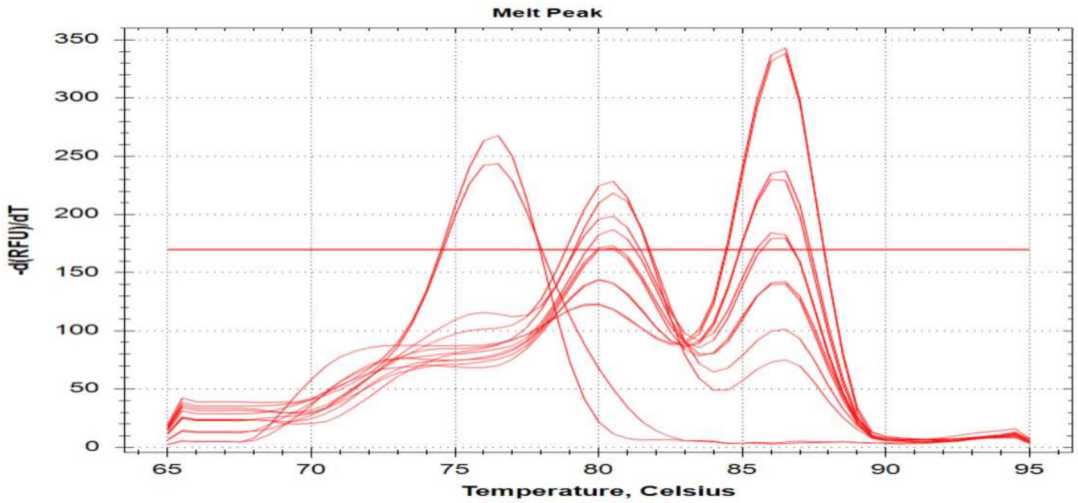
Here, a single NTC sample was amplified but it is far enough away from the cDNA samples that it can disregarded.

g. Check the melt curve. This is created in CFX Maestro. The melt curve occurs at the end of the qPCR reaction and will show if a single product is being formed. One peak should be created which confirms one product is being amplified.



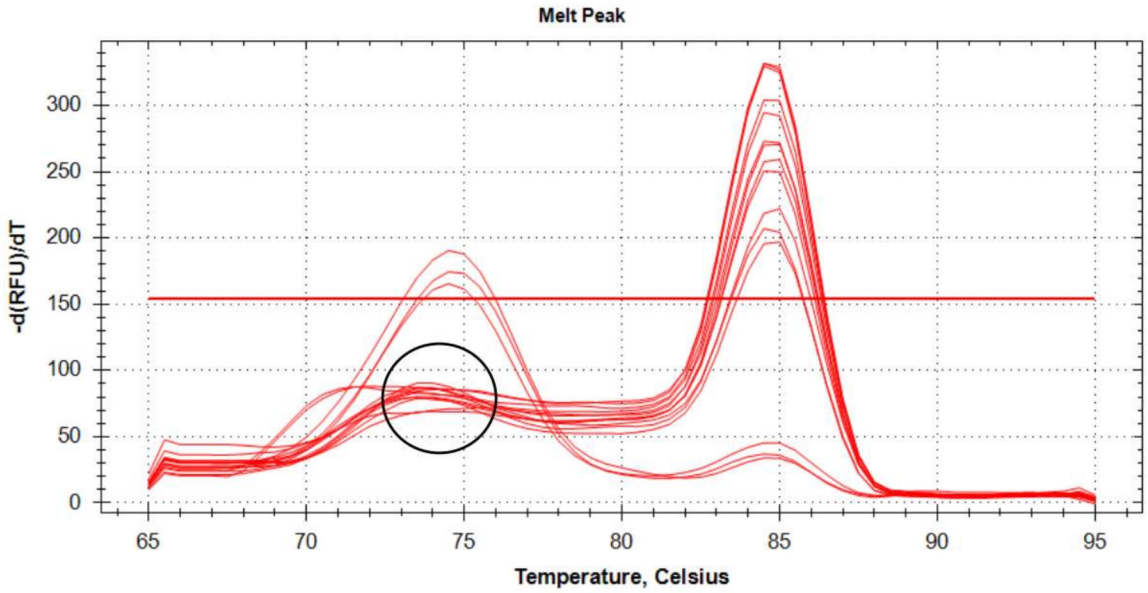
The orange signal is the NTC. Since there is no product being formed, no peak is created. The melt curve measures the negative slope of the relative fluorescent expression against the temperature (°C). As the product becomes denatured, the fluorescent signals decrease which results in the peak. Once there is no signal read, the signal plateaus at the bottom.

If there is non-specific binding a melt curve may look like:



Here a single cDNA sample produces two different amplicons at the same time.

If there are primer dimers, a melt curve may look like:



The NTC samples are amplified at a temperature much lower than the amplicon temperature and produced a shorter peak. This is only a concern if the Cq values of the NTC wells are less than 10 Cq values away from the experimental Cq values or if cDNA samples also begin producing a peak under the primer dimer peak as shown in circle.

8. Experimental data

Running experiments should be performed with at least 3 biological replicates. Each biological replicate should have technical triplicates, meaning a total of at least nine wells for each cell condition. For each experimental condition, a NRT and NTC should be included in triplicate. Biological replicates of NRT can be pooled together so that each cDNA condition only has 3 wells for NRT instead of 9.

a. Each primer pair has a different efficiency. As a result, primer pair data cannot be compared to each other. For example, if the efficiency for one primer pair is 90 while another is 110, the second primer pair will produce a much larger signal than the first, even though this might not be reflective of the amount of cDNA in the well.
b. Between plates, there is variation due to the light, plate or even the machine. As a result, if a single primer pair is being used, try to contain all cDNA samples to one plate. However, if this is not possible, an inter-plate calibrator must be used. This is a cDNA condition that is repeated on multiple plates. During analysis, this sample will be used to normalize the results between plates.

9. Reference genes

Reference genes are used to normalize the results to a single set of values. Since qPCR is used to measure the difference in expression or relative quantity of cDNA between conditions, it is important to standardize the results to reference genes. Reference genes are also important for reducing any variation between plates since all data is compared to the reference genes. Reference genes should be stable – meaning that their expression does not change between

different experimental conditions. According to the MIQE guidelines at least two reference genes should be used.

a. Test the reference genes experimentally like all other genes, including a NRT and NTC.
b. To test the stability of reference genes, multiple different software and techniques must be used including geNorm and T tests or ANOVA on the RNA expression levels.
c. Once reference genes have been created and analyzed for their stability, the same reference gene plate can be used continuously for all other plates and even different cDNA samples. However, before performing experimental reactions, enough cDNA should be created for all reactions.