U5 snRNA- specific proteins studies, RNA-protein interaction, assembly of the spliceosomal Sm complex, and function implications of splicing in *Cyanidioschyzon merolae*

by

Fatimat Almentina Ramos Shidi

B.Sc., Federal University of Alfenas, Brazil, 2015

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

UNIVERSITY OF NORTHERN BRITISH COLUMBIA

September 2018

Abstract

Splicing is an interesting step in the processing of the precursor messenger RNA (pre-mRNA) that involves removal of the non-coding sequences (introns) and ligation of the coding sequences (exons). Fifty percent of genetic diseases exert their effects through errors in splicing (López-Bigas et al. 2005). Therefore, a better understanding of this process can ease the development of cures for these diseases through genetic therapy. I proposed to investigate splicing in *Cyanidioschyzon merolae* that possesses a simpler spliceosome comprised of four snRNPs and 68 splicing proteins. I present successful expression of the Dib1 protein, and co-expression and purification of the Sm complex. I was able to prove formation of the ring-shaped Sm complex by electron microscopy analysis and binding of the complex to U2, U4, and U5 snRNAs. This work also initiated an investigation of splicing as a vital process for *C. merolae* by blockage of this mRNA maturation step with morpholino oligonucleotides.

Table of Contents

Abstract i
Table of Contents iii
List of Tables v
List of Figures vii
Acknowledgment x
1. Chapter One - Introduction 1
1.1 Processing of precursor messenger RNA: Splicing 1
1.2 *Cyanidioschyzon merolae:* a suitable model organism for splicing studies 4
1.3 U5 snRNP complex 5
1.4 General thesis objectives 16
2. Chapter Two - U5 snRNA reconstitution 19
2.1 Introduction 19
2.2 Materials and Methods 24
2.2.1 Preparation of *C. merolae* genomic DNA 24
2.2.2 Construction of expression vectors for ligation independent cloning 24
2.2.3 Amplification of the genomic sequences of U5 snRNA-specific proteins by polymerase chain reaction (PCR) 27
2.2.4 Insertion of the protein genes into the vector by LIC 31
2.2.5 Sequencing of amplified genes 33
2.2.6 Construction of a vector for co-expression of the U5-specific proteins 33
2.2.7 Expression of the U5-specific proteins 34
2.3 Results 37
2.3.1 Construction of expression vectors for ligation independent cloning 37
2.3.2 Amplification of the genomic sequences of U5 snRNA-specific proteins by polymerase chain reaction (PCR) 38
2.3.3 Insertion of the protein genes into the vectors 40
2.3.4 Construction of vectors for co-expression of the U5-specific proteins 42
2.3.5 Expression and solubility tests 48
2.4 Discussion 54
3. Chapter Three - Structural and functional studies of C. merolae Sm complex 59

3.1 Introduction 59
3.2 Materials and Methods 62
3.2.1 Two-step purification of recombinantly expressed Sm complex 62
3.2.2 Characterization of the purified Sm complex by Mass Spectrometry 63
3.2.3 Biophysical characterisation of the purified Sm complex by Electron Microscopy 64
3.2.4 Binding assays 64
3.3 Results 66
3.3.1 Two-step purification of recombinantly expressed Sm complex 66
3.3.2 Characterization of purified Sm complex by Mass spectrometry 68
3.3.3 Biophysical characterisation of the purified Sm complex by Electron Microscopy 69
3.3.4 Binding Assays 70
3.4 Discussion 85
4. Chapter Four - An investigation of splicing relevance and 5' splice site recognition in *Cyanidioschyzon merolae* 89
4.1 Introduction 89
4.2 Materials and Methods 94
4.2.1 C. merolae cell growth for assessment of doubling time 94
4.2.2 Treatment of cells with MO and vivo-MO 94
4.2.3 Delivery of MO by electroporation 97
4.3 Results 98
4.3.1 Assessment of *C. merolae* cell growth 98
4.3.2 Treatment of *C. merolae* cells with MO and vivo-MO 101
4.3.3 Delivery of MO by electroporation 110
4.4 Discussion 111
5. Chapter Five - Concluding remarks 114
References cited 116
Appendix 1 123
Appendix 2 155

List of Tables

Table 1 DNA oligonucleotide sequence of the primers used for amplification of protein genes. 30

Table 2 Thermocycler set-up for amplification of protein genes. 31

Table 3 Presentation of the size of the protein genes in *C. merolae*..... 39

Table 4 Construction of expression vectors. 41

Table 5 Construction of expression vectors containing U5-specific protein genes. 47

Table 6 Presentation of the molecular weight of the proteins in *C. merolae*..... 48

Table 7 Presentation of the CAI and CDF calculated by GenScript, based on the DNA sequence of the *C. merolae* proteins..... 58

Table 8 Presentation of the data collected after mass spectrometry analysis. 69

Table 9 Assessment of binding of U2 snRNA to the Sm complex by filter binding..... 71

Table 10 Assessment of binding of U4 snRNA to the Sm complex by filter binding..... 71

Table 11 Assessment of binding of U5 snRNA to the Sm complex by filter binding..... 72

Table 12 Assessment of binding of U4 snRNA to the Snu13 by filter binding..... 73

Table 13 Assessment of binding of U4 snRNA to the *S. cerevisiae* Nph2 by filter binding..... 73

Table 14 Comparison between the binding of folded and refolded U2 snRNA to the *C. merolae* Sm complex by filter binding. 74

Table 15 Comparison between the binding of folded and refolded U4 snRNA to the *C. merolae* Sm complex by filter binding. 74

Table 16 Assessment of binding of U2 snRNA to the *C. merolae* Sm complex by filter binding. 75

Table 17 Assessment of binding of U4 snRNA to the *C. merolae* Sm complex by filter binding 75

Table 18 Assessment of binding of U4 snRNA to the *C. merolae* Sm complex by filter binding. 76

Table 19 Binding parameters: A fluorescent oligonucleotide (ro64) was used as a negative control when performing FP. 88

Table 20 DNA oligonucleotide sequence of the MO and vivo-MO designed for binding to U2 snRNA: DNA sequences are shown from 5` to 3`..... 97

Table 21 Summary of electroporation conditions..... 98

Table 22 Summary of the doubling times of control and MO treated *C. merolae* cells..... 110

Table 23 Oligonucleotide sequences of the primers used for sequencing of Prp8, Brr2 and Snu114 genes. 123

Table 24 Summary of the expression of the proteins using different constructs..... 155

List of Figures

Figure 1-1 Illustration of the 2 SN2 transesterification reactions. 3

Figure 1-2 Proposed binding of U5 snRNA to the 5`splice site. 5

Figure 1-3 Comparison of the predicted *C. merolae* U5 snRNA to *S. cerevisiae* U5 snRNA. 7

Figure 1-4 Crystal structure and architecture of Prp8 in *S. cerevisiae*. 8

Figure 1-5 Crystal structure and architecture of *S. cerevisiae*`s Brr2. 9

Figure 1-6 Crystal structure of the *S. cerevisiae* Snu114 10

Figure 1-7 Crystal structure of the human Dib1 11

Figure 1-8 Comparison of the structure of the Sm complex and Sm-like complex. 12

Figure 1-9 Structure-based sequence alignment of the *C. merolae* Sm proteins. 13

Figure 1-10 S. cerevisiae U1 snRNP Core formation. 14

Figure 1-11 Predicted sequence of the Sm sites in *C. merolae* U2, U4 and U5 snRNA. 15

Figure 2-1 T4 DNA polymerase reaction. 21

Figure 2-2 Construction of expression vectors for co-expression of proteins. 22

Figure 2-3 Vectors construction for insertion of protein genes by LIC. 26

Figure 2-4 Complementarity of amplified genes to the modified vectors. 29

Figure 2-5 Construction of expression vectors oligonucleotides. 38

Figure 2-6 Amplification of protein genes. 39

Figure 2-7 Assessment of insertion of protein genes into expression vectors. 41

Figure 2-8 Construction of the U5-specific proteins co-expression vector. 43

Figure 2-9 First step of construction of Sms-containing vector 43

Figure 2-10 Second and third steps of construction of Sms-containing vector. 44

Figure 2-11 Last step of construction of Sms-containing vector. 45

Figure 2-12 Step-wise construction of the vector containing all U5-specific protein genes. 47

Figure 2-13 IPTG induction of the U5-specific proteins. 49

Figure 2-14 Induction of expression of Prp8 in Rosetta pLysS. 50

Figure 2-15 Expression and Purification of Snu114 fused to MBP 51

Figure 2-16 Induction of expression of Brr2 using two expression systems. 52

Figure 2-17 Induction of expression of Dib1 by IPTG induction and auto-induction 53

Figure 2-18 Co-expression and purification of the Sm proteins. 54

Figure 3-1 Purification of the recombinantly co-expressed Sm complex 68

Figure 3-2 Electron microscope of the Sm complex. 70

Figure 3-3 Investigation of U5, U4 and U2 snRNA stability. 73

Figure 3-4 Assessment of binding of U4 snRNA to the Sm complex by filter binding 77

Figure 3-5 Assessment of binding of U4 snRNA to the Sm complex by filter binding. 78

Figure 3-6 Assessment of binding of U4 snRNA to the Sm complex by filter binding. 79

Figure 3-7 Assessment of binding of U6 and U4 snRNA to the Sm complex by EMSA. 80

Figure 3-8 Assesment of binding of the Sm complex to the U4 Sm site by FP. 81

Figure 3-9 Assessment of binding of U2 snRNA to the Sm complex by EMSA. 82

Figure 3-10 Assessment of binding of U5 snRNA to the Sm complex by EMSA. 83

Figure 3-11 Assessment of binding of the Sm complex to the U2 and U5 Sm site by FP. 84

Figure 4-1 MO structure. 91

Figure 4-2 The mechanism of MO delivery by Endo-Porter. 93

Figure 4-3 Region of binding of the MO oligonucleotide. 96

Figure 4-4 Assessment of optimal wavelength for measurement of *C. merolae* culture optical density. 99

Figure 4-5 Microscope images of *C. merolae* cells at different pHs. 100

Figure 4-6 Cell growth of *C. merolae* cells. 101

Figure 4-7 Microscopic images of the *C. merolae* cells treated with control MO and Endo-Porter. 102

Figure 4-8 Microscope images of the *C. merolae* cells treated with control BPS MO and Endo-Porter at higher pH values..... 103

Figure 4-9 Delivery of MO to *P. pastoris* and *C. merolae* at pH 7. 104

Figure 4-10 Cell growth of *C. merolae* treated with the MO that targets the branch point binding site of U2 snRNA. 106

Figure 4-11 Cell growth of *C. merolae* cells treated with vivo-MO. 107

Figure 4-12 Growth of *C. merolae* cells treated with two concentrations of vivo-MO and guanidinium. 108

Figure 4-13 24 hours treatment of *C. merolae* cells with 5 μM of vivo-MO. 109

Figure 4-14 Electroporation of *C. merolae* cells for introduction of MO to cytosol. 111

Acknowledgment

First, I would like to thank my supervisor, Dr. Stephen Rader, for giving me the opportunity to be part of the Rader lab in 2014, and for allowing me to continue my research in 2015 as a master`s student. All his guidance, along with Dr. Martha Stark`s guidance, were key to improve my research skills throughout my graduate research. Also, I would like to thank my graduate thesis committee members, Dr.Kerry Reimer and Dr. Brent Murray, for helping me with my projects. In addition, I want to thank the collaborators of this project, Dr. Calvin Yip for the electron microscopy analysis and for helping with the interpretation of the results, Dr.Marlene Oeffinger for the utilization of the mass spectrometer for analysis of my samples, and Dr. Ute Kothe for allowing me to conduct the filter binding experiments in her laboratory and for teaching me the technique. Also, special thanks to the lab members for the feedbacks throughout my graduate work.

In addition, huge thanks to my parents and my brother, Rasheed Shidi, Andiara Ramos, and Ibraheem Shidi, for encouraging me to come to Canada from Brazil five years ago, and for continuing to encourage me to follow my dreams. I also want to thank all my friends from Brazil and Canada for supporting me all these years.

1. Chapter One - Introduction

1.1 Processing of precursor messenger RNA: Splicing

Occurring in eukaryotic cells, splicing is an interesting step in the processing of the precursor messenger RNA (pre-mRNA) that involves removal of the non-coding sequences (introns) and ligation of the coding sequences (exons). This mechanism is comprised of two transesterification reactions involving nucleophilic substitutions (SN2) coordinated by a megadalton complex called the spliceosome (Wahl et al. 2009). The spliceosome is a dynamic ribonucleoprotein (RNP) that includes five small nuclear RNAs (snRNAs) and over 200 proteins in humans. Each spliceosome subunit - U1, U2, U4, U5 and U6 snRNAs - associates with complex-specific proteins forming small ribonucleoproteins (snRNPs) (Wahl et al. 2009). In addition to these snRNA-specific proteins, the Sm complex associated with all snRNAs, except for U6 snRNA which is bound to the Lsm complex (Wahl et al. 2009). Notably, pre-mRNA splicing is an essential step in gene expression in Eukaryotes. Ninety percent of the human genes contain introns, and splicing is thought to give rise to much of the protein diversity in humans (Sakharkar et al. 2004). Therefore, it is not surprising that a significant number of diseases are linked to defects in splicing. For instance, mutations in the SMN protein cause the human disorder spinal muscular atrophy (Lefebvre et al. 1995). Indeed, 50% of genetic diseases exert their effects through errors in splicing (López-Bigas et al. 2005).

In *Saccharomyces cerevisiae*, the Sm core proteins and U1 snRNA specific-proteins associate with U1 snRNA forming a snRNP, which is responsible for the recognition of the 5`splice site. This interaction is known to be the first step in the assembly of the precursor spliceosome (Zhang et al. 2015). Furthermore, the branch proteins BBP-MUD2 recognise the branch point sequence

(BPS) of the pre-mRNA, and U2 snRNA associates to the Sm complex and specific-proteins forming the U2 snRNP subunit (Dunn et al. 2014). The U2 snRNP binds to Prp5 for the association with the BPS, followed by the release of BBP-MUD2 from the BPS by Sub2. It allows base pairing of U2 snRNP to the intronic region (Dunn et al. 2014). Furthermore, the U4/U6.U5 tri-snRNP formation initiates by the association of U6 snRNA to the LSm complex and Prp24, binding of U4 snRNA to U4 snRNA specific-proteins and the Sm complex. Next, the formation of U4/U6 di-snRNP is catalysed by Prp24, and U5 snRNA associates with the Sm complex and U5 snRNA specific-proteins forming the U5 snRNP (Dunn et al. 2014). The U5 snRNP joins U4/U6 di-snRNP forming the U4/U6.U5 tri-snRNP, which finally associates with the pre-spliceosome (pre-catalytic B complex) (Yan et al. 2015).

Two reactions follow the spliceosome activation. However, before the first reaction occurs, the spliceosome is rearranged by the release of U1 snRNP driven by an ATP-dependent helicase, Prp28 (U5 snRNP component), that disrupts the binding of U1 to the 5`splice site (Zhang et al. 2015; Staley & Guthrie 1999; Stevens et al. 2001). The U6 snRNP then replaces the U1 snRNP (Zhang et al. 2015). The DExD/H-box RNA helicase Brr2 (U5 snRNP component), unwinds the U4/U6 snRNA duplex resulting in the release of U4 snRNA and the binding of U2 snRNA, which initiates interaction with the U6 snRNA (Zhang et al. 2015; Nguyen et al. 2014). Following these rearrangements, conversion of the pre-catalytic B to the activated B complex results in recruitment of the nineteen complex (NTC) for its stabilisation (Yan et al. 2015; Chan et al. 2003). The translation of the activated B complex into the catalytically competent B complex is driven by the ATP-dependent Prp2 that, in cooperation with a G-patch protein Spp2, promotes remodelling of the spliceosome (Warkocki et al. 2015). Prp2 is responsible for the displacement of nine of the eleven proteins that interact with U2 snRNA and allows the first transesterification reaction to

occur (Liu & Cheng 2012). The reactive 2`-hydroxyl of the adenosine in the branch point sequence nucleophilic attacks the phosphorous atom at the guanine nucleotide at the 5`-end of the intron (Hang et al. 2015). Consequently, the 5` exon is released, and the intron lariat-3` -exon is formed (Figure 1-1) (Hang et al. 2015). The transesterification reactions require two Mg^{2+}. During the first reaction, one Mg^{2+} activates the 2`-hydroxyl of the adenosine in the branch point sequence, and the other $Mg^{2+}$ stabilises the 3`-OH of the 3`-end nucleotide I of the 5`-exon (Hang et al. 2015). Upon completion of this step, the spliceosome is re-rearranged by the ATP-dependent protein, Prp16 (Schwer & Guthrie 1992). During the second step, the 3` hydroxyl at the 3' end of the released 5`- nucleophilic exon attacks the phosphorous atom of the guanine nucleotide at the 5`-end of the 3`-exon. It results in binding of the two exons and release of the intron lariat (Figure 1-1) (Hang et al. 2015). This second reaction requires two Mg^{2+}. The first Mg^{2+} activates the nucleophile, and the second stabilises the leaving group (Hang et al. 2015). Prp22, an ATP-dependent protein, then releases the mature mRNA by unwinding the U5 snRNP/exon junction (Tsai et al. 2005).

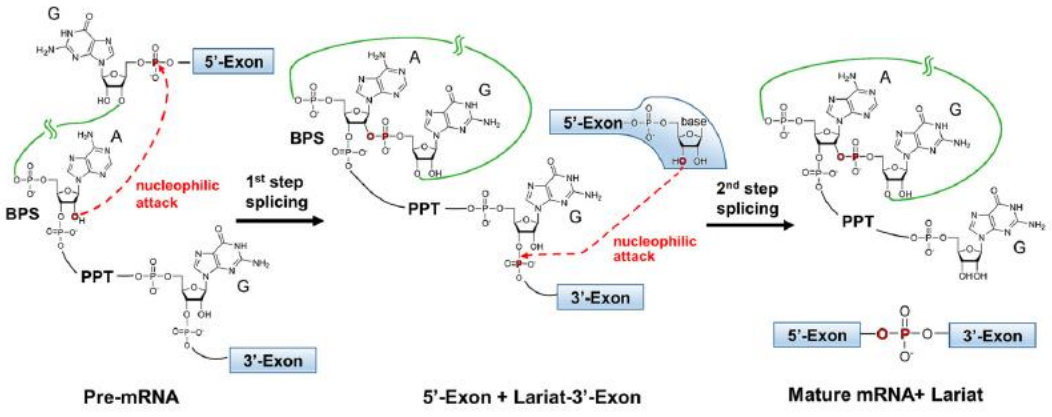


Figure 1-1 Illustration of the 2 SN2 transesterification reactions: The first reaction shows the release of the 5`-exon and formation of an intron lariat-3`-exon. The second reaction shows the release of an intron and ligation of the 3`-exon to the 5`-exon forming a mature mRNA (Hang et al. 2015).

Finally, the post-catalytic spliceosome undergoes disassembly by disassociation of the U5, U2, U6 snRNPs as well as the NTC and the intron lariat. During the disassembly, the NTR complex is recruited, and in an ATP manner, the helicase Ppr43 associates to the Ntr1 and Ntr2. For subsequent rounds of splicing, recycling of the subunits of the post-catalytic reaction is allowed by this arrangement (Graveley 2001).

1.2 *Cyanidioschyzon merolae:* a suitable model organism for splicing studies

Most investigations of mRNA processing are done in yeast cells, mainly S*accharomyces cerevisiae*, due to their simplicity in relation to human cells. However, for this study, *Cyanidioschyzon merolae*, a unicellular red alga, is proposed as a suitable alternative for splicing studies. *C. merolae* was the first complete algal genome to be sequenced revealing a similar number of genes compared to some yeasts (Higashiyama et al. 2004). *C. merolae* belongs to a class of acidothermophilic alga, *Cyanidiophyceae*, that inhabits thermal acidic environments (pH 1.5 and temperature of 45ºC). This primitive class has the cell morphology, reproduction, and biochemical components that suggests a link to both cyanobacteria and rhodophyta. Therefore, *Cyanidiophyceae* have been proposed to be primitive among the eukaryotes (Seckbach 2012). Through evolution, it has been observed that the less evolved species contain a more reduced amount of DNA keeping only the most critical machineries for sustaining life (Seckbach 2012). For instance, compared to more evolved organisms this algal spliceosome has less subunits. The Rader lab characterized the spliceosome in *C. merolae* identifying four snRNPs and 68 splicing proteins (Stark et al. 2015). Surprisingly, U1 snRNP, a relevant spliceosome subunit, is missing in this alga. As mentioned previously, U1 recognises the 5`splice site of the mRNA allowing initiation of spliceosome formation; therefore, the absence of this snRNP suggests a different

assembly of the spliceosome in this alga. U5 snRNA has been hypothesised to play U1`s role due to the complementarity of the 5`end of U5 snRNA to all annotated 5` splice sites in *C. merolae* (Figure 1-2).

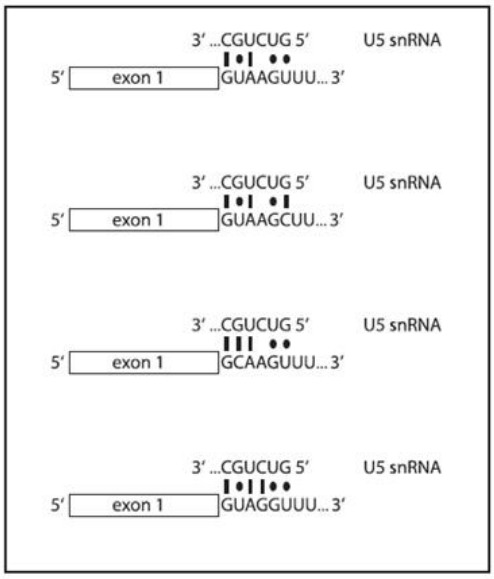
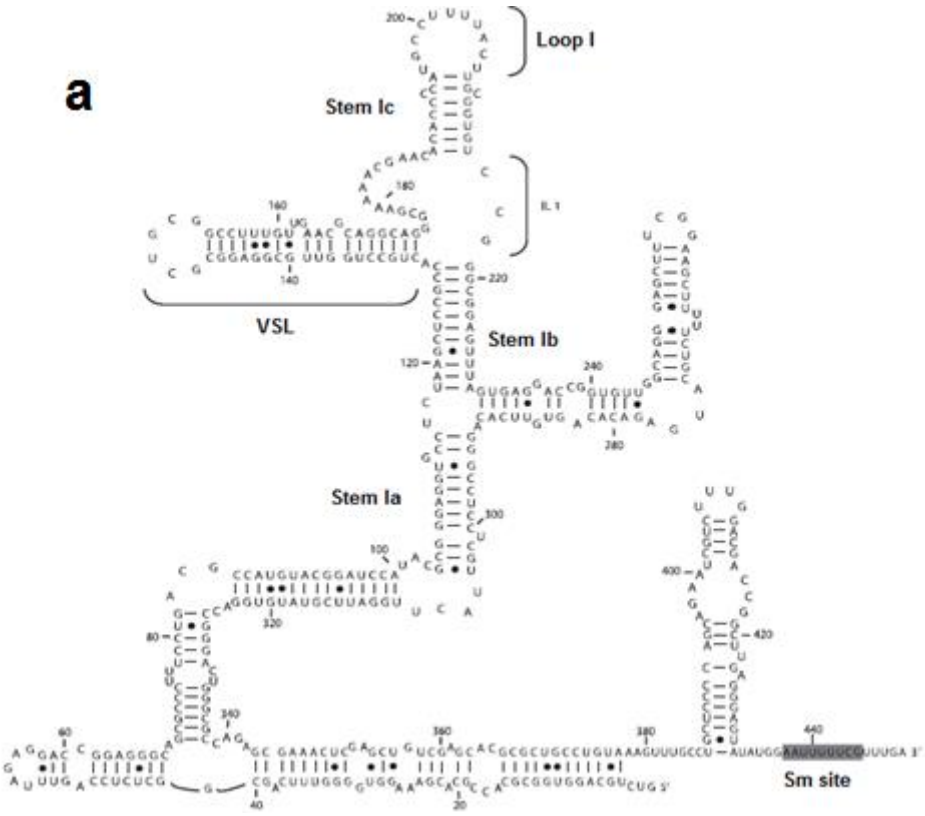


**Figure 1-2 Proposed binding of U5 snRNA to the 5`splice site:** Figure from Stark et al. (2015) presenting the annotated sequences of the 5`splice site significantly complementary to the U5 snRNA. This suggests that U5 is a valid candidate for recognition of the 5`splice site, playing U1`s role in *C. merolae*.

1.3 U5 snRNP complex

In yeast, the U5 snRNP has eight specific proteins and the Sm complex. However, only half of these U5-specific proteins - Prp8, Snu114, Brr2 and Dib1 - were bioinformatically identified (i.e. orthologous genes) in *C. merolae* (missing Aar2, Prp28, Prp6, and Lin1) (Stark et al. 2015). U5 snRNP is one of the subunits of the pre-assembled spliceosome complex U4/U6.U5 snRNP and is also a subunit of the catalytic spliceosome. U2 and U6 snRNA form the catalytic centre of the spliceosome, and the loop I of U5 snRNA can be found close to that core located at the bottom of the catalytic spliceosome (Yan et al. 2015; Hang et al. 2015). During the second

transesterification reaction, the loop I of U5 snRNA aligns and approaches the exon 1 to exon 2 allowing nucleophilic attack of exon 1 to the 3' splice site (Nguyen et al. 2015). Surprisingly, *C. merolae*`s U5 snRNA has a more extended sequence with unique 5` and 3` ends, maintaining conserved region from nucleotides 112-282, when compared to the *S. cerevisiae* U5 snRNA (Figure 1-3). Intriguingly, the 5` end (GUCUGC) is complementary to all annotated 5` splice sites, which presumably explains the absence of U1 in *C. merolae*, as the U5 would initiate assembly of the spliceosome by recognising the 5`splice site.



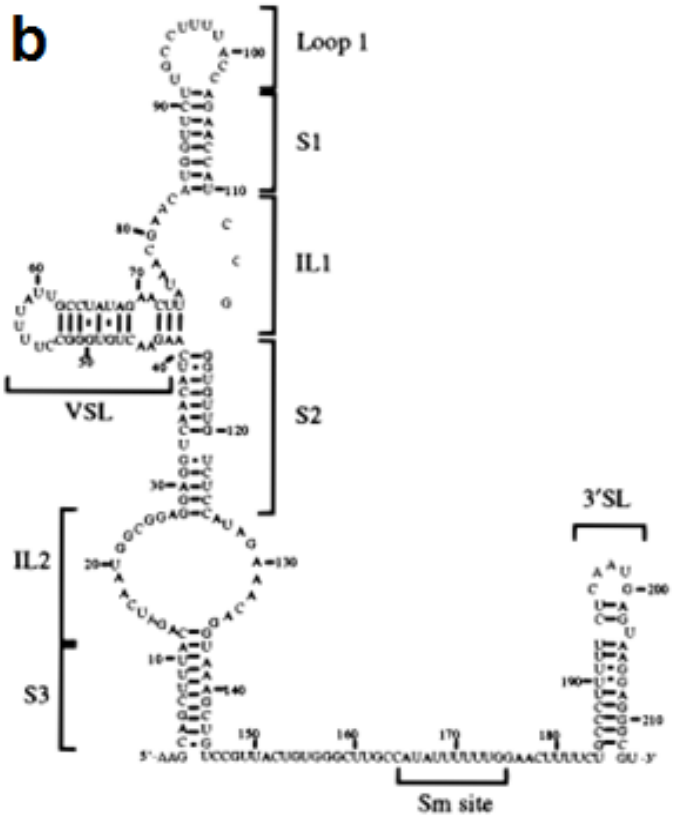


Figure 1-3 Comparison of the predicted *C. merolae* U5 snRNA to *S. cerevisiae* U5 snRNA: a) Stark et al. (2015) predicted the secondary structure of *C. merolae* U5 snRNA which shows a more extended structure when compared to *S. cerevisiae*. The highlighted box shows the predicted Sm site based on its uridyl-rich sequence. b) Frank et al. (1992)`s secondary structure of the *S. cerevisiae* U5 snRNA showing predicted features by Dix et al. (1998). These figures show that the location of the Sm site is unique in *C. merolae*, since the Sm site does not have a stem-loop in the 3`end. It suggests a different Sm assembly. IL, internal loop; S, stem; SL, stem-loop; VSL, variable stem-loop.

The proteins that associate to U5 snRNA also play essential roles in splicing before and after activation of the spliceosome. The core protein Prp8, the DExD/H-box family helicase Brr2 and the EF2-like GTPase Snu114 are essential for activation and formation of the core of the spliceosome. In yeast, Prp8 is a large protein comprised of six known domains: reverse transcriptase-like domain, thumb/X, linker, endonuclease-like domain, RNase-like domain, Jab/MPN domain, and an N terminal domain (Nguyen et al. 2015).

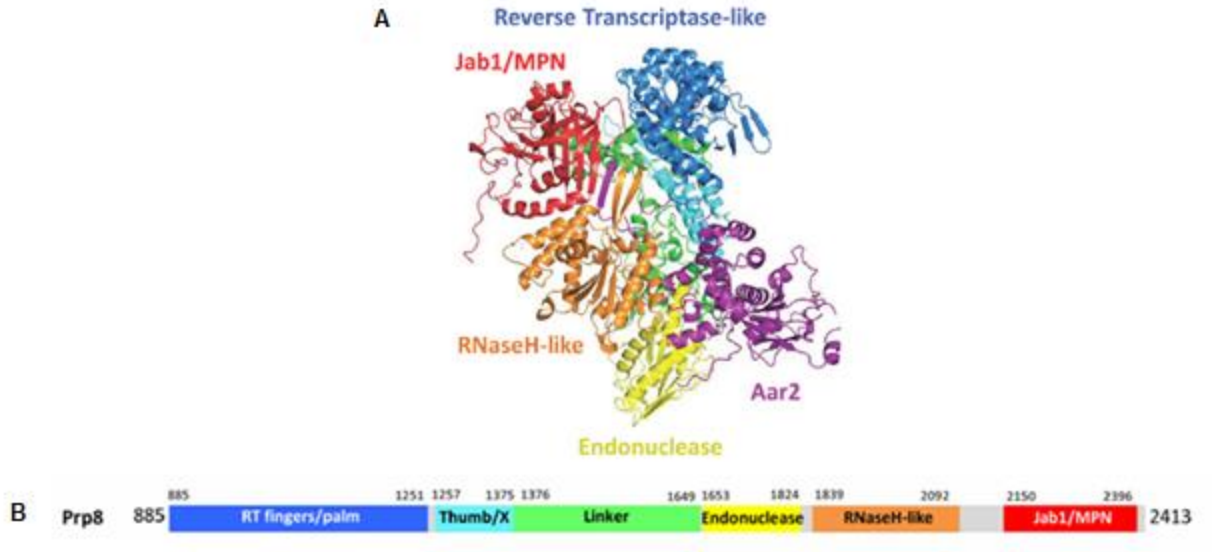


Figure 1-4 Crystal structure and architecture of Prp8 in *S. cerevisiae*: a) Crystal of Prp8 shows its association with a U5 assembly factor, Aar2, that is absent in *C. merolae*. The tri-dimensionality shows the Prp8 domains Jab1/MPN (red), RNaseH-like (orange), Endonuclease (yellow), Reverse Transcriptase-like (blue), Thumb/X (light blue) (Galej et al. 2013). b) The figure presents the architecture of the yeast Prp8 domains from residue 885 to 2413 (Galej et al. 2013).

Prp8 is at the core of the spliceosome having close contact with critical RNA residues (Galej et al. 2013). This U5-specific protein crosslinks to critical U6 snRNA residues, to the exon-binding loop I of U5 snRNA, and to three sites of the mRNA (3` splice site, branch point and 5` splice site). Prp8 mutations can suppress splicing-related mutations (Galej et al. 2013). For instance, the 3`splice site, branch point and 5` splice site mutants can be suppressed by mutations on the thumb/X and endonuclease domains. Mutants on the reverse transcriptase-like and endonuclease domains can minimise the effects of U4 mutants (Galej et al. 2013). The active site cavity is located between the RNase H-like and reverse transcriptase domain (Galej et al. 2013). The active site is also proposed to cover the N-terminal and thumb/X-linker, where U2 snRNA, U6 snRNA and the intron lariat are located; therefore, Prp8 is a crucial core protein (Yan et al. 2015).

Interacting with the Jab/MPN domain is the helicase Brr2 that plays an important role in unwinding of the U4/U6 snRNA duplex (Nguyen et al. 2015). This process of unwinding also relies on other U5-specific proteins, such as Snu114, that is involved in regulation of Brr2 (Nguyen et al. 2013). *In vivo*, Brr2 was crosslinked to loop 1 of U5 snRNA and close to the 5` and 3` splice sites (Hahn et al. 2012). Brr2 is comprised of an N-terminal domain and two consecutive helicase cassettes, in which each cassette has a helicase core N-RecA-1 and N-RecA-2 (Nguyen et al. 2013). Therefore, each one is comprised of six domains: two RecA domains, WH, Ratchet, helix-loop-helix (HLH), and FN3 (Nguyen et al. 2013). Each set of Ratchet (comprised of HLH and FN3 domains), is named Sec63 (Nguyen et al. 2013; Figure 1-5).

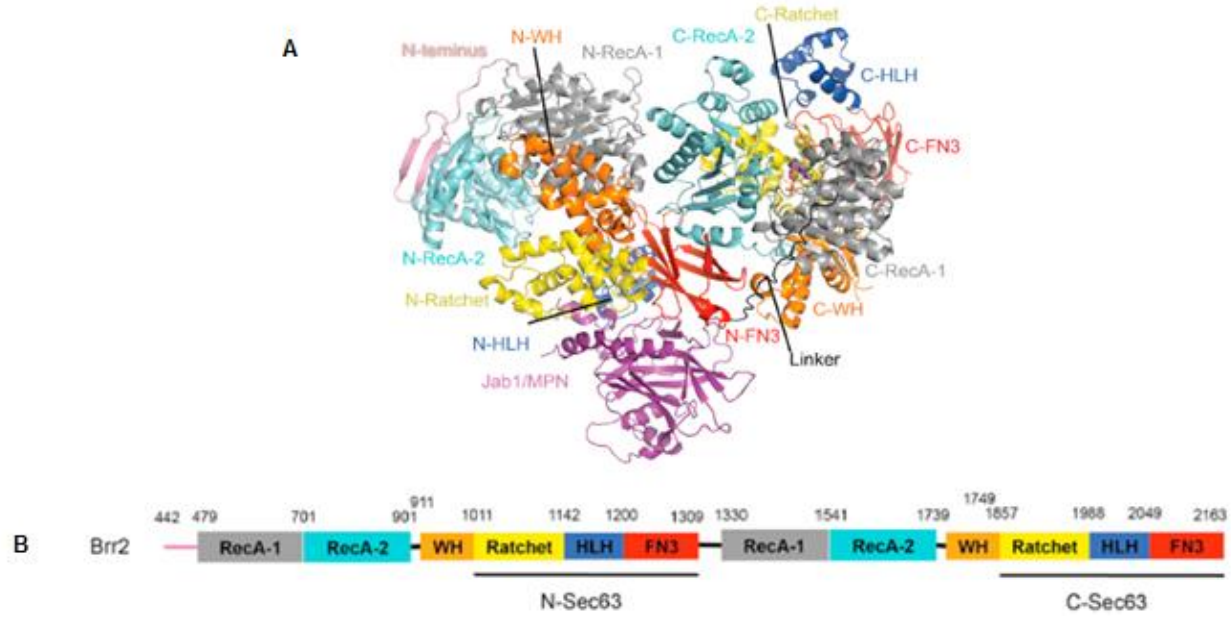


Figure 1-5 Crystal structure and architecture of *S. cerevisiae*`s Brr2: a) The crystal structure of Brr2 shows its association with the Prp8 domain Jab1/MPN. The tri-dimensionality shows the Brr2 domains RecA 1 (grey) and 2 (light blue), WH (orange), Ratchet (yellow), helix-loop-helix (HLH) (blue), and FN3 (red). b) The architecture of the yeast Brr2 helicases showing the two helicases comprised of six domains RecA 1 and 2, WH, Ratchet, HLH and FN3 (Nguyen et al. 2013).

The N-terminal domain has an unclear function, but it has been proposed to be essential for retention of U5 and U6 snRNP during and after spliceosomal activation (Zhang et al. 2015). In

yeast, in the early stages of Prp8 maturation, it associates with Aar2 in the cytoplasm preventing binding of Brr2 to Prp8. During maturation of U5 snRNA, Prp8 replaces Aar2, where Prp8 is found associated to the ratchet and FN3 domains of the N-terminal region of Brr2 (Nguyen et al. 2013). However, Aar2 is not present in *C. merolae* suggesting a different mechanism of maturation of Prp8 (Stark et al. 2015). In humans, mutations near this region of interaction cause the disease type 13 retinitis pigmentosa (Nguyen et al. 2013; Boon et al. 2007).

Snu114 is a GTPase comprised of five domains (Figure 1-6). GTPases are known to allow structure rearrangements of ribonucleoproteins, such as ribosomes (Brenner & Guthrie 2006). Snu114 shares similar structure with the translation elongation factor EF2, which catalyzes translocation of tRNA and mRNA (Brenner & Guthrie 2006). Mutations of Snu114 have been shown to affect spliceosome activation by increasing levels of U4 snRNA through changes in Brr2 functionality (Brenner & Guthrie 2006; Small et al. 2006). These modifications can also affect the interaction of U5 snRNA to Prp8 and Brr2 and disassembly of the spliceosome preventing the release of the excised intron and dissociation of the snRNAs (Brenner & Guthrie 2006; Small et al. 2006).

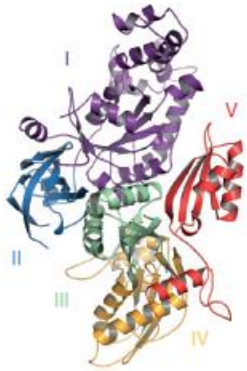


Figure 1-6 Crystal structure of the *S. cerevisiae* Snu114: Tri-dimensionality structure of Snu114 arranged in five domains as the eukaryotic translation elongation factor 2 (Nguyen et al. 2015).

Mutations in the guanine-binding pocket have shown a switch of specificity from guanines to xanthines XDP repressing disassembly of the spliceosome. Since its functionality can be recovered by addition of GDP, it suggests that Snu114 can regulate disassembly of the spliceosome (28). GDP is also known to inhibit U4/U6 unwinding. An assay involving XDP and GDP showed that mutations on snu114 prevent inhibition of unwinding when XDP is added. It suggested that Snu114 regulates Brr2 by obstruction of U4/U6 unwinding (28).

Dib1 is the smallest protein that associates to U5 snRNA. It is an ortholog of the protein *Schizosaccharomyces pombe* Dim1, which plays a relevant role in cell cycle progression (Reuter et al. 1999). Mutations in Dim1 affect cell viability by causing splicing defects that prevent cell cycle progression (Stevens et al. 2001). Previously, Dib1 has been suggested to be an essential splicing protein since its depletion results in accumulation of pre-U3 RNA (Reuter et al. 1999; Stevens et al. 2001). The crystal structure of the human homolog of Dib1 (also called Snu16) shows its similarity to thioredoxins in humans, having a thioredoxin-like fold (Figure 1-7a and b).

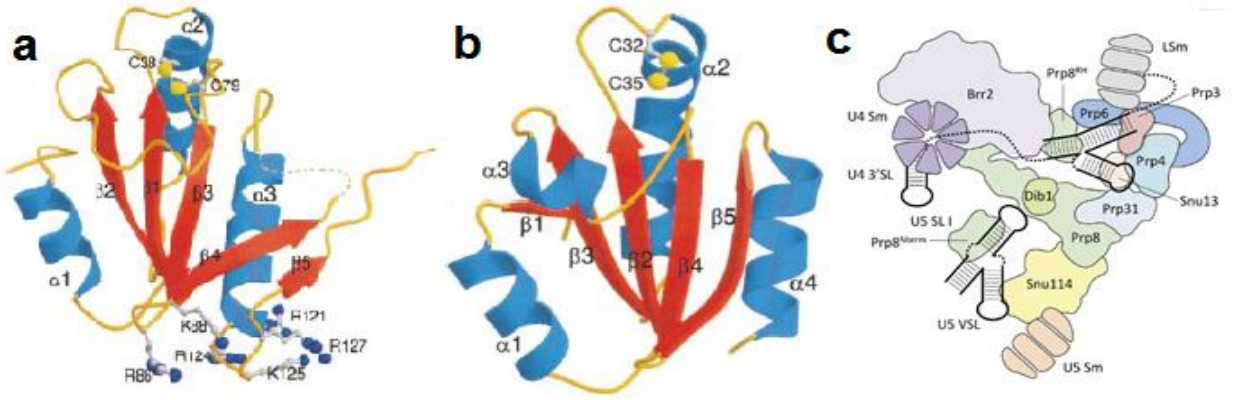


Figure 1-7 Crystal structure of the human Dib1: a) Tri-dimensional image of the human Dib1 showing significant similarity to the thioredoxin b) human thioredoxin structure (Reuter et al. 1999). c) *S. cerevisiae* Tri-snRNP model showing Dib1 located in the centre (Nguyen et al. 2015).

A recent reconstruction of the tri-snRNP by single-particle cryo-electron microscopy (cryo-EM) has shown interactions of Dib1 with the RT thumb/X domain of Prp8 and the loop I of U5 snRNA. Indeed, it is observed that Dib1 is in the centre of the tri-snRNP (Figure 1-7c).

In addition to the U5-specific proteins, a protein complex named Sm complex also interacts with U5 snRNA. In humans, the Sm complex is known to bind to all snRNPs (U1, U2, U4, and U5), except U6 that associates with a compound from the same family of proteins, Sm-like (Lsm). The Sm proteins belong to a large family of Sm and Lsm proteins, that are known to be highly conserved among different organisms and to form a doughnut-shape (Figure 1-8).

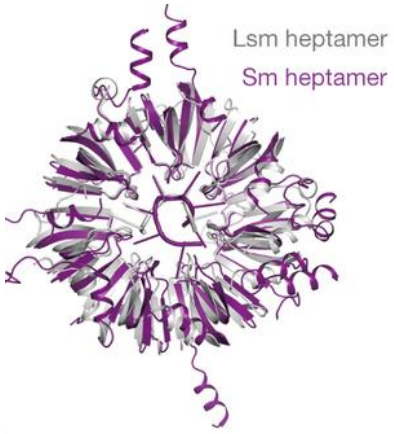


Figure 1-8 Comparison of the structure of the Sm complex and Sm-like complex: Tri-dimensional structures of the Sm complex (purple) and the Lsm complex (grey) show a similarity of the ring-shaped pentamer (Zhou et al. 2014).

The proteins from this family share a conserved Sm motif comprised of two conserved regions called Sm1 and Sm2 that are linked by a non-conserved sequence (Hermann et al. 1995); Séraphin 1995) (Figure 1-9). The Sm motif has been proven to be involved in the interaction among the Sm proteins encoding for the same folding domain in every Sm protein (Urlaub et al. 2001).

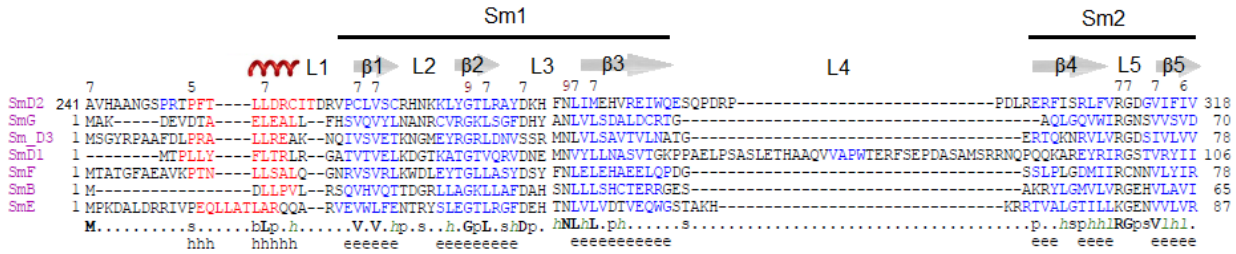


Figure 1-9 Structure-based sequence alignment of the *C. merolae* Sm proteins: Alignment of the red alga Sm protein sequences also present a conserved Sm motif comprised of Sm1 and Sm2. As observed in other organisms, the secondary structure of the Sm proteins has an alpha-helix (residues highlighted in red) linked by a non-conserved sequence to beta sheets (residues highlighted in blue). Sm1, three beta-sheets, is linked to Sm2, two beta-sheets, forming the Sm motif.

The spliceosomal Sm complex consists of seven proteins - Sm F, Sm E, Sm G, Sm D3, Sm B, Sm D1, and Sm D2. This protein complex is crucial for biogenesis and recruitment of snRNA particles. In the cytoplasm, the snRNA binds to the Sm complex forming a snRNP core particle termed Sm core RNP. In the absence of the snRNA, the Sm complex forms three heteromeric sub core complexes – Sm E-F-G, Sm D1-D2, and Sm B-D3. *In vitro* studies have shown that the purified Sms can bind to an oligonucleotide that contains some similarity to the consensus Sm site (Raker et al. 1996). This binding occurs in a stepwise manner when a pentameric complex is formed by binding of Sm E-F-G and Sm D1-D2 resulting in a unique substrate for coupling of Sm B-D3 (Raker et al. 1999; Figure 1-10). Indeed, the presence of the stem-loop 3` of the Sm site and the narrow hole of the ring (Kambach et al. 1999) explain the step-wise assembly of the Sm protein (Figure 1-3 and 1-10). *In vitro* analysis of the interaction of the Sm protein to the Sm site has shown that the core of the Sm proteins assembles in uridyl-rich sequences. The presence of 5' adenosines downstream this uridyl region has been confirmed to play an essential role in Sm protein association (Jones & Guthrie 1990; Jarmolowski & Mattaj 1993). Since the Sm complex can bind to any RNA or oligonucleotide that has the consensus Sm site, the Sm heterodimers bind to the SMN complex *in vivo*, followed by the importation of the SMN-Sm complex into the nucleus

ensuring binding to the snRNAs (Fischer et al. 1997; Liu et al. 1997; Meister et al. 2001; Pellizzoni et al. 2002).

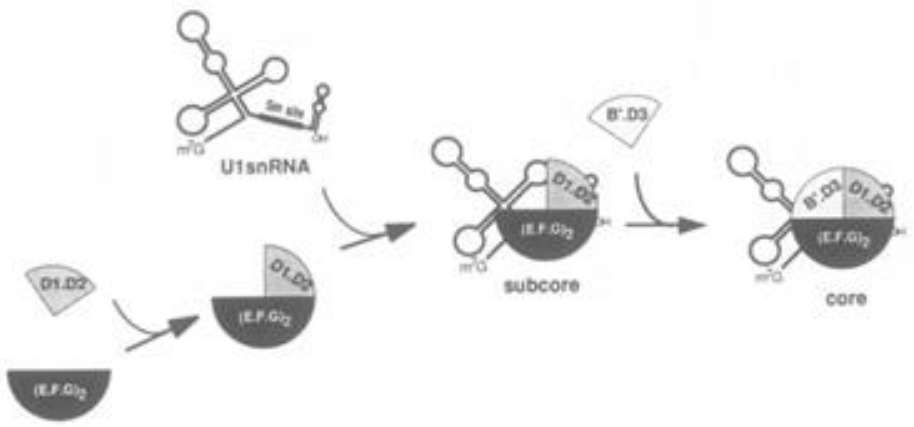


Figure 1-10 S. cerevisiae U1 snRNP Core formation: The binding of two heteromeric sub core complexes, Sm E-F-G, Sm D1-D2, forming a pentamer followed by binding of the U1 snRNA. By formation of the sub-core, the dimer Sm B-D3 joins the other heteromeric complexes forming a U1 snRNP core (Raker et al. 1996).

The Rader lab identified bioinformatically all seven proteins in *C. merolae* and hypothesises that the Sm proteins bind to U2, U4, and U5 snRNA due to the similarity of these sequences to the consensus Sm binding site (AU4-6G) (Branlant et al. 1982; Figure 1-11).

Surprisingly, the predicted Sm site located in the 3`end of the *C. merolae* snRNA lacks a loop. Therefore, the absence of the stem-loop 3` of the Sm site in the *C. merolae* snRNAs implies a different snRNP core formation (Figure 1-3 and 1-11a) (Stark et al. 2015). Presumably, pre-assembly of the Sm ring occurs before binding of the Sm complex to the Sm site. The absence of the SMN proteins, that are known to help in the recognition and stable interaction of the Sm E-F-G and Sm D1-D2, also suggests a different Sm proteins assembly (Zhang et al. (2011).

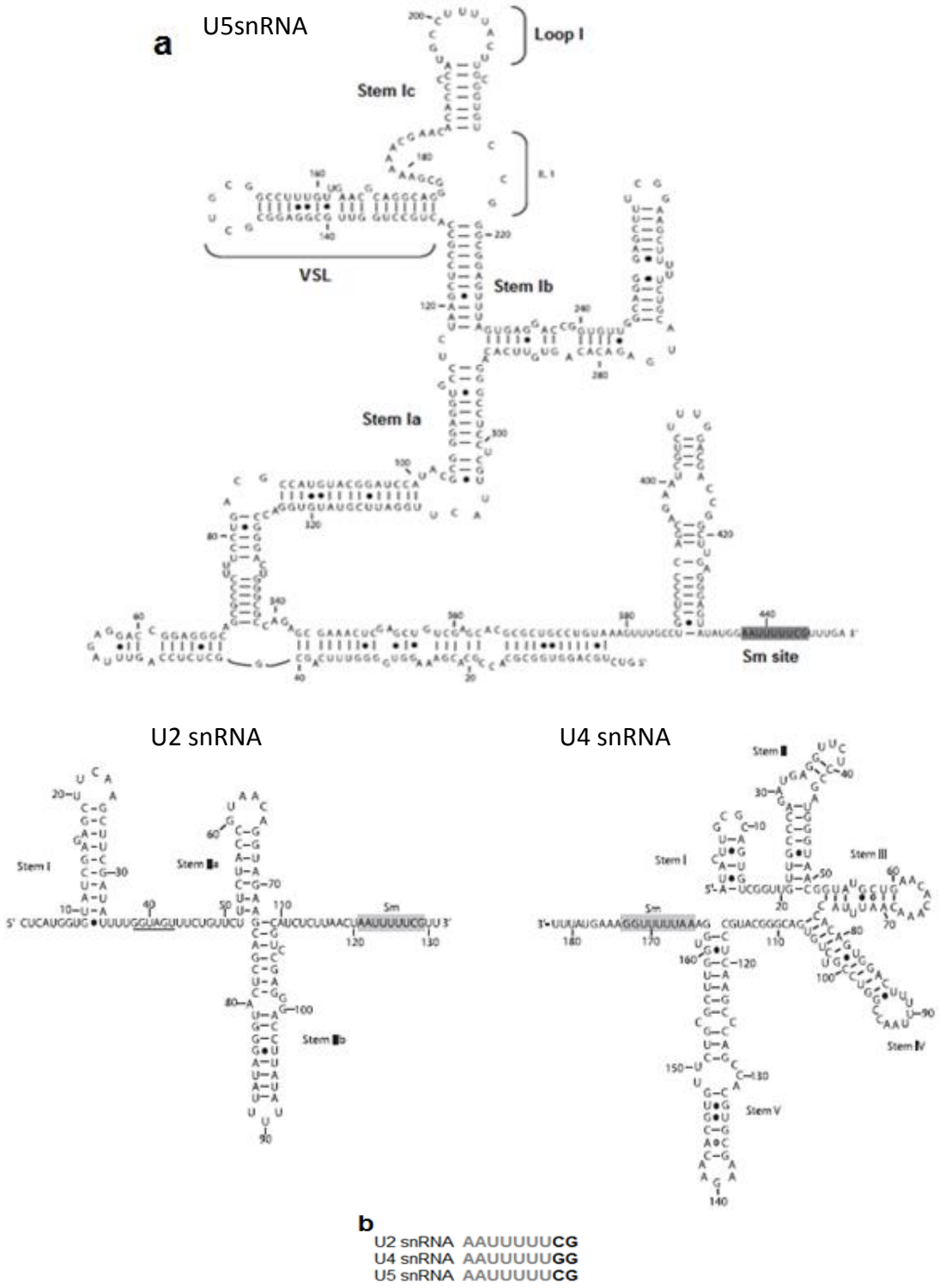


Figure 1-11 Predicted sequence of the Sm sites in *C. merolae* U2, U4 and U5 snRNA: a) The secondary structure predicted by Stark et al. (2015) presents the Sm site in the 3`end lacking a stem-loop suggesting binding of a pre-assembled Sm ring to the Sm site. b) Due to presence of a uridyl-rich sequence that is conserved in most organisms, it is proposed these Sm binding sites in U2, U4 and U5 snRNA. The highlighted nucleotides show that U2 and U5 snRNA share the Sm site sequence.

1.4 General thesis objectives

In the past three years, many advances have been made regarding structural studies of the different assembly steps of the spliceosome. These have been enabled by the development of new technologies, such as cryo-electron microscope. This approach has great advantage over X-ray crystallography, since the crystal formation step is skipped (Callaway 2015). Since 2015, several publications have revealed the spliceosome at different assembly stages at a higher resolution. For instance, Nguyen et al. (2015) reconstituted the *Saccharomyces cerevisiae* tri-snRNP presenting a cryo-electron microscope image of the U4/U6.U5 tri-snRNP complex at a resolution of 5.9 angstroms. This structure allowed for a better understanding of the snRNAs and proteins distribution. In less than a year, Ruixue Wan et al. (2016) revealed the tri-snRNP at an higher resolution, 3.8 angstroms, supporting Nguyen et al. (2015)`s tri-snRNP structure. Both works were crucial for the investigation of the spatial distribution of the snRNAs and their proteins, linking their structure to the role of each snRNP. Yan et al. (2015) published images of the spliceosome at different assembly steps presenting a 3.6 angstroms cryo-electron microscope using the model organism *Schizosaccharomyces pombe*. This publication revealed the spatial distribution of U2, U5 and U6 and its associated proteins, which was relevant to understand the positioning of these spliceosome components at the active centre of the spliceosome.

Therefore, a reconstitution of *C. merolae* U5 snRNP, for both functional and structural analyses of this splicing component, is one of the main aims of the chapter two. As mentioned, much has been learnt in the past three years by looking at the three-dimensional spatial distribution of snRNA and proteins. Therefore, I propose to assemble the U5 snRNP *in vitro*. This objective is intended to be achieved by co-expression and co-purification of the U5-associated proteins and *in vitro* transcription of U5 snRNA. It is expected that the snRNA will bind to the proteins allowing

structural and functional studies of this splicing component. Although we have seen a significant advance in structural studies of the spliceosome, several questions are challenging to answer due to the complexity of the spliceosome, such as the structure of the U2 snRNP and the roles played by each U2 snRNA-specific protein. Here, it is proposed that the investigation in a `paucity` spliceosome would be less challenging. However, only a few expression strategies have previously been developed in *C. merolae*; therefore, in this chapter two, expression and purification strategies are the initial goal.

In chapter three, I focus on the investigation of a splicing complex that is known to bind to U5, the Sm complex, by functional and structural approaches. As it has been stated previously, in other organisms, that the Sm proteins are known to associate with U1, U2, U4 and U5 snRNAs. Therefore, in *C. merolae* it is proposed that the Sm complex bind to U2, U4 and U5, since U1 is absent in this organism. Also, the snRNAs predicted secondary structure suggests a Sm site at the 3' end of each snRNA. However, no substantial evidence of the binding of the Sms to these snRNAs has been shown. Therefore, a second objective of this dissertation is to co-express and co-purify the Sm proteins, which would assemble and allow the investigation of the structure and functionality of this complex. Previously, it has been seen that a similar *C. merolae* splicing protein complex, Lsm proteins, formed a doughnut-shape, when expressed and purified together; therefore, it is expected to see the same ring formation by electron microscope. Successful reconstitution would allow for binding analysis of the Sm complex to the snRNAs, revealing the same functionality observed in other organisms. The predicted secondary structures of U2, U4 and U5 do not show a stem-loop 3` of the Sm site, which suggests that in *C. merolae* the Sm proteins pre-assemble prior to binding to the snRNAs. Further evidence that supports this pre-assembling step, it is the absence of the SMN complex that is involved in assembly of the Sm proteins around

the Sm site in a step-wise manner. Therefore, it is initiated an investigation of the Sm proteins functionality, structure and assembly in *C. merolae*.

In chapter four, a crucial area that will be addressed is the intron recognition within *C. merolae*, since the snRNP that plays this role – U1 – is missing. The complementarity between U5 snRNA and the intron suggests that U5 may have co-opted this intron-recognition role in *C. merolae*. The Rader lab has been able to confirm that splicing occurs in this alga; however, there is no evidence that this process is vital (Stark et al. 2015). For that, a novel strategy of splicing blockage, morpholino oligonucleotides, is proposed to help this question. This DNA has been previously used to target snRNAs in other organisms, thus preventing splicing. Therefore, this method is intended to address splicing's relevance in *C. merolae* by prevention of the initial step of the spliceosome assembly. A morpholino oligo that has a higher affinity to U2 snRNA and prevents binding of this snRNA to the branch point site will be used. Presumably, binding of the morpholino to the snRNA will cause a decrease in cell growth or death of the cells, if splicing is in fact essential in this red alga. Since this method has never been attempt in *C. merolae*, different strategies will be attempted that could achieve successful delivery of the morpholino to the cell, thus optimizing the efficiency of the protocol, and as well assessing splicing blockage by RT-PCR. Furthermore, if this technique addresses the question successfully, a morpholino complementary to the 3' end of U5, that is proposed to recognise the intron region, will be designed to address the intron-recognition hypothesis.

2. Chapter Two - U5 snRNA reconstitution

2.1 Introduction

The U5 snRNP has been thoroughly investigated, mainly in yeast, and most cryo-EM structures of the spliceosome have been able to identify all U5-associated proteins. Structural studies of this component have shown that the U5 snRNA-specific proteins are strategically located in the catalytic region of the spliceosome (Yan et al. 2015; Hang et al. 2015). Indeed, most proteins that associate to this particle play essential roles during splicing activation and the two catalytic reactions. For instance, Brr2 is a helicase that plays a crucial role in activation of the spliceosome by unwinding the U4/U6 di-snRNA (Nguyen et al. 2013). Although much has been learnt about the spliceosome in the past three years, several questions still need to be answered (Yan et al. 2015). Therefore, this chapter will describe my investigation into the U5 snRNP spliceosome component in *C. merolae*. Interestingly, the Rader lab has presented a reduced spliceosome, where half of the proteins that associated with U5 snRNA are missing. Although Prp8, Snu114, Brr2, Dib1, and the Sm complex were identified, not much is known about the structure of these proteins, and consequently, their function in *C. merolae*. Thus, different methodologies will be presented to co-express and purify all the proteins that associate to U5. It is proposed that, by co-purification of the U5-specific proteins and presumably self-assembly, an increase in protein yield would assist in any structural investigations of the U5 snRNP. Reconstitution of this spliceosome subunit would lead to a better understanding of its functionality. Furthermore, this work would help to address U5`s role in *C. merolae*. As stated previously, it is hypothesised that U5 plays U1`s role in this organism due to its 3`end complementarity to all annotated 5` splice sites.

A cloning strategy performed by Dunn to clone the *C. merolae* Lsm proteins into one vector was used to construct a co-expression vector (Dunn, 2010). By constructing a Lsm-containing vector, Dunn was able to co-express the protein complex in bacteria. Self-assembly of the proteins allowed co-purification of the Lsm complex and higher protein yield. The design of this protocol was essential for this work as it was deemed the most suitable for reconstituting the U5 snRNP.

Ligation-independent cloning (LIC) is a method that has been discovered to be very efficient (Schmid-Burgk et al. 2014). Relatively, conventional methods of cloning are less efficient and more time-consuming due to the consecutive digestions, ligations of DNA to the vector, and transformations. Besides, the number of cleavage sites available decrease proportionally to the number of insertions and the size of DNA segments (Schmid-Burgk et al. 2014). Therefore, for this work, LIC is the best method for introduction of all U5-specific protein genes into a vector for further co-expression. This cloning strategy does not require restriction enzymes and DNA ligase for insertion of DNA segments into the vector. Cloning of long genes, such as the Prp8 gene, is challenging due to the short single-stranded sequence that is used for ligation of the gene to the vector. Therefore, LIC requires the formation of a 12-20 single-stranded overhang that is created by the T4 DNA polymerase segment. This enzyme can delete nucleotides in the absence of dNTPs facilitating insertion of genes. As exemplified in figure 2-1, in the presence of only one dNTP, this enzyme adds back the presented nucleotide, dATPs in this case, allowing the formation of a single-stranded segment (Dyson & Durocher 2007).

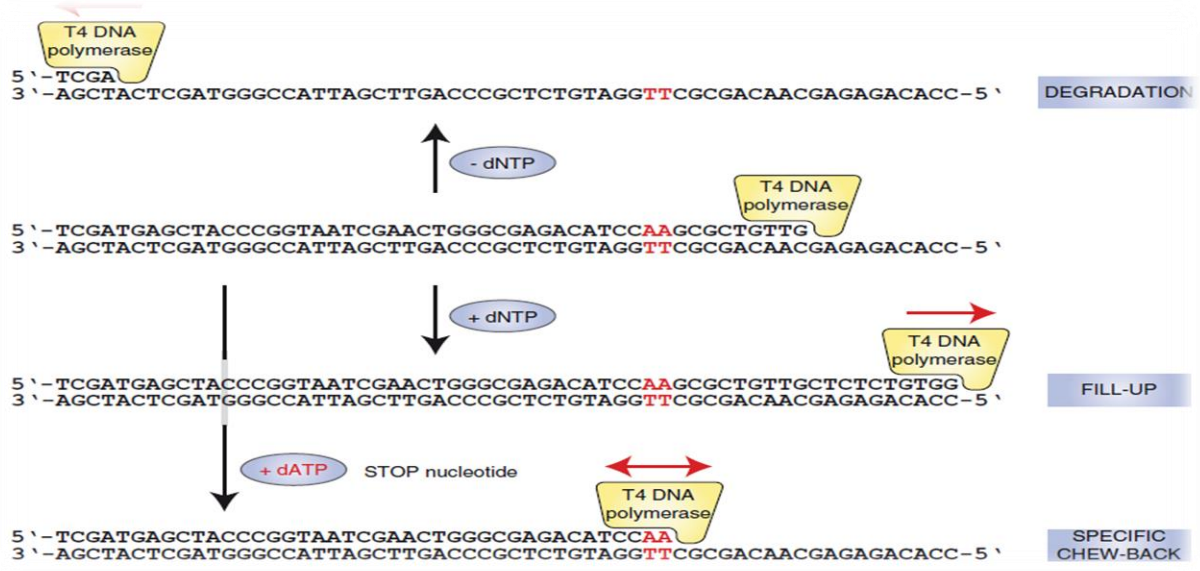


Figure 2-1 T4 DNA polymerase reaction: The presence of all four dNTPs in the reaction results in the fill-up of nucleotides in the 5`→3` direction. However, in the presence of only one dNTP in reaction, dATPs for instance, the enzyme removes the nucleotides until it finds a dTTP in the single-stranded fragment. Therefore, this mechanism of degradation of the double-stranded segment is used to create long overhangs (Schmid-Burgk et al. 2014).

By the creation of a single-stranded segment both in the vector and in the gene, it is possible to hybridise the vector to the gene due to its single-stranded ends complementarity. The genes are inserted into an expression vector called pQLink that contains the restriction sites SwaI and PacI. These restriction sites allow ligation of several vectors and insertion of all genes into one vector (Addgene plasmid 13667 and 13670; 82) (Figure 2-2).

This chapter describes the process of constructing an expression vector containing all the corresponding genes of the proteins that associate with the U5 snRNA. To express these proteins many different vectors were employed to see which was most optimal. In addition to pQLink, the pMCSG23 vector was found to allow for insertion of genes by ligation independent cloning and was used as an alternative expression vector. For instance, the Prp8 and Snu114 protein genes were inserted into pMCSG23 vector fused to a maltose binding protein (MBP). Expression of the protein

attached to MBP prevents its aggregation enhancing protein solubility. Therefore, this second vector was used when attempting to express large proteins individually.

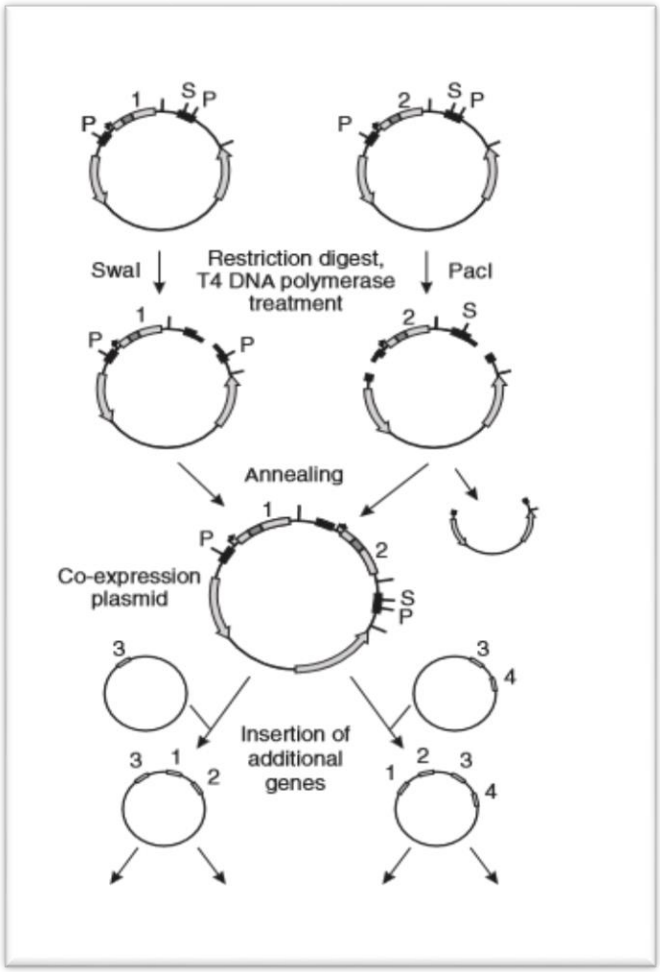


Figure 2-2 Construction of expression vectors for co-expression of proteins: The presence of the restriction sites SwaI and PacI allow insertion of genes into one vector. SwaI digestion enables linearization of the first vector, and PacI double-digestion of the second vector enables the release of the gene. Therefore, T4 DNA polymerase treatment of the SwaI digested vector and the PacI digested insert permits the creation of complementary overhangs for ligation of the insert into the vector. Consecutive SwaI and PacI digestions and T4 DNA treatments allow insertion of several genes into one vector (Scheich et al. 2007).

Induction of protein expression was performed by Isopropyl β-D-1-thiogalactopyranoside (IPTG), auto-induction and methanol-induction. Auto-induction is an alternative method that has produced favourable results when protein expression via IPTG induction was found to be

challenging. This technique involves saturating the culture which automatically facilitates use of lactose for induction of protein expression by depletion of inhibitory factors. Therefore, since induction happens during saturation of the culture, it is not necessary to monitor cell growth, thus making it a more convenient strategy to express proteins, as described by Studier (2005). Indeed, saturation of cultures in non-inducing media enables retention of the vector allowing for storage of the culture in the refrigerator for weeks. In addition, protein yield has been found to increase when compared to IPTG induction.

In addition to the use of *Escherichia coli* as a host organism for protein expression by IPTG and auto-induction, protein expression was also performed via methanol-induction in *Pichia pastoris*. Some proteins are not efficiently expressed in bacteria due to the absence of translation modifications. Indeed, post-translation modifications, such as glycosylation, can compromise protein folding affecting protein stability and functionality (Burgess & Deutscher 2009). Consequently, expression of challenging proteins in Eukaryotic organisms, such as *P.pastoris*, is advantageous. Thus, protein expression of Brr2 was also attempted in yeast by methanol induction. *P. pastoris* cells are methylotrophic organisms capable of using methanol as their primary carbon source due to the presence of alcohol oxidase (AOX) genes. Therefore, integration of protein genes of interest to the AOX locus allows dramatical transcriptional induction when cells are grown in a methanol-containing medium (Burgess & Deutscher 2009). Thus, another vector containing an AOX promoter was used (pPICZA – Qiagen), and insertion of a protein gene to the vector allows for integration of the vector to the 5` AOX1 region of the host cell. Integration enables the recombinant yeast to metabolise methanol. Therefore, the presence of methanol will activate the AOX gene inducing expression of the protein in a methanol-dependent manner (Burgess &

Deutscher 2009). The Brr2 gene was chosen to be inserted into pPICZA, as an alternative strategy for expression of large proteins.

By attempting, troubleshooting, and eventually optimizing the different strategies described above, I successfully expressed eight of the eleven proteins that associate with U5. The approaches described in this chapter will be valuable to future studies that investigate these proteins and will contribute new methods and techniques that can be used generally towards protein expressions.

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 for Environmental Studies in Tsukuba, Japan (mcc.nies.go.jp/). *C. merolae* genomic DNA was prepared by Martha Stark as previously described (Stark et al. 2015).

2.2.2 Construction of expression vectors for ligation independent cloning

The construction of the vectors for insertion of U5-specific proteins genes was done by modification of the pQLinkN and pQLinkH vectors by introducing multiple genes into one vector by LIC (Scheich et al. 2007). These vectors were previously modified by Dunn, which a 40 base pairs sequence was inserted to the vector by double digestion with EcoRI and BamHI. Therefore, for this work, modified pQLinkN and pQLinkH (pQLinkN^{mod} and pQLinkHmod) were used to design a vector that would allow insertion of each gene via LIC, followed by a combination of all genes into one vector. In regard to pQLinkH^{mod}, an oligonucleotide containing the PmlI restriction

site, a ribosome binding site (RBS), seven histidines (HIS_7-tag) and a Tobacco Etch Virus (TEV) sequence was inserted (Figure 2-3a). The HIS_7-tag was inserted for purification of the protein expressed, and the TEV cleavage site was added to enable cleavage of the tag after purification. Conversely, only the restriction site PmlI was inserted into $pQLinkN^{mod}$ as it would allow for digestion of the vector for later insertion of the genes by LIC (Figure 2-3b). Introduction of the oligonucleotides was accomplished by EcoRI and HindIII restriction digest followed by ligation of annealed oligonucleotides into the vector. The oligonucleotides oSDR 1070/1071 and 1072/1073 replaced the EcoRI-HindIII fragment removed from $pQLinkH^{mod}$ and $pQLinkN^{mod}$ (Figure 2-3).

In order to carry out the above above reactions, five micrograms of both vectors were digested for 3 hours using EcoRI and HindIII restriction enzymes (30U/µg of DNA) followed by gel purification of the vector. The oligonucleotides were annealed by addition of T4 DNA ligase buffer (New England Biolabs), left at room temperature for 10 minutes, heated at 85^{o} C for 5 minutes and cooled down at room temperature for one hour. Ligation of the annealed oligonucleotides (50 fmols) to the digested vectors (150 fmols) was performed by addition of T4 DNA ligase. After the ligation reaction, the DH5α bacterial strain was transformed with modified $pQLinkN^{mod}$ and $pQLinkH^{mod}$. Cells were screened on LB plates and ampicillin and incubated for 18 hours. Colonies were selected, and vectors were isolated by using a plasmid DNA mini kit (Omega bio-tek). Insertion confirmation of the oligonucleotide was done by digestion of vectors with PmlI. The modified $pQLinkN^{mod}$ was named pSR617, and the modified $pQLinkH^{mod}$ was named pSR627.

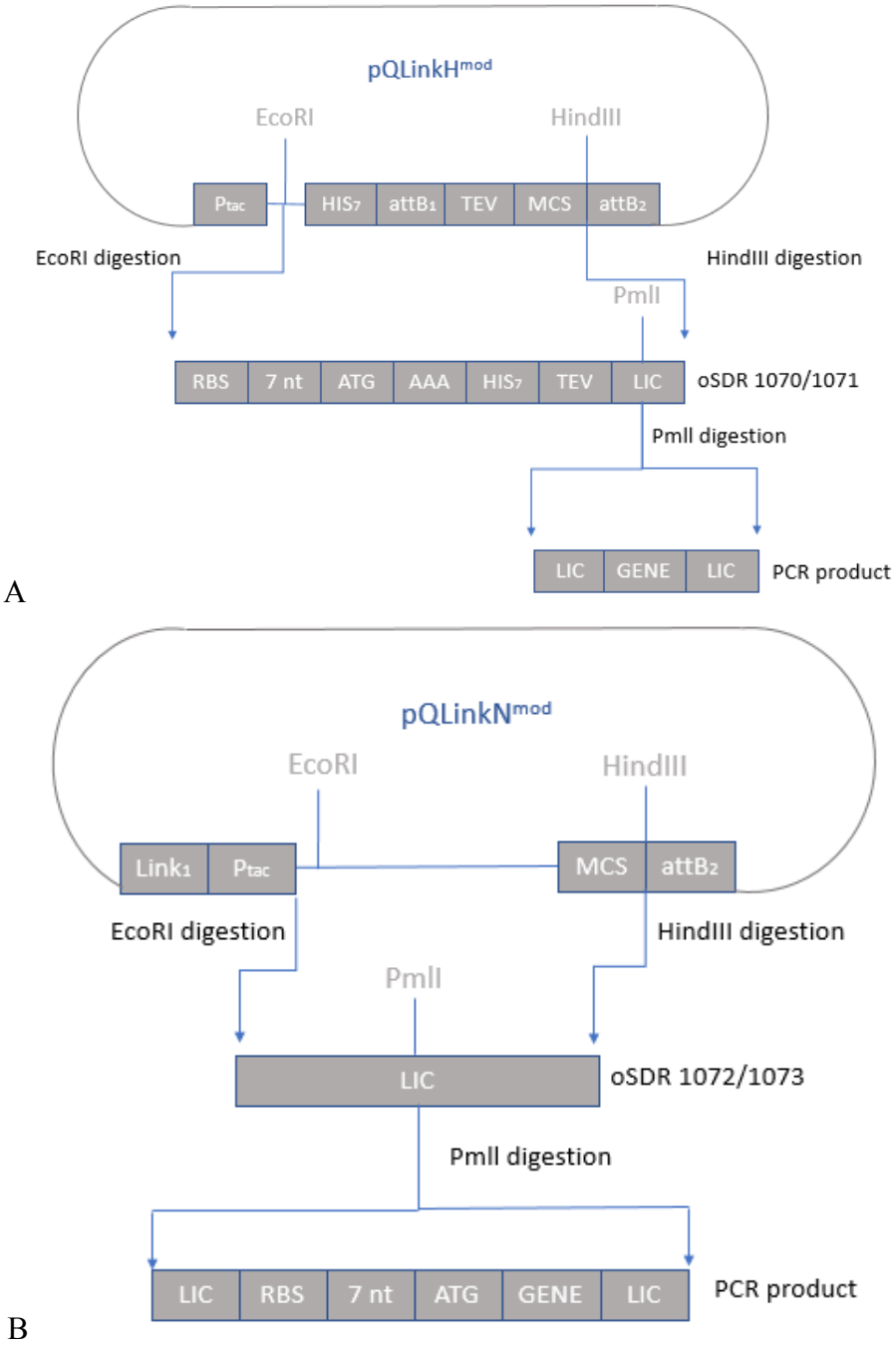


Figure 2-3 Vectors construction for insertion of protein genes by LIC: a) The pQLinkH^{mod} vector was remodified by EcoRI-HindIII double-digestion for insertion of the oligonucleotide oSDR 1070/1071. Since this vector is going to be used for expression of Sm E and Sm F in frame with a HIS_7 tag, oSDR 1070/1071 contains an RBS, seven nucleotides, a start codon (ATG), a lysine (AAA), seven histidines and a TEV site. A sequence downstream from the TEV site, here called the LIC sequence, was added to allow LIC. The LIC contains a PmlI site, for insertion of the PCR product that will also have a complementary LIC sequence. Therefore, after T4 DNA polymerase treatment, both vector and PCR product will have complementary overhangs. b) The pQLinkN^{mod} vector was remodified as pQLinkH^{mod}; however, the annealed oligonucleotide oSDR 1072/1073 only contains the PmlI-containing LIC sequence.

Table 1 DNA oligonucleotides used to modify pQLinkNmod and pQLinkHmod: DNA oligonucleotides used to modify pQLinkN^{mod} and pQLinkH^{mod}: DNA sequences are shown from 5` to 3`. The oligonucleotides oSDR1070 and 1071 and oSDR1072 and 1073 were annealed for insertion into pQLinkN^{mod} and pQLinkH^{mod}, respectively. The RBS sequence is in blue, the start codon in pink, the seven histidines is in green, the TEV site is in red, and in bold is the PmlI restriction site sequence.

Oligonucleotide name	Vector of insertion	Sequence
oSDR1070	pQLinkN^{mod}	GAATTCAGGAGAAATTAACTATGAAACATCACC ATCACCATCACCATGAGAATCTGTACTTCCAAT CC**CACGTG**GGAAGTGGATAACCAGCTT
oSDR 1071	pQLinkN^{mod}	CTTAAGTCCTCTTTAATTGATACTTTGTAGTGG TAGTGGTAGTGGTAGTGTTAGACATGAAGGTTA GGGTGCACCCTTCACCTATTGGTCGAA
oSDR 1072	pQLinkH^{mod}	GAATTCCGTACTTCCAATCC**CACGTG**GGAAGTG GATAACGGTAAGCTT
oSDR 1073	pQLinkH^{mod}	CTTAAGGCATGAAGGTTAGGGTGCACCCTTCAC CTATTGCCATTCGAA

2.2.3 Amplification of the genomic sequences of U5 snRNA-specific proteins by polymerase chain reaction (PCR)

As explained previously, two vectors were constructed for insertion of all genes into one vector and co-expression of U5-specific proteins. The vector pSR627 allows purification of the protein since it was designed to have a HIS_7-tag in frame with the protein`s gene sequence. Therefore, it is expected that by co-expression of the proteins they will self-assemble as one complex, making it necessary for only one protein to be tagged for purification of the complex. For this, two proteins were chosen to have a HIS_7-tag: Sm E and Sm F. The other genes were selected to be inserted into pSR617. For insertion of the genes into the vectors, primers were designed that would amplify the genes with 5` and 3`ends complementary to the LIC sequence of the vectors (Figure 2-4). The primers were designed as seen in table 2. An essential guanidine nucleotide (dGTP) is observed in all primers before the gene-specific sequence for formation of a 15 nucleotides overhang that will be complementary to the vector overhang (Figure 2-4). In addition, the genes inserted into pSR617

have a forward primer containing an RBS, seven nucleotides, and a start codon. For the Sm E3 and F genes that will be entered to pSR627, a different 5`end of the forward primer is seen. It contains only a dGTP upstream from the gene.

For insertion of Prp8 into a pMCSG23 plasmid, which was an alternative strategy for expression of Prp8 having an MBP tag, a primer was designed that also allows LIC (Table 2). For insertion of Brr2 into pPICZA, different primers were used as traditional ligation was performed for the introduction of the gene into the vector. Thus, the designed primers contained a restriction site at the 5` and 3`end of the gene, resulting in the KpnI and NotI restriction sites flanking the Brr2 gene. As well, the Kozak consensus sequence was inserted downstream from the KpnI restriction site, as it optimises initiation of translation in eukaryotic cells (Table 2).

Genes were amplified from *C. merolae* genomic DNA by polymerase chain reaction (PCR). For amplification of all genes, but Prp8 and Dib1 genes, Q5 high fidelity DNA polymerase was used. For Prp8 and Dib1, the enzyme 5X Q5 high GC enhancer buffer was used for amplification of the genes since both genes had a high GC content, 60 % and 62% respectively. For each 50 µl PCR reaction, 10 µl of 5X Q5 reaction buffer, 1 µl of 10 mM dNTPs, 2.5 µl of 10 mM reverse and forward primers, 2 µl of 1 ng/ul of genomic *C. merolae*, 1 µl of Q5 high fidelity DNA polymerase, 10 µl of 5X Q5 high GC enhancer (Prp8 and Dib1) and Milli Q water were added. The thermocycler was programmed for 35 cycles of denaturation, annealing and extension. The details of this PCR set-up are presented in table 3.

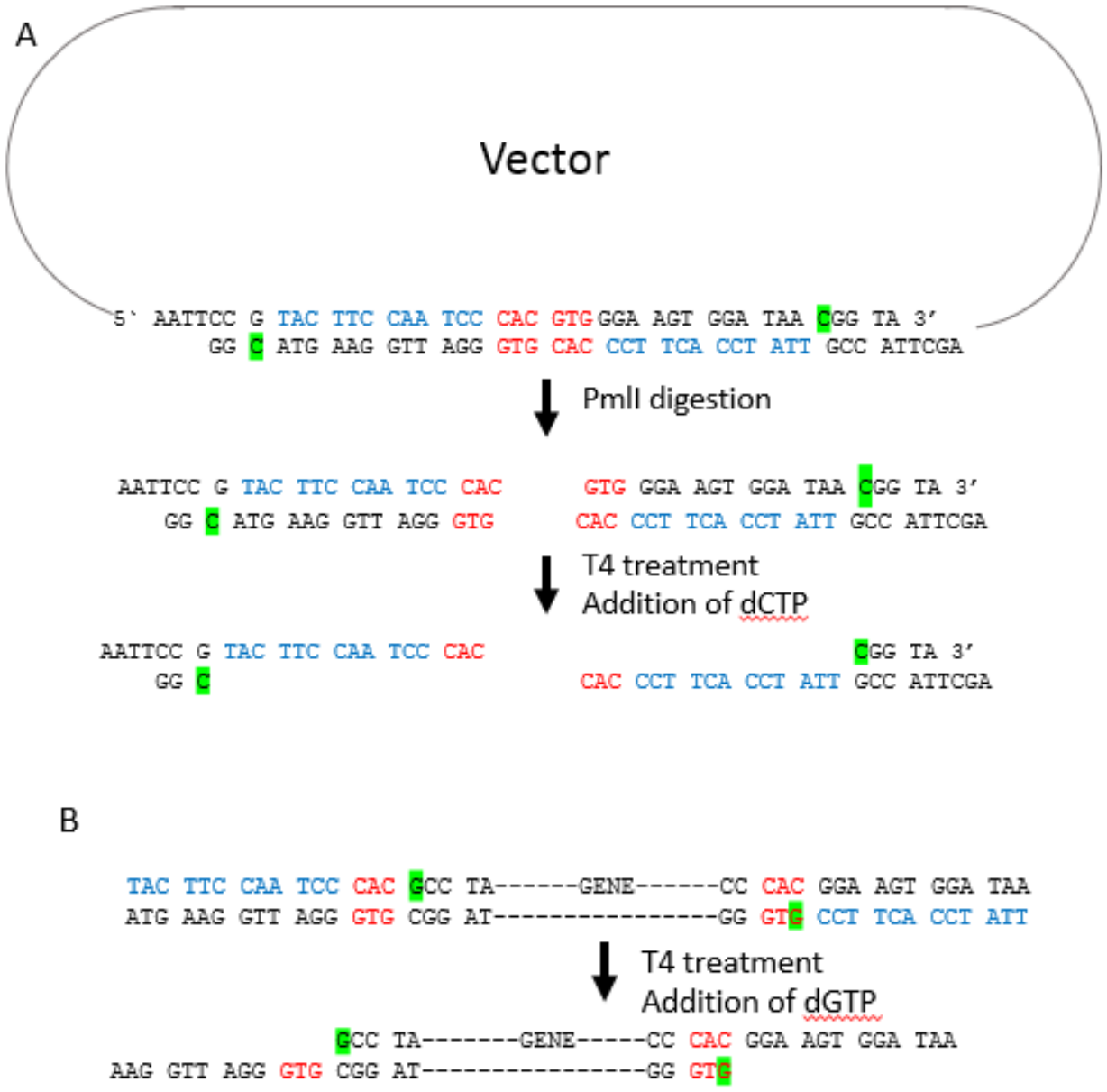


Figure 2-4 Complementarity of amplified genes to the modified vectors: a) PmlI digestion of the vector enables its linearization. The digested vector is treated with T4 DNA polymerase removing all nucleotides until it reaches a citosine, since dCTPs are in reaction. b) T4 treatment of PCR products with dGTP allows removal of nucleotides and; therefore, formation of overhang complementary to vector.

Table 1 DNA oligonucleotide sequence of the primers used for amplification of protein genes: DNA sequences are shown from 5` to 3`. In bold are the LIC sequences. The dGTP at the 3`end will be used for the creation of the overhang. The first oligonucleotide listed for each gene is the forward primer and the second is the reverse primer. Sm E and F will be in frame with a $HIS_7$ tag, so two forward primers for these protein`s genes were designed. The RBS sequence is in blue, the start codon is in pink, underlined is the KpnI and NotI restriction sites, and highlighted in green is the kosak consensus sequence.

oSDR #	Protein	Primers
1099	Prp8	**TACTTCCAATCCCACG**AGGAGAAATTAACTATGCCCAAACGTGCG
1100	Prp8	**TTATCCACTTCCCACG**TCAAGTTCCCTCTTCGAT
1101	Snu114	**TACTTCCAATCCCACG**AGGAGAAATTAACTATGAGTTCAGCGTTTCG
1102	Snu114	**TTATCCACTTCCCACG**TCAGAGGTCGGTCCC
1103	Brr2	**TACTTCCAATCCCACG**AGGAGAAATTAACTATGCCTCAGGAACCT
1104	Brr2	**TTATCCACTTCCCACG**TCAGATACTCGGATCCGC
1080	Dib1	**TACTTCCAATCCCACG**AGGAGAAATTAACTATGGACAGTGCACCG
1081	Dib1	**TTATCCACTTCCCACG**CTAGAGTCGGAACGG
1225	Sm B	**TACTTCCAATCCCACG**AGGAGAAATTAACTATGGATCTTCTGCCTGTGC
1226	Sm B	**TTATCCACTTCCCACG**TCATTCAGATGCGGCAGTTTTC
1227	Sm D3	**TACTTCCAATCCCACG**AGGAGAAATTAACTATGAGCGGGTATCGACC
1228	Sm D3	**TTATCCACTTCCCACG**TCACACGTTCCGCCG
1229	Sm D2	**TACTTCCAATCCCACG**AGGAGAAATTAACTATGCCTCCAGTTGATCAGC
1230	Sm D2	**TTATCCACTTCCCACG**TTACGGCTGTGCGCG
1231	Sm D1	**TACTTCCAATCCCACG**AGGAGAAATTAACTATGACCCCCTTGCTTTATTTC
1232	Sm D1	**TTATCCACTTCCCACG**TCAGTGTCTCTCTTTCTGATATCG
1233	Sm E	**TACTTCCAATCCCACG**AGGAGAAATTAACTATGCCGAAGGACGCTC
1235	Sm E	**TTATCCACTTCCCACG**TCACTCCCGAGTCGC
1236	Sm F	**TACTTCCAATCCCACG**AGGAGAAATTAACTATGACTGCGACTGGTTTCG
1237	Sm F	**TTATCCACTTCCCACG**TCAAGATAGAGGCGGGAC
1238	Sm G	**TACTTCCAATCCCACG**AGGAGAAATTAACTATGGCAAAAGACGAGGTCG
1239	Sm G	**TTATCCACTTCCCACG**TCAGCTGCTAAGCGAAGAAC
1240	Sm E-HIS	**TACTTCCAATCCCACGCA**CCGAAGGACGCTCTGG
1241	Sm F-HIS	**TACTTCCAATCCCACGCA**ACTGCGACTGGTTTCGC
1300	Prp8-MBP	**TACTTCCAATCCAATGCA**CCCAAACGTGCGTTTTTC
1301	Prp8-MBP	**TTATCCACTTCCAATG**CTAAGTTCCCTCTTCGATCG
1380	Brr2-pPICZA	CGGATCGGTACCGCCATGGTGCCTCAGGAACCTGAACTAGAA
1381	Brr2-pPICZA	AAGCTGGCGGCCGCTGATACTCGGATCCGCGGT

Table 2 Thermocycler set-up for amplification of protein genes: PCR reactions were set-up for 35 cycles of denaturation, annealing and extension, and the DNA polymerase Q5 *high fidelity* was used for the amplification of the genes.

PCR steps	Prp8	Snu114	Dib1	Brr2	Sms
Initial denaturation	98ºC 2 minutes	98ºC 2 minutes	98 ºC 2 minutes	98 ºC 30 seconds	98 ºC 30 seconds
Denaturation	98 ºC 10 seconds	98 ºC 10 seconds	98 ºC 10 seconds	98 ºC 10 seconds	98 ºC 10 seconds
Annealing	60 ºC 30 seconds	55 ºC 30 seconds	72 ºC 30 seconds	59 ºC 30 seconds	64 ºC 30 seconds
Extension	72 ºC 3.5 minutes	72 ºC 2.5 minutes	72 ºC 20 seconds	72 ºC 2.75 minutes	72 ºC 20 seconds
Final extension	72 ºC 10 minutes	72 ºC 2 minutes	72 ºC 2 minutes	72 ºC 2 minutes	72 ºC 12 minutes

2.2.4 Insertion of the protein genes into the vector by LIC

As seen in figures 2-3 and 2-4, digestion and T4 treatments of the vectors and PCR products needed to be performed prior to ligation of the vector to the PCR product. Therefore, the vectors were digested with PmlI and then a 15 nucleotides overhang is created by treatment of the digested vector with T4 DNA polymerase. Six micrograms of pSR617 and pSR627 were digested for three hours with five µl of PmlI (20 U/µl), and the digested vector was run on a 0.5% Agarose Ethidium Bromide gel for three hours at 100 volts followed by gel extraction. After purification of the digested vectors, creation of the overhang was done by treating 200 nanograms of each vector with 2 µl of dGTP (25mM), 2 µl of 10X T4 buffer, 1 µl of DTT (100 mM), 0.4 µl of (1U) T4 DNA polymerase (New England Biolabs), and Milli Q water to make a final 20 µl reaction volume. The PCR products were also treated with T4 DNA polymerase; however, before T4 treatment, removal of all dNTPs was necessary. Removal of dNTPs during the T4 DNA polymerase treatment allows

for only one dNTP to be present in the T4 DNA polymerase reaction, and thus for the creation of an overhang. Therefore, a PCR clean-up was done using an Omega kit. To create the overhang complementary to the vectors, dCTPs were added to the reaction instead of dGTPs. In regard to Dib, Prp8, and Brr2 genes, 200 fmols of PCR product were treated with 2 µl of dCTP (25mM), 2 µl of 10X T4 buffer, 1 µl of DTT (100 mM), 0.4 µl of (1U) T4 DNA polymerase (New England Biolabs) and Milli Q water was added for creation of a 20 µl reaction volume. For Sm B/B', Sm D3, Sm D2, Sm D1, Sm E3, Sm F, Sm G genes, 250 fmols of PCR products were mixed with 0.5 µl of dCTP (100 mM), 1 µl of 10X T4 buffer, 0.5 µl of DTT (100 mM), 0.4 µl de (1U) T4 DNA polymerase (New England Biolabs) and added Milli Q water to create a total volume of 10 µl.

The T4 DNA polymerase (New England Biolabs) the reactions were incubated at 25^{o} C for 30 minutes followed by inactivation at 70^{o} C for 20 minutes. After creation of the overhangs in both vectors and PCR products, 25 fmols of the T4 treated vector were mixed with 75 fmols of the T4 treated PCR products and incubated at room temperature for 2 minutes. After incubation, 50 µl of DH5α bacteria were transformed with the vectors and incubated for 18 hours. Confirmation of the insertion of genes into pSR617 and pSR627 was done via either colony PCR or digestion.

For insertion of the Prp8 gene into pMCSG23, the same procedure described above was performed since this plasmid also allows LIC. Insertion of Snu114 to pMCSG23 was completed by another lab member, Mona Amin. For insertion of Brr2 into pPICZA, the vector and the PCR product were double digested with KpnI (20/ µl) and NotI (20/ µl), followed by gel purification with a Qiagen QIAquick gel extraction kit. After gel purification, 11 fmols of plasmid was ligated to 40 fmols of the insert by ligation with T4 DNA ligase.

2.2.5 Sequencing of amplified genes

After confirmation of the presence of genes into the vector either by colony PCR or digestion, vectors were sent to the genetics facility at the University of Northern British Columbia (UNBC) for sequencing of the genes inserted. They were checked for the absence of mutations in the gene and confirmation of the correct sequence of the ribosome binding site, start and stop codons using the program CodonCode aligner®.

2.2.6 Construction of a vector for co-expression of the U5-specific proteins

For reconstitution of the *C. merolae* U5 snRNP, co-express and co-purify all U5-specific proteins was performed. For that, all genes were inserted into pSR617, and Sm E3 and F genes were inserted into pSR627 via LIC. Therefore, in this section, all genes will be combined into one vector, pSR627, by LIC (Figure 2-8). As seen in figure 2-2, the combination of genes into one vector is facilitated by the presence of the PacI and SwaI restriction sites. Indeed, consecutive PacI and SwaI digestions of the vectors and T4 treatments. The creation of overhangs (as seen described by Scheich et al. 2007 – Figure 2-2) were done to combine all genes into pSR627, which will contain either Sm E3 or Sm F in frame with HIS_7-tag. For combination of Prp8, Dib1, Brr2, and Snu114 into one gene, 200 nanograms of each gene were added to 1 µl of 10X smart cut buffer (New England Biolabs), 1 µl of 10X BSA, 0.5 µl (5 U) of restriction enzyme (either PacI or SwaI) and added Milli Q water to reach a total volume of 10 µl. PacI digests were incubated at 37° C, and SwaI digests were incubated at 25° C for three hours. After digestion, the enzymes were heat inactivated at 65° C for 20 minutes and digests were treated with 1µl of 1M Tris HCl (pH 8), 0.2 µl of 1M $MgCl_2$, 1µl of 1X BSA, 1 µl of 0.1M DTT, 0.5 µl 100mM of dCTP (PacI digests), and dGTP (SwaI digests), 0.5 µl T4 DNA polymerase (New England Biolabs) and Milli Q for a total

volume of 15 µl. Each reaction was incubated at 25ºC for 30 minutes for T4 DNA polymerase activation, and then it was inactivated at 65º C for two minutes. T4 treated vectors and inserts were then combined and incubated at 65º C for 5 minutes, and slowly cooled to room temperature to allow annealing of the insert into the vector. Two microliters of EDTA were added to the reaction and transformed into 50 µl of DH5α. For the combination of the Sm proteins into one vector, the vectors underwent an overnight restriction digest with PacI or SwaI. Also, five microliters of the digested vector were T4 treated by addition of 0.5 µl of 100 mM DTT, 0.5 µl 100mM of dCTP (PacI digests) and dGTP (SwaI digests), 0.4 µl T4 DNA polymerase, and Milli Q water for a 10 µl total volume. Reactions were incubated at 22º C for 30 minutes and 75º C for 20 minutes. One microliter of both the T4 treated vector and insert reactions were combined, and annealing was permitted to happen at room temperature for five minutes. For confirmation of insertion of genes into vectors, colony PCR or digestion of vectors was performed.

2.2.7 Expression of the U5-specific proteins

2.2.7.1 Expression

The expression of the proteins in bacteria (Rosetta pLysS strain) was attempted by inducing with both IPTG and auto-induction. This *E. coli* strain carries genes for rare tRNAs allowing expression of Eukaryotic proteins. When inducing protein expression through the addition of 1 mM IPTG (Amresco), cells were grown in 10 ml of either Luria Bertani (LB) or 2xYT. Media were supplemented with 34 mg/ml of chloramphenicol to select for plasmids carrying tRNA genes, 100 mg/ml Spectinomycin to for select Snu114 and Prp8-containing pMCSG23, and 50 mg/ml Carbenicillin to select for pQLink vectors. Ten millilitres of culture were added to a 200 ml Erlenmeyer flask for better aeration and incubated at 37º C and 300 rpm until an OD_{600} of 0.4-0.6

was reached. Cells were induced by addition of 1 mM IPTG (1:1000 of the total culture volume). Cultures were incubated shaking at 37° C for 1-4 hours. For auto-induction of proteins, Rosetta pLysS cells were first grown in 1.5 ml of MDG non-inducing media, as described by Studier (2005) in a 50 ml Erlenmeyer supplemented with the same antibiotics described above. The culture was incubated at 37° C and 300 rpm for 24 hours and cell density was checked at OD_{600}. For an OD_{600} 10, 1:1000 of the total volume of the auto-inducing media of non-induced culture (1 OD_{600} unit) was added to 10 ml of ZYM-5052 auto-inducing media, as described by Studier (2005). Media was supplemented with antibiotics and added to a 125 ml Erlenmeyer flask and incubated at 37° C and 300 rpm for 24 hours.

When attempting to express Brr2 in yeast, the instructions from the Invitrogen protocol and Lin-Cereghino et al. (2005) for preparation and transformation of competent Pichia cells were followed. Thus, X-33 strain competent cells were transformed with 3 µg of PmeI digested vector (zeocin-resistant) and plated on YPD supplemented with the zeocin antibiotic. After two days, colonies from YPD plates were chosen and inoculated in a 125 ml Erlenmeyer containing 10 ml of buffered glycerol-complex medium supplemented with antibiotics, as described by Weidner et al. (2010). Cells were grown at 30° C and 300 rpm for 24 hours. The culture was harvested and centrifuged for 5 minutes at 3000 g at room temperature, and the cell pellet was resuspended to achieve an OD_{600} of 1, with buffered methanol-complex medium containing 0.5% of methanol, as described by Weidner et al. (2010). The culture was returned to the incubator and supplemented with methanol every 24 hours for four days.

2.2.7.2 Solubility tests

Expressed proteins were tested for solubility, since purification conditions require soluble proteins. One OD_{600} unit of culture was centrifuged at 17,000 x G at a cold temperature for ten seconds. The supernatant was discarded, and the pellet was suspended in 20 µl of BugBuster protein extraction reagent (Novagen), 1 µl of Benzonase (0.5 units/µl), and 1 µl of lysozyme (0.4 mg/ml). Cells lysis was done at room temperature for 30 minutes and then centrifuged at 17,000 x G for five minutes. The supernatant contained the soluble fraction (protein, if soluble), and the pellet contained the insoluble fraction.

2.2.7.3 Small-scale purification

Batch binding of proteins to Ni^{2+} -NTA (Thermo Scientific) or amylose (New England Biolabs) resin was performed by harvesting and centrifuging cells for 10 minutes at 3000 rpm at 4ºC in a JLA-8.1000 rotor (Beckman Coulter Avanti HP-20 XPI). The resulting cell pellets were washed once by addition of 1.5 ml of buffer A1 (20 mM HEPES-NaOH, pH 7.5, 500 mM NaCl, 20 mM imidazole, 5 mM β-mercaptoethanol) and repeated centrifugation. The cell pellet was snap frozen in liquid nitrogen and stored at -80ºC. Resuspension of cell pellet was performed by the addition of 1.5 ml of buffer A1 and sonicated four times on ice in ten seconds bursts at five W with ten seconds breaks. Sonicated samples were centrifuged for ten minutes at 1,300 rpm at 4ºC. After centrifugation, a Ni^{2+} -NTA or amylose resin was prepared, as described by the manufacturer`s protocol. The soluble sample was transferred to a resin-containing column and centrifuged for 2 minutes at 700G at room temperature, followed by two washes with buffer A1. A third washing was done using buffer A2 (20 mM HEPES-NaOH, pH 7.5, 500 mM NaCl, 60 mM imidazole, 5 mM β-mercaptoethanol). Protein was then eluted from the resin by addition of buffer B1 (20 mM

HEPES-NaOH, pH 7.5, 500 mM NaCl, 500 mM imidazole, 5 mM β -mercaptoethanol) when using Ni^{2+} -NTA resin, or buffer B2 (20 mM HEPES-NaOH, pH 7.5, 500 mM NaCl, 10 mM maltose) when using amylose resin.

2.3 Results

2.3.1 Construction of expression vectors for ligation independent cloning

For reconstitution of U5 snRNP, an expression vector from pQLinkN and pQLinkH expression vectors was constructed that would allow for insertion of the U5-specific protein genes into one vector by LIC. A adaptor sequence that is complementary to the 5` and 3` ends of the amplified genes was inserted. The $pQLinkN^{mod}$ and $pQLinkH^{mod}$ vectors were first double-digested with EcoRI and HindIII, and the EcoRI-HindIII fragment was replaced with oSDR 1070/1071 and oSDR 1072/1073 respectively. Figure 2-5a shows the linearization of $pQLinkH^{mod}$ and $pQLinkN^{mod}$ and removal of a 105 and a 174 base pairs fragments, respectively.

Followed by double-digestion of the vectors, oSDR 1070/1071 and oSDR 1072/1073 were ligated into $pQLinkH^{mod}$ and $pQLinkN^{mod}$ and DH5α cells were transformed using these vectors. The vectors were digested with PmlI for confirmation of insertion of oligonucleotides since pQLink vectors do not contain a PmlI restriction site. As observed in figure 2-5b, the introduction of oligonucleotide was confirmed by linearization of the vectors.

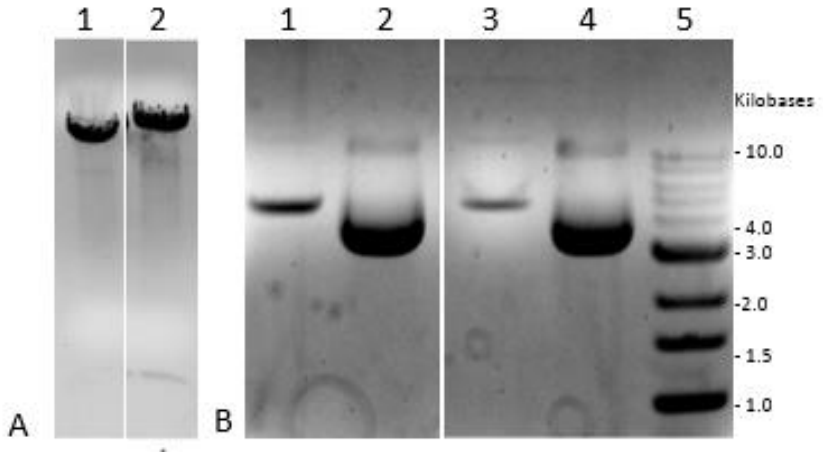


Figure 2-5 Construction of expression vectors: a) A 0.7% Agarose Ethidium Bromide gel presenting the EcoRI-HindIII double digestion of pQLinkN^{mod} (lane 1) and pQLinkH^{mod} (lane 2). b) A 1% Agarose Ethidium Bromide gel presenting digestion of the vector with PmlI. Lanes 1 and 3 show control samples; therefore, pQLinkH^{mod} and pQLinkN^{mod}, respectively, after PmlI digestion. As expected, the vectors do not linearise since PmlI restriction site is absent. Lanes 2 and 4 show pQLinkH^{mod} and pQLinkN^{mod}, respectively, after insertion of PmlI-containing oligonucleotides. Insertion is confirmed by linearization of the vectors running on gel around 4 kilobases as expected, since vector is ~ 4.7 kilobases long.

2.3.2 Amplification of the genomic sequences of U5 snRNA-specific proteins by polymerase chain reaction (PCR)

After construction of the vectors allowing for insertion of genes, each gene containing 5` and 3`ends complementary to the sequences inserted into the vectors was amplified. Figure 2-6 shows on gel amplification of Dib1, Brr2, Prp8, Snu114, Sm B, Sm D2, Sm D1, Sm E, Sm F, Sm G, where successful amplification of genes is confirmed (Table 4). When amplifying genes for insertion into pSR617, pSR627 and pMCSG2, the genes were 46 base pairs longer since the 5` and 3` ends contain sequences for insertion of genes into the vector by LIC.

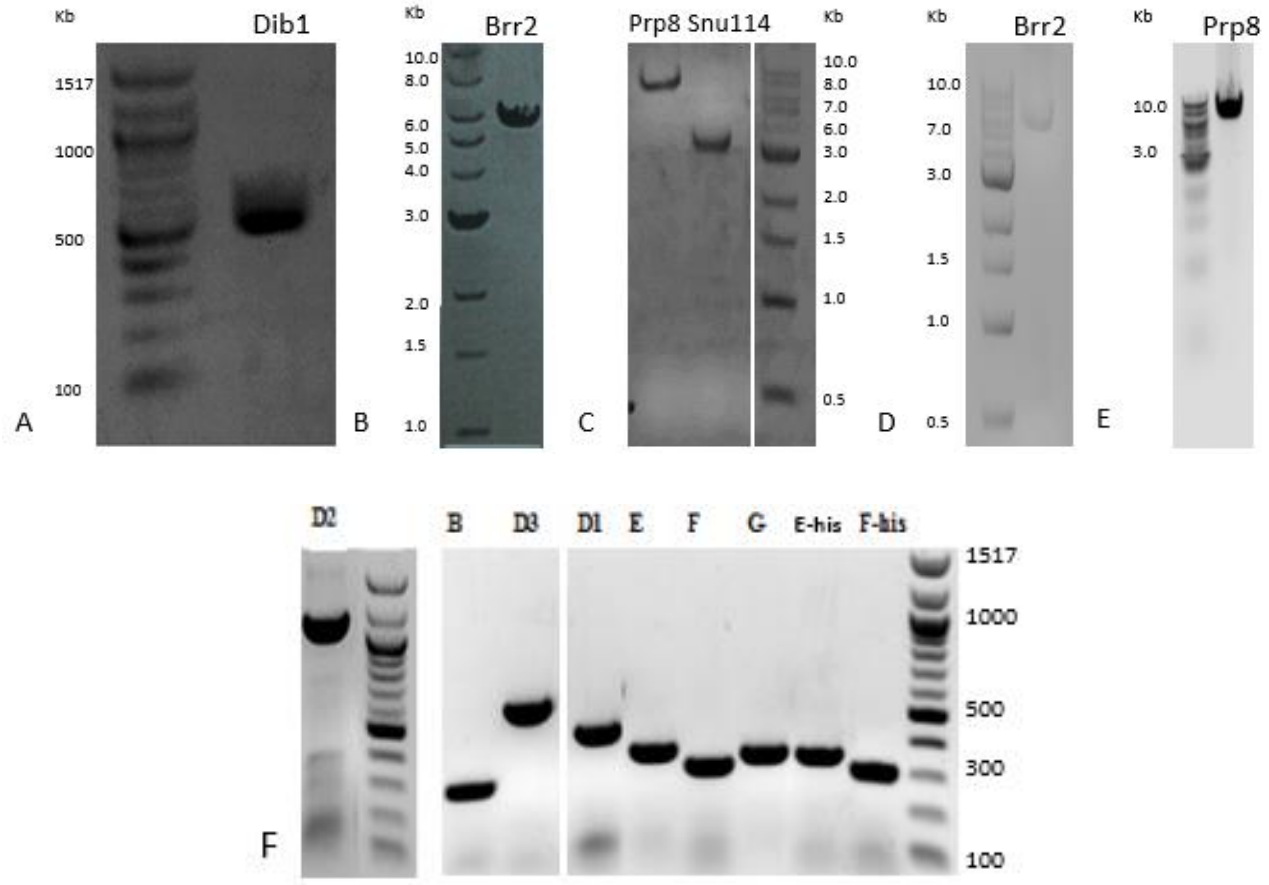


Figure 2-6 Amplification of protein genes: A 1% Agarose Ethidium Bromide gel presenting amplification of Dib1 gene (452 bp) (A) Brr2 gene (5515 bp) (B) Prp8 (7234 bp) and Snu114 (3373 bp) genes (C) Brr2 gene for insertion in Ppicza (D) Prp8 gene for insertion in pMCSG23 (E), and SmB (289 bp), SmD3 (559 bp), SmD2 (1045 bp), SmD1 (451 bp), SmE (364 bp), SmF (319), SmG (346 bp) (F). SmE-HIS and SmF-HIS represent the SmE and F genes that will be inserted into pSR627.

Table 3 Presentation of the size of the protein genes in *C. merolae*.

Protein	Gene size (base pairs)
Prp8	**7188**
Snu114	**3327**
Brr2	**5469**
Dib1	**426**
SmF	**270**
SmE	**315**
SmG	**300**
SmD3	**510**
SmB	**240**
SmD1	**402**
SmD2	**996**

2.3.3 Insertion of the protein genes into the vectors

After construction of expression vectors and amplification of the genes, the genes were inserted into pSR617, pSR627, or pMCSG2 by LIC. A successful insertion was assessed by either digestion or colony PCR. Brr2 was inserted into pPICZA by traditional ligation. For confirmation of insertion of Dib1 into pSR617, non-gene specific primers were used that annealed to a sequence flanking the gene. Therefore, if the gene was not inserted into the vector, a smaller PCR product (~1 kb) would be amplified. Successful insertion of the gene was confirmed by the presence of a 1.5 kb fragment on the gel (Figure 2-7a). For assessment of insertion of Snu114 gene into pSR617, different primers were used that would also anneal to the vector flanking the region of introduction of the gene. Therefore, it was expected to see a PCR product of about ~4 kb, instead of a 223 bp product (Figure 2-7b). For confirmation of the introduction of Prp8 into pSR617, a gene-specific primer was used as well as a primer that binds to the vector; therefore, if insertion of the gene occurred a ~1 kb PCR product would be expected to be observed (Figure 2-7c).

Detection of the insertion of Prp8 gene in the pMCSG2 vector was confirmed by digestion with NdeI, and the band sizes were expected to be 8704 and 3459 base pairs long (Figure 2-7d). For detection of Sm B, Sm D3, Sm D2, Sm D1, Sm E3, Sm F, and Sm G into pSR617, Sm E3-HIS into pSR627, and Sm F-HIS into pSR627, a colony PCR was done using gene-specific primers. As seen in figure 2-7e, the presence of each gene was assessed by amplification of the genes. Insertion of Brr2 in pSR617 was confirmed by linearization of the vector with SalI. A 10 kb fragment was observed on the gel when Brr2 gene was successfully inserted into pSR617, and a 4 kb fragment is seen when Brr2 was not added (Figure 2-7f). For detection of insertion of Brr2 into pPICZA, the vector was digested with ApaI and expected band sizes were 4707 and 4088 base

pairs long. As seen in figure 2-7g, successful insertion of the gene was observed. Confirmation of insertion of the genes into the vector were performed by sequencing of each vector (Appendix 1).

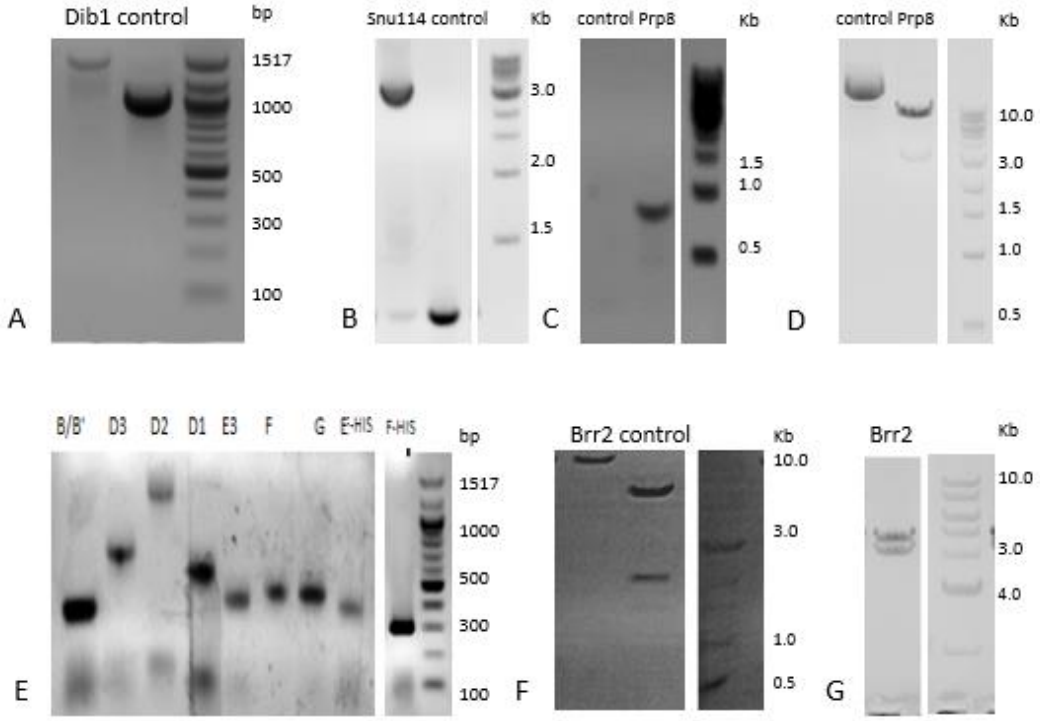


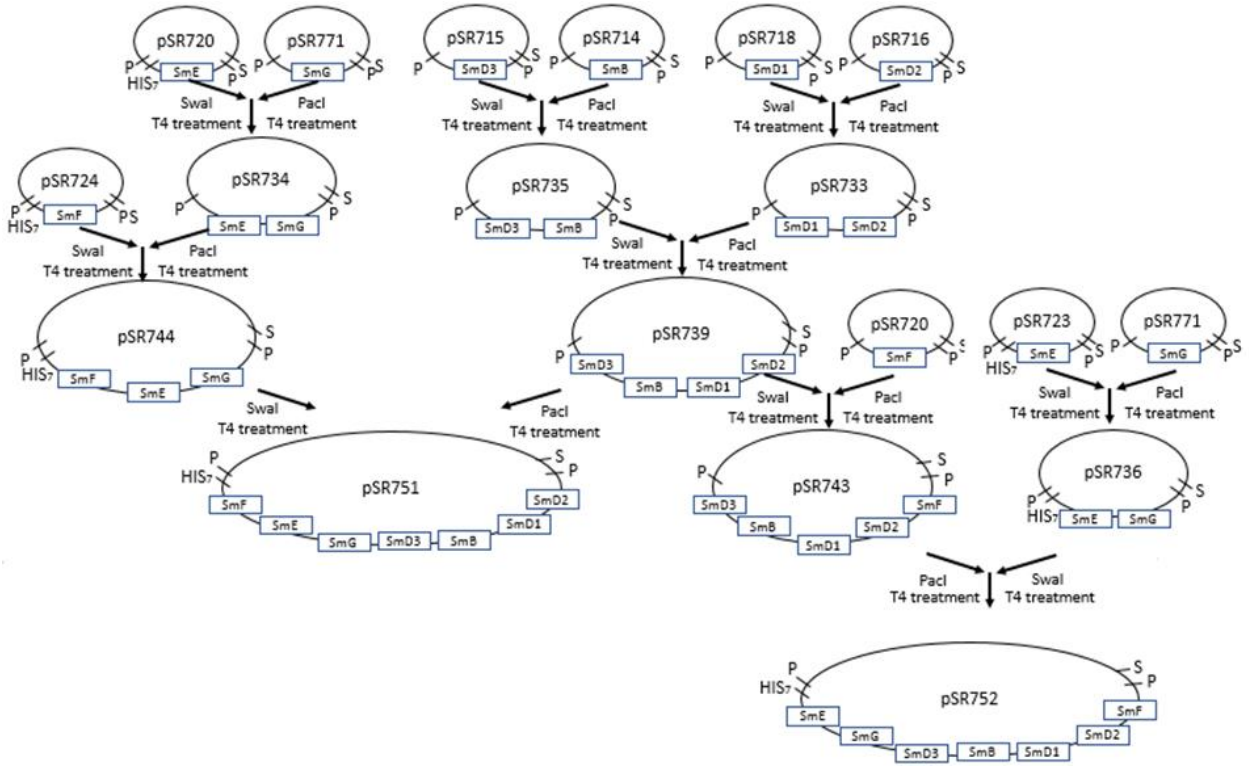
Figure 2-7 Assessment of insertion of protein genes into expression vectors: A 1% Agarose Ethidium Bromide gels presenting confirmation of insertion of all protein genes into the vectors. a) Amplification of ~1.5 kb fragment flanking Dib1 gene confirms the introduction of the gene into pSR617.b) Amplification of ~4 kb fragment flanking Snu114 ensures the presence of gene into pSR617. c) Amplification of Prp8 using gene-specific primers shows the presence of Prp8 gene into pSR617, where a ~1kb fragment is amplified. d) NdeI digestion of vector confirms insertion of Prp8 gene into pMCSG2 presenting ~8 and ~3 kb fragments. e) Presence of all the Sms B, D3, D2, D1, E, F and G is observed since proteins genes were amplified. Sm E-his and Sm-F on the gel represents the Sms E and F that were inserted into pSR627. f) Linearization by SalI digestion confirms the presence of Brr2 into pSR617 since a ~10 kb fragment is observed on the gel. g) ApaI digestion of the vector presents the expected 5 and 4 kb fragments confirming insertion of Brr2 into pPICZA.

Table 4 Construction of expression vectors: After insertion of the genes, each vector was given a name (here called pSR). The Snu114 gene was inserted into pMCSG2 by another lab member. (*) represents the insertion of the gene into pMCSG2. (**) represents the insertion of the gene into pPICZA. (***) represents the insertion of the gene into pQLinkH.

Gene	pSR #	Gene	pSR#	Gene	pSR#	Gene	pSR#
Dib1	634	Brr2**	855	**Sm D3**	715	Sm B	714
Snu114	647	Brr2	656	**Sm D2**	716	**Sm G**	721
Prp8	655	Snu114*	767	**Sm D1**	717	**Sm E*** **	723
Prp8*	797	**Sm F**	720	**Sm E**	719	**Sm F*** **	724

2.3.4 Construction of vectors for co-expression of the U5-specific proteins

After insertion of each protein gene into the pSR617 and pSR627 (one vector with Sm E his tagged and other with Sm F his tagged), the combination of all genes into pSR627 containing either Sm E or Sm F was initiated. Indeed, these two proteins were in frame with six histidines allowing for further purification of the protein complex. LIC step-by-step and confirmation of gene insertions were performed by either colony PCR or restriction digestion of vectors.



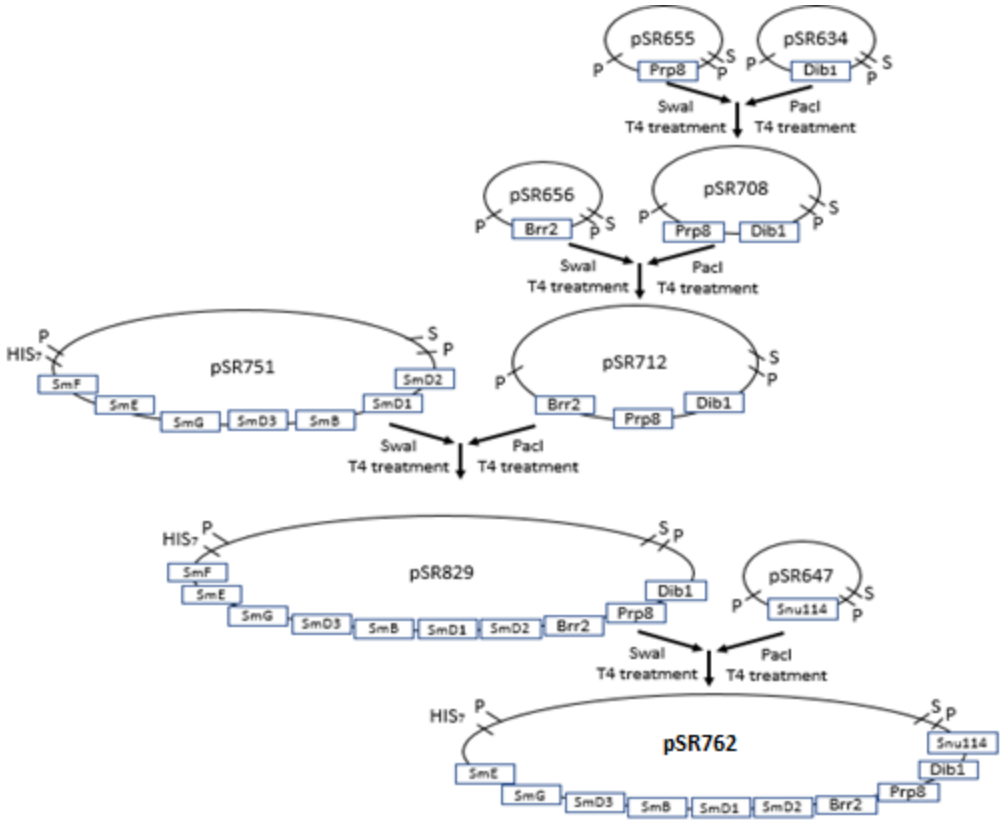


Figure 2-8 Construction of the U5-specific proteins co-expression vector: This scheme presents a step-by-step construction of a co-expression vector. Addition of inserts was done by consecutive SwaI and PacI digestions of vectors followed by T4 DNA polymerase treatments, allowing for the creation of the overhangs and ligation of the inserts into one vector. Each gene contained its own promotor and ribosome binding site.

As seen on the scheme (Figure 2-8), Sms D1/D2, E/G, D3/B and E-His/G vectors were constructed, and assessment of gene insertions was done by colony PCR (Figure 2-9).

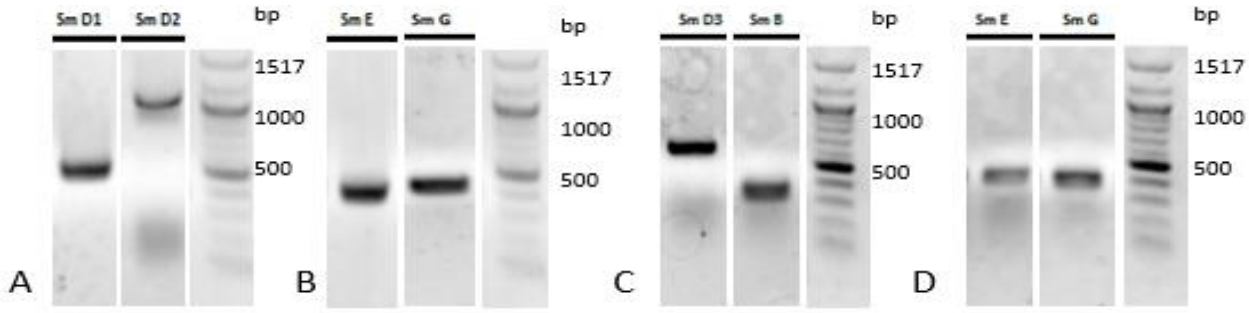


Figure 2-9 First step of construction of Sms-containing vector: Presentation of the amplified genes on a 1% Agarose Ethidium Bromide gel. a) Confirmation of insertion of SmD1 and Sm D2 into pSR617. b) Confirmation of insertion of SmE and SmG into pSR617. c) Confirmation of insertion of SmD3 and SmB into pSR617. d) Confirmation of insertion of SmD3 and SmB into pSR617.

The second step was to combine SmD3/SmB to SmD1/SmD2, and insertions were confirmed by colony PCR (Figure 2-10a). The third part was comprised of two LICs. The vector containing SmD3/B/D1/D2 was combined with Sm F (pSR617), and Sm E3/G was combined with Sm F (pSR627). Construction of the Sm D3/B/D1/D2/F vector was confirmed by colony PCR using Sm F and B primers (Figure 2-10b). To confirm F-HIS/E/G construction, SmF, E and G primers were used (Figure 2-10c).

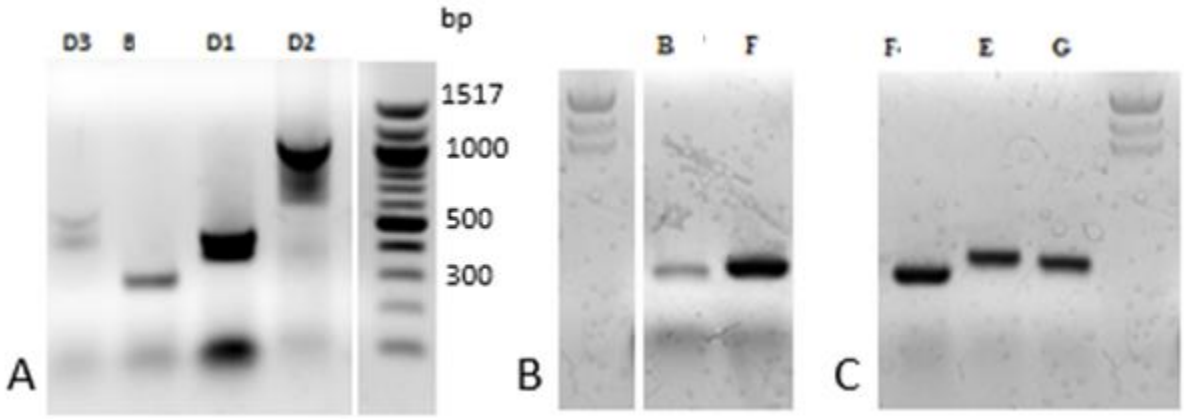


Figure 2-10 Second and third steps of construction of Sms-containing vector: Presentation of the amplified genes on a 1% Agarose Ethidium Bromide gel. a) Confirmation of the construction of a Sm D3/SmB/SmD1/SmD2-containing expression vector by amplification of the genes. b) Confirmation of insertion of Sm D3, B, D1 and D2 genes into Sm F-containing vector (pSR617) by amplification of SmB and F genes. c) Confirmation of introduction of SmE and G into a SmF-HIS-containing vector by amplification of the genes. Markers on figures b and c are not clear; however, insertion is confirmed.

At the last step, insertion of all seven Sms into pSR627 was performed, where Sm E and Sm F were in frame with six histidines. The construct containing SmD3/B/D1/D2/F was ligated to the SmE3-his/G construct, the construct containing SmF-HIS/E3/G was ligated to the SmD3/B/D1/D2 construct. For confirmation of these ligations, SmE and D3 genes were amplified from the SmE3-his/G/D3/B/D1/D2/F-containing vector (Figure 2-11a), and Sm F-HIS and B genes from the Sm F-HIS/E3/G/D3/B/D1/D2-containing vector (Figure 2-11b). A second confirmation of the

presence of all the Sm genes into the vectors is presented in figures 2-11c and d. SmE3-his/G/D3/B/D1/D2/F-containing vector was named pSR752, and Sm F-HIS/E3/G/D3/B/D1/D2-containing vector was named pSR751 (Table 6).

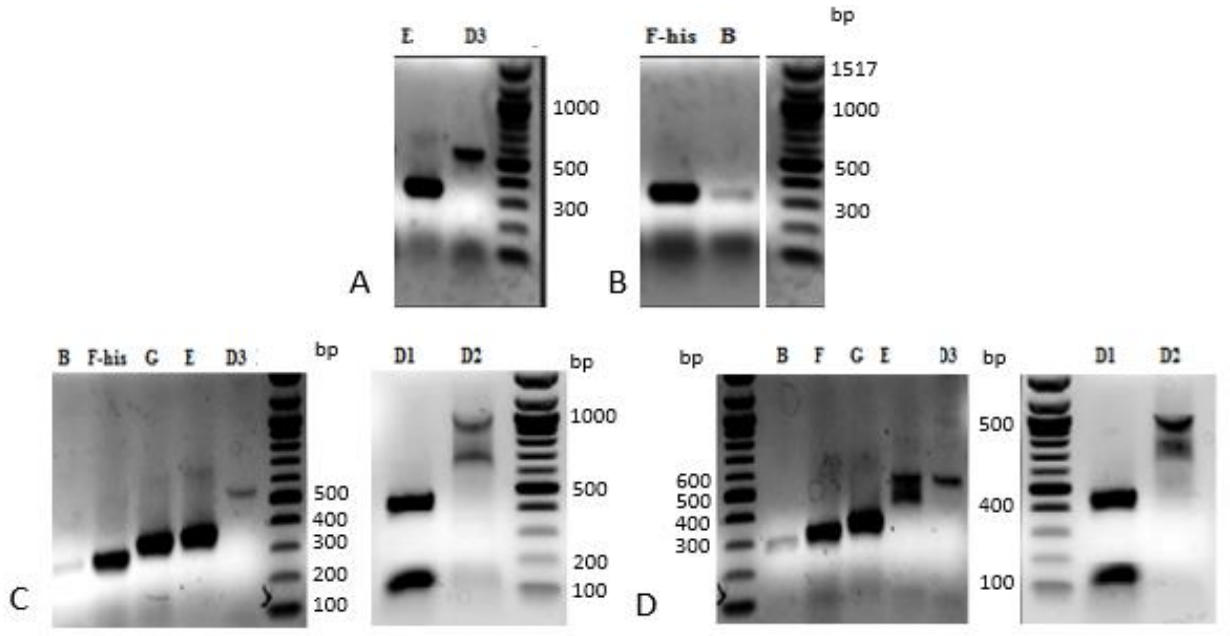
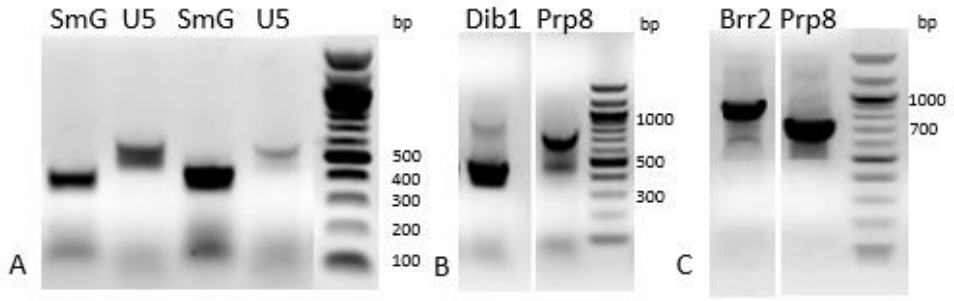


Figure 2-11 Last step of construction of Sms-containing vector: Presentation of the amplified genes on a 1% Agarose Ethidium Bromide gel. a) Amplification of the SmE and D3 genes confirmed the introduction of the SmD3, B, D1, D2, and F genes into the SmE-his/G construct. b) Amplification of the SmF and B genes confirmed the presence of the SmD3, B, D1, and D2 genes in the SmF-HIS/E/G construct. c) Construction of the F-HIS/E/G/D3/B/D1/D2-containing vector was confirmed by amplification of all the genes. d) Construction of the SmE3-his/G/D3/B/D1/D2/F-containing vector was confirmed by amplification of all the genes.

After construction of vectors containing all seven Sm genes, the U5 snRNA gene was combined to both vectors, pSR751 and 752. The U5 snRNA-containing vector (pSR660) was previously constructed by Kirsten Reimer, a former lab member. Insertion of U5 gene to both vectors was confirmed by amplification of SmG and U5 genes (Figure 2-12a). In the last step of constructing a vector containing all U5-specific protein genes, Prp8, Dib1, and Brr2 were inserted into pSR753 and 755 (Table 6). The Prp8 gene was first combined to the Dib1-containing vector,

followed by insertion of Brr2 to this vector. The first ligation was confirmed using Prp8 and Dib1 primers (Figure 2-12b), and for the second construct (pSR712) Prp8 and Brr2 primers were used (Figure 2-12c), Prp8 primers that would amplify only 718 base pairs of the gene were used; therefore, the full-length is not seen on the gel.

The last step was to combine the Brr2/Prp8/Dib1 construct to pSR753 and 755. The pSR712 construct to pSR753 was successfully ligated; however, ligation of pSR712 to pSR753 failed multiple times. Figure 2-12d shows amplification of U5 and Dib1genes to confirm the ligation of pSR712 to pSR753. As previously described, it was deemed appropriated to have two strategies to purify the complex. The first approach was to have all U5-specific protein genes in one vector having either a Sm E or Sm F tag. Another construct would not have the Snu114 gene, and this gene would be inserted into a different plasmid in frame with an MBP gene. Bacteria would be transformed with both vectors and a two-step purification using both HIS and MBP tags would allow purification of the complex. Thus, the Snu114 gene was the last gene to be inserted into the final construct (pSR762). Confirmation of insertion of that gene to this vector was performed by amplification of Prp8, Brr2, Snu114, U5, and Sm G (Figure 2-12e). For amplification of Prp8, Brr2 and Snu114, primers were used that partially amplified the gene, with bands expected to be 718, 958, and 1421 base pairs.



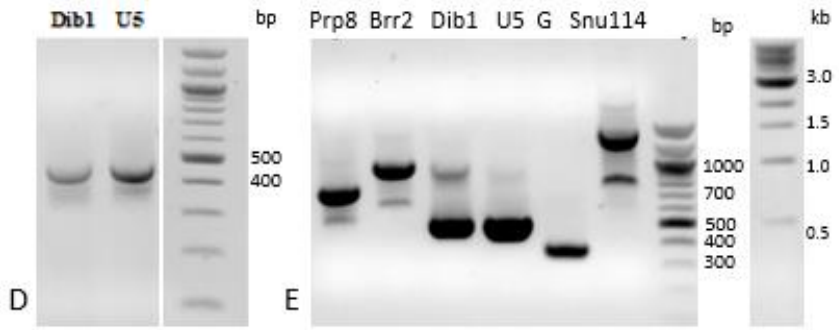


Figure 2-12 Step-wise construction of the vector containing all U5-specific protein genes: Presentation of the amplified genes on a 1% Agarose Ethidium Bromide gel. a) Amplification of the SmG e U5 genes for confirmation of insertion of U5 to Sms-containing vectors. Amplification of U5 was performed using oSDR1125 and oSDR 1126 primers; therefore, an expected 410 bp fragment is observed. Lanes 1 and 2 display the insertion of U5 to pSR751 and pSR752, respectively. b) Combination of Prp8 and Dib1 genes into one vector was confirmed by amplification of Prp8 and Dib1 genes, and bands of 718 and 426 bp were observed. c) Confirmation of insertion of the Brr2 gene into the Prp8/Dib1-containing vector is confirmed by amplification of the Prp8 and Brr2 genes, and the correspondent bands of 958 and 718 bp are seen. d) Insertion of Prp8, Brr2, and Dib1 genes into the U5/Sms-containing vector is confirmed by the presence of amplification of the Dib1 and U5 genes. e) Insertion of Snu114 into the final construct was confirmed by the presence of Ppr8, Brr2, Dib1, U5, SmG, and Snu114 genes.

Table 5 Construction of expression vectors containing U5-specific protein genes: Each vector used for the construction of expression vector was named pSR followed by a number.

Gene-containing construct	pSR#
Sm D1/D2	733
Sm D3/B	735
Sm E/G	734
Sm E-HIS/G	736
Sm D3/B/D1/D2	739
Sm D3/B/D1/D2/F	743
Sm F-HIS/ E/G	744
Sm E-HIS/G/ D3/B/D1/D2/F	752
Sm F-HIS/ E/G/ D3/B/D1/D2	751
Sm E-HIS/G/ D3/B/D1/D2/F/U5	755
Sm F-HIS/ E/G/ D3/B/D1/D2/U5	753
Sm Prp8/Dib1	708
Sm Brr2/ Prp8/Dib1	712
Sm F-HIS/ E/G/ D3/B/D1/D2/U5/Brr2/ Prp8/Dib1	829
Sm F-HIS/ E/G/ D3/B/D1/D2/U5/Brr2/ Prp8/Dib1/Snu114	762

2.3.5 Expression and solubility tests

Once construction of expression vectors containing the genes of the proteins that associate to U5 snRNA was accomplished, co-expression of the proteins using the vectors pSR767 and pSR829 was attempted. The vector pSR767 was constructed by Mona Amin, a former lab member and contains the Snu114 gene in frame with the MBP gene and six histidines. Expression the protein by IPTG induction in bacteria was confirmed based on the proteins molecular weights (Table 7). As seen in figure 2-13a, protein expression in both soluble and insoluble fraction could not be seen. Protein expression after 1, 2, 3, and 4 hours of induction was assessed and no differences between the non-induced sample and IPTG induced samples were observed (Figure 2-13b).

Table 6 Presentation of the molecular weight of the proteins in *C. merolae*.

Protein	Approximate molecular Weight (kDa)
Prp8	**274**
Snu114	**122**
Brr2	**205**
Dib1	**16**
SmF	**10**
SmF +HIS_7 tag	**12**
SmE	**12**
SmE +HIS_7 tag	**14**
SmG	**11**
SmD3	**19**
SmB	**9**
SmD1	**15**
SmD2	**36**

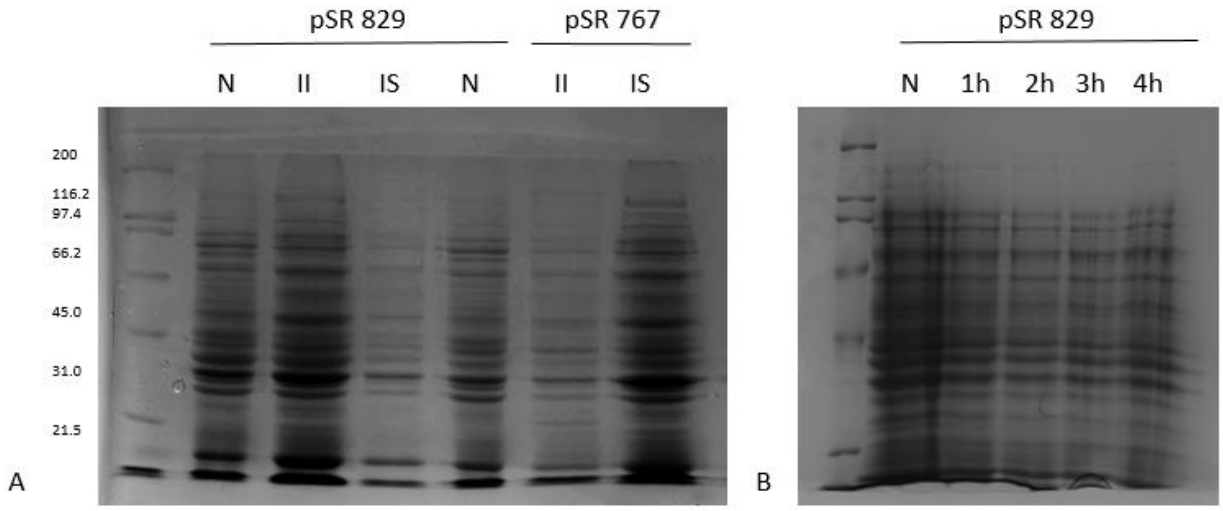


Figure 2-13 IPTG induction of the U5-specific proteins: 8% SDS PAGE gel presenting one of the attempts to express all U5-specific proteins in Rosetta pLysS. When using the vector pSR829 the proteins Prp8 (274 kDa), Snu114 (122 kDa), Brr2 (205 kDa), Dib1 (16 kDa), SmF (10 kDa), SmF with His_7 (12 kDa), SmE (12 kDa), SmE with His_7 (14 kDa), SmG (11 kDa), SmD3 (19 kDa), SmB (9 kDa), SmD1 (15 kDa), SmD2 (16 kDa) were expected to be co-expressed. When using the vector pSR767 the protein Snu114 with MBP (164.5 kDa) was expected to express. a) No changes between non-induced and induced lanes on gel confirms no expression of proteins after 4 hours induction by addition of 1mM IPTG. b) The gel shows no expression of proteins after 1, 2, 3 and 4 hours induction by addition of 1mM IPTG. (N) non-induced, (II) induced, and insoluble material, (IS) induced and soluble material.

The negative results for co-expression of these proteins requires changing the initial strategy. It was deemed most suitable to check for expression of each protein individually since vectors containing each protein genes were available (Table 5). Protein expression was tested using different constructs, such as pSR751, pSR752, and pSR708. Protein expression was attempted by by either IPTG or auto-induction. As seen on acrylamide gels, expression of Prp8 either individually (pSR655 – Figure 2-14a) or in the presence of Dib1 (pSR708 – Figure 2-14b) was not observed, since it was expected to be seen on the gel a protein of 274 kDa. For this protein, Prp8 was inserted into a different vector fused to MBP (pSR797), and the resulting construct was used. However, expression of Prp8 when induced by IPTG induced was also not observed (Figure 2-14c).

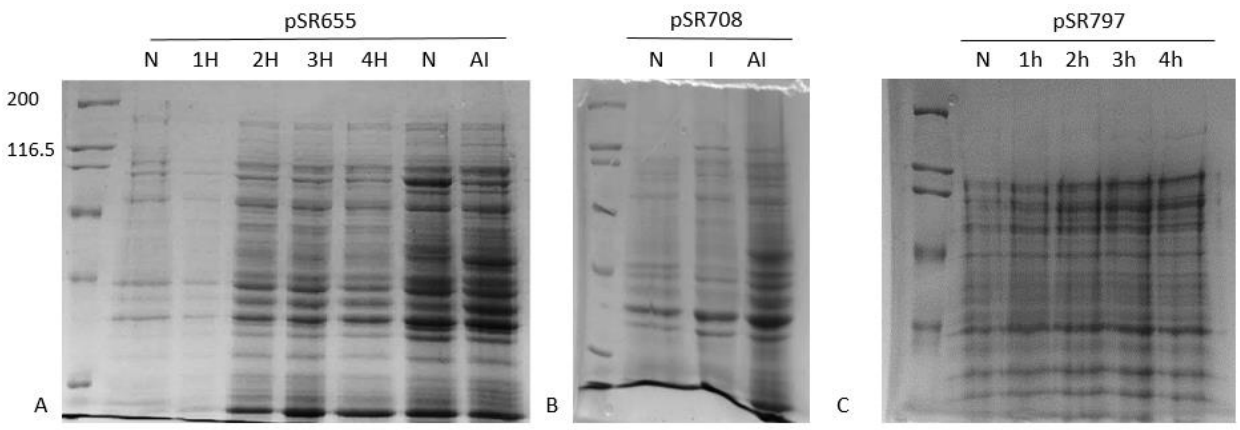


Figure 2-14 Induction of expression of Prp8 in Rosetta pLysS: A 8% SDS PAGE gel presenting different attempts to express Prp8 in Rosetta pLysS. a) No changes between non-induced and induced lanes on the gel confirmed that Prp8 (274 kDa) was not expressed by either IPTG induction or auto-induction. Expression of Prp8 by addition of 1 mM IPTG was checked after 1, 2, 3 and 4 hours of induction. b) No expression of Prp8 (274 kDa) was also observed after to co-expressing Prp8 and Dib1 by either IPTG induction or auto-induction. c) No expression is observed when expressing Prp8 joined to MBP (289.5 kDa) after 1, 2, 3 and 4 hours of induction. (N) non-induced, (AI) auto-induced, (I) IPTG induced.

Expression of Snu114 was also checked using the vector pSR767. This vector (constructed by Mona Amin, a former lab member) contains the Snu114 gene in frame with the MBP gene and six histidines. Snu114 was presumably expressed after adding 1 mM IPTG to the culture since a protein below 200 kDa was observed on the gel (Figure 2-15a). Due to the low protein expression, four different expression conditions were attempted to improve the yield (Figure 2-15b). First, inductions at both 1 mM IPTG and 0.1 mM IPTG were attempted. The second strategy was to grow the culture overnight in non-inducing media followed by dilutions of the cultures and an induction with 1 mM IPTG or auto-induction. In addition to this, auto-induction was also tried after the cells had been grown for 24 hours in non-inducing media. As observed on the gel, a band close to the size of the Snu114 protein is observed; however, any improvement in protein expression is presented (Figure 2-15b).

To confirm expression of Snu114, protein purification was performed by batch-binding after protein expression by a 1 mM IPTG induction (Figure 2-15c). Since the construct allows the

protein to have two tags, His and MBP, purification of Snu114 was attempted using two resins: nickel and amylose. Batch binding utilising a nickel resin is appropriated with His tag, and MBP tag have an affinity to amylose resin. Although a band below 200 kDa is observed in induced, soluble and flow through lanes, no protein in the elution sample is seen. Therefore, the protein found is not binding to either nickel or amylose resins.

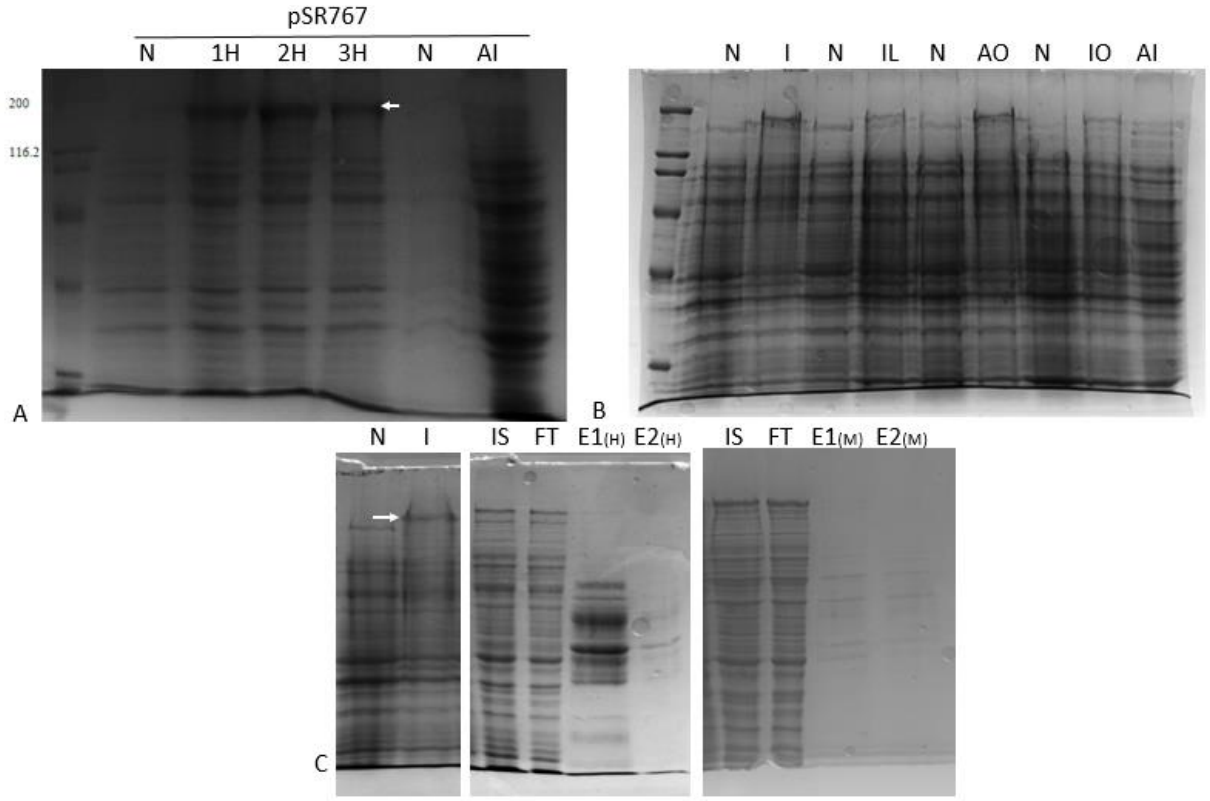


Figure 2-15 Expression and Purification of Snu114 fused to MBP: A 8% SDS PAGE gel presenting attempts to express Snu114 in Rosetta pLysS. a) Induction of expression of Snu114 with 1 mM IPTG after 1, 2, 3 hours shows expression of a ~200 kDa protein (white arrow). Auto-induction does not show any protein highly expressed around 160 kDa. b) Different attempts to increase the yield of protein expressed shows no improvement in protein expression. c) Purification of Snu114 by batch-binding indicates that the protein observed on induced, soluble and flow-through lanes (white arrow) does not bind to either nickel or amylose resins. (N) non-induced, (AI) auto-induced, (I) 1 mM IPTG induced, (IL) 0.1 mM IPTG induced, (AO) non-induced cultures grown overnight and auto-induced, (IO) non-induced cultures grown overnight and 1 mM IPTG induced, (IS) soluble material, (FT) flow-through, ($E1_H$) first elution with 500 mM imidazole, ($E2_H$) second elution with 500 mM imidazole, ($E1_M$) first elution with 10 mM maltose, ($E2_M$) first elution with 10 mM maltose.

Attempts to express Brr2 were made in two different expression systems: *E. coli* and *P. pastoris*. First, Brr2 expression in bacteria by both IPTG and auto-induction was tried. As seen on the gel, a band is observed on the induced lane (white arrow in Figure 2-16a); however, it was expected to be above the 200 kDa mark, since Brr2 is a 205 kDa protein. Brr2 expression was also tried in another expression system. Yeast was transformed with pSR855 and Brr2 expression was checked after induction with methanol at the 24, 48, 72, 96 hours time points. Expression of Brr2 on the gel could not be observed; however, on all gels the presence of a methanol-induced protein with a molecular weight of ~70 kDa could be observed according to the marker used (Figure 2-16b).

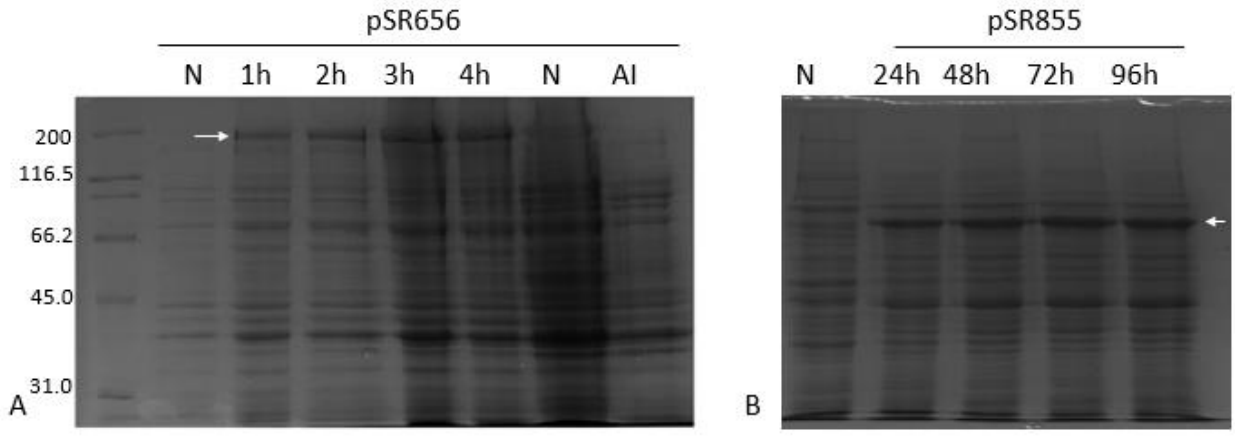


Figure 2-16 Induction of expression of Brr2 using two expression systems: a) A 8% SDS PAGE gel presenting attempts to express Brr2 in Rosetta pLysS. No proteins are observed at 1, 2, 3 and 4 hours after induction via 1 mM IPTG and auto-induction. b) No expression of Brr2 is found when expression was methanol-induced in X-33 strain for 24, 48, 72 and 96 hours. However, expression of a ~70 kDa protein (white arrow) was observed. (N) non-induced, (AI) auto-induced.

For expression of Dib1, bacteria were transformed using two different constructs pSR708 and pSR745 (construct contains Dib in frame with a HIS_7 tag, which was made by Maya De Vos, a former lab member). Both IPTG and auto-induction were tried. When attempting to co-express Dib1 with Prp8, a ~16 kDa protein band was not observed on the induced lanes (Figure 2-17a).

However, when trying to induce expression of Dib1 alone with a His_7 tag by addition of 1 mM of IPTG, the presence of a band near the molecular weight expected was observed (~18 kDa including His tag) (Figure 2-17b). Surprisingly, Dib1 is not expressed by auto-induction. Followed by strong expression of Dib1, the solubility of the protein was checked. As seen in figure 2-17c, the presence of the protein in both soluble and insoluble fractions is displayed.

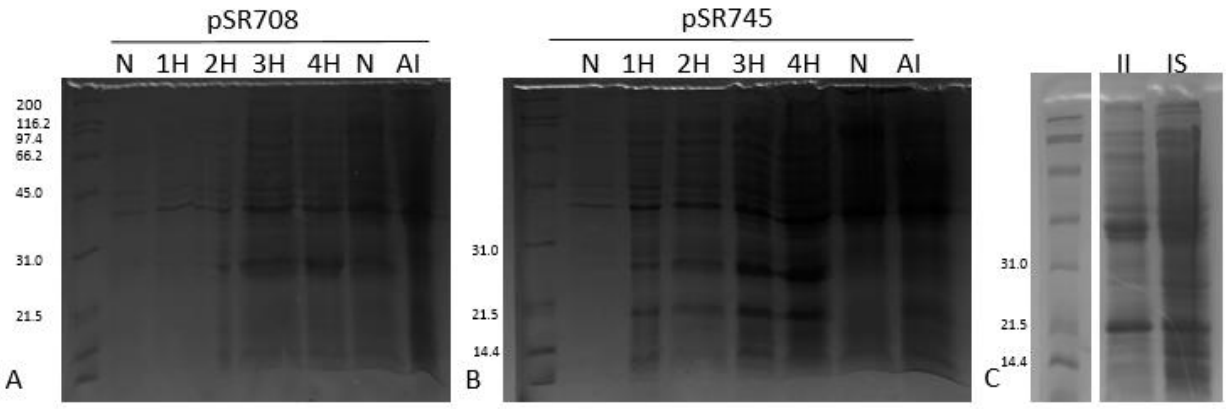


Figure 2-17 Induction of expression of Dib1 by IPTG induction and auto-induction: A 12% SDS PAGE gel presenting the expression of Dib1 in Rosetta pLysS. a) Dib1 is not induced by either IPTG or auto-induction when co-expressed with Prp8. A ~16 kDa protein is not observed on the gel. b) Dib1 expressed individually attached to a HIS_7 tag with IPTG. c) Solubility test shows the presence of Dib1 in both insoluble and soluble material. (N) non-induced, (AI) auto-induced, (II) induced and insoluble material, (IS) induced and soluble material.

Successful co-expression of the Sm proteins by both IPTG and auto-induction was observed on gels. Both constructs, having Sm E and Sm F HIS_7 tag, showed an excellent protein expression using both methods. However, auto-induction presented a better protein expression (Figure 2-18a). Furthermore, co-purification of the proteins using both construction containing tagged SmE or SmF proteins was performed by batch-binding. In figure 2-18b, bands corresponding to the sizes of the Sms are observed on eluted fractions. Since not only the tagged proteins are seen on gels,

but other Sm proteins, it suggests that the proteins are interacting with each other allowing co-purification. However, these results do not guarantee the presence of all seven proteins.

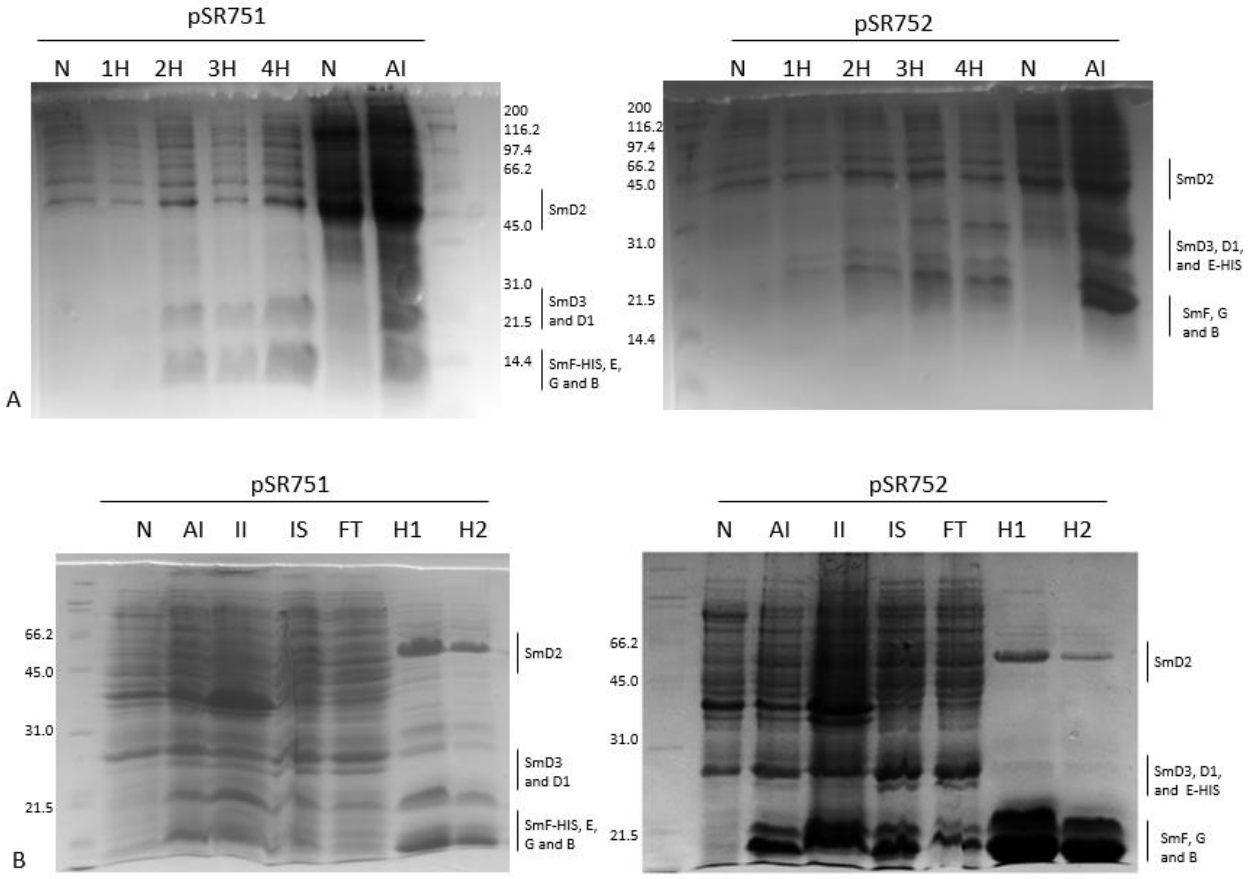


Figure 2-18 Co-expression and purification of the Sm proteins: A 8% SDS PAGE gel presenting the expression of Sm proteins in Rosetta pLysS. a) Expression of proteins using both pSR751 and 752 is observed by IPTG and auto-induction. After 4 hours of induction a better expression of protein is seen; however, auto-induction presents a higher protein expression. b) Purification of proteins using nickel resin shows binding of the Sm to the resin suggesting the interaction of the proteins. (N) non-induced, (AI) auto-induced, (II) insoluble material, (IS) soluble material, (FT) flow-through, (H1) first elution with 500 mM imidazole, (H2) second elution with 500 mM imidazole.

2.4 Discussion

This chapter describes different attempts to reconstitute the *C. merolae* U5 snRNP. Indeed, the first strategy proposed in this section was to clone all protein genes into an expression vector for co-expression and purification of the complex. Although insertion of the genes into the

expression vector was successful, the different attempts to co-express all eleven proteins together failed. Two different approaches were attempted to express these proteins in bacteria. IPTG induction and auto-induction and different protein expression conditions were also attempted since recombinant expression of these *C. merolae* proteins has never been done before. An analysis of protein expression at different time courses was also performed. Protein expression assessment was done after 1, 2, 3, and 4 hours after addition of IPTG to cells, since time-life and stability of proteins in the culture conditions are unknown. Failure of all these attempts made it necessary to develop an alternative approach. Therefore, expression of Brr2, Prp8, Snu114 and Dib1 individually by both IPTG and auto-induction was performed. In addition, it was checked for expression of Snu114 and Prp8 by fusion of these proteins to MBP.

At first, it was believed that Snu114 fused to MBP was being expressed, albeit it in a low yield. Therefore, several conditions to increase protein expression levels were tried. However, a considerable increase in protein yield was not observed. Batch binding was attempted since the protein had both His and MBP tags, so nickel and amylose resins were used (Figure 2-15b). As presented in the results section, purification of the protein using the nickel resin was not successful. The expected band is not seen on the elution fraction although it is present in both soluble and flow-through fractions. This result suggests that the band observed on induction lane on the gel was not Snu114. Failure of purification of the protein by batch-binding using the amylose resin confirmed that the protein observed was not the tagged Snu114. Since the same band was seen when trying to express Brr2, it suggests that the protein found on the gel could be some bacterial protein.

As expression of Brr2, Snu114 and Ppr8 failed in *E. coli*, it was proposed that expression of these proteins in another organism could be an alternative to solve the expression problems.

Therefore, expression of these proteins in yeast was attempted, and the Brr2 gene was inserted into pPICZA for expression in *P. pastoris*. Unfortunately, expression of Brr2 in this organism also failed. Surprisingly, the presence of a protein of approximately 70 kDa was observed. The presence of that band raised some questions regarding the failure in integration, disruption of the AOX1 gene, expression of a truncated Brr2, and expression of yeast proteins that were methanol-induced. Since cells were able to survive in media supplemented with Zeocin, it suggests that the linearised vector was integrated into the genome because the cells are resistant to Zeocin. However, it was not assessed if integration of the gene occurred at the AOX1 locus. The X-33 strain has a Mut^+ phenotype, and the presence of the AOX1 enables normal growth in methanol-containing media. According to the literature, crossover integration into the AOX1 locus frequently happens (50-80% frequency), permitting survival of cells in methanol-containing culture (Li et al. 2007). However, there is a 10-20% chance that integration of the gene has disrupted the AOX1 gene forcing the cell to rely on a weak AOX2 gene. This event also allows survival of cells in the methanol-containing media, resulting in a Mut^s phenotype (Li et al. 2007). Therefore, it suggests that failure of Brr2 expression could be due to the disruption of the AOX1 gene, and the presence of AOX2 on the X-33 strain was enabling survival of the cell in methanol-enriched media. However, on all the gels, expression of an unknown protein is observed when methanol is added to the media, which implies that the AOX1 is not disrupted and is controlling transcription (Figure 2-16b).

If the AOX1 gene is in fact not interrupted, the protein present on the gel could be a truncated form of Brr2 or some yeast protein that is induced by methanol. Indeed, by searching for genes that are present in methylotrophic organisms on UniProt, it was found that not only the AOX 1 and 2 genes are induced by methanol, but also DAS1 and 2, which encode for dihydroxyacetone

synthase proteins. Indeed, all these genes encode for proteins between 60 to 80 kDa, which could explain the presence of a protein being expressed in methanol-containing media. However, further exploration would be necessary to understand why Brr2 is not being expressed in this organism, or if a truncated form of Brr2 is being expressed. Another approach that should be consider to ensure integration of the gene at the right locus is to design primers for the amplification of the 5 ` end of the AOX 1 gene along with the Brr2 gene. Screening of more colonies on zeocin is also an alternative to increase the chances of finding colonies that have an intact AOX1 gene.

Expression of Brr2, Prp8 and Snu114 failed. The reasons that all the attempts to express these proteins in *E. coli* failed are not confirmed. Expression of large proteins has been found to be challenging due to their complex folding and lack of stability. Therefore, one proposed reason for the failure of expression of these large proteins is the presence of rare codons. Indeed, *in vivo* and *in vitro* biochemical studies have shown that codons can be related to translation efficiency and mRNA stability. Thus, expression of some proteins in *E. coli* can be compromised when codons needed for translation of protein are rare (Boël et al. 2016). In fact, large proteins increase the chances of having rare codons involved in translation. In spite of the fact that the Rosetta pLysS strain possesses genes for rare tRNAs, translation efficiency and mRNA stability could be presumably compromised. A bioinformatical tool called GenScript was used to verify the presence of rare codons on these three large proteins. This tool considers two parameters: codon adaptation (CAI) and codon frequency distribution (CFD). The first parameter is related to the distribution of codon usage frequency along to the length of the protein gene to be expressed in a specific host. CAI values below 0.8 represent a reduced expression of the gene in a specified host. The second parameter considers the percentage of distribution of codons. CFD values below 30% express a reduced efficiency of translation in the chosen host. Thus, this tool was used to check for CAI and

CDF values of these three genes, when expressed in *E. coli* cells. All genes presented low CAI and CDF values suggesting a low translation efficiency (Table 8). Therefore, for future attempts to express these proteins in either bacteria or yeast, it would be useful to do codon optimisation. It might increase translation levels and RNA stability making protein expression more promising. Another alternative to solve protein expression and solubility issue would be to express truncations of each protein, this would allow stability and interactions among proteins, as described in previous publications.

Table 7 Presentation of the CAI and CDF calculated by GenScript, based on the DNA sequence of the *C. merolae* proteins.

Protein	CAI	CDF
Brr2	0.62	15%
Snu114	0.63	13%
Prp8	0.65	13%

Only Dib1 was expressed individually by both IPTG and auto-induction, and also displayed high solubility. The Sm proteins also presented strong expression by auto-induction. It was not surprising that co-expression of the Sms would result in solubility of the complex. It has been reported previously by Kambach et al. (1999) and Zaric et al. (2005) that expression of the Sm and LSm proteins individually has shown significant instability of the proteins due to the hydrophobicity of the β4 and β5 strands. Indeed, instability of the Sms was solved by co-expression of Sm pairs burying the hydrophobic strands of each other from solvent (Kambach et al. 1999). Therefore, the results regarding expression and solubility of the Sm complex enabled the continued investigation of the Sm complex in *C. merolae*, which will be described in chapter 3.

To conclude, the attempts presented in this chapter will be an excellent reference to propose new strategies to reconstitute the U5 snRNP or to express each of the proteins individually.

3. Chapter Three - Structural and functional studies of C. merolae Sm complex

3.1 Introduction

As previously discussed, the Sm complex has been well investigated in other organisms. This splicing complex is comprised of seven Sm proteins that make up three distinct subunits: Sm E-F-G, Sm D1-D2, and Sm B-D3 (Raker et al. 1996). Binding of the subunits to the snRNAs occurs in a step-wise manner coordinated by a protein complex called SMN (Fischer et al. 1997; Liu et al. 1997; Meister et al. 2001; Pellizzoni et al. 2002). This protein is responsible for the formation of a snRNP core particle called the Sm core RNP. Indeed, the Sms have been found to play a crucial role in both biogenesis and recruitment of recruitment of snRNA particles. Although this protein complex has been found bioinformatically in *C. merolae*, nothing is known about its structure and functionality in the organism (Stark et al. 2015). It is likely that the uridyl-rich sequences in U2, U4 and U5 snRNA, similar to the consensus Sm binding site, facilitate the binding of the Sm complex to these snRNAs. Since the SMN protein complex is not present in *C. merolae,* it raises some questions regarding assembly of the Sm proteins. The absence of this assembly factor suggests self assembly of the Sm proteins prior to binding to the snRNAs. As well, the absence of a stem-loop 3` of the Sm site supports this idea, since the complex could move along the 3`end of the snRNA in order to bind to the Sm site. The main objective of this chapter will be to investigate the function, structure and assembly of the Sm complex in *C. merolae*.

In the previous chapter, the construction of vectors containing all seven Sm genes were described. Two vectors were constructed, with each one containing either Sm E or F in frame with a HIS_7 tag for further co-purification of the complex. After accomplishing this, the Sm proteins were successfully expressed through auto-induction and displayed a high degree of solubility. Here, a two-step purification of the protein complex will be described. The first step involves

nickel affinity chromatography (IMAC). Due to the great affinity of the HIS_7 tag to nickel, a nickel column can be used to separate the protein complex from bacterial proteins. Retrieval of the protein from the nickel column can be done by adding a high concentration of imidazole. Since imidazole is a histidine competitor, it replaces the His-tagged protein by binding to the column allowing elution of the protein. A second purification was applied called size exclusion chromatography. Since this method involves purification of particles by size, it was applied to separate fully assembled complexes from partially assembled complexes. To do this, the IMAC purified protein is applied to a column comprised of pores, which allows small particles to be retained momentarily, while the big particles run freely in the column resulting in premature elution.

After purification of the complex, it was necessary to confirm the presence of all seven Sm proteins in the purified sample. Therefore, the Sm complex was characterized by mass spectrometry (MS). Briefly, this method involves digestion of the protein by proteases, such as trypsin, and the fragmented peptides are ionized and run through a magnetic field. This allows separation of the peptides due to their variety of masses and charges. The proteins are characterized in samples looking for the unique amino acids sequence of each protein. By performing this technique, it was possible to confirm the presence of all seven Sm proteins in the purified sample.

In other organisms, the Sm complex forms a ring shape, which was observed in the LSm complex. Previously, Dunn was able to confirm by Electron Microscopy (EM) that *C. merolae*`s LSms form a complex with a hole in the center (Dunn, 2010). Therefore, the investigation of the functionality of the purified Sm was initiated by the analysis of its assembly. A random interaction of the Sms would suggest a non-functional complex. Samples of the recombinantly expressed and purified complex were sent for EM analysis. In brief, the shape of the protein was assessed through exposure to an electron beam. Electron microscopy results confirmed that the protein complex was

assembled in a doughnut shape, which suggested that the recombinantly expressed and purified Sm complex from *C. merolae* was functional.

In addition, evidence in support of the functionally of the Sm complex was confirmed by binding of the Sm complex to U2, U4 and U5. A filter binding assay, an electrophoretic mobility shift assay (EMSA), and a fluorescence polarization (FP) assay are all described in this chapter and were conducted to assess function of the Sm complex. By performing filter binding, detection of a RNA-protein interaction is enabled by filtration of the RNA-protein mixture through a nitrocellulose filter. Due to the affinity of the protein to the filter, it is possible to measure RNA-protein binding by utilisation of a radioactively labelled RNA. Thus, an increase in signal in the filter should be detected if RNA is interacting with the protein, since the RNA can only be retained on a filter when interacting with the protein. This method was performed in the University of Lethbridge at the Kothe laboratory. The filter binding results did not detect binding of U2 and U5 to the Sm proteins, but it presented promising results regarding binding of U4 to the complex. Therefore, it was necessary to find an alternative method to further investigation of binding of the snRNAs to this protein complex.

The second method, EMSA, assesses RNA-protein binding by running the RNA-protein mixture on a polyacrylamide gel. Binding of the RNA to the protein results in changes of RNA mobility that can be observed on the gel. Gel shifts can be detected by utilisation of radioactively labelled RNA, as bound RNA will not run as far on a gel when compared to free RNA. Since it is expected to see an increase of shifted free RNA with increased protein concentration, the intensity of the bound and unbound RNA bands can be used for binding measurements. By performing this method, it was possible to assess binding of full-length U2, U4, and U5 to the Sm complex and

calculate the equilibrium binding constant, Kd, for U4 and U2. Inconsistencies in the binding of U5 to the Sm complex, made necessary the utilization of a third method: FP.

FP assesses binding of the protein to the RNA by utilisation of a fluorescently labelled RNA. By calculating the anisotropy, a property that relates perpendicular and parallel polarized light, it is possible to determine binding of the protein to the RNA. In brief, if a protein interacts to a fluorescently labelled RNA, an increase in anisotropy will occur since the rotational freedom of the RNA decreases (high polarization). Therefore, it will result in a higher difference between parallel and perpendicular polarization. By knowing this, anisotropy can be related to binding of the protein to the RNA, and consequently, to the percentage of RNA bound. However, for FP there is a limitation on the length of the RNA. Due to this, fluorescent RNA was designed that covers only the U5, U4 and U2 sm sites. This binding assay confirmed binding of all three snRNAs to the Sm complex. This chapter will present a structural, functional and assembly investigation of the recombinantly expressed Sm complex.

3.2 Materials and Methods

3.2.1 Two-step purification of recombinantly expressed Sm complex

The Sm proteins expressed by auto-induction, as described in chapter 2, were co-purified by nickel affinity and size exclusion chromatography. In order to prepare the protein for nickel affinity chromatography, cells were harvested by undergoing centrifugation for 10 minutes at 3000 rpm at 4^{o}C in a JLA-8.1000 rotor (Beckman coulter Avanti HP-20 XPI), and the resulting cell pellet was washed once by addition of buffer A1 (20 mM HEPES-NaOH, pH 7.5, 500 mM NaCl, 20 mM imidazole, 5 mM β-mercaptoethanol). A second centrifugation of the sample was performed, and the cell pellet was snap frozen in liquid nitrogen and stored at -80^{o}C. The

resuspension of the cell pellet was performed by adding 5 ml of buffer A1 for every gram of cell pellet and sonicating the mixture five times in one-minute bursts at 5-8 W with one-minute breaks on ice in between. After sonication, streptomycin sulphate (Sigma) was added for a 1% w/v to remove nucleic acids. Cell fragments were cleared by centrifugation for 30 minutes at 25,000 g at 4ºC in a JA-25.50 rotor (Beckman coulter Avanti HP-20 XPI). The soluble sample was filtered through a 0.45 µm syringe filter and passed over a HisTrap HP Ni sepharose column (GE Healthcare). The column was equilibrated in five column volumes of buffer A1. The sample was washed in 15 column volumes of buffer A2 (20 mM HEPES-NaOH, pH 7.5, 500 mM NaCl, 60 mM imidazole, 5 mM β-mercaptoethanol) and eluted in eight column volumes of buffer B1 (20 mM HEPES-NaOH, pH 7.5, 500 mM NaCl, 500 mM imidazole, 5 mM β -mercaptoethanol). The Sm complex was then loaded at a 0.1 ml/min flow rate onto a size exclusion column (Superdex 200 10/300 GL, GE Healthcare) and equilibrated in buffer A1 without imidazole. Peak fractions were collected, pooled, and concentrated using a YM-30 Centriprep centrifugal filter unit (Millipore). The Superdex 200 column was calibrated using gel filtration standards (BioRad), with the following sizes: thyroglobulin (670 kDa), gamma globulin (158 kDa), ovalbumin (44 kDa), myoglobin (17 kDa), and vitamin B12 (1.4 kDa). Protein aggregates were separated in the void volume of the column (7.65 mL). Protein concentration was determined by Thermo Scientific™ Nanodrop™.

3.2.2 Characterization of the purified Sm complex by Mass Spectrometry

Assessment of the purity of the Sm complex was performed by in-solution digestion of purified proteins followed by MS. Preparation of the sample for MS analysis was performed by Martha Stark, and the sample was analysed as described by Reimer et al. (2017).

3.2.3 Biophysical characterisation of the purified Sm complex by Electron Microscopy

The purified Sm complex was prepared for EM analysis by the concentration of the sample using a YM-30 Centriprep centrifugal filter unit (Millipore). Concentrated protein sample was shipped to Dr Calvin Yip at University of British Columbia, who obtained EM images of the Sm complex, as described by Reimer et al. (2017).

3.2.4 Binding assays

U2, U4, U5 and U6 samples were prepared by *in vitro* transcription (IVT) and purified by gel purification as described by Reimer et al. (2017). Before end-labelling of the snRNAs, IVT snRNAs were dephosphorylated using Shrimp Alkaline Phosphatase (SAP) (New England Biolabs). Free phosphates were removed by purifying with a G-25 spin column (Santa Cruz Biotechnology), as described by the manufacturer. End-labelling of the snRNAs was done using T4 polynucleotide kinase (PNK) (New England Biolabs) and ^{32}P-γATP. Unincorporated ^{32}P-γATP was removed using a G-25 spin column. IVT snRNAs used in filter binding experiments were not dephosphorylated prior to end-labelling, since the PNK manufacturer`s protocol assured phosphate group exchange between 5`-P-RNA and ATP. EMSA reactions were 20 µl containing 12 mM HEPES-NaOH, pH 7.5, 1.5 mM MgCl2, 100 mM NaCl, 10% glycerol, 0.1% Triton X-100, 5 µg *E. coli* tRNA, 2.5 µg of BSA, and 2.5 µl of SUPERase• In™ RNase Inhibitor (20 U/µL). FP reactions were 100 µl containing the same reagents needed for EMSA, except for the RNase inhibitor. Filter binding reactions were 25 µl containing the same reagents as described for EMSA, except for glycerol and RNase inhibitor. For EMSA and filter binding, ^{32}P-RNA was added to make a final concentration of 10 nM and 8 nM, respectively. For FP experiments, fluorescein-labelled U4 Sm site oligo (ro66, IDT) and U2/U5 Sm site oligo (ro67, IDT) were added to reach a final concentration of 15 nM.

Filter binding reactions were incubated for 30 minutes at room temperature, and then filtered through the nitrocellulose membrane (0.2 μm, Whatman, Maidstone, United Kingdom). The membrane was rapidly washed with 1 ml of pre-cooled buffer. The filter was placed in scintillation cocktail for 30 minutes to enhance the radioactive signal, and radioactivity measured on the membrane using a liquid scintillation counter. Data were fitted using Kaleidagraph (Synergy Software) and measured in triplicate. Three different equations (listed below), Hill equation (Equation 1), and equations described by Buenrostro et al. (2014) (Equation 2) and Kuriyan et al. (2012) (Equation 3) were used to fit the data and generate *Kd* values:

Equation 1: $\theta = \frac{((a)[\text{protein}]^n)}{(Kd+[\text{protein}]^n)}$, where θ is the percentage of RNA bound, maximum asymptote, *Kd* is the equilibrium binding constant, and n the Hill coefficient.

Equation 2: $\theta = \frac{\text{a}}{(1+\frac{Kd}{protein})} + a$, where θ is the fraction of RNA bound, a is the maximum asymptote, and *Kd* is the equilibrium binding constant.

Equation 3: $\theta = \frac{[\text{protein}]}{([\text{protein}] + Kd)}$ and $\log(\frac{\theta}{1-\theta}) = \log(\frac{protein}{Kd})$, where θ is the fraction of RNA bound, and *Kd* is the equilibrium binding constant.

EMSA reactions were incubated for 30 minutes at room temperature, then loaded directly onto a 6% native polyacrylamide gel with CHES running buffer and electrophoresed at 200 V. The gels were run at 4 C, and for U2, U4, and U6 were run for 50 minutes; whereas, U5 was run for 1.5 hours. Radioactive EMSAs were imaged on a phosphor imager screen overnight and visualised with a Cyclone Phosphor Imager and OptiQuant software (Perkin Elmer). Data were fitted using Kaleidagraph (Synergy Software) and measured in triplicate. The modified Hill equation (Equation 4) below was used to adjust the data and generate *Kd* values:

Equation 4: $\theta = \frac{a-b}{(1+\frac{Kd}{[protein]^n})} + b$, where θ is the percentage of RNA bound, a is the maximum asymptote, b is the minimum asymptote, *Kd* is the equilibrium binding constant, and n the Hill coefficient.

When assessing binding of U2, U4, and U5 Sm sites to the Sm complex, FP was performed, and the anisotropy was measured using a Synergy 2 Multi-Mode reader (BioTek) with black 384-well microplates (Nunc Thermo Scientific). Data were fitted using Kaleidagraph (Synergy Software) and measured in triplicate. The modified Hill equation (Equation 4) above was used to adjust the data and generate K_d values.

3.3 Results

3.3.1 Two-step purification of recombinantly expressed Sm complex

Purification of the recombinantly expressed Sm complex was performed in two steps as summarised in Figure 3-1a. First, the batch binding purifications were compared when using the constructs pSR751 (SmE-HIS_7) or pSR752 (SmF-HIS_7), to determine the best protein yield and purification. Figure 2-18 shows a much cleaner purification of the Sms when SmE is tagged, thus suggesting better accessibility of the His_7-tag to the nickel resin. Therefore, this construct was used to investigate structure and function of the Sm complex. At the first step, the proteins were successfully purified by IMAC, and the presence of a single peak on the FPLC chromatogram suggested an elution of the complex (Figure 3-1b). Samples were collected and run on a 15% SDS-PAGE gel as seen in figure 3-1c. To further purify the complex, size exclusion chromatography was performed to assure the presence of fully assembled complexes. Surprisingly, two peaks were observed in chromatogram coming out at 11.41 ml and 14.29 ml. This suggests separation of two complexes (Figure 3-1d). According to the Superdex200 standard, the first complex should have

a molecular weight higher than 158 kDa, and the second complex should have a molecular weight of ~ 44 kDa. Based on the sum of the molecular weights of the seven Sm proteins, the complex is expected to have a molecular weight of ~112 kDa. Therefore, the first peak (P1) showed an earlier elution of the complex than expected for the intact ring. Both fractions were collected and loaded onto a 15% SDS-PAGE gel, and the observed bands were close to the Sm proteins molecular weight (Figure 3-1e). Surprisingly, both peaks presented the same bands, although the higher band is faded for the second peak sample. It was suggested that this protein of approximately 50 kDa could be SmD2 since this is the largest protein of the Sm complex. However, SmD2 has a lower molecular weight, ~36 kDa, and the presence of this protein on the second peak would not be expected since the purified complex has a molecular weight of ~44 kDa. Therefore, it was necessary to identify of the proteins present on the first peak by mass spectrometry to confirm the presence of all seven Sms on the purified complex.

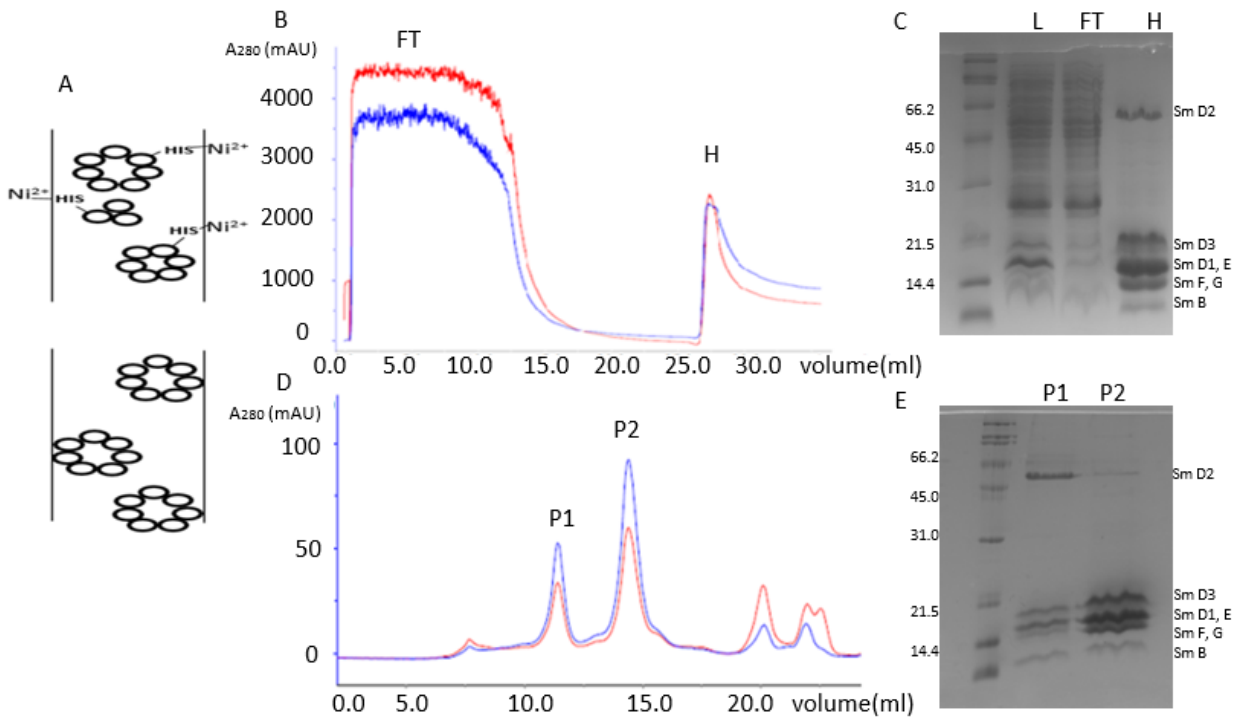


Figure 3-1 Purification of the recombinantly co-expressed Sm complex: a) A Two-step purification of the scheme. The first step allows purification of the HIS_7-containing compounds by binding of the tag to Ni^{2+}. The second step enables separation of small complexes from fully-assembled complexes. b) IMAC chromatogram shows the A280 trace in blue (protein) and the A260 trace (nucleotides) in red. It is presented on chromatogram the flow-through (FT) and the eluted sample (H). c) Collected FT and H samples were run on 15% SDS-PAGE gel, and the observed bands correspond to Sm proteins. d)A size exclusion chromatogram showing the A280 trace in blue (protein) and the A260 trace (nucleotides) in red. The first peak (P1) is observed around 11.41 ml suggesting a ~ 158 kDa complex. The second peak comes around 14.29 ml suggesting a ~ 44 kDa complex. e) Collected P1 and P2 samples were run on 15% SDS-PAGE gel, where the first peak shows a more intense 50 kDa band**.**

3.3.2 Characterization of purified Sm complex by Mass spectrometry

Successful identification of the unique peptides of all seven Sm subunits and fair coverage of the protein sequences by mass spectrometry confirmed the presence of proteins in the co-purified sample in Table 9. Therefore, the purified protein complex, as presented on the first peak of chromatogram (Figure 3-1d), contained all expected proteins. Indeed, the presence of SmD2 was confirmed in sample (coverage of 63%), although it was not expected to be running higher (slower) on the 15% SDS-PAGE gel.

Table 8 Presentation of the data collected after mass spectrometry analysis of the recombinantly co-expressed and purified Sm complex (peak 1 from size exclusion chromatography). Highlighted in grey, are the peptides identified.

Description	Identification probability	Coverage	# Unique peptides	# AA	MW (kDa)
SmF	100%	44%	4	40	~10
MTATGFAEAV KPTNLLSALQ GNRVSVRLKW DLEYTGLLAS YDSYFNLELE HAEELQPDGS SLPLGDMIIR CNNVLYIRDL RSTVPVPPLS					
SmE	100%	90%	15	94	~12
MPKDALDRRI VPEQLLATLA RQQARVEVWL FENTRYSLEG TLRGFDEHTN LVLVDTVEQW GSTAKHKRRT VALGTILLKG ENVVLVRSLG MPTQRKEVTH SATRE					
SmG	100%	95%	10	95	~11
MAKDEVDTAE LEALLFHSVQ VYLNANRCVR GKLSGFDHYA NLVLSDALDC RTGAQLGQVW IRGNSVVSVD LLRDVNADRT EPPTGTGSVA DDPVGSSLSS					
SmD3	100%	27%	4	46	~19
MSGYRPAAFD LPRALLREAK NQIVSVETKN GMEYRGRLDN VSSRMNLVLS AVTVLNATGE RTQKNRVLVR GDSIVLVVLP EALEDAPQLD VLLQVKQARK AAMHVNNTDR KSRGAGRSEA DVHERSGAST LPLPQSESQP QLKRTRVFLS GNAETVQRTK EGGDSNRRNV					
SmB	100%	80%	7	64	~9
MDLLPVLRSQ VHVQTTDGRL LAGKLLAFDA HSNLLLSHCT ERRGESAKRY LGMVLVRGEH VLAVITPRIT ETEQKTAASE					
SmD1	100%	76%	10	102	~15
MTPLLYFLTR LRGATVTVEL KDGTKATGTV QRVDNEMNVY LLNASVTGKP PAELPSASLE THAAQVVAPW TERFSEPDAS AMSRRNQPQQ KAREYRIRGS TVRYIILPES LNLESALKET RKFSPRTRYQ KERH					
SmD2	100%	63%	17	209	~36
MPPVDQPTAL EAGAVAGLTV AQLRRELAAR EAPTSGRKAE LQKRLLDLLG VKLEQEARDE DSSVAPGATQ GEAGRATNLG DATTTSSAQQ QEQQQEQQQE QQQEQQQEQQ QEQQQEQQQE QKLAQTLDPA ALSPSPIQSS AYPQSTTTTQ RRKRRWAEPA SAPPTAPRKR RPLDAHDTHL DQAGATPAAS ELSAAAEAST SYQTLIAATT PATTQSIPNS SESAASALKP AVHAANGSPR TPFTLLDRCI TDRVPCLVSC RHNKKLYGTL RAYDKHFNLI MEHVREIWQE SQPDRPPDLR ERFISRLFVR GRGVIFIVRP CVSATSTARA QP					

3.3.3 Biophysical characterisation of the purified Sm complex by Electron Microscopy

Following the identification of the recombinantly expressed and co-purified complex, a biophysical method was performed to characterise the shape of the assembled complex. Co-purification of the complex suggested assembly of the seven Sms; however, it did not confirm how the proteins are interacting with each other. Therefore, EM was performed on the IMAC purified complex to assure formation of a globular complex instead of a randomly assembled complex. At a magnification of 48,000 (Figure 3-2a) and 98,000 (Figure 3-2b), the formation of a few complexes assembled in ring shape were observed. These results suggest that the co-purified and expressed proteins are interacting in a functional form. A comparison between the Lsm (92 kDa)

and Sm (112 kDa) complexes supports a bigger Sm ring (Figure 3-2c). However, it is observed that more Lsm rings when corelative to Sm complexes, which suggests a significant instability of the Sm complex.

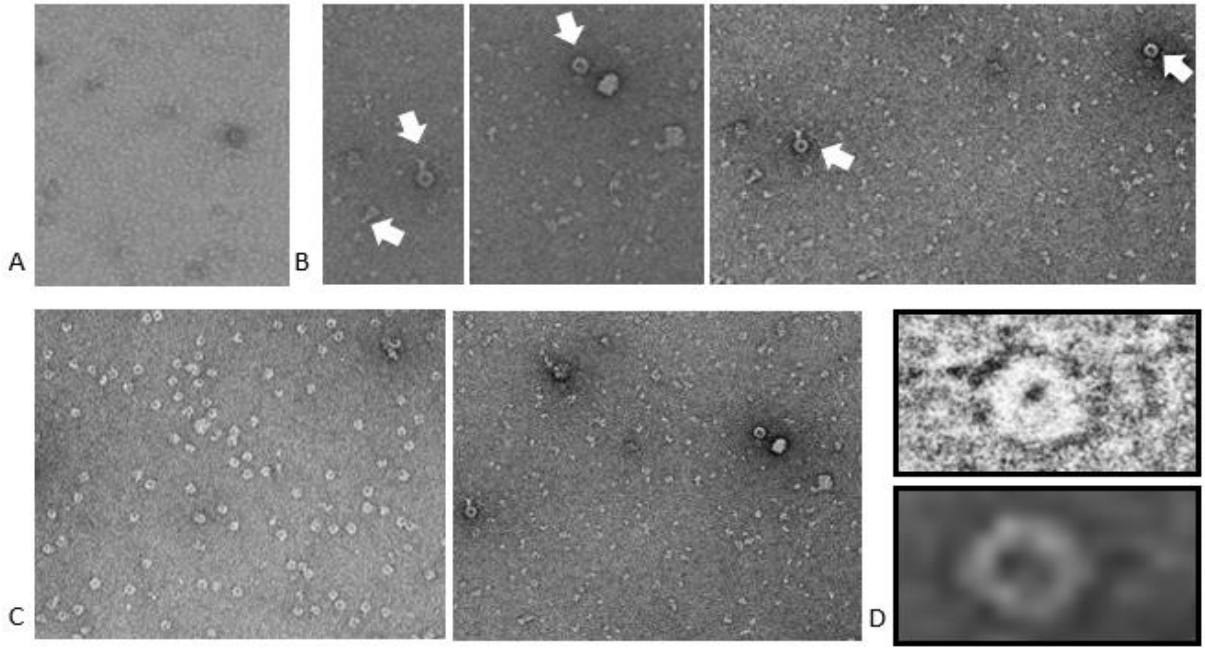


Figure 3-2 Electron microscope of the Sm complex: a) An EM image of the Sm complex at a magnification of 48,000 times. b) EM images of the Sm complex at a magnification of 98,000 times, where the ring-shaped Sm complexes are indicated by the white arrow. c) A comparison between the Lsm complex (92 kDa, left) and the Sm complex (112 kDa, right) EM results appear to present a comparable size, as expected. d) Comparison between the Lsm (top) and Sm (bottom) rings.

3.3.4 Binding Assays

In addition to the EM results that suggested the formation of a functional Sm complex, the functionality of the recombinantly co-expressed and purified Sm proteins was confirmed by binding of the complex to snRNAs. It has been reported in other organisms that the Sm complex plays a critical role binding to U1, U2, U4, and U5 forming the Sm core RNP. Therefore, three binding assays were performed to assess the binding of this complex to U2 (Table 10), U4 (Table 11), and U5 (Table 12) in *C. merolae*. After performing filter binding, I expected to observe a more radioactive nitrocellulose membrane as the concentration of protein increased, and 100% of RNA binding to the protein when the protein concentration is high. Indeed, since the protein interacts

with the nitrocellulose, interaction of the ^{32}P-end-labeled RNA with the protein resulted in an increase of RNA retained on the membrane. The first results did not show a significant rise in the signal measured on the nitrocellulose filter, when increasing the protein concentration from 0 to 2000 nM (Tables 10, 11, and 12). For instance, approximately 10% and 5% of the U4 and U5 snRNA was bound, respectively, at a protein concentration of 2000 nM.

Table 9 Assessment of binding of U2 snRNA to the *C. merolae* Sm complex by filter binding: The percentage of RNA bound was calculated based on the radioactivity (4577 dpm) of 8 nM of U2 snRNA added to each reaction (n=1).

Protein, nM	dpm	% of RNA bound
0	33	0.72
25	47	1.03
50	42	0.92
100	80	1.75
125	138	3.02
150	59	1.29
175	94	2.05
200	65	1.42
500	134	2.93
2000	230	5.03

Table 10 Assessment of binding of U4 snRNA to the *C. merolae* Sm complex by filter binding: The percentage of RNA bound was calculated based on the radioactivity (4251 dpm) of 8 nM of U4 snRNA added to each reaction (n=1).

Protein, nM	dpm	% of RNA bound
0	38	0.89
25	64	1.51
50	28	0.66
100	101	2.38
125	88	2.07
150	99	2.33
175	82	1.93
200	98	2.31
500	243	5.72
2000	455	10.70

Table 11 Assessment of binding of U5 snRNA to the *C. merolae* Sm complex by filter binding: The percentage of RNA bound was calculated based on the radioactivity (3319 dpm) of 8 nM of U5 snRNA added to each reaction (n=1).

Protein, nM	dpm	% of RNA bound
0	49	1.48
25	41	1.24
50	86	2.59
100	86	2.59
125	86	2.59
150	67	2.02
175	74	2.23
200	76	2.29
500	122	3.68
2000	188	5.66

The low binding of the snRNAs to the protein complex raised questions regarding RNA degradation and protein precipitation. Analysis of labelled snRNAs on 6% urea PAGE gel presented a fair amount of labelled U2 and U4 and possibly degraded U5 (Figure 3-3). In addition, the figure shows that the samples are not running well on the gel since a single band, representing each labelled snRNA, is not observed. Since two proteins that are known to bind to U4 snRNA in other organisms were available, their binding to U4 was assessed. The *C. merolae* Snu13 protein binds to U4 (Black et al. 2016); therefore, it is likely binding to the yeast Nhp2 since it is a Snu13 homolog. Surprisingly, measurement of the membrane radioactivity at 0 and 1000 nM of protein suggests no binding of CmSnu13 to U4 (Table 3-13). Unexpected results are also observed when assessing binding of yNph2 to U4 (Table 3-14). As presented in table 14, ~25% of RNA is bound to the protein (1054 dpm) at Sm concentration of 1000 nM; however, the counts surprisingly decrease as the protein concentration increases to 5000 nM.

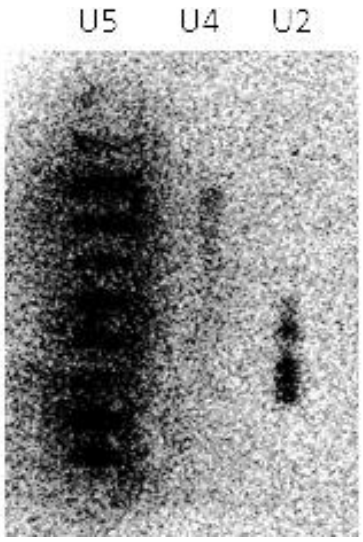


Figure 3-3 Investigation of U5, U4 and U2 snRNA stability: 6% urea PAGE gel presenting ^{32}P-end-labelled RNA.

Table 12 Assessment of binding of U4 snRNA to the *C. merolae* Snu13 by filter binding: The percentage of RNA bound was calculated based on the radioactivity (4251 dpm) of 8 nM of U4 snRNA added to each reaction (n=1).

Protein, nM	dpm	% of RNA bound
0	34	0.80
50	27	0.64
200	47	1.11
1000	38	0.89

Table 13 Assessment of binding of U4 snRNA to the *S. cerevisiae* Nph2 by filter binding: The percentage of RNA bound was calculated based on the radioactivity (4251 dpm) of 8 nM of U4 snRNA added to each reaction (n=1).

Protein, nM	dpm	% of RNA bound
0	34	0.80
1000	1054	24.79
2000	865	20.35
5000	639	15.03

Since the U4 gel analysis and binding of U4 to yNh2 and Snu13 were not consistent, not much could be concluded from these divergent results. Therefore, it was suggested that the RNA labelling was not efficient, which would explain the low counts observed. In addition, the little

binding of the RNA to the protein could be due to the wrong folding of the RNA. Therefore, the concentration of RNA was increased from 8 to 15 nM added. U2 and U4 were unfolded at 70ºC and allowed to refold at room temperature. A comparison between the snRNAs used previously to refolded RNA is presented in tables 3-15 and 3-16. Two times and four times increase in binding was observed for U2 and U4 snRNA to the protein complex, this improvement only represents 5.66 and 4.28% of the U2 and U4 bound, respectively.

Table 14 Comparison between the binding of folded and refolded U2 snRNA to the *C. merolae* Sm complex by filter binding: The percentage of RNA bound was calculated based on the radioactivity (7197 dpm) of 15 nM of U2 snRNA added to each reaction (n=1).

	Protein, nM	dpm	% of RNA bound
FOLDED	0	32	0.44
	150	203	2.82
	200	90	1.25
	2000	234	3.25
REFOLDED	0	36	0.50
	150	66	0.92
	200	104	1.45
	2000	407	5.66

Table 15 Comparison between the binding of folded and refolded U4 snRNA to the *C. merolae* Sm complex by filter binding: The percentage of RNA bound was calculated based on the radioactivity (9385 dpm) of 15 nM of U4 snRNA added to each reaction (n=1).

	Protein, nM	dpm	% of RNA bound
FOLDED	0	62	0.66
	150	48	0.51
	200	51	0.54
	2000	114	1.21
REFOLDED	0	66	0.70
	150	83	0.88
	200	78	0.83
	2000	402	4.28

These results raised questions regarding the protein stock concentration. Therefore, the concentration of the purified Sm complex was assessed by the utilisation of a Nanodrop™ spectrophotometer (Thermo Fisher Scientific). A lower concentration of the protein was observed than expected, suggesting protein precipitation. Therefore, the protein concentrations was corrected and repeated refolding of the RNA was followed by filter binding. As observed in table 17, an improvement in binding of U2 to the Sm complex was not observed, since the maximum percentage of RNA bound is 4.67%. However, an increase in the binding of U4 snRNA to the Sms occurred, with bound RNA bound reaching 12.42% at 2000 nM (Table 18).

Table 16 Assessment of binding of U2 snRNA to the *C. merolae* Sm complex by filter binding: Filter binding was performed after correction of the concentration of the protein stock sample. The percentage of RNA bound was calculated based on the radioactivity (7197 dpm) of 15 nM of U2 snRNA added to each reaction (n=1).

Protein, nM	Dpm	% of RNA bound
0	41	0.57
150	113	1.57
200	147	2.04
2000	336	4.67

Table 17 Assessment of binding of U4 snRNA to the *C. merolae* Sm complex by filter binding: Filter binding was performed after correction of the concentration of the protein stock sample. The percentage of RNA bound was calculated based on the radioactivity (9385 dpm) of 15 nM of U4 snRNA added to each reaction (n=1).

Protein, nM	dpm	% of RNA bound
0	48	0.51
150	160	1.70
200	215	2.29
2000	1166	12.42

Since a substantial change in binding of U4 to the Sms was seen after refolding and correction of the protein concentration, filter binding was repeated increasing the binding reaction

volume from 25 to 50 µl. This change facilitates spreading of the binding reaction over the nitrocellulose membrane resulting in a higher coupling of protein-snRNA to the membrane. However, results showed a higher background noise that could be due to the higher reaction volume making it challenging to wash off unbound snRNA from the membrane. The data collected by filter binding presented improvement in binding of U4 to the Sm complex since at 3500 nM of protein ~17% of snRNA was bound. However, a higher binding was expected, and a considerable deviation was observed between the three trials (Table 19). The average data (n=3) was fitted using three different equations, and an estimation of the dissociation constant, K_d, was obtained for the full-length U4. Curves were fitted with the assumption that the final binding is 100%. First, the data were fitted using equation 2 plotting the fraction of RNA bound against the concentration of protein (Figure 3-4). As already expected, a high K_d was generated, 8000 ± 5000 nM, due to the weak data. As observed in the graph, a plateau was not reached when 3500 nM of the protein was added suggesting a low affinity of the snRNA to the protein.

Table 18 Assessment of binding of U4 snRNA to the *C. merolae* Sm complex by filter binding: The table presents the average of the signal (dpm) collected from three trials. The percentage of RNA bound was calculated based on the radioactivity (7086 dpm) of 10 nM of U4 snRNA added to each reaction.

Protein, nM	dpm (average)	Standard deviation	% bound
0	108.33	56.09	0.00
50	101.33	17.15	-0.12
100	131.66	11.95	0.41
125	177.33	66.21	1.22
150	146.33	19.77	0.67
175	164.00	67.00	0.98
200	180.50	29.50	1.27
500	290.00	72.00	1.91
2000	785.00	241.39	11.94
3500	1093.00	0.00	17.37

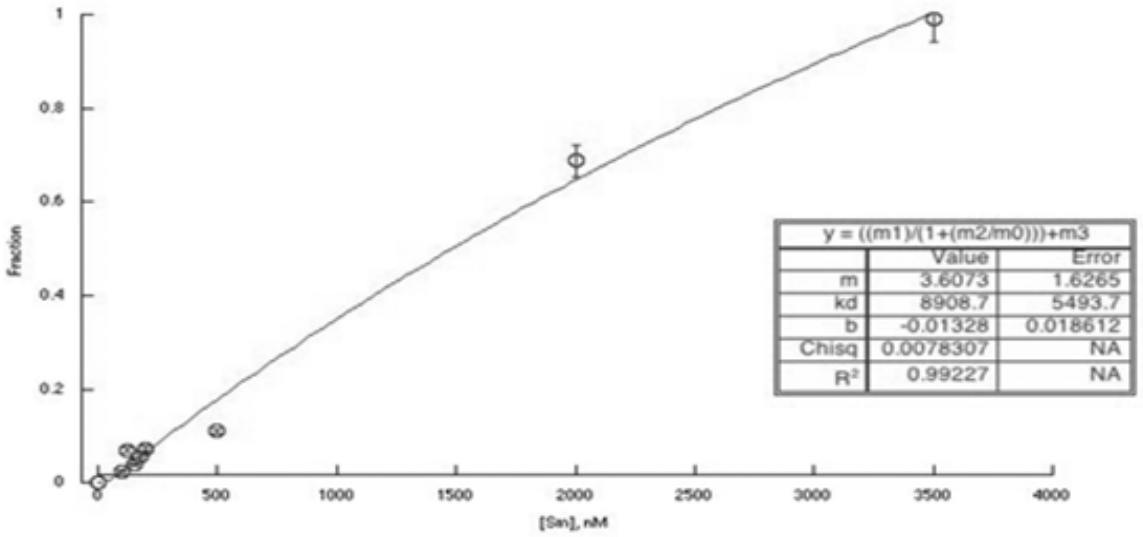


Figure 3-4 Assessment of binding of U4 snRNA to the Sm complex by filter binding: Equation 2 was utilised for generation of the K_d**.**

When plotting data using equation 1, it was fitted with data for the generation of the Hill coefficient (n) and the K_d. As presented in figure 3-5, a high K_d of 40000 ± 3600 nM is also seen for that data fit and a Hill coefficient of 1, suggesting pre-assembly of the Sm complex (Figure 3-5a). Thus, when fixing n=1, a K_d of 10000 ± 7000 nM (Figure 3-5b) was obtained. A third equation was also used to fit the data, which seemed to be more suitable to fit the curve since the maximum fraction does not fluctuate (fmax = 1), reaching a more reasonable plateau value (Figure 3-6a). The calculation of the K_d was done by plotting the graph using logarithm axis since this facilitates visualization of the data, spreading the more informative points out (Figure 3-6b). Therefore, since log of *Kd* is the intercept of the line on the horizontal axis, the *Kd* is 700 nM.

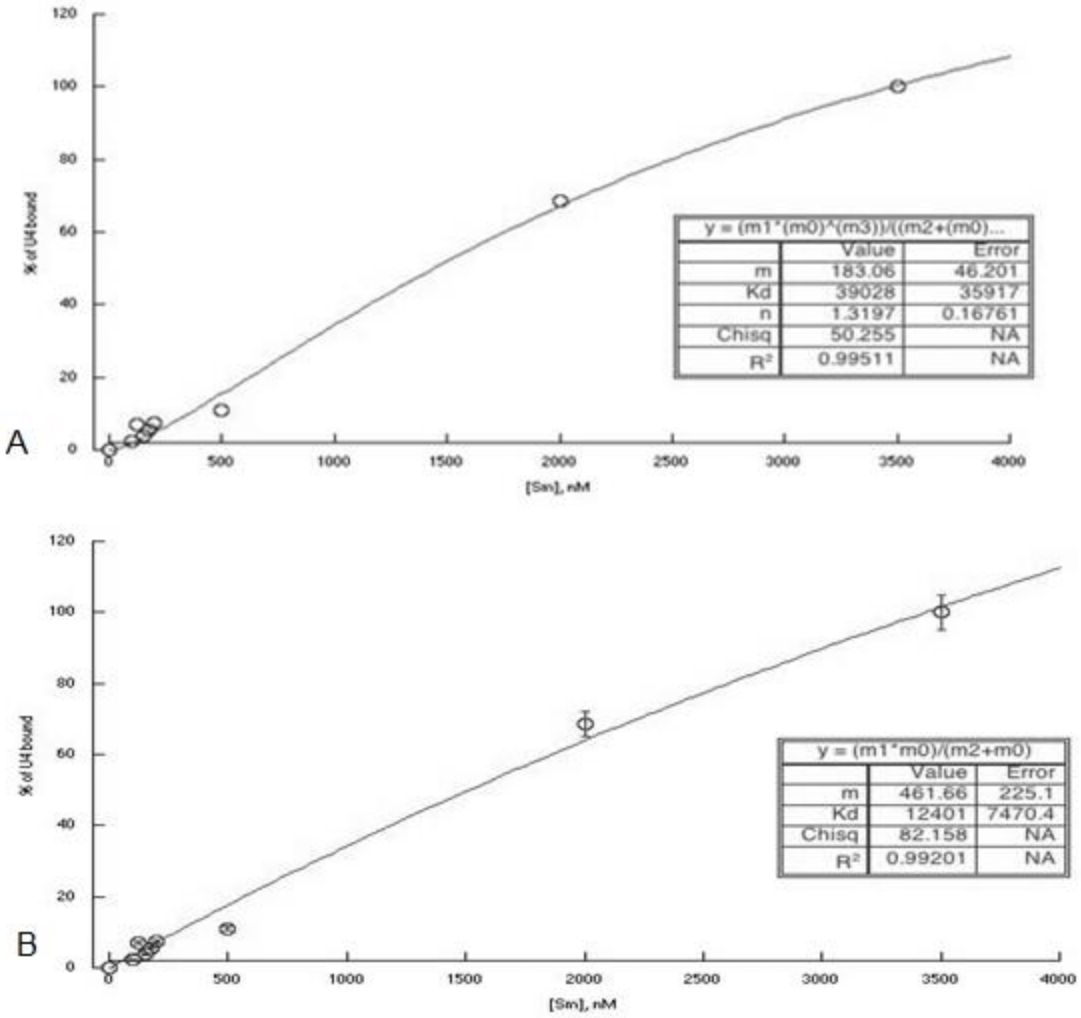


Figure 3-5 Assessment of binding of U4 snRNA to the Sm complex by filter binding: a) Equation 1 was utilised for generation of the K_d and n values. b) Generation of the K_d using equation 1 considering n=1.

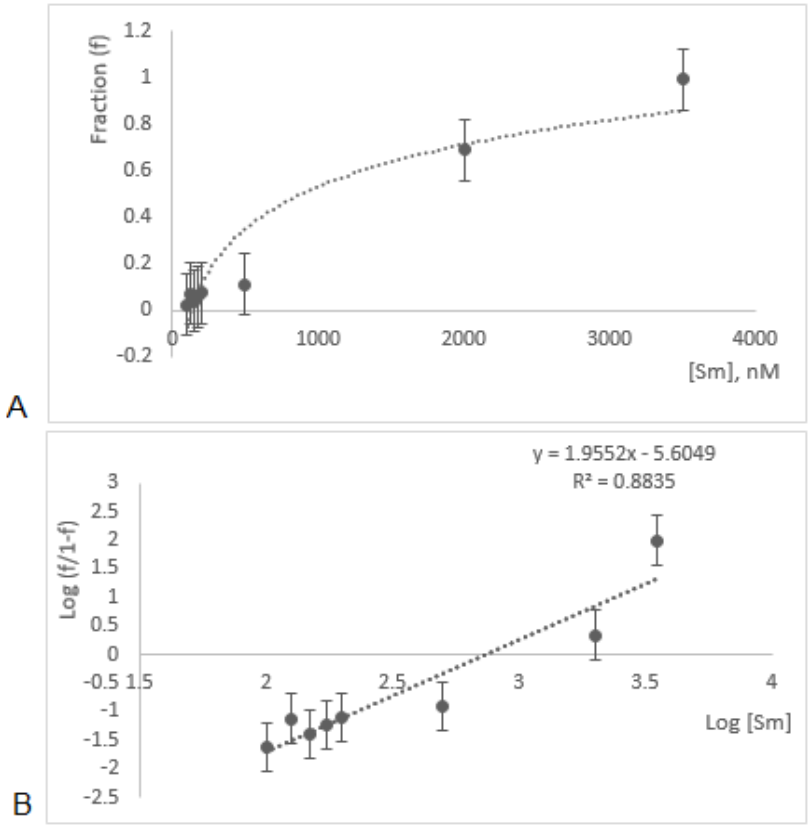


Figure 3-6 Assessment of binding of U4 snRNA to the Sm complex by filter binding: a) Normal binding isotherm using equation 3. b) Data from the graph (a) plotted using logarithm axes, which log of K_d is equal to the intercept (2.87); therefore, the K_d is 700 nM.

Due to the inconsistent results and low binding affinity observed when performing filter binding, binding of U2, U4, and U5 to the Sm complex was assessed by EMSA and FP. As a negative control, binding of U6 to the protein complex was assessed by EMSA. As expected, U6 does not bind to the Sm complex (Figure 3-7a). For assessment of binding of U4 snRNA to the Sms, EMSA was performed using a full-length *in vitro* transcribed U4. Binding of U4 to the Sm complex was confirmed since a shift from free-RNA to bound-RNA was observed on the native gel (Figure 3-7b). Data collected from the four trials was utilised to generate the K_d using equation 4 since this equation considers more parameters to better fit the curve. The K_d for the full-length

U4 was calculated to be 170 ± 6 nM, and the line fit gave an n value of 3 ± 0.3 (Figure 3-7c). Surprisingly, this hill coefficient value suggests that pre-assembly of the Sm complex before binding to the RNA does not occur.

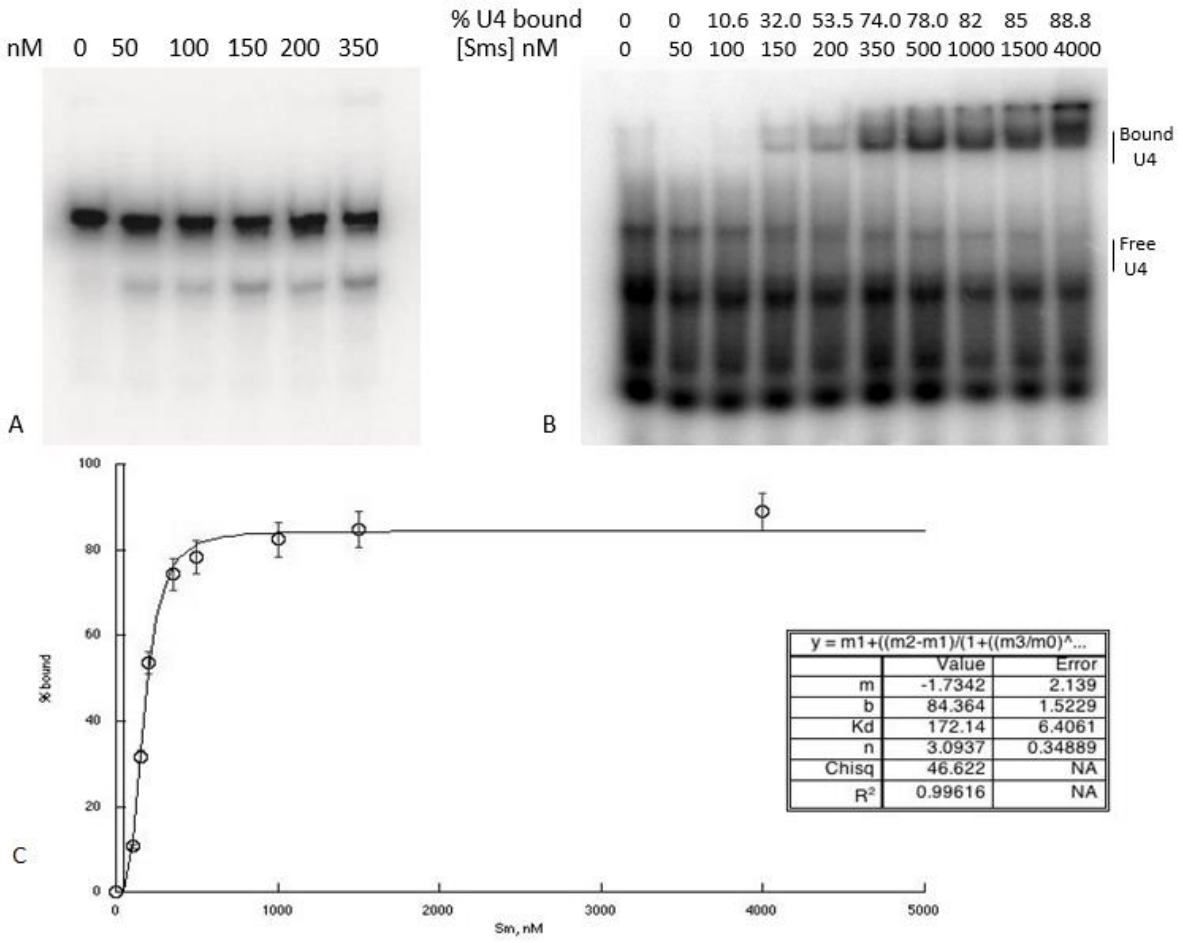


Figure 3-7 Assessment of binding of U6 and U4 snRNA to the Sm complex by EMSA: a) 6% native polyacrylamide gel presenting no binding of radio-labelled U6 snRNA to the Sm complex. b) 6% native polyacrylamide gel presenting binding of radio-labelled U4 snRNA to the Sm complex, as a shift between unbound and bound U4 is observed. c) The % bound of U4 was graphed against the concentration of protein using equation 4.

To further investigate this binding, FP using a fluorescent oligonucleotide of the predicted Sm site in U4 was performed. Triplicate data were collected and used to generate the K_d using

equation 4. The K_d for the U4 Sm site was 400 ± 20 nM, and line fit gave an n value of 3 (Figure 3-8a). When fitting both EMSA data and FP data considering n=1, the K_d was calculated to be 200 ± 80 nM and 365 ± 20 nM, respectively, suggesting that affinity of the Sms to the Sm site doubles when bound to full-length U4 (Figure 3-8b).

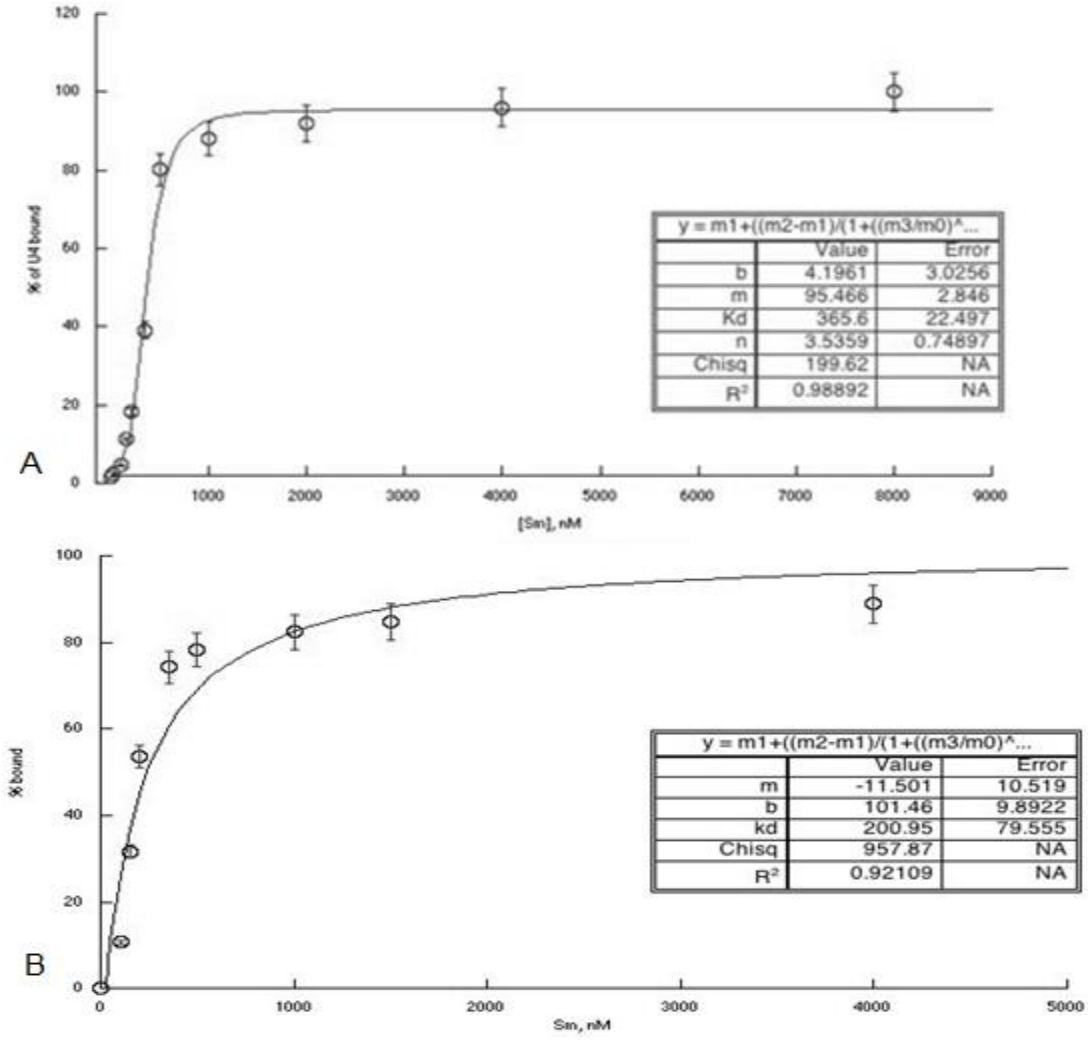


Figure 3-8 Assesment of binding of the Sm complex to the U4 Sm site by FP: a) The % bound of the U4 Sm site was graphed against the concentration of protein, which the K_d was generated using equation 4. b) Generation of the K_d using equation 1 considering n=1.

Binding of U2 and U5 to the purified Sm complex was also assessed by EMSA and FP. The binding of full-length U2 to the Sm is observed on the native gel. However, it is seen at a

lower affinity of U2 to the complex when compared to U4, since protein interaction initiates at a concentration of 100 nM (Figure 3-9a). By fitting the data collected from four trials, using equation 4, the K_d was calculated to be 4600 ± 2000 nM with an n value of 0.9 ± 0.2 nM, indicating a lower affinity compared to U4 (Figure 3-9b). The Hill coefficient value (n=1) suggests that the Sm complex is pre-assembling prior to binding to the snRNA.

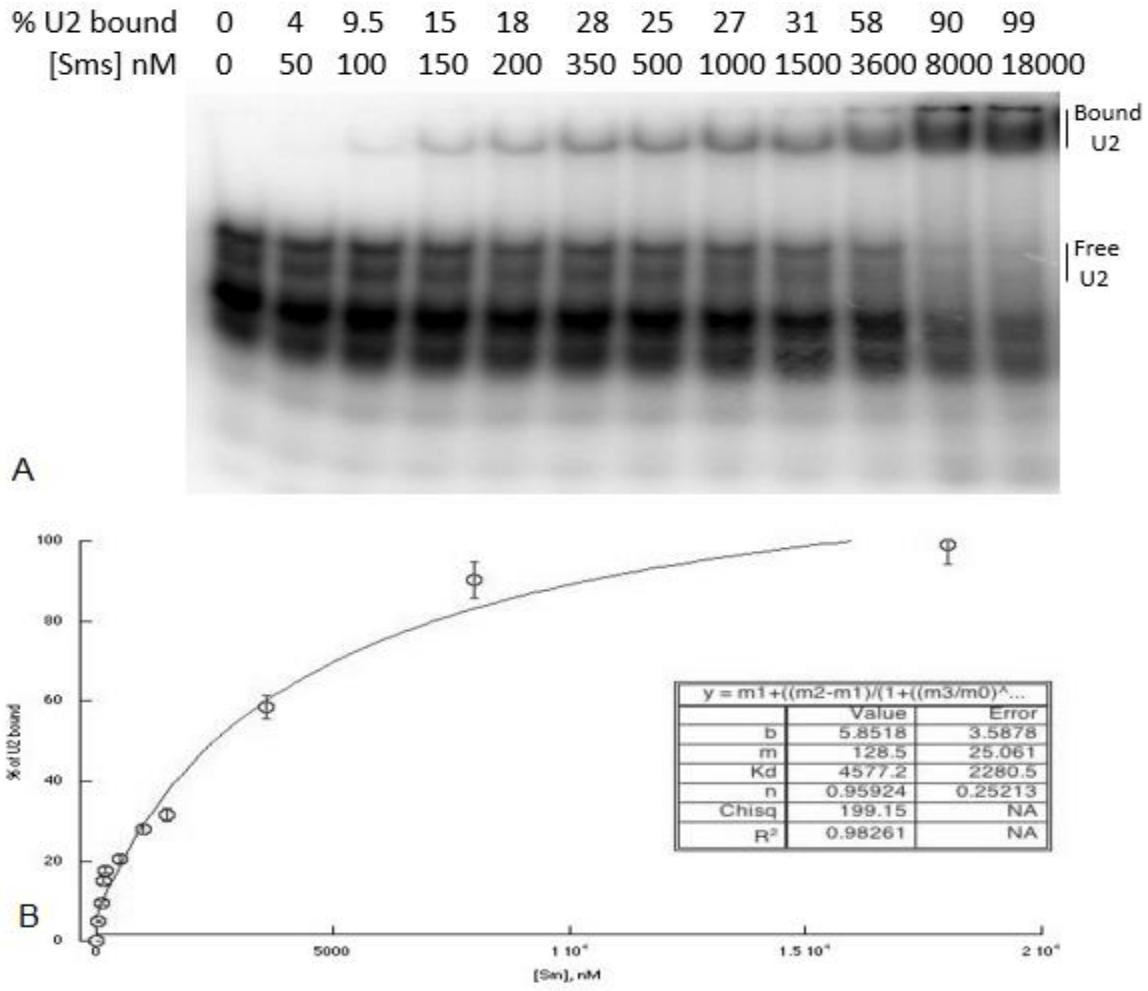


Figure 3-9 Assessment of binding of U2 snRNA to the Sm complex by EMSA: a) 6% native polyacrylamide gel presenting binding of radio-labelled U2 snRNA to the Sm complex. b) The % bound of U2 was graphed against the concentration of protein using equation 4.

Performing of EMSA using full-length U5 also presented the interaction between U5 and the Sm complex; however, this binding was observed on the native gel twice and was unable to be reproduced for unknown reasons (Figure 3-10a). Although enough data to support binding of U5 to the Sm complex was not collected, a graph was plotted using data from two trials to estimate the K_d. A K_d of 185 ± 6 nM and n = 2.8 ± 0.2 were estimated (Figure 3-10b). These data suggest a cooperative event occurring.

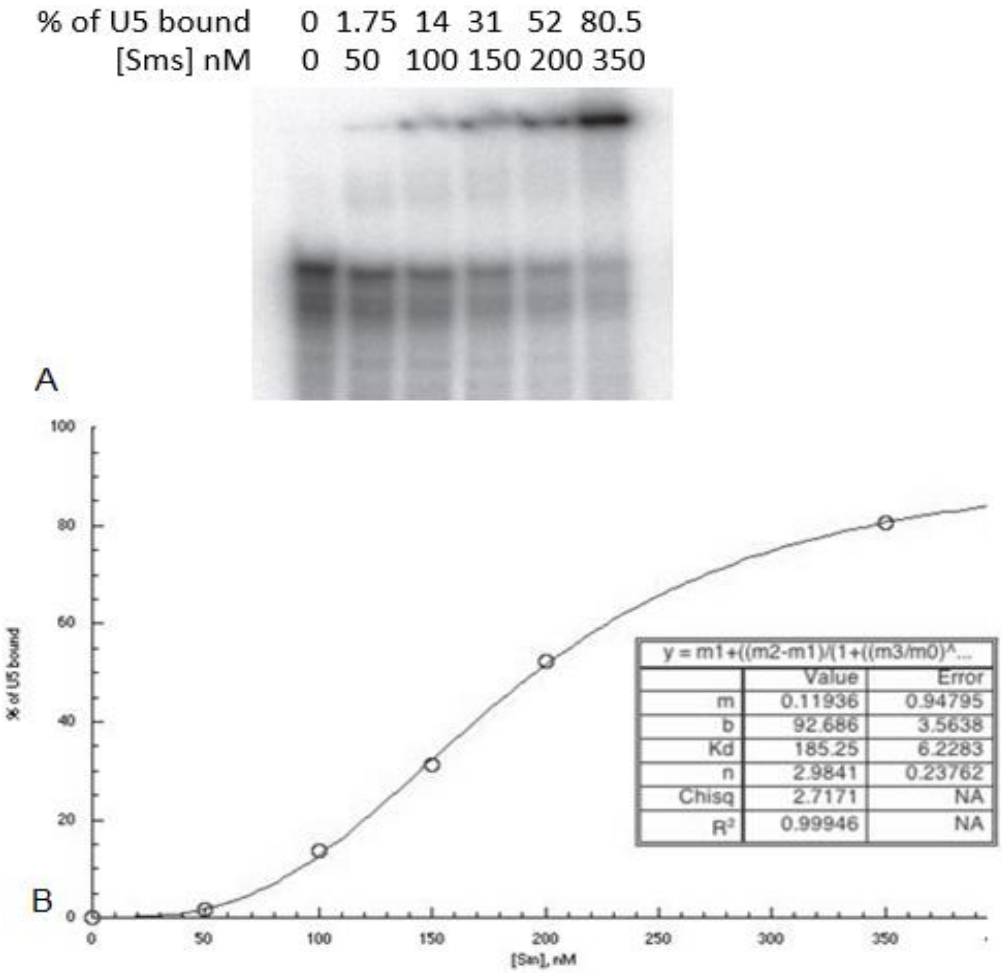


Figure 3-10 Assessment of binding of U5 snRNA to the Sm complex by EMSA: a) 6% native polyacrylamide gel presenting binding of radio-labelled U5 snRNA to the Sm complex. b) The % bound of U5 was graphed against the concentration of protein using equation 4.

Since the EMSA data suggested a low affinity of the Sms to U2 and inconsistent binding of U5 to the Sms, a different binding assay was performed to confirm these results. As seen in figure 1-11, U2 and U5 share the same Sm site sequence enabling the usage of the same oligonucleotide to verify binding of the Sms by FP. As seen in figure 3-11, binding of the Sms to the Sm site is confirmed, and calculation of K_d from triplicate data is done by using equation 4. For the U5 and U2 Sm site, the K_d was calculated to be 430 ± 20 nM, and line fit gave an n value of 4.7 ± 0.9.

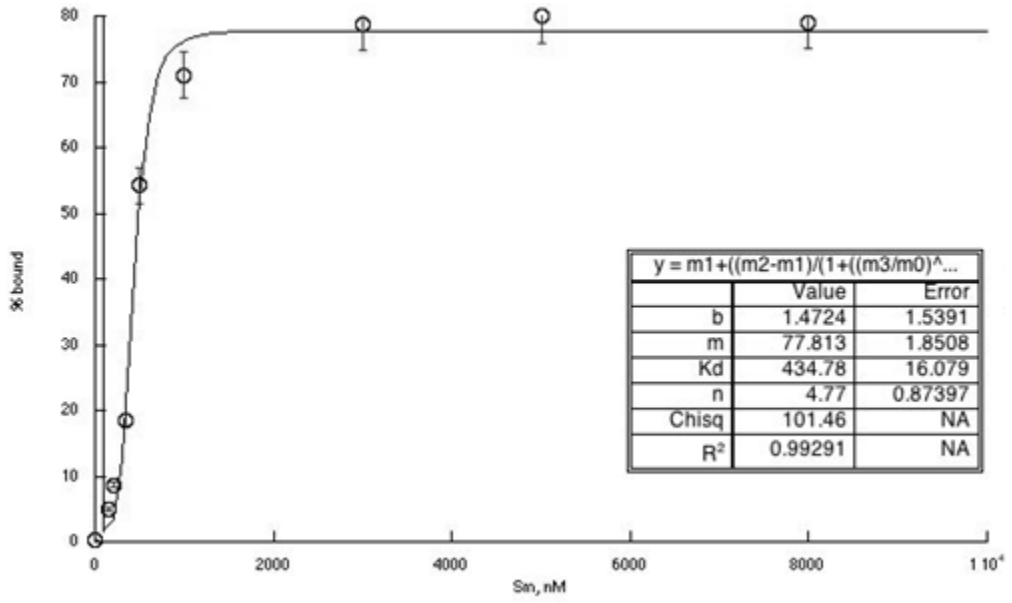


Figure 3-11 Assessment of binding of the Sm complex to the U2 and U5 Sm site by FP: The % bound of the U2 and U5 Sm site was graphed against the concentration of protein, which the K_d was generated using equation 4.

3.4 Discussion

Co-expression of the seven Sm proteins proved to be highly advantageous since it enabled co-purification of the complex. As stated previously, expression of the Sms individually is difficult due to exposure of the hydrophobic β strands. Therefore, insolubility of the compound can be solved by co-expression of proteins since the interaction of the Sms has been shown to bury the hydrophobic strands (Kambach et al. 1999). Presumably, this explains the presence of the Sm proteins in both soluble and insoluble fractions (Figure 2-18). The relevant MS results confirmed that the co-purified complex is comprised of all seven Sms. However, this method does not address the stoichiometry of the purified Sm complex. Presence of an insoluble fraction could be due to an abundant expression of some of the proteins. Thus, the proteins in excess would not assemble into complexes, resulting in precipitation of proteins since the hydrophobic β strands are exposed. An investigation of the Sm proteins present in the insoluble fraction would be necessary to identify which proteins are highly expressed.

The EM results demonstrated the formation of a few rings confirming the functional interaction of the Sm proteins. These results are coherent with the successful co-purification of the complex and presence of all seven proteins in the purified sample. Nevertheless, it is surprising that not many rings are observed in the EM results. Presumably, the proteins that have lower expression are limiting the amount of Sm complexes formed. In addition, instability of the complex *in vitro* might suggest that *in vivo* the complex interacts with some unknown *C. merolae* protein to increase its stability. In other organisms, the SMN protein complex is responsible for assembly and stability of the Sm complex for the formation of the Sm core (Pellizzoni et al. 2002). However, the absence of this protein in *C. merolae* raises questions regarding its instability in the cell. Co-immunoprecipitation of the Sms using an antibody against free snRNA proteins could address this

issue by investigation of the proteins that are interacting with this complex. Also, the absence of an assembly factor suggests that the Sm complex pre-assembles before binding to the snRNA. This hypothesis is also supported by the location of the Sm site in the 3`end of U2, U4 and U5 lacking a stem-loop 3` that region. Thus, it would allow the pre-formed Sm complex to slide into the RNA single-stranded 3`-end.

Surprisingly, the binding assays do not support this hypothesis. The calculated Hill coefficient value suggests a cooperative event occurring (n>1) for both the full-length snRNAs and the Sm sites. Cooperative binding indicates that the Sm complex is not pre-assembled, and the assembly takes place on the Sm site without any protein`s assistance. Another hypothesis is that the protein complex is being retained at the bottom of the reaction tubes. Indeed, an increase in protein concentration would cause an abrupt change in the percentage of RNA bound as represented by the sigmoidal binding curve. To allow for the homogenous binding reaction, a high concentration of *E. coli* tRNA and BSA was added to the binding reactions. However, even changes of the binding reaction, the n value supports a cooperative binding of the Sm subcomplexes.

Further investigation of the snRNP core assembly would be necessary to address these questions. In other organisms, the Sm proteins are known to form stable subunits: SmE.F.G, SmE.F.G.D1.D2, SmBD3 and SmE.F.G.D1.D2.B.D3. Therefore, the study of the stable heteromeric complexes formed in *C. merolae* would be crucial to address the assembly of the Sm complex. For instance, it has also been identified that SmE/F/G does not interact with SmB/D3 in the absence of dimer SmD1/D2 (Raker et al. 1996). Indeed, immunoprecipitation analysis confirmed no formation of a pentamer comprised of SmE.F.G.B.D3 suggesting that SmE.F.G requires SmD1.D2 for binding to SmB.D3. *In vitro* analysis of stable sub core formation by co-

immunoprecipitation showed that U1 could stably bind to SmE.F.G and SmE.F.G.D1.D2 implying that these complexes are Sm core intermediates (Raker et al. 1996). Therefore, further investigation of stable Sm subcomplexes formation could be performed by co-immunoprecipitation analysis. *In vivo* identification of subcomplexes formation would suggest a step-wise assembly of the Sm complex. In addition, investigation of interaction of SmE.F.G and SmE.F.G.D2.D1 to the snRNAs *in vitro* could be assessed by EMSA. Stable interaction of the subcomplexes with the snRNAs would indicate the formation of sub-core intermediates as seen in other organisms. It is expected to observe a weak interaction of the SmE.F.G trimer and the SmE.F.G.D2.D1 pentamer to the snRNA, if the Sm complex assembles before binding to the snRNAs.

Notably, the full-length U2 presents a higher binding affinity to the Sm complex when compared to U4`s *Kd* and the estimated U5`s *Kd* (Table 19). Although the EMSA data of U5 binding to the Sms is not reliable, it was surprising to see that U2 and U5 do not present a similar *Kd* of the full-length snRNA since these snRNAs share the same Sm site sequence. Not much can be concluded about these results since there are not enough replicates of U5`s EMSA results. However, if the estimated *Kd* is correct, the differences in *Kd* could be due to the secondary structure of the snRNAs. As seen in figure 1-11, a comparison of the predicted secondary structure between U2 and U5 shows the presence of a more extended sequence downstream to the U5 Sm site which might facilitate binding of the Sm complex to U5. Indeed, the presence of only two uridines on the 3` end of the Sm site would ease degradation of the 3`end of U2. An increase in affinity of the Sm proteins to the full-length snRNA compared to Sm site is presented in both EMSA and FP results (Table 20). As shown in figure 1-11, the U4 Sm site differs from the U2 and

U5 Sm site; therefore, the presence of the Sm sites consensus sequence might increase the affinity of the Sm complex to U4.

Table 19 Binding parameters: A fluorescent oligonucleotide (ro64) was used as a negative control when performing FP.

Construct	K_d (nM)	Hill coefficient
Full length U4	170 ± 6	3.0 ± 0.3
Full length U2	4600 ± 2000	0.9 ± 0.2
Full length U5	185 ± 6	2.8 ± 0.2
U4 Sm site (ro66)	360 ± 20	3.0 ± 0.7
U2/U5 Sm site (ro67)	430 ± 20	4.7 ± 0.9
Control (ro64)	>10000	n/d

4. Chapter Four - An investigation of splicing relevance and 5' splice site recognition in *Cyanidioschyzon merolae*

4.1 Introduction

As stated previously in the first chapter, *C. merolae* was proposed to be a good candidate to study splicing due to its reduced spliceosome. A recent work from the Rader Lab on this alga has demonstrated that this organism lacks several splicing components present in most organisms (Stark et al 2015). For instance, a relevant splicing component, U1 snRNP, has been found to be absent in *C. merolae*. In other organisms, splicing initiates when U1 snRNA recognizes the 5`splice site on the pre-mRNA followed by binding of U2 snRNA to the branch point site (BPS). Surprisingly, the U5 snRNA has its 5' end sequence complementary to all 5' splice sites in *C. merolae* (Figure 1-2). Thus, it suggests that U5 initiates spliceosome assembly by recognition of 5`splice sites. In this chapter, the use of morpholino (MO) and vivo-morpholino (vivo-MO) oligonucleotides is explored. These oligonucleotides are capable of blocking RNA-RNA base-paring interactions, thus making them suitable for investigating the proposed U5-5` splice site interaction. Presumably, the blockage of the binding of U5 to the 5`splice site should cause a decay in cell growth or death of the cells; however, splicing has never been proven to be essential in *C. merolae*. Thus, in order to investigate 5`splice site recognition site using this method, it must first be determined whether or not splicing is vital for this alga. To investigation this, I attempt to block the binding of U2 snRNA to the BPS of the pre-mRNA, and therefore block the processing of pre-mRNA. Blockage of this interaction should result in cell death if splicing is crucial for *C. merolae* cells.

MOs are DNA oligonucleotides widely used for blocking splicing due to their high sequence specificity when compared to short interfering RNAs (siRNAs) and Phosphorothioate-linked DNA

(S-DNA) (Summerton, 2007). For instance, Draper et al. (2001) describe efficient blockage of pre-mRNA splicing by usage of MOs complementary to the exon/intron junction resulting in exon skipping in zebrafish. The MO structure has been redesigned to distinguish it from the DNA structure since it is comprised of a 6-membered sugar ring (morpholino ring) rather than a deoxyribose ring (Figure 4-1a). In addition, the negatively charged phosphate linkages present in DNA and RNA are replaced by non-ionic phosphorodiamidate linkages.

The MO backbone presents several advantages, such as sequence specificity. In addition, its non-ionic structure prevents electrostatically binding to proteins and decreases binding to extracellular and cellular structures (Summerton, 2007). Furthermore, the prevention of degradation of MO by RNAse cleavage is achieved by the presence of the morpholino rings. The fact that MOs cannot be degraded biologically is advantageous since they are unable to form degradation products that might be toxic to the cell. Regarding the MO sequence, it is usually comprised of 25 base pairs and a high GC content that increases target affinity. Its sequence content and length enable binding to secondary structures. For instance, the oligonucleotide length increases nucleation of pairing and the probability of binding to single-stranded regions of the RNA secondary structure. In human cells type, a small number of morpholino oligonucleotides bind to intracellular, membrane and extracellular proteins by no Watson-Crick base-paring in S-DNA and siRNA oligonucleotides. (Summerton, 2007).

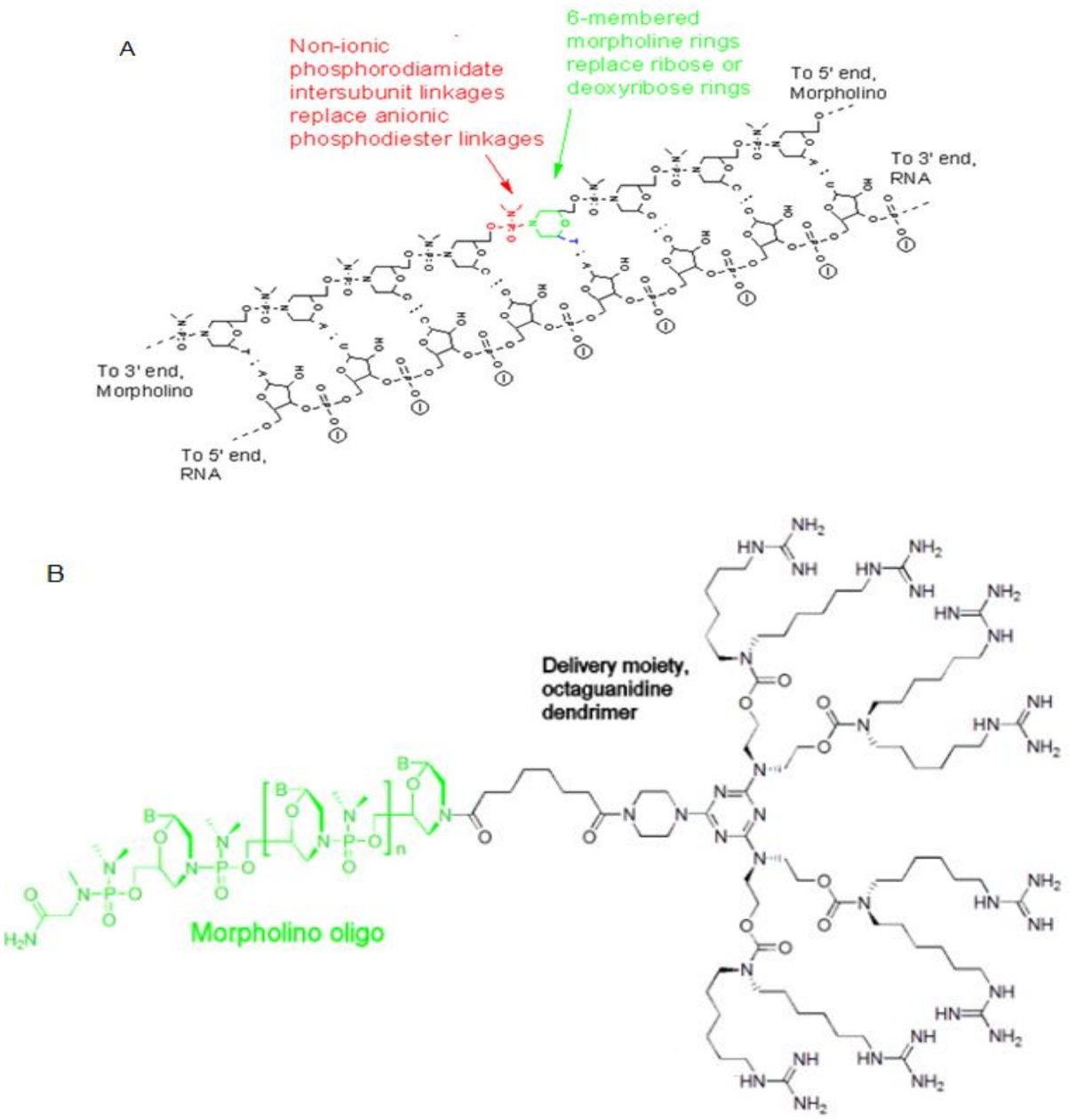


Figure 4-1 MO structure: a) The MO backbone presents some advantages over DNA oligonucleotides. The morpholino rings prevent RNAse cleavage and the non-ionic phosphorodiamidate linkages prevent binding of to proteins. b) The presence of an octaguanidine attached to the MO oligonucleotide and interaction with cell membrane allows delivery of the MO to the cells. These images were drawn at Gene Tools, LLC by Jon D. Moulton.

Among the different approaches to deliver the MO into the cells there is a novel peptide called Endo-Porter (Figure 4-2). This delivery system enables transport of the MO from the membrane into the cytosol of the cell by an endocytosis-mediated process. This method has several

advantages: avoids plasma membrane damage by adsorption, does not require interaction of the cargo with the endo-porter, and delivers high concentrations of cargos that exceed 70 kDa into adherent and non-adherent cells (Summerton, 2005). Notably, the Endo-Porter has been proven to successfully deliver MO oligonucleotides, peptides and proteins into different cells types, such as zebrafish and mammalian cells. This mechanism of transporting molecules to the cytosol occurs by interaction of the Endo-Porter with the cell membrane enabling its endocytosis along with any substance present in the media. Acidification of the endosome results in poly-cationic formation of the endo-porter, permitting exit of the endo-porter from the endosome by permeabilization of the membrane (Figure 4-2). Due to this acid-induced permeabilization of the endosome, the Endo-Porter allows any substance co-endocytosed to be transported from the endosome to the cytosol of cell (Summerton, 2005). Indeed, the presence of a fluorochrome attached to the MO allows assessment of its delivery to the cells by microscopy analysis. Vivo-MOs can also be used for the same purpose; however, they are more advantageous since they do not require the addition of the endo-porter for delivery. Vivo-MOs are morpholino oligos attached to a delivery moiety comprised of an octaguanidine dendrimer (Figure 4-1a). The structure of the vivo-MO enables delivery of the morpholino to the cytosol of the cell by interaction of the guanidinium head groups with phosphates of membrane phospholipids (Morcos et al. 2008). However, none of these methods has been tried in *C. merola*e cells.

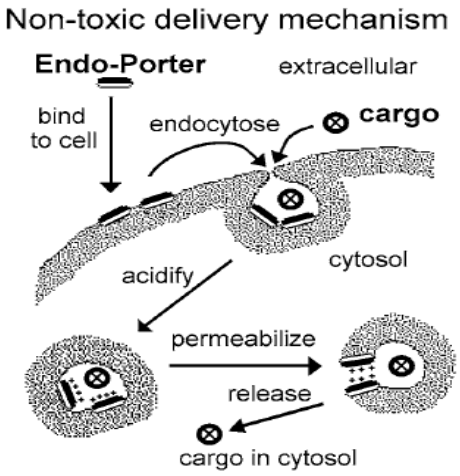


Figure 4-2 The mechanism of MO delivery by Endo-Porter: The binding of the Endo-Porter to the membrane allows its endocytosis along with the MO. In the cytosol, acidification of the endosome results in permeabilization of the endosome membrane and release of the MO (Summerton, 2005).

In addition, the transport of the MO oligonucleotides into the *C. merolae* cytosol was attempted by electroporation. This method allows introduction of DNA into the cells by application of an electrical field, which increases permeability of the cell membrane (Potter & Heller 2011). Electroporation has been shown to efficiently deliver genetic material into algal cells, such as *Chlamydomonas reinhardtii*. Unfortunately, not much information regarding electroporation of *C. merolae* cells is available. Although comparison between *C. reinhardtii* and *C. merolae* genomes presented a great conservation of cell wall biosynthesis genes, electron microscope studies suggests the absence of a cell wall in *C. merolae* (Kuroiwa et al. 1994; Misumi et al. 2005). Therefore, the delivery of genetic material from the membrane to the cytosol of the cells by electroporation is presumably less challenging. Thus, in this chapter, different attempts to block splicing through the delivery of MOs to the cytosol of *C. merolae* cells, and the resulting impact this process has on the cell survival were investigated.

4.2 Materials and Methods

4.2.1 C. merolae cell growth for assessment of doubling time

C. merolae (10D strain) cells were obtained from the Microbial Culture Collection at the National Institute for Environmental Studies in Tsukuba, Japan (mcc.nies.go.jp/). Cells were grown in MA2 media in a cultivation chamber exposed to white light at 42.0^{o} C and air was supplied that was supplemented with 2.0% CO_2 (Kobayashi et al. 2010). The assessment of the optimal wavelength for optical density (OD) measurements was performed with a spectrophotometer. Absorbance of the culture at different wavelength was measured, a graph of absorbance versus wavelength was plotted. The survival of the cells at pH values higher than 3 was also tested since morpholino oligos are unstable at pH 3 according to the manufacturer. Cells were grown at pH values of 4 and 4.65 and visualized with an Olympus BX61 fluorescence microscope using the following Semrock (IDEX Health and Science) filters: FITC-3540C, TxRED-4040C, DAPI-5060C for 5 days. Cell survival was checked at higher pHs such as 6, 6.5, and 7, and MA2 media was supplemented with 0.4 mM of sorbitol. The addition of sorbitol to MA2 media allows the cells to survive at this high pH. For assessment of cell culture doubling time, the OD of three 50 mL cell cultures (at 624 and 750 nm) was checked every 8 hours for 5 days at pH 4, and graph of OD versus time was plotted.

4.2.2 Treatment of cells with MO and vivo-MO

To block the binding of U2 snRNA to the BPS, a MO oligonucleotide complementary to the BPS binding sequence of U2 was designed. The BPS MO was purchased from Gene Tools along with a standard control that is not complementary to U2 (Table 21; Figure 4-3). The BPS MO and control contain blue and green emitting fluorescence tags, respectively, for assessment of

introduction of MO into the cell. When attempting to deliver the MO by the Endo-Porter delivery system, *C. merolae* cells were cultured at pH values of 4, 4.65, 6, 6.5 and 7. One millilitre of cell culture at an OD_{750} between 0.8-1.0 was treated with 10 µM of morpholino and 2, 4 and 6 µM of Endo-Porter for 1 to 48 hours and exposed to white light at 42.0° C. In addition, the p200 vector that has been used previously to transform *C. merolae* cells was attempted to be delivered by endo-porter.

When delivering 20 µg of the p200 vector using the Endo-Porter, cells were treated with 6 µM of Endo-Porter. The p200 vector is comprised of a sequence that encodes the GFP protein, which expression of GFP is controlled by the heat-shock CMJ101C promoter (Sumiya et al. 2014). Therefore, transcription occurs when the cells are exposed to elevated temperature, and production of GFP is assessed by microscope visualization. Treated cells were added to a 50 ml culture and incubated at 36.0° C overnight, instead of at 42.0° C, to prevent transcription (Sumiya et al. 2014). After overnight growth, cells were heat shocked for 1 hour at 50.0° C for transcription of GFP mRNA, and protein expression was checked under a microscope.

Hashimoto et al. (2016) describes the treatment of a parasite with MO without usage of any delivery system, and incubation of cells with MO resulted in the delivery of oligonucleotides to the cytosol. Thus, this procedure was replicated, and the cell growth of treated and untreated cells was checked for 8 hours in triplicate. A 50 ml culture was incubated for 2 days until an OD_{750} of 0.2 was achieved, and culture volume was reduced to 250 µL followed by treatment of cells with 10 µM of MO for 8 hours. Control cultures were treated with water. After 8 hours of treatment, cultures were resuspended to their initial volume and left in the incubator for approximately 3 days. Changes in the cell growth of the culture were checked daily by assessment

of cell density. The assessment of splicing blockage was done by Radu Pasca, who performed RT-PCR of treated samples.

Blockage of binding of U2 snRNA to the BPS was also attempted by treating cells with vivo-MO oligonucleotides that are complementary to the BPS binding sequence of U2. Vivo-MOs were purchased from Gene Tools flanking the 3` and the 5`ends of the BPS (Table 21; Figure 4-3). Since neither oligos was fluorescent, the introduction of the oligonucleotides was checked by monitoring cell growth. A 150 ml culture was incubated for 36 days until an OD_{750} of 0.2 was achieved, and the culture volume was then divided into three 50 ml cultures. Volume was reduced to 1 liter by centrifugation and treated with 10 µM of MO for 12 hours. The same procedure was performed to control cultures; however, cells were treated with water. After 12 hours of treatment, cultures were resuspended to their initial volume and left in the incubator for approximately 3 days. Changes in cell growth were checked daily through assessment of cell density. Radu Pasca assessed splicing blockage performing RT-PCR on collected samples.

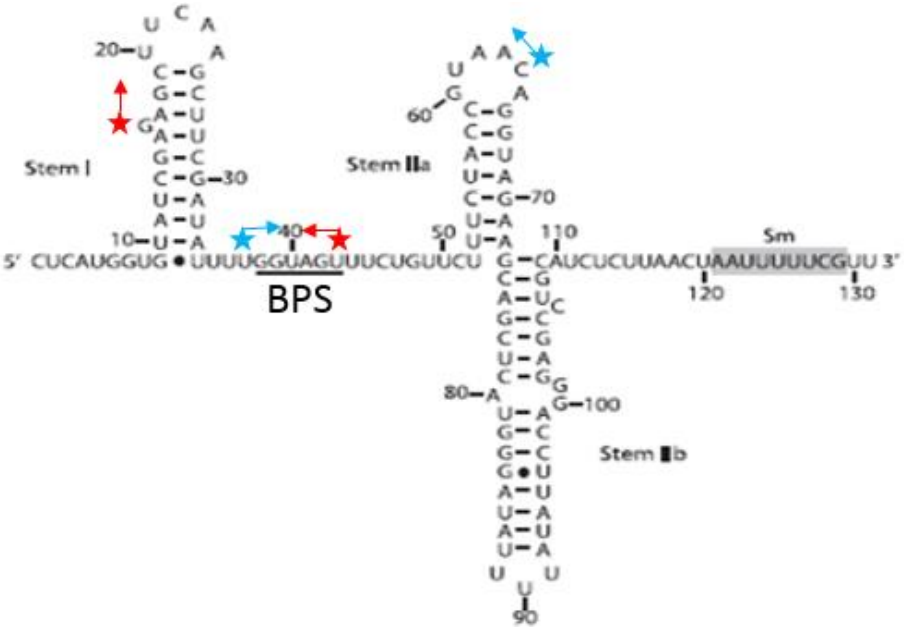


Figure 4-3 Region of binding of the MO oligonucleotide: The stars represent the 5` and 3` ends of the sequence that the designed MO is binding to. The blue stars cover the region of binding of the 5` flank Vivo-MO, and the red stars cover the 3` flank MO and vivo-MO.

Table 20 DNA oligonucleotide sequence of the MO and vivo-MO designed for binding to U2 snRNA: DNA sequences are shown from 5` to 3`.

Oligo	Oligo sequence
Vivo MO BPS (5'-flank)	ACTACCAAAATATCGAAGCTTGAAGCTC
Vivo MO BPS (3'-flank)	GTTACGGTAGAAAGAACAGAAACTACCA
MO BPS and 3'-flank blue fluorochrome	GTTACGGTAGAAAGAACAGAAACTACCA
Standard Control oligo	CCTCTTACCTCAGTTACAATTTATA

4.2.3 Delivery of MO by electroporation

Delivery of MO was attempted under several electroporation conditions. The MA2 media (Kobayashi et al. 2010) was compared to the optimal media for electroporation of protoplasts (Potter & Heller 2011), and the MA2 media was modified accordingly. As described by Potter & Heller (2011), the salt concentration in the electroporation buffer can change the efficiency of electroporation. Therefore, the phosphate concentration was increased from 8 to 10 mM and the $CaCl_2$ concentration was increased from 1 to 5 mM. Sorbitol was added to the media to allow cell growth of cells at higher pH values. Since there is no available protocol for the preparation of *C. merolae* cells for electroporation, a protocol was designed based on the preparations described by Potter & Heller (2011). A 15 ml culture was grown and harvested at an OD_{750} of 0.8, as *C. merolae* cells present a more efficient transformation at an early stage of growth (Sumiya et al. 2014). Half of the initial cultures were spun down (approximately 2 million cells), at 3000rpm for 10 minutes at room temperature. Cells were resuspended on 1 ml of modified MA2 media. Electroporation was performed in a Gene Pulser Xcell electroporation system (Biorad) as described by manufacturer. A summary of the attempted conditions attempted are presented in table 22.

Table 21 Summary of electroporation conditions

OD_{750}	Voltage (V)	Capacitance (uF)	Temperature	DNA mass (ug)	Pulses	DNA carrier	Sample volume (ul)	# cells
0.8	200	800	20-25°C and 0-5 °C	1, 10, and 20	1	n/d	40	10 x 10^6/ml
0.8	300	500	20-25°C and 0-5 °C	1, 10, and 20	1	n/d	40	10 x 10^6/ml
0.8	1200	25	0-5 °C	1, 10, and 20	1	n/d	40	10 x 10^6/ml
0.8	1500	n/d	0-5 °C	1 and 0.5	1	n/d	40	10 x 10^6/ml
0.8	2000	n/d	0-5 °C	1 and 0.5	2	n/d	40	10 x 10^6/ml
0.8	3000	n/d	0-5 °C	1 and 0.5	1	n/d	40	10 x 10^6/ml
1.5-3	1500	20	0-5 °C	1	1	n/d	40	4 x 10^6/ml
1.5-3	1200	25	0-5 °C	1	1	n/d	40	4 x 10^6/ml
0.3	2000	10	0-5 °C	0.5, 1, and 2	1	n/d	40	4.5x10^5/ml
1.5-3	1800-2300	10	0-5 °C	2.5	1	Salmon sperm DNA	250	1 x 10^8/ml
1.5-3	250	n/d	20-25°C	0.4	2	n/d	40	4 x 10^6/ml

4.3 Results

4.3.1 Assessment of *C. merolae* cell growth

First, the best wavelength to measure the cell density of *C. merolae* cultures was determined. After measurement of the absorbance of the cultures at different wavelengths, the resulting values were plotted on a graph of absorbance versus wavelength (Figure 4.4). The graph presents the highest absorbance readings at approximately 430, 624, and 684 nm. However, most *C. merolae* publications present measurement of cell cultures at 750 nm. Indeed, the United States Environmental Protection Agency (EPA - 2002) recommends measuring algae growth at a wavelength of 750 nm. It was suggested, when using spectrophotometric absorbance, to construct calibration curves that relate absorbance and cell density. However, some authors state that the correlation of cell density to the light absorbance of chlorophyll better represents the cell growth of algae (Rodrigues et al. 2011; Hersh and Crumpton, 1987; Fargasová, 1996; Rojícková-Padrtová

et al. 1998). The measurement of the absorbance of chlorophyll is around 664 nm, which is consistent with the peak seen on graph (between 624 and 684 nm on the below graph). Although, the highest peak is the 430 nm, it was decided to measure the cell growth of the cells at both 750 and 624 nm, since according to literature cell growth is better correlated to cell density absorbance at these wavelengths.

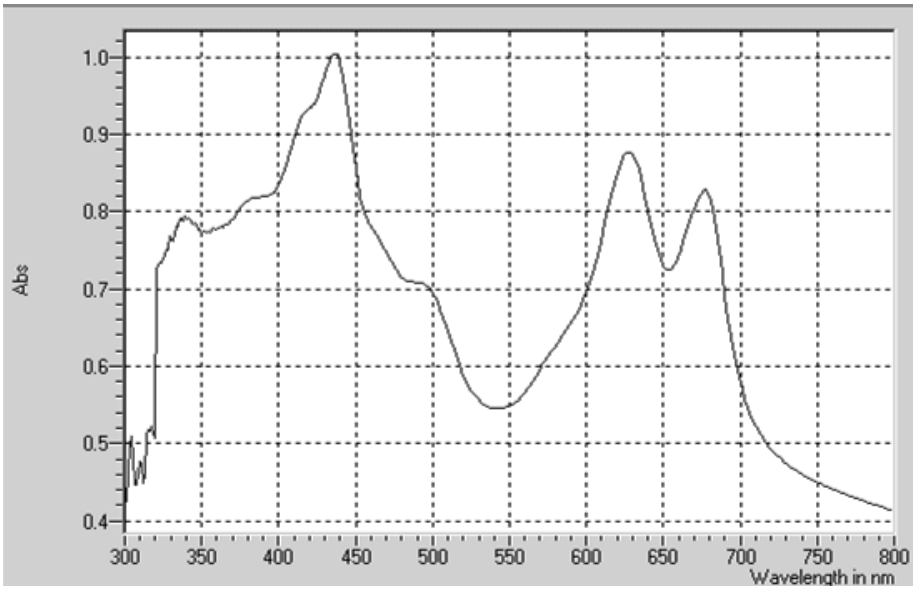


Figure 4-4 Assessment of optimal wavelength for measurement of *C. merolae* culture optical density. Graph presents three possible optimal wavelengths at approximately 430, 624, and 684 nm.

When checking for survival of the cells at pH values of 4.0, 4.65, 6.0, 6.5 and 7.0, the cells were analysed under a microscope at different time points (Figure 4.5). At a pH of 4.0 and 4.65 the cells survived for over 96 hours. At a pH of 6.0, 6.5 and 7.0 the cells died in 2 hours. However, when the media was supplemented with sorbitol, known to reduce osmotic stress, the cells survived at these pHs for over 24 hours (Figure 4.5g, Figure 4.5h and Figure 4.5i).

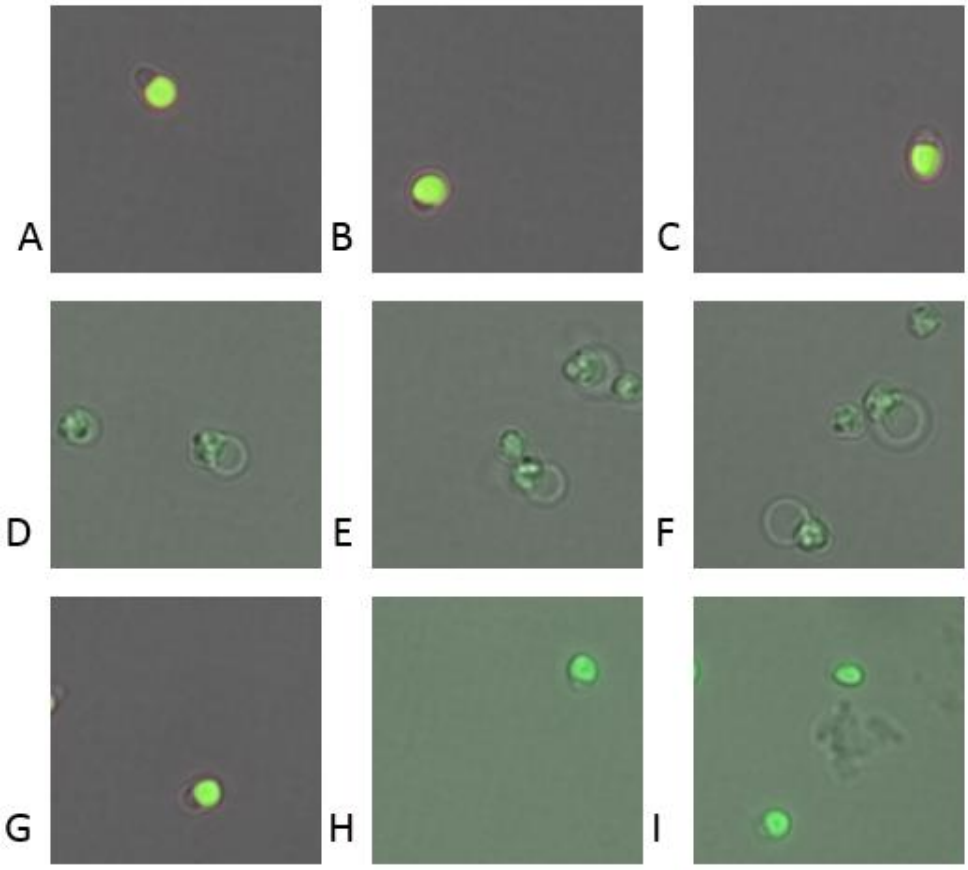


Figure 4-5 Microscope images of *C. merolae* cells at different pHs: Merged bright field and autofluorescence (FITC channel -green). Cells cultured at 4.0 (b) and 4.5 (c) for 96 hours presented similar morphology as the cells cultured at pH 3.0 (a). When growing cells at pH values of 6.0 (d), 6.5 (e) and 7.0 (f) for 2 hours, a different morphology of the cells was observed probably due to the stressed environment. When sorbitol was added to the media, cells survived at pH values of 6.0 (g), 6.5 (h) and 7.0 (i) for over 24 hours.

Calculation of the *C. merolae* doubling time at pH value of 4 was achieved through measuring of the cell density of three cultures for 5 days. The relationship between the average of the logarithm of the optical density (OD) at both wavelengths (624 and 750 nm) was plotted against time. As observed in figure 4-6, when measuring the cell density at 624 nm, the R^2 value was slightly higher compared to 750 nm. It suggests a better relationship between OD and time at 624

nm. By analysis of the growth curve, *C. merolae* cell density was calculated and found to double every 16.3 ± 0.005 hours at OD_{750} and 16.8 ± 0.004 hours at OD_{624}.

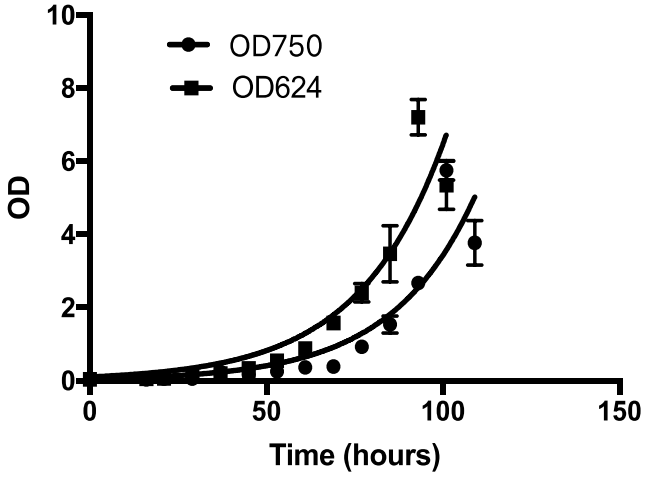


Figure 4-6 Cell growth of *C. merolae* cells: Relationship between the average (n=3) optical density at 750 and 624 nm and time (in hours).

4.3.2 Treatment of *C. merolae* cells with MO and vivo-MO

The standard control MO (10 μM) was delivered by treatment of cells (OD_{750} 0.8) with different concentrations of Endo-Porter (2, 4 and 6 μM) at pH 3. The Endo-Porter has been proven to successfully deliver the MO oligonucleotides to the cytosol mammalian cells via endocytosis; however, it has never been used to deliver oligonucleotides to algal cells. The assessment of the delivery of MO to the cytosol was done by microscope analysis. Since the MO is fluorescently labelled, bright green spots were expected to be observed in the cytosol of the cell under the microscope. Unfortunately, it was concluded that the MO was not delivered (Figure 4-7). Microscopic images of the cells treated with 2 μM (Figure 4-7b) and 4 μM (Figure 4-7c) of Endo-Porter presented similar cell morphology to the untreated cells (Figure 4-7a) after 24 hours. However, 6 μM of Endo-Porter was found to be toxic to cells after 16 hours of treatment (Figure 4-7d). The delivery of the standard control MO by treatment of cells (OD_{750} 0.65) with 4 μM of

Endo-Porter, was again attempted. Cells were checked under the microscope after 16, 24, 48 and 66 hours (Figure 4-7e). Delivery of MO to the cytosol at these time points was not observed.

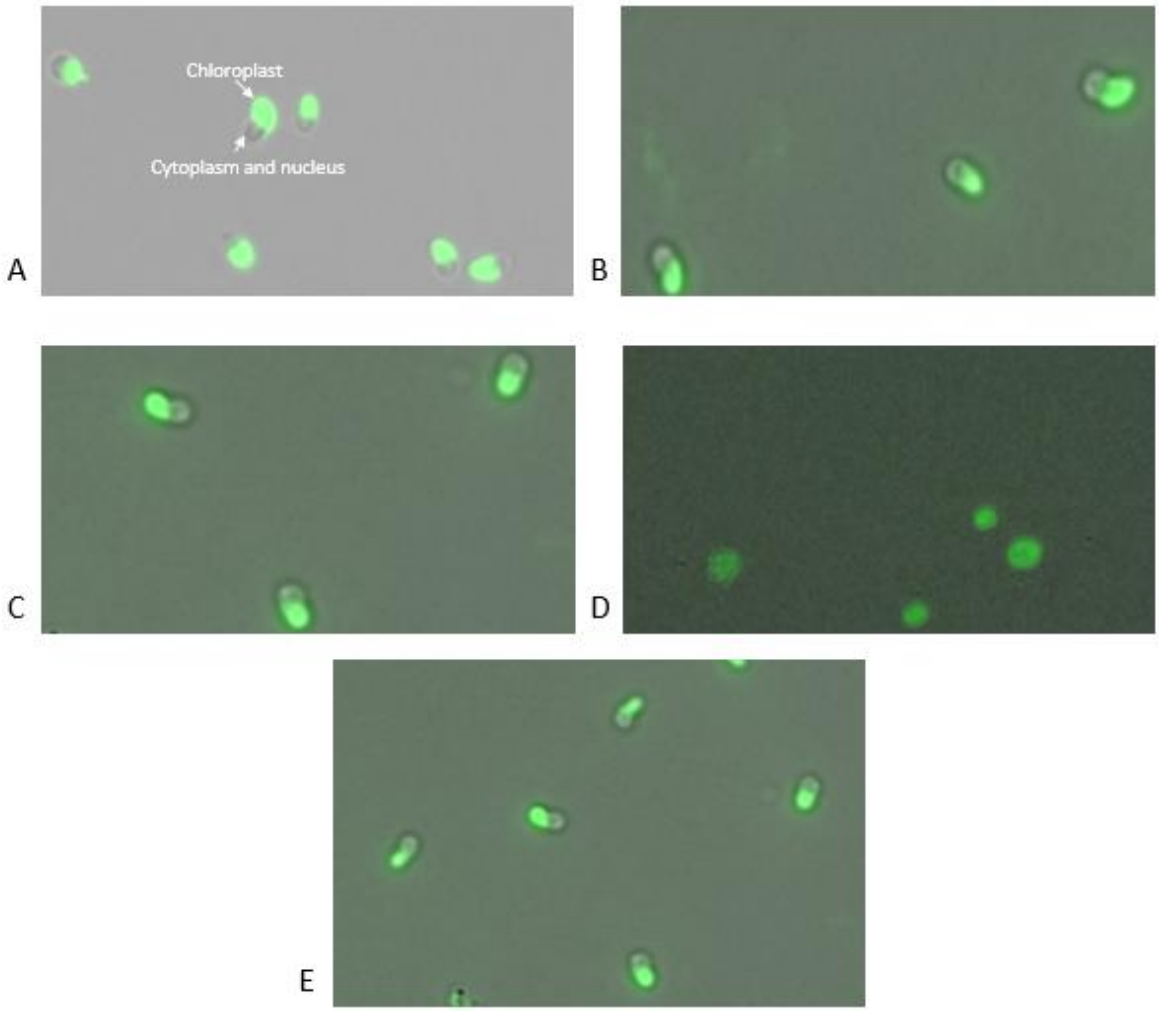


Figure 4-7 Microscopic images of the *C. merolae* cells treated with control MO and Endo-Porter: Merged bright field and autofluorescence (FITC channel - green). a) untreated cells. Arrows indicated the chloroplast, nucleus and cytoplasmic regions. The presence of green fluorescence in the cytoplasm region of the cells when cells were seen when treated for 24 hours with 10 µM control MO and 2 µM (b) or 4 µM (c) of Endo-Porter. d) Cells treated with 6 µM of Endo-Porter presented a different morphology suggesting toxicity at higher concentration of Endo-Porter. e) After treatment of cells with 10 µM control MO and 4 µM of Endo-Porter for 66 hours, delivery of MO to the cytosol of cells was still not observed.

Since the first attempts were done at pH 3, a repeated delivery of the MO was attempted at a higher pH. According to the manufacturer, MO oligonucleotides are unstable below pH 3.

Therefore, since it was confirmed previously that *C. merolae* cells survived at pH 4 and 4.65, the delivery of the BPS MO and control to the cells (OD_{624} 0.8) was reattempted at pH 4.5 via Endo-Porter. Microscopy images confirmed no delivery of BPS MO after 24 hours of treatment since no blue fluorescence was observed in the cell (Figure 4-8). Further, it was decided to treat the cells at much higher pH values: 6, 6.5 and 7. At these pH values, the delivery of p200 vector and BPS MO by addition of Endo-Porter was attempted. Microscopic images of the cells after 24 hours of treatment indicated no delivery of vector and MO (Figure 4-8b).

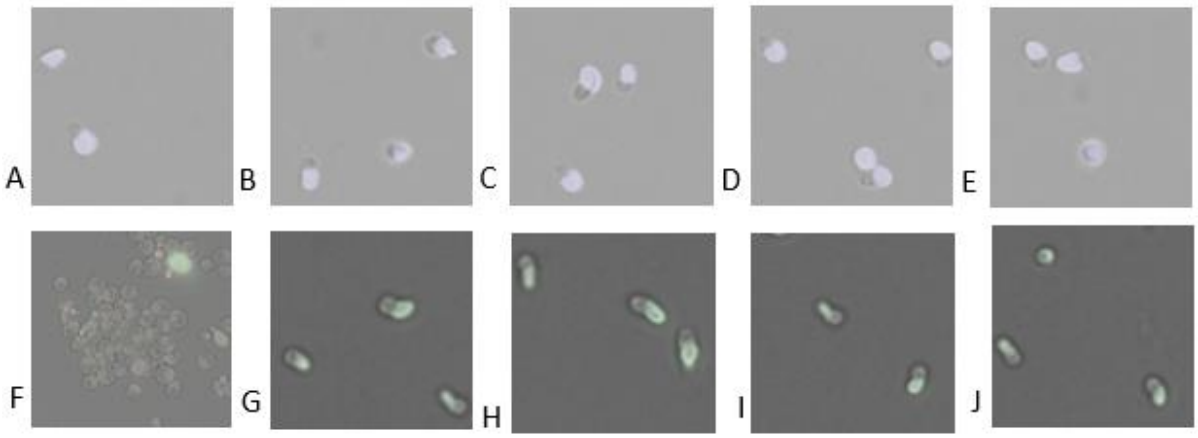


Figure 4-8 Microscope images of the *C. merolae* cells treated with control BPS MO and Endo-Porter at higher pH values: Merged bright field and autofluorescence for visualization of BPS MO (DAPI channel - blue) and GFP (FITC channel -green). (a) untreated cells. The presence of blue fluorescence in the cytoplasmic region of the cells when cells were treated for 24 hours with 10 µM control MO and 4 µM at pHs (b) 4.5, (c) 6.0, (d) 6.5, and (e) 7.0. was not observed. (f) Positive control (the cells transformed by via PEG with p200 vector showing GFP fluorescence). (g) Failure of delivery of p200 via Endo-Porter in the cells at pH values 6.0, (h) 6.5 and (i) 7.0.

Since all attempts to deliver MO to *C. merolae* via endocytosis failed, it was hypothesised that a different mechanism of endocytosis in this alga could be making this method of delivery challenging. Indeed, the Endo-Porter has never been used in alga and not much is known about endocytosis in *C. merolae*. Therefore, this procedure was attempted in yeast cells. This cell type

can be easily visualized under the microscope and were available; therefore, *P. Pastoris* cells were treated with Endo-Porter and MO. *C. merolae* and *P. pastoris* cells were treated at pH 7 for 20 hours, and microscopy images of the cells showed no delivery of MO (Figure 4-9).

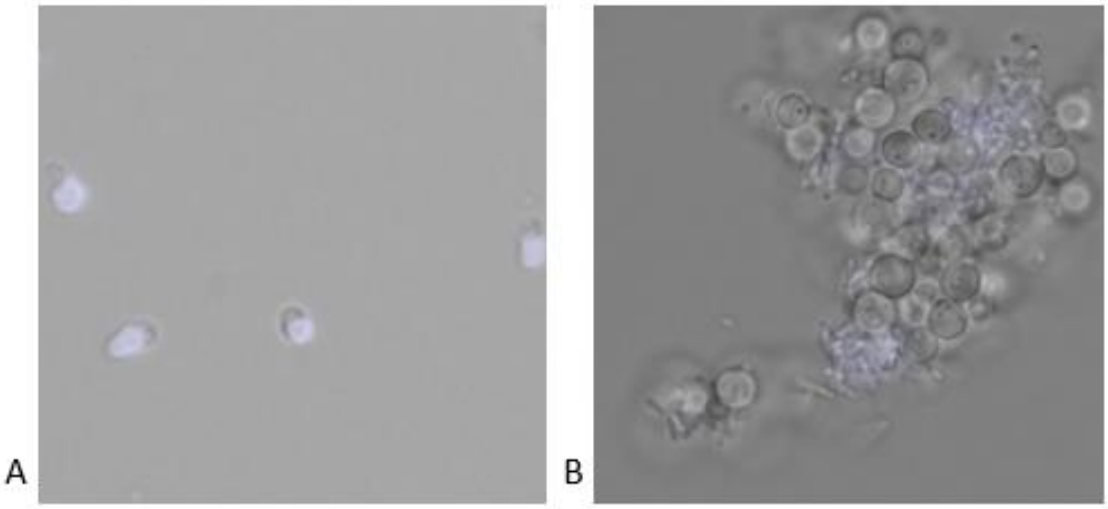
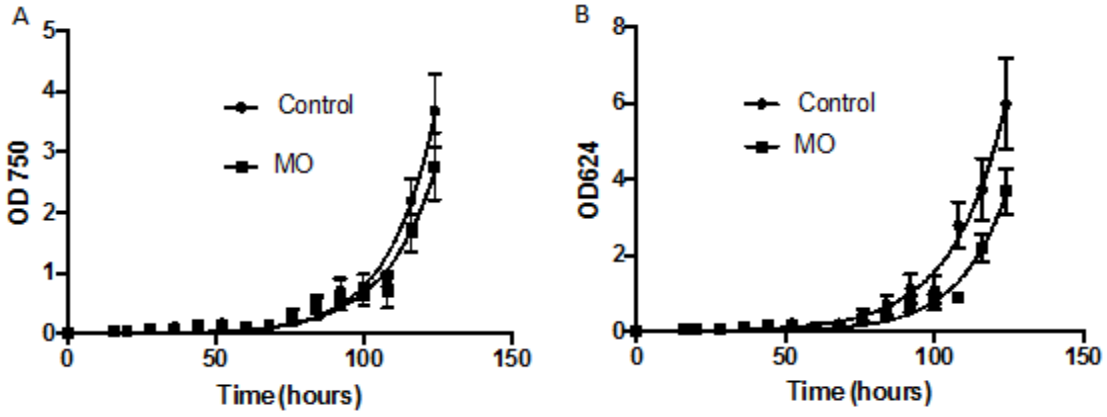


Figure 4-9 Delivery of MO to *P. pastoris* and *C. merolae* at pH 7: Merged bright field and autofluorescence for visualization of BPS MO (DAPI channel - blue). Failure of delivery of BPS MO oligonucleotide to the cytosol of (a) *C. merolae* and (b) *P. pastoris* is observed.

Hashimoto et al. (2016) have described successful introduction of MO in protozoa cells without addition of any reagent to assist delivery of the MO. Therefore, the same protocol was attempted in *C. merolae*. The introduction of the MO to the cytosol and its binding to U2 snRNA were expected to result in changes in cell growth. Thus, the cell growth of treated and untreated cells was compared, and cell growth was plotted against time. Cell density of the cells was assessed at both of 750 and 624 nm wavelengths. Graphs were plotted on logarithmic scale showing exponential growth of cells for both control and treated cells (Figure 4-10). No significant changes were observed in the growth of treated and untreated cells. For a better analysis of data and calculation of doubling time, a linear graph of the log of the OD against time was plotted. Treated

and untreated cells did not present significant changes in doubling time for both graphs using cell densities measured at OD_{624} and OD_{750} ($p > 5\%$). At a wavelength of 750 nm, a doubling was observed after 10.93 ± 0.004 hours for the control cultures and 12.07 ± 0.004 for the treated cultures (Figure 4-10a). At a wavelength of 624 nm, a doubling time of 12.42 ± 0.003 hours was calculated for the control and 10.93 ± 0.004 hours for the treated cells (Figure 4-10b). It was observed that resuspension of cells after treatment of cells caused a decrease in cell density of the culture, which was clearly observed on the graphs after 60 hours. Therefore, the data points were divided into two trendlines (before and after 60 hours) and differences in doubling times were checked between the control and the MO treated cells (Figure 4-10c). The slope of the equation using the data collected before and after 60 hours differs as a drop of cell density is caused by resuspension of cells. However, when comparing the slope value of untreated and treated cells, a significant change in cell growth after 60 hours is not detected. Indeed, the doubling times of cells after 60 hours does not change dramatically since after 60 hours, the control and treated cells present doubling times of 10.8 ± 0.005 and 11.8 ± 0.006 hours respectively ($p > 5\%$). It is concluded that the MO is either not causing changes in cell growth or it is not being delivered since the difference of doubling between control and treated cells is almost the same before 60 hours, 20.41 ± 0.004 and 20.28 ± 0.007 hours, respectively.



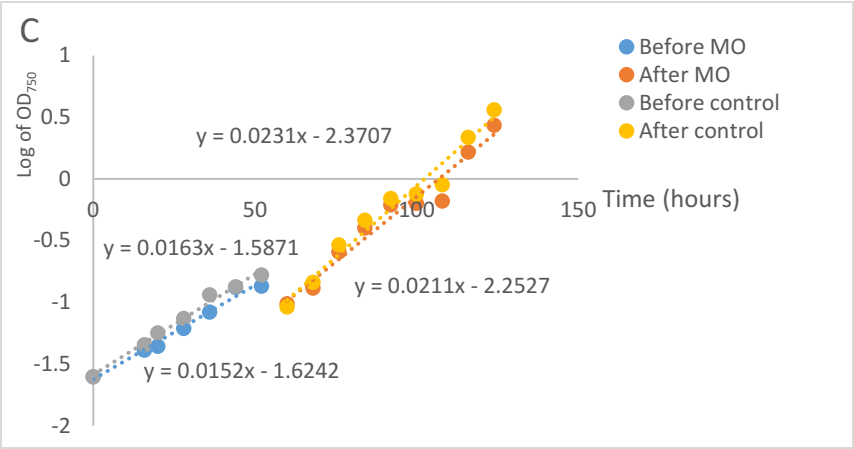


Figure 4-10 Cell growth of *C. merolae* treated with the MO that targets the branch point binding site of U2 snRNA: a) Cell growth of untreated and treated cells presenting the relationship between the average (n=3) of optical density at 750 nm and time (in hours). b) Cell growth of untreated and treated cells presenting the relationship between the average (n=3) of optical density at 624 nm and time (in hours). c) Relationship between the average (n=3) of logarithm of the optical density at 750 nm and time (in hours). Equations represent cell growth of treated and control cells before and after 60 hours.

As all methods to deliver the MO to the cells failed, it was decided to try a different approach, vivo-MO, that does not rely on the addition of any delivery reagent. This type of oligonucleotide is comprised of a MO attached to a delivery moiety that is capable to interact to the cell membrane allowing the MO to enter the cells. The cells were treated with vivo-MOs flanking the 3` and 5` ends of the BPS sequence of U2. Since vivo-MOs are not fluorescently labelled, blockage of U2`s binding to the BPS was assessed by variations in cell growth. First, cells were treated with 10 µM of MO (in duplicate), and the cell growth of the treat cells was compared to untreated cells (Figure 4-11). Graphs showed a decrease in cell growth after the addition of vivo-MOs to the cells. A significant increase in doubling time between controls and treated cells was observed. Control cultures presented a doubling time of 18.1 ± 0.038 hours, and cells treated with vivo-MOs flanking the 3`end and 5 `end of the BPS presented doubling times of 49.86 ± 0.014 and 47.94 ± 0.011 hours respectively ($p < 5\%$).

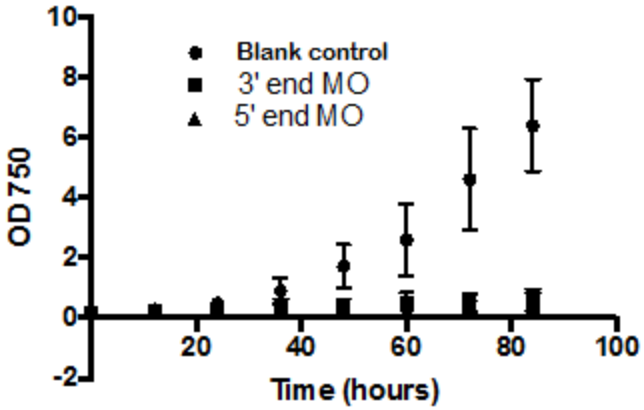


Figure 4-11 Cell growth of *C. merolae* cells treated with vivo-MO: Growth of cells treated with 10 µM of vivo-MOs presenting the relationship between the average (n=2) of optical density at 750 nm and time (in hours). Trendlines represent each equation. Graph presents a decrease in growth when both vivo-MOs were added.

Experiments checking for the toxicity of guanidinium were repeated since vivo-MO has been proven to be toxic at some concentrations depending on the cell type according to Gene Tools. Since vivo-MOs contain guanidinium groups, the cell growth rate difference observed could be due to the presence of the guanidinium. Therefore, cells were treated with 8 and 80 µM of guanidinium to check it`s toxicity. In addition, a comparison of cell growth between cells treated with 10 µM and 1 µM of the vivo-MO (flanking the 5` of the BPS) was performed to confirm that the oligonucleotide was not toxic to the cells. Unfortunately, these results were not consistent with previous results, since no significant changes in cell growth of treated cells was observed in the graphs (Figure 4-12; $p > 5\%$). The doubling times for control cells, treated cells with 1 µM Vivo-MO, treated cells with 10 µM Vivo-MO and treated cells with 8 µM guanidinium were 16.83 ± 0.004, 16.79 ± 0.005, 17.15 ± 0.040, and 17.08 ± 0.044 hours, respectively (Figure 4-12a). The doubling times for control cells, treated cells with 1 µM Vivo-MO, treated cells with 10 µM Vivo-MO and treated cells with 80 µM guanidinium were 16.83 ± 0.040, 17.59 ± 0.040, 19.1 ± 0.036, and 16.02 ± 0.043 hours, respectively (Figure 4-12b). Therefore, the addition of guanidinium (8

and 80 µM) is presumably not toxic to the cells based on the cell growth and doubling times. In addition, both treatments of cells with 10 µM vivo-MO (Figures 4-12a and b) confirm that at this concentration the MO are not toxic to the cells since cell growth does not change compared to control and cells treated with 1 µM vivo-MO.

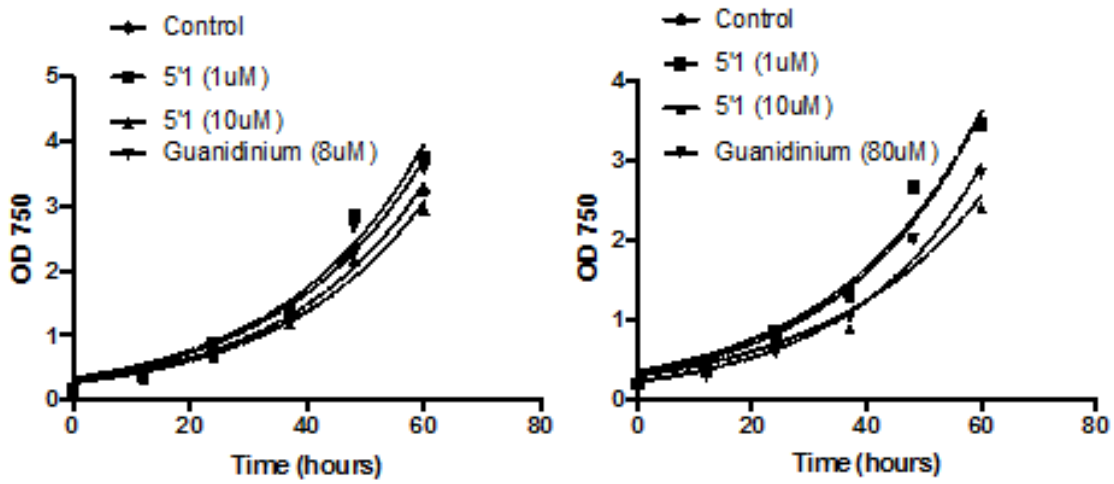


Figure 4-12 Growth of *C. merolae* cells treated with two concentrations of vivo-MO and guanidinium: Growth of cells treated with 1 µM and 10 µM of the vivo-MO flanking the 5`end of the BPS was assessed twice in duplicates (a) and (b). Treatment of cells with 8 µM (a) and 80 µM (b) of guanidium was also performed in duplicate. Cell growths present no significant difference between control cells and treated cells with both vivo-MO and guanidinium.

Since the treatment of cells with vivo-MO was presenting inconsistent results, it was not possible to conclude that the vivo-MO was affecting cell growth due to binding to U2 (Table 23). It was expected that by increasing the incubation temperature cells would be more stressed favouring acceptance of foreign genetic material, therefore a treatment was conducted at 50°C. 5 µM of vivo-MOs were added to cells and assessed the growth for 24 hours. A drastic decrease in the growth of cells was observed as indicated on figures 4-13a and b. Conversely, the control cultures presented changes in cell density but were able to recover after incubation at 42°C (after 150 hours of growth). To confirm that the decrease in cell growth was being caused by splicing prevention, Radu Pasca (undergraduate lab member) performed RT-PCR. Gene-containing introns were amplified and showed a clear difference in size between pre-mRNA and mRNA.

Surprisingly, RT-PCR results did not show changes between control and vivo-MO treated cells, suggesting that the MO was not affecting splicing (Figure 4-13c)

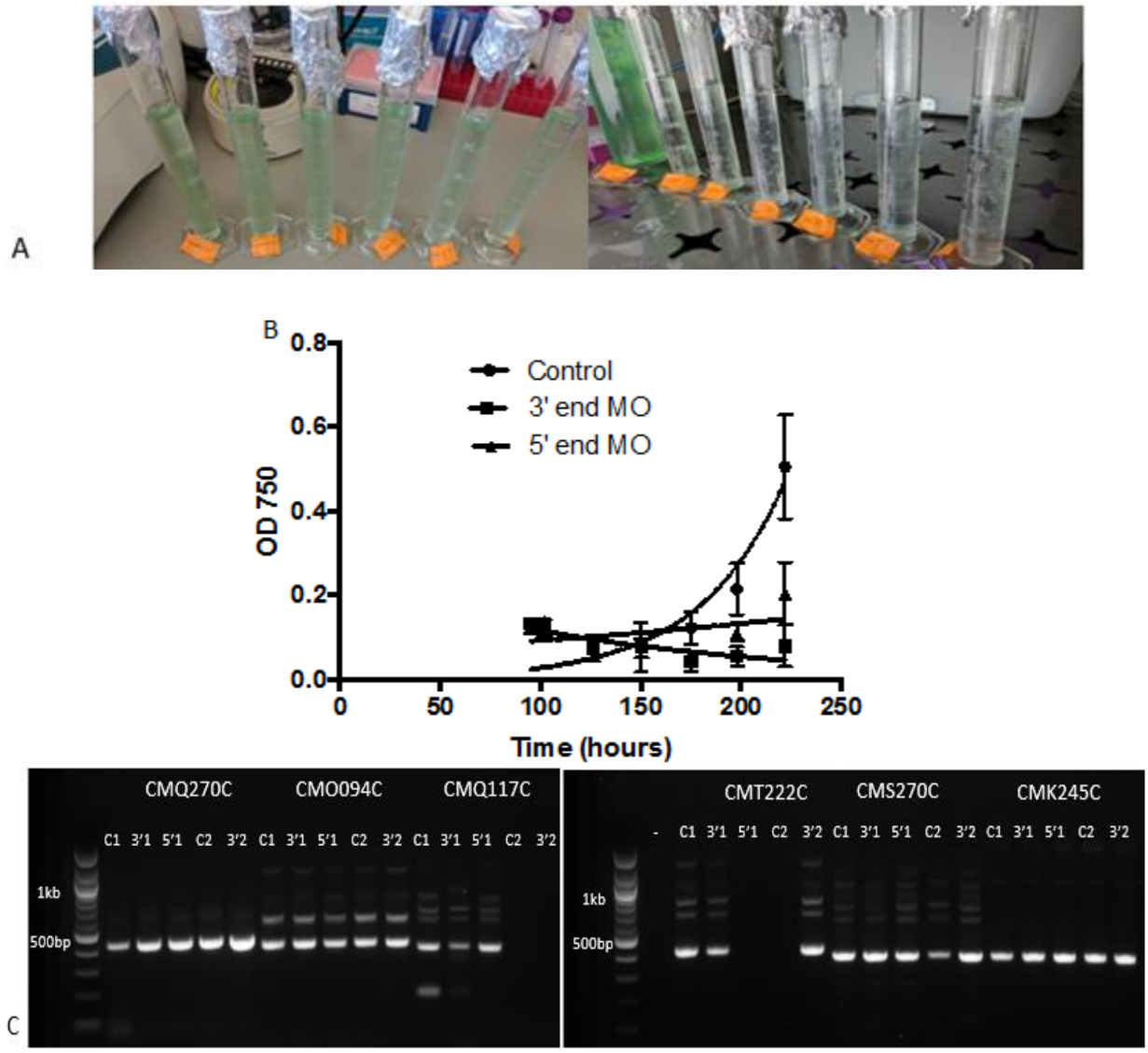


Figure 4-13 24 hours treatment of *C. merolae* cells with 5 µM of vivo-MO: a) A clear colour change of the cultures before (left) and after (right) treatment of cells was observed as a dramatic decrease in cell growth. b) Cell densities of the cells collected over time present an increase in cell density of the initial cultures incubated at 42ºC. As the cells are concentrated and split into control and vivo-MO treated cultures for 24 hours incubation at 50ºC, a decrease in cell density was observed. However, control cultures recover at 150 hours when cells are incubated at 42º C. The same is not observed for treated cells. c) RT-PCR analysis of untreated and treated cells shows that processing of mRNA is not being prevented. CMQ270C, CMO094C, CMQ117C, CMT222C, CMS270C and CMK245C genes were amplified since clear size difference between pre-mRNA and mRNA PCR products had been previously observed. C – indicates the control cultures; 3`- indicates the cultures treated with vivo-MO flanking the 3`end of the BPS; 5`- indicated the cultures treated with vivo-MO flanking the 5`end of the BPS.

Table 22 Summary of the doubling times of control and MO treated *C. merolae* cells. The correlation coefficients are greater than 0.92. Comparative statistics, ANOVA, was applied for comparison of control with treated cells.

Figure	Cell treatment	Doubling time (in hours) + Standard error	Significantly different (ANOVA)
4.13	Control	10.93 ± 0.004	No (p > 5%)
	MO	12.07 ± 0.004	
	Control (before 60 hours of growth)	10. 8 ± 0.005	
	MO (before 60 hours of growth)	11.8 ± 0.006	
	Control (after 60 hours of growth)	20.41 ± 0.004	
	MO (after 60 hours of growth)	20.28 ± 0.007	
4.11	Control	18.1 ± 0.038	Yes (p < 5%)
	10 µM Vivo-MO flanking the 3`end of the BPS	49.86 ± 0.014	
	10 µM Vivo-MO flanking the 5`end of the BPS	47.94 ± 0.011	
4.12 a	Control	16.83 ± 0.004	No (p > 5%)
	1 µM of Vivo-MO flanking the 5`end of the BPS	16.79 ± 0.005	
	10 µM of Vivo-MO flanking the 5`end of the BPS	17.15 ± 0.040	
	8 µM of guanidinium	17.08 ± 0.044	
4.12 b	Control	16.83 ± 0.040	No (p > 5%)
	1 µM of Vivo-MO flanking the 5`end of the BPS	17.59 ± 0.040	
	10 µM of Vivo-MO flanking the 5`end of the BPS	19.1 ± 0.036	
	80 µM of guanidinium	16.02 ± 0.043	

4.3.3 Delivery of MO by electroporation

In order to electroporate MO to the cytosol of *C. merolae* cells, several electroporation conditions were attempted and are described in table 22. Unfortunately, all attempts to electroporate MO to the cells failed. Microscopy images of the cells after electroporation confirmed no introduction of MO to the cells since no blue fluorescence was detected in cell

cytosol. As seen on figure 4-14, electroporation of cells would either cause death of cells or result in no delivery of MO to cells.

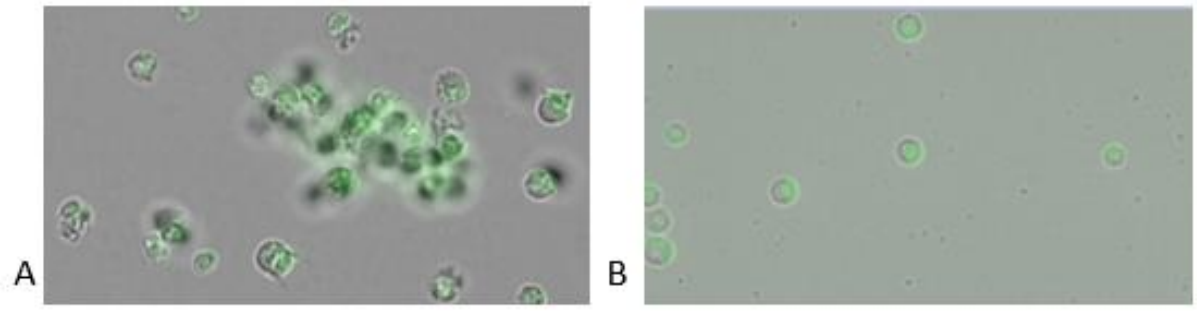


Figure 4-14 Electroporation of *C. merolae* cells for introduction of MO to cytosol: a) Cell death after electroporation at high voltage and low capacitance. b) Failure of delivery of MO after electroporation at low voltage and high capacitance.

4.4 Discussion

This chapter explored different methods to investigate splicing essentiality in *C. merolae*. In the early stages of assembly of the spliceosome, U2 snRNA binds to the BPS; therefore, it was proposed that blocking this early step should prevent splicing. It was expected that prevention of splicing would cause either death of cells or a decrease in cell growth. Thus, blockage was attempted through the use of MOs since these oligonucleotide analogues are known to efficiently block splicing in other organisms. In addition, MOs are capable of stably binding to secondary structures. However, this method of blocking splicing has never been tried in *C. merolae,* which made it necessary to explore different methods to deliver the MO oligonucleotides to the cells.

First, it was attempted to use a novel delivery system, Endo-Porter. To explore the best approach to efficiently deliver MO to the cells by this method, the concentrations of Endo-Porter and pH values of growth media were varied. As presented, all attempts to deliver MO to cells by Endo-Porter failed as the microscopy results did not confirm presence of the fluorescently labelled

MO in the cytosol of the cells. The reasons why the Endo-Porter failed to deliver the MO to the cells are unknown. Since the Endo-Porter is an endocytosis-mediated method, failure of delivery of MO might suggest a different mechanism of endocytosis in *C. merolae* cells. Unfortunately, little is known about endocytosis in *C. merolae*. However, the presence of the components that are part of the endocytosis machinery in algae have been observed. In addition, the interaction of these components for occurrence of endocytosis has been investigated in Chlamydomonas (Rappoport & Simon 2003). A comparison of the endocytosis system among eukaryotes shows that *C. merolae* lacks some of the endocytosis components, such as Rab11 and 7, which should result in absence of some pathways (Jékely 2008). Since it is proposed that *C. merolae* is comprised of a reduced endocytosis system, it might suggest a minimal endocytic activity in this alga. It is unknown what effects endocytosis in *C. merolae*. Presumably, the *C. merolae* cells are more likely to accept heterologous genetic material under stress conditions. Also, changes in pH media, for instance, should alter the chances of the Endo-Porter to enter the cell. Indeed, the endo-porter needs a basic environment to enter the cells followed by acidification of the endosome to be released in the cytosol (Summerton, 2005). Therefore, the reason that the Endo-porter failed to deliver the MO at pH 4 should be explained. However, when increasing the pH of the media (6, 6.5 and 7), delivery of MO to the cells was not observed. Presumably a basic environment does not favour endocytosis in *C. merolae*. Yeast cells were also treated with Endo-Porter. Failure of delivery was observed suggesting that the use of Endo-Porter can be a challenging method in cell types.

Since MO has been successfully introduced into protozoa cells without any additional reagent (Hashimoto et al. 2016), treating *C. merolae* cells with MO using the same approach was attempted. Microscopy results also showed no MO in the cytoplasm of the cells. This suggests that protozoa cells are more acceptable to foreign genetic material than *C. merolae cells*. It was also

attempted to block splicing by treatment of cells with vivo-MO. Since this oligonucleotide is attached to a delivery moiety, it was expected that through the interaction of the guanidinium head groups to phosphates of membrane phospholipids, the vivo-MO would enter the cells. Some of the results presented a decrease in cell growth. However, repetitive treatment of the cells with vivo-MO presented a deviation of results. Survival of cells at both 1 μM and 10 μM of vivo-MO suggest that the delivery moiety is not toxic to cells. Indeed, since the delivery moiety is comprised of a octaguanidine, it was decided to add guanidinium to the cells. The addition of guanidinium (8 and 80 μM) also presented no cell growth changes. Promising results were observed when cells were treated with 5 μM of vivo-MO; however, the RT-PCR results (provided by Radu Pasca) confirmed that the MO was not preventing splicing. Therefore, it was concluded that the changes in cell growth was not caused by blockage of splicing.

Electroporation was another method that failed to deliver MO to the *C. merolae* cells. Due to the lack of knowledge regarding electroporation in *C. merolae*, several conditions were tried (Table 22). Surprisingly, the absence of the cell wall does not facilitate the introduction of MO since all attempts to electroporate MO to the cells failed. Further investigation of the cell membrane of *C. merolae*, or the use of another types of electroporation equipments would be necessary to investigate these negative results.

In conclusion, it is still unknown if splicing is vital for *C. merolae* cells. The use of MO did not present to be the best method to address this question since delivery of this oligonucleotide has proven to be challenging. In the future, new approaches can be attempted to address these questions, such as by blockage of expression of essential spliceosome core proteins. This should cause death of cells if splicing is vital to *C. merolae*.

5. Chapter Five - Concluding remarks

To conclude, this thesis explores different approaches to investigate processing of mRNA in *C. merolae* by focusing on U5 snRNP. This ribonucleoprotein subunit is comprised of most of the core proteins of the spliceosome playing an important role in both early and late stages of the spliceosome assembly process. In addition to this, it has been proposed that U5 snRNA could be replacing U1 snRNA, recognizing the 5`splice site of the mRNA. Therefore, this thesis describes different techniques to investigate the U5 snRNA associated proteins and methods that could be performed to investigate the 5`splice site recognition.

First, an investigation of the U5 snRNP structure and function by co-expression of the proteins that associate to U5 was performed. A plasmid carrying all protein genes of interest was successfully constructed. Since co-expression of the U5 proteins failed, it was decided to express each protein individually at different expression conditions. Although the largest proteins, Prp8, Brr2, and Snu114, failed to express, the smallest U5 snRNP`s protein, Dib1 was expressed. In addition, the Sm complex proteins, Sm B, F, D1, D2, E, G, and D3 were co-expressed.

The successful co-expression of the Sm complex allowed functional and assembly investigation of the complex in *C. merolae*. The third chapter presents a two-step co-purification of the Sm complex. Any additional steps to assemble the Sm complex were required, and the mass spectrometry analysis confirmed the presence of all seven Sm in the co-purified sample. Functional assembly of the Sm complex was first assured by electron microscope, showing formation of a few rings in solution. These results are consistent with the ring formation of this complex in other organisms demonstrating that the assembly of the recombinantly co-purified complex is not occurring randomly. In addition, functionality of the complex was investigated by binding to U2, U4 and U5, since this complex is known to bind to these snRNAs in organisms for biogenesis.

EMSA and FP binding assays confirmed binding of the Sm complex to U4 and U2. U5 presented difficulties to reproduce binding of to the complex by EMSA. However, FP was performed using an oligonucleotide of the proposed U5 Sm site, which happens to be the same Sm site present in U2. Binding of the U5 and U2 Sm site to the Sm complex was observed. Therefore, both electron microscope and binding results confirmed that the recombinantly co-purified Sm complex is functional. Interestingly, the structure of U2, U4 and U5 snRNA and absence of an Sm assembly factor, SMN, suggests that the Sm complex pre-assembles prior to binding to the snRNAs. However, both EMSA and FP binding curves indicate that Sm proteins are cooperatively binding to U2, U4 and U5 (hill coefficient higher than 1). These results raise questions regarding the assembly of the Sm complex in absence of any additional proteins. Indeed, further exploration of the Sm`s assembly will be necessary to confirm formation of Sm dimers and trimers prior to binding to the snRNAs.

The fourth chapter focused on the investigation of the relevance of the splicing process to *C. merolae*. To achieve this, blocking the binding of the U2 snRNA to the BPS was attempted. A novel oligonucleotide was used, MO, that had complementarity to the U2 BPS binding region. Two methods to deliver the MO to the cytosol of the cells were attempted: the endo-porter delivery reagent and electroporation. Unfortunately, both methods failed to deliver the MO. Therefore, a third method was performed, vivo-MO. Incubation of cells with vivo-MO showed a disturbance in cell growth; however, the results were not consistent in all trials. Indeed, RT-PCR results confirmed that splicing was not being blocked. Preventing the expression of core proteins would be a great alternative method to investigate if splicing is crucial in *C. merolae*.

References cited

Black, C. S., Garside, E. L., MacMillan, A. M., & Rader, S. D. (2016). Conserved structure of Snu13 from the highly reduced spliceosome of Cyanidioschyzon merolae. *Protein Science*, *25*(4), 911-916.

Boël, G., Letso, R., Neely, H., Price, W. N., Wong, K. H., Su, M., Luff J. D., Valecha, M., Everett, Acton, T. B. & Xiao, R. (2016). Codon influence on protein expression in *E. coli* correlates with mRNA levels. *Nature*, *529*(7586), 358.

Boon, K. L., Grainger, R. J., Ehsani, P., Barrass, J. D., Auchynnikava, T., Inglehearn, C. F., & Beggs, J. D. (2007). prp8 mutations that cause human retinitis pigmentosa lead to a U5 snRNP maturation defect in yeast. *Nature structural & molecular biology*, *14*(11), 1077-1083.

Branlant, C., Krol, A., Ebel, J. P., Lazar, E., Haendler, B., & Jacob, M. (1982). U2 RNA shares a structural domain with U1, U4, and U5 RNAs. *The EMBO journal*, *1*(10), 1259-1265.

Brenner, T. J., & Guthrie, C. (2006). Assembly of Snu114 into U5 snRNP requires Prp8 and a functional GTPase domain. *RNA*, *12*(5), 862–871.

Buenrostro, J. D., Araya, C. L., Chircus, L. M., Layton, C. J., Chang, H. Y., Snyder, M. P., & Greenleaf, W. J. (2014). Quantitative analysis of RNA-protein interactions on a massively parallel array reveals biophysical and evolutionary landscapes. *Nature biotechnology*, *32*(6), 562-568.

Burgess, R. R., & Deutscher, M. P. (2009). *Guide to protein purification*. Academic Press, 463.

Callaway, E. (2015). The revolution will not be crystallized. *Nature*, *525*(7568), 172.

Chan, S. P., Kao, D. I., Tsai, W. Y., & Cheng, S. C. (2003). The Prp19p-associated complex in spliceosome activation. *Science*, *302*(5643), 279-282.

Dix, I., Russell, C. S., O'keefe, R. T., Newman, A. J., & Beggs, J. D. (1998). Protein-RNA interactions in the U5 snRNP of *Saccharomyces cerevisiae*. *Rna*, *4*(12), 1674-1686.

Draper, B. W., Morcos, P. A., & Kimmel, C. B. (2001). Inhibition of zebrafish fgf8 pre-mRNA splicing with morpholino oligos: A quantifiable method for gene knockdown. *genesis*, *30*(3), 154-156.

Dunn, E. A. (2010). U6 snRNA secondary structure in free U6 snRNPs. doi:https://doi.org/10.24124/2010/bpgub646

Dunn, E. A., & Rader, S. D. (2014). Pre-mRNA splicing and the spliceosome: Assembly, catalysis, and fidelity. In *Fungal RNA Biology* (pp. 27-57). Springer, Cham.

Dyson, M. R. & Durocher, Y. (2007). *Expression Systems*. Bloxham, Banbury: Scion Publishing Ltd, 2(14).

Fargašová, A. (1996). Inhibitive effect of organotin compounds on the chlorophyll content of the green freshwater alga Scenedesmus quadricauda. *Bulletin of environmental contamination and toxicology*, *57*(1), 99-106.

Fischer, U., Liu, Q., & Dreyfuss, G. (1997). The SMN–SIP1 complex has an essential role in spliceosomal snRNP biogenesis. *Cell*, *90*(6), 1023-1029.

Frank, D. N., Roiha, H., & Guthrie, CH. (1994). Architecture of the U5 small nuclear RNA. *Molecular and Cellular Biology*, *14*(3), 2180-2190.

Galej, W. P., Oubridge, C., Newman, A. J., & Nagai, K. (2013). Crystal structure of Prp8 reveals active site cavity of the spliceosome. *Nature, 493*(7434), 638-643. D oi:10.1038/nature11843

Graveley, B. R. (2001). Alternative splicing: increasing diversity in the proteomic world. *TRENDS in Genetics*, *17*(2), 100-107.

Hahn, D., Kudla, G., Tollervey, D., & Beggs, J. D. (2012). Brr2p-mediated conformational rearrangements in the spliceosome during activation and substrate repositioning.*Genes & Development, 26*(21), 2408.

Hang, J., Wan, R., Yan, C., & Shi, Y. (2015). Structural basis of pre-mRNA splicing. *Science, 349*(6253), 1191-1198. doi:10.1126/science.aac8159

Hashimoto, M., Nara, T., Mita, T., & Mikoshiba, K. (2016). Morpholino antisense oligo inhibits trans-splicing of pre-inositol 1, 4, 5-trisphosphate receptor mRNA of Trypanosoma cruzi and suppresses parasite growth and infectivity. *Parasitology international*, *65*(3), 175-179.

Hermann, H., Fabrizio, P., Raker, V. A., Foulaki, K., Hornig, H., Brahms, H., & Lührmann, R. (1995). snRNP Sm proteins share two evolutionarily conserved sequence motifs which are involved in Sm protein-protein interactions. *The EMBO journal*, *14*(9), 2076.

Hersh, C. M., & Crumpton, W. G. (1987). Determination of growth rate depression of some green algae by atrazine. *Bulletin of environmental contamination and toxicology*, *39*(6), 1041-1048.

Jarmolowski, A., & Mattaj, I. W. (1993). The determinants for Sm protein binding to Xenopus U1 and U5 snRNAs are complex and non-identical. *The EMBO journal*, *12*(1), 223-232.

Jékely, G. (Ed.). (2008). *Eukaryotic Membranes and Cytoskeleton: Origins and Evolution* (Vol. 607). Springer Science & Business Media.

Jones, M. H., & Guthrie, C. (1990). Unexpected flexibility in an evolutionarily conserved protein-

RNA interaction: genetic analysis of the Sm binding site. *The EMBO journal*, *9*(8), 2555-2561.

Kambach, C., Walke, S., Young, R., Avis, J. M., de la Fortelle, E., Raker, V. A., Lührmann, R., Li, J. & Nagai, K. (1999). Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell*, *96*(3), 375-387.

Kelemen, O., Convertini, P., Zhang, Z., Wen, Y., Shen, M., Falaleeva, M., & Stamm, S. (2013). Function of alternative splicing. *Gene*, *514*(1), 1-30.

Kobayashi Y., Ohnuma M., Kuroiwa T., Tanaka K., & Hanaoka M. (2010) The basics of cultivation and molecular genetic analysis of the unicellular red alga *Cyanidioschyzon merolae*. *Endocytobiosis Cell Res*, 20, 53-61.

Kuroiwa, T., Kawazu, T., Takahashi, H., Suzuki, K., Ohta, N., & Kuroiwa, H. (1994). Comparison of ultrastructures between the ultra-small eukaryote Cyanidioschyzon merolae and Cyanidium caldarium. *Cytologia*, *59*(2), 149-158.

Kuriyan, J., Konforti, B., & Wemmer, D. (2012). *The molecules of life: Physical and chemical principles*. Garland Science.

Lefebvre, S., Bürglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., Benichou, B., Cruaud, C., Millasseau, P., Zeviani, M. & Le Paslier, D. (1995). Identification and characterization of a spinal muscular atrophy-determining gene. *Cell*, *80*(1), 155-165.

Li, P., Anumanthan, A., Gao, X. G., Ilangovan, K., Suzara, V. V., Düzgüneş, N., & Renugopalakrishnan, V. (2007). Expression of recombinant proteins in Pichia pastoris. *Applied biochemistry and biotechnology*, *142*(2), 105-124.

Lin-Cereghino, J., Wong, W. W., Giang, W., Luong, L. T., Vu, J., Johnson, S. D., & Lin-Cereghino, G. P. (2005). Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast Pichia pastoris. *Biotechniques*, *38*(1), 44.

Liu, H. L., & Cheng, S. C. (2012). The interaction of Prp2 with a defined region of the intron is required for the first splicing reaction. *Molecular and cellular biology*, *32*(24), 5056-5066.

Liu, Q., Fischer, U., Wang, F., & Dreyfuss, G. (1997). The spinal muscular atrophy disease gene product, SMN, and its associated protein SIP1 are in a complex with spliceosomal snRNP proteins. *Cell*, *90*(6), 1013-1021.

López-Bigas, N., Audit, B., Ouzounis, C., Parra, G., & Guigó, R. (2005). Are splicing mutations the most frequent cause of hereditary disease? *FEBS Letters, 579*(9), 1900-1903. doi:10.1016/j.febslet.2005.02.047

Meister, G., Bühler, D., Pillai, R., Lottspeich, F., & Fischer, U. (2001). A multiprotein complex

mediates the ATP-dependent assembly of spliceosomal U snRNPs. *Nature cell biology*, *3*(11), 945-949.

Misumi, O., Matsuzaki, M., Nozaki, H., Miyagishima, S. Y., Mori, T., Nishida, K., ... & Kuroiwa, T. (2005). Cyanidioschyzon merolae genome. A tool for facilitating comparable studies on organelle biogenesis in photosynthetic eukaryotes. *Plant physiology*, *137*(2), 567-585.

Morcos, P. A., Li, Y., & Jiang, S. (2008). Vivo-Morpholinos: a non-peptide transporter delivers Morpholinos into a wide array of mouse tissues. *Biotechniques*, *45*(6), 613-614.

Nguyen, T. H. D., Galej, W. P., Bai, X. C., Savva, C. G., Newman, A. J., Scheres, S. H., & Nagai, K. (2015). The architecture of the spliceosomal U4/U6. U5 tri-snRNP. *Nature*, *523*(7558), 47-52.

Nguyen, T. H. D., Li, J., Galej, W. P., Oshikane, H., Newman, A. J., & Nagai, K. (2013). Structural basis of Brr2-Prp8 interactions and implications for U5 snRNP biogenesis and the spliceosome active site. *Structure (London, England : 1993), 21*(6), 910-919. doi:10.1016/j.str.2013.04.017

Pellizzoni, L., Yong, J., & Dreyfuss, G. (2002). Essential role for the SMN complex in the specificity of snRNP assembly. *Science*, *298*(5599), 1775-1779.

Potter, H., & Heller, R. (2011). Transfection by electroporation. *Current Protocols in Neuroscience*, A-1E.

Raker, V. A., Hartmuth, K., Kastner, B., & Lührmann, R. (1999). Spliceosomal U snRNP core assembly: Sm proteins assemble onto an Sm site RNA nonanucleotide in a specific and thermodynamically stable manner. *Molecular and Cellular Biology*, *19*(10), 6554-6565.

Rappoport, J. Z., & Simon, S. M. (2003). Real-time analysis of clathrin-mediated endocytosis during cell migration. *Journal of cell science*, *116*(5), 847-855.

Raker, V. A., Plessel, G., & Lührmann, R. (1996). The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle in vitro. *The EMBO journal*, *15*(9), 2256.

Reimer, K. A., Stark, M. R., Aguilar, L. C., Stark, S. R., Burke, R. D., Moore, J., Fahlman, R. P., Yip, C. K., Kuroiwa, H., Oeffinger, M. & Rader, S. D. (2017). The sole LSm complex in Cyanidioschyzon merolae associates with pre-mRNA splicing and mRNA degradation factors. *RNA*, *23*(6), 952-967.

Reuter, K., Nottrott, S., Fabrizio, P., Lührmann, R., & Ficner, R. (1999). Identification, characterization and crystal structure analysis of the human spliceosomal U5 snRNP-specific 15 kD protein. *Journal of molecular biology*,*294*(2), 515-525.

Rodrigues, L. H. R., Raya-Rodriguez, M. T., & Fontoura, N. F. (2011). Algal density assessed by

spectrophotometry: A calibration curve for the unicellular algae Pseudokirchneriella subcapitata. *Journal of Environmental Chemistry and Ecotoxicology*, *3*(8), 225-228.

Rojíčková-Padrtová, R., Maršálek, B., & Holoubek, I. (1998). Evaluation of alternative and standard toxicity assays for screening of environmental samples: selection of an optimal test battery. *Chemosphere*, *37*(3), 495-507.

Sakharkar, M. K., Chow, V. T. K., & Kangueane, P. (2004). Distributions of exons and introns in the human genome. *In Silico Biology, 4*(4), 387.

Scheich, C., Kümmel, D., Soumailakakis, D., Heinemann, U., & Büssow, K. (2007). Vectors for co-expression of an unrestricted number of proteins. *Nucleic acids research*, *35*(6), e43.

Schmid-Burgk, J. L., Xie, Z., & Benenson, Y. (2014). Hierarchical ligation-independent assembly of PCR fragments. *DNA Cloning and Assembly Methods*, 49-58.

Schwer, B., & Guthrie, C. (1992). A conformational rearrangement in the spliceosome is dependent on PRP16 and ATP hydrolysis. *The EMBO Journal*, *11*(13), 5033-5039.

Séraphin B. (1995). Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs. EMBO J., 14, 2089–2098.

Small, E. C., Leggett, S. R., Winans, A. A., & Staley, J. P. (2006). The EF-G-like GTPase Snu114p regulates spliceosome dynamics mediated by Brr2p, a DExD/H box ATPase. *Molecular Cell, 23*(3), 389-399. doi:10.1016/j.molcel.2006.05.043

Staley, J. P., & Guthrie, C. (1999). An RNA switch at the 5′ splice site requires ATP and the DEAD box protein Prp28p. *Molecular cell*, *3*(1), 55-64.

Stark, M. R., Dunn, E. A., William S. C. Dunn, Grisdale, C. J., Daniele, A. R., Matthew R. G. Halstead. Rader, S. D. (2015). Dramatically reduced spliceosome in cyanidioschyzon merolae. *Proceedings of the National Academy of Sciences, 112*(11), E1191-E1200. doi:10.1073/pnas.1416879112

Stevens, S. W., Barta, I., Ge, H. Y., Moore, R. E., Young, M. K., Lee, T. D., & Abelson, J. (2001). Biochemical and genetic analyses of the U5, U6, and U4/U6 x U5 small nuclear ribonucleoproteins from Saccharomyces cerevisiae. *Rna*, *7*(11), 1543-1553.

Studier, F. W. (2005). Protein production by auto-induction in high-density shaking cultures. *Protein expression and purification*, *41*(1), 207-234.

Seckbach, J. (Ed.). (2012). *Evolutionary pathways and enigmatic algae: Cyanidium caldarium (Rhodophyta) and related cells* (Vol. 91). Springer Science & Business Media.

Summerton, J. E. (2007). Morpholino, siRNA, and S-DNA compared: impact of structure and

mechanism of action on off-target effects and sequence specificity. *Current topics in medicinal chemistry*, *7*(7), 651-660.

Summerton, J. E. (2005). Endo-Porter: A Novel Reagent for Safe, Effective Delivery of Substances into Cells. *Annals of the New York Academy of Sciences*, *1058*(1), 62-75.

Sumiya, N., Fujiwara, T., Kobayashi, Y., Misumi, O., & Miyagishima, S. Y. (2014). Development of a heat-shock inducible gene expression system in the red alga Cyanidioschyzon merolae. *PloS one*, *9*(10), e111261.

Tsai, R. T., Fu, R. H., Yeh, F. L., Tseng, C. K., Lin, Y. C., Huang, Y. H., & Cheng, S. C. (2005). Spliceosome disassembly catalyzed by Prp43 and its associated components Ntr1 and Ntr2. *Genes & development*, *19*(24), 2991-3003.

Urlaub, H., Raker, V. A., Kostka, S., & Lührmann, R. (2001). Sm protein–Sm site RNA interactions within the inner ring of the spliceosomal snRNP core structure. *The EMBO Journal*, *20*(1-2), 187–196. http://doi.org/10.1093/emboj/20.1.187

Wan, R., Yan, C., Bai, R., Wang, L., Huang, M., Wong, C. C., & Shi, Y. (2016). The 3.8 Å structure of the U4/U6. U5 tri-snRNP: Insights into spliceosome assembly and catalysis. *Science*, *351*(6272), 466-475.

Wahl, M. C., Will, C. L., & Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell*, *136*(4), 701-718.

Warkocki, Z., Schneider, C., Mozaffari-Jovin, S., Schmitzová, J., Höbartner, C., Fabrizio, P., & Lührmann, R. (2015). The G-patch protein Spp2 couples the spliceosome-stimulated ATPase activity of the DEAH-box protein Prp2 to catalytic activation of the spliceosome. *Genes & Development, 29*(1), 94-107.

Weidner, M., Taupp, M., & Hallam, S. J. (2010). Expression of recombinant proteins in the methylotrophic yeast Pichia pastoris. *Journal of visualized experiments: JoVE*, (36).

Yan, C., Hang, J., Wan, R., Huang, M., Catherine C L Wong, & Shi, Y. (2015). Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science, 349*(6253), 1182-1191. doi:10.1126/science.aac7629

Zaric, B., Chami, M., Rémigy, H., Engel, A., Ballmer-Hofer, K., Winkler, F. K., & Kambach, C. (2005). Reconstitution of two recombinant LSm protein complexes reveals aspects of their architecture, assembly, and function. *Journal of biological chemistry*, *280*(16), 16066-16075.

Zhang, L., Li, X., Hill, R. C., Qiu, Y., Zhang, W., Hansen, K. C., & Zhao, R. (2015). Brr2 plays a role in spliceosomal activation in addition to U4/U6 unwinding. *Nucleic Acids Research, 43*(6), 3286.

Zhang, R., So, B. R., Li, P., Yong, J., Glisovic, T., Wan, L., & Dreyfuss, G. (2011). Structure of a key intermediate of the SMN complex reveals Gemin2's crucial function in snRNP assembly. *Cell*, *146*(3), 384-395.

Zhou, L., Hang, J., Zhou, Y., Wan, R., Lu, G., Yin, P., Yan, C & Shi, Y. (2014). Crystal structures of the Lsm complex bound to the 3′ end sequence of U6 small nuclear RNA. *Nature*, *506*(7486), 116.

Appendix 1

Sequencing results

Alignment of the sequenced protein genes (inserted into each vector) with the protein gene sequences using the pairwise BLAST alignment.

Table 23 Oligonucleotide sequences of the primers used for sequencing of Prp8, Brr2 and Snu114 genes. The other genes were sequenced using the same primers presented in table 2. (*) represents the gene inserted into pMCSG2. (**) represents the gene inserted into pPICZA. (***) represents the gene inserted into pQLink.

Gene	oSDR #	Primers	Coverage (bp)
Brr2* **	1084	CCATTTGTCGAGAAATCATAAAAAATTTATTTGCTTTGTG	1-720
	1145	TGAGTCATCGCGACTT	712-1444
	1146	ATTGGAGTCCATCACG	1312-2050
	1147	ACGCTGCGGACAATAGA	1974-2753
	1148	ACACGTTGCGTTTCGA	2664-3436
	1149	TGACCCCAAACACTTTG	3219-4008
	1150	TGGGCAACGGCTTTGC	3905-4634
	1151	AAGCGCGTCCTAAGAA	4564-5256
	1085	CAACCGAGCGTTCTGAACAAATCCAG	5257-5469
Brr2 **	1382	ATTGACAAGCTTTTGA	2-721
	1145	TGAGTCATCGCGACTT	709-1384
	1146	ATTGGAGTCCATCACG	1382-2094
	1147	ACGCTGCGGACAATAGA	1974-2748
	1148	ACACGTTGCGTTTCGA	2665-3371
	1149	TGACCCCAAACACTTTG	3292-3905
	1150	TGGGCAACGGCTTTGC	3905-4619
	1151	AAGCGCGTCCTAAGAA	4577-5342
	1383	TCAATGATGATGATGA	5133-5366
Snu114* **	1084	CCATTTGTCGAGAAATCATAAAAAATTTATTTGCTTTGTG	1-713
	1152	GTGACGCTGATGGATG	676-1519
	1153	TGCTGGTCAATGTGCA	1339-2173
	1164	GCTTGCACCGTGTTT	1965-2775
	1085	CAACCGAGCGTTCTGAACAAATCCAG	2605-3327
Prp8* **	1084	CCATTTGTCGAGAAATCATAAAAAATTTATTTGCTTTGTG	1-696
	1207	ACGAGCGAATCCAACAG	501-1325
	1155	ATTTGCTCTATGCACC	1310-2137

	1156	ATTTGCTCTATGCACC	1966-2743
	1157	AGGTCTCTATCGTTAC	2629-3408
	1158	CGATATTGTACAGATTCGC	3262-4054
	1159	CAAAGACATGCGTTATACG	3916-4673
	1160	CAGCAACTGCAGGGAT	4552-5346
	1161	ATTTCATGCACGACGG	5214-5997
	1162	GCAGCCAAAATCTATTGGA	5859-6646
	1085	TGGACCCGTTGAAGAC	6566-7188
Prp8 *	1283	TGATCAACGCCGCCAGC	4-662
	1154	TAACGATGCAGACGCG	661-1438
	1155	ATTTGCTCTATGCACC	1302-2021
	1156	AGGTCTCTATCGTTAC	1963-2765
	1157	CGATATTGTACAGATTCGC	2605-3264
	1158	CAAAGACATGCGTTATACG	3253-4062
	1159	CAGCAACTGCAGGGAT	3910-4589
	1160	ATTTCATGCACGACGG	4552-5312
	1161	GCAGCCAAAATCTATTGGA	5212-5886
	1342	AATCGCTTGGCGAGAT	5759-6502
	1284	GCAGCGGTTTCTTTACC	6433-7188

Prp8 gene inserted into pQLinkN (The RBS sequence is in blue and the LIC sequence is in bold; pSR655)

```
Query  38   TACTTCCAATCCCACGAGGAGAAATTAACT 88
            ||||||||||||||||||||||||||||||
Sbjct        ------------------------------


Query  89   ATGCCCAAACGTGCGTTTTTCGGACGCGACGAAGAGGCAGATCAGACGCTGGACCTGAAG  148
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGCCCAAACGTGCGTTTTTCGGACGCGACGAAGAGGCAGATCAGACGCTGGACCTGAAG  60

Query  149  CGAAAGCGGCCACGCCGTGCTTTCGGAGCGGATCAGCCATTTTTCAAACCGTATACCTCC  208
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   CGAAAGCGGCCACGCCGTGCTTTCGGAGCGGATCAGCCATTTTTCAAACCGTATACCTCC  120

Query  209  CCCGATATTTTCCGCAAACTAGCGAGTCTTGCACGGATCGAGTTGGAGCGAAGCGAAAAT  268
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  CCCGATATTTTCCGCAAACTAGCGAGTCTTGCACGGATCGAGTTGGAGCGAAGCGAAAAT  180

Query  269  GACCTGGAAAACAGTGCCCACCGTGGTAACCGAGACCGACATATCTCCCGGAACACAGAT  328
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GACCTGGAAAACAGTGCCCACCGTGGTAACCGAGACCGACATATCTCCCGGAACACAGAT  240

Query  329  GCACTGACACTGAACGCGCTCCGCTACCTTCCGCACGCGGTGTACAAATGGCTGGAACAC  388
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCACTGACACTGAACGCGCTCCGCTACCTTCCGCACGCGGTGTACAAATGGCTGGAACAC  300

Query  389  ATGCCGGCCCCTTGGGAGCCAACGCGGTTTGTACCCGTGATTTTCCATCACACTGGTGCG  448
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  ATGCCGGCCCCTTGGGAGCCAACGCGGTTTGTACCCGTGATTTTCCATCACACTGGTGCG  360

Query  449  TTGGCTTTTATCGAGGGTTCACGTCGAGTCCCCGAGGTCGTTCATCGGGCCCAATGGGCG  508
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  TTGGCTTTTATCGAGGGTTCACGTCGAGTCCCCGAGGTCGTTCATCGGGCCCAATGGGCG  420

Query  509  CGTTGGTGCGCCTACCTCGACCGTCGTCGTCATGAGGCACACGTCGCCTCGACAAGCAGC  568
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  CGTTGGTGCGCCTACCTCGACCGTCGTCGTCATGAGGCACACGTCGCCTCGACAAGCAGC  480

Query  569  GGAAAGCGGCAGCGAGATCTGACGCGTTCCTTCGTACGTTTGCAAGTGCCGGCGTTCGAC  628
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  481  GGAAAGCGGCAGCGAGATCTGACGCGTTCCTTCGTACGTTTGCAAGTGCCGGCGTTCGAC  540
```

```
Query  629  GACGATGAGCCGTCACCGAGCTTCCGCGAGTTTGTCTACCATCGAACGCTGCCGACACCA  688
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  541  GACGATGAGCCGTCACCGAGCTTCCGCGAGTTTGTCTACCATCGAACGCTGCCGACACCA  600

Query  689  TTTGCTAACGATGCAGACGCGGCAGCTGAATACGAGCGAATCCAACAGCTCGCCGCGAAT  748
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  601  TTTGCTAACGATGCAGACGCGGCAGCTGAATACGAGCGAATCCAACAGCTCGCCGCGAAT  660

Query  749  TTGCACACACCGGAGGCGCGTGCTGCGCTCTACCGG  784
            |||||||||||||||||||||||||||||||||||
Sbjct  661  TTGCACACACCGGAGGCGCGTGCTGCGCTCTACCGG  696
```

```
Query  1     GACGCGTTCCTTCGTACGTTTGCAAGTGCCGGCGTTCGACGACGATGAGCCGTCACCGAG  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  501   GACGCGTTCCTTCGTACGTTTGCAAGTGCCGGCGTTCGACGACGATGAGCCGTCACCGAG  560

Query  61    CTTCCGCGAGTTTGTCTACCATCGAACGCTGCCGACACCATTTGCTAACGATGCAGACGC  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  561   CTTCCGCGAGTTTGTCTACCATCGAACGCTGCCGACACCATTTGCTAACGATGCAGACGC  620

Query  121   GGCAGCTGAATACGAGCGAATCCAACAGCTCGCCGCGAATTTGCACACACCGGAGGCGCG  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  621   GGCAGCTGAATACGAGCGAATCCAACAGCTCGCCGCGAATTTGCACACACCGGAGGCGCG  680

Query  181   TGCTGCGCTCTACCGGATGAGTTTGCCGTGGGCGCCCACACCCGAACGATCTGGATCCAAG  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  681   TGCTGCGCTCTACCGGATGAGTTTGCCGTGGGCGCCCACACCCGAACGATCTGGATCCAAG  740

Query  241   TCAGTTCTATCTGCGAAATCGCTTGACATTGCGACGAGTGCATCGACTCGAGCATGAGCG  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  741   TCAGTTCTATCTGCGAAATCGCTTGACATTGCGACGAGTGCATCGACTCGAGCATGAGCG  800

Query  301   TGCAAACCTTCGACGTAAGAGCCTACACGCCGGCTCTTTGCACAGCGAGGGCGCGCATCT  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  801   TGCAAACCTTCGACGTAAGAGCCTACACGCCGGCTCTTTGCACAGCGAGGGCGCGCATCT  860

Query  361   GCTCCAGGAGGACTGGGACGATTTCACGGAGCTCGGTATGCATTCCATCCTGCCTGGTCA  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  861   GCTCCAGGAGGACTGGGACGATTTCACGGAGCTCGGTATGCATTCCATCCTGCCTGGTCA  920

Query  421   ACGTCGCGCTCCAGAACTGCACCAGGCTTATCCGTACCTGTACGACGCATTCGATGAACG  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  921   ACGTCGCGCTCCAGAACTGCACCAGGCTTATCCGTACCTGTACGACGCATTCGATGAACG  980

Query  481   CACCCCGAACCCAGAGCAACCGCTGTGGCACCACCATGCGCCGCGGCCGCTCATGTGGCC  540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  981   CACCCCGAACCCAGAGCAACCGCTGTGGCACCACCATGCGCCGCGGCCGCTCATGTGGCC  1040

Query  541   GCCAACACCACCACCGCACCCGCTCTGGCCGAGTGCGTGGAGCTGGCAGCCTACGCTAGC  600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1041  GCCAACACCACCACCGCACCCGCTCTGGCCGAGTGCGTGGAGCTGGCAGCCTACGCTAGC  1100

Query  601   ACCGCTCCTGCACAGCAGTGCACAAGGCGCTGGCGAGCGCATCATCCCAGACACCAGCGA  660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1101  ACCGCTCCTGCACAGCAGTGCACAAGGCGCTGGCGAGCGCATCATCCCAGACACCAGCGA  1160

Query  661   AGGCGCCAGCGCTCAGCGTACGCGCTACCGCGCTCATCCGGCGCTGGGCGATTTCGATCT  720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1161  AGGCGCCAGCGCTCAGCGTACGCGCTACCGCGCTCATCCGGCGCTGGGCGATTTCGATCT  1220

Query  721   GGACGCGTGCGGCGAACGCGTCGGAGCCTGGCTCGATTTGCTCTATGCACCGCGTCCATA  780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1221  GGACGCGTGCGGCGAACGCGTCGGAGCCTGGCTCGATTTGCTCTATGCACCGCGTCCATA  1280

Query  781   TAAGGGCAGATGTGTTCGAAAACGGAAGCGGATGCAGGATATCGA  825
             |||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1281  TAAGGGCAGATGTGTTCGAAAACGGAAGCGGATGCAGGATATCGA  1325
```

```
Query  36    GGATGCAGGATATCGACATGGAACGCGAGTTGTACTGGCAACTGGAGCGCGTCCATTACA  95
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1310  GGATGCAGGATATCGACATGGAACGCGAGTTGTACTGGCAACTGGAGCGCGTCCATTACA  1369

Query  96    TGCGCCCGAAAAGTACTGCCGCAGTCCACAGCTTGTTGAAACGACGCTGTCGACAAGCAG  155
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1370  TGCGCCCGAAAAGTACTGCCGCAGTCCACAGCTTGTTGAAACGACGCTGTCGACAAGCAG  1429

Query  156   AAGCACAGCGTCTTCGCTACCTGCGAAAAGAACGTGGTCGCGGGCGCAAGCAATCGACTT  215
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1430  AAGCACAGCGTCTTCGCTACCTGCGAAAAGAACGTGGTCGCGGGCGCAAGCAATCGACTT  1489

Query  216   TGGCGATCCAGTCGGCCTCTGGAGACGCTGGCGATGAAGCTGAGAAGCAGCACCGACGGC  275
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1490  TGGCGATCCAGTCGGCCTCTGGAGACGCTGGCGATGAAGCTGAGAAGCAGCACCGACGGC  1549

Query  276   CGGATCTGCTGAGGACGTTGCGGCAATCGGGAttttttATCGCACCAGTATGGATTGGT  335
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1550  CGGATCTGCTGAGGACGTTGCGGCAATCGGGATTTTTTTATCGCACCAGTATGGATTGGT  1609
```

```
Query  336   TGGAGGTCGGCATCTGGCTGTGCGACGCTGCACGATCGATGTTTTTGCTACTTTTGAGAC  395
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1610  TGGAGGTCGGCATCTGGCTGTGCGACGCTGCACGATCGATGTTTTTGCTACTTTTGAGAC  1669

Query  396   GAAAACGATTCTTCTTTCTGAGCATGGACTACAATTTTCAGATCGTGCCGTTGCGTACGC  455
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1670  GAAAACGATTCTTCTTTCTGAGCATGGACTACAATTTTCAGATCGTGCCGTTGCGTACGC  1729

Query  456   TGACGACCAAAGAACGAAAGCAGTCTCGCTTCGGGAATGCGTTTCATTTGATGCGCGAAT  515
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1730  TGACGACCAAAGAACGAAAGCAGTCTCGCTTCGGGAATGCGTTTCATTTGATGCGCGAAT  1789

Query  516   GGATGCGTATGGTGAAGCTGATTGTGGATGTGCATCTGCGGTATCGGGCGGGTCTCGGCG  575
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1790  GGATGCGTATGGTGAAGCTGATTGTGGATGTGCATCTGCGGTATCGGGCGGGTCTCGGCG  1849

Query  576   CGGATGCCATTCAACTCGCCGACAGCATCACGTACATCGAATCACATATCGGTGAACTGA  635
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1850  CGGATGCCATTCAACTCGCCGACAGCATCACGTACATCGAATCACATATCGGTGAACTGA  1909

Query  636   CAGGTCTCTATCGTTACAAGTACCGGGTTATGCGCCAAATCCATGCGACCAAGGACCTCA  695
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1910  CAGGTCTCTATCGTTACAAGTACCGGGTTATGCGCCAAATCCATGCGACCAAGGACCTCA  1969

Query  696   AGCATCTTATGTATCAGCGTTTCGGATGGTCGGGTCCCGGCAAGGCCTGCTGGCAACCCC  755
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1970  AGCATCTTATGTATCAGCGTTTCGGATGGTCGGGTCCCGGCAAGGCCTGCTGGCAACCCC  2029

Query  756   TCTGGCGGCAATGGGTGCACCTACTTCGAGGCCTGATGCCCTTGCTCGAACAGTGGTTGG  815
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2030  TCTGGCGGCAATGGGTGCACCTACTTCGAGGCCTGATGCCCTTGCTCGAACAGTGGTTGG  2089

Query  816   GTAACCTGCTGAATCGCCATTTTGAAGGACGCGAGCCGCTGCGCATGG  863
             ||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2090  GTAACCTGCTGAATCGCCATTTTGAAGGACGCGAGCCGCTGCGCATGG  2137


Query  27    CTCAAGCATCTTATGTATCAGCGTTTCGGATGGTCGGGTCCCGGCAAGGCCTGCTGGCAA  86
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1966  CTCAAGCATCTTATGTATCAGCGTTTCGGATGGTCGGGTCCCGGCAAGGCCTGCTGGCAA  2025

Query  87    CCCCTCTGGCGGCAATGGGTGCACCTACTTCGAGGCCTGATGCCCTTGCTCGAACAGTGG  146
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2026  CCCCTCTGGCGGCAATGGGTGCACCTACTTCGAGGCCTGATGCCCTTGCTCGAACAGTGG  2085

Query  147   TTGGGTAACCTGCTGAATCGCCATTTTGAAGGACGCGAGCCGCTGCGCATGGCGCAACGC  206
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2086  TTGGGTAACCTGCTGAATCGCCATTTTGAAGGACGCGAGCCGCTGCGCATGGCGCAACGC  2145

Query  207   ACCGTCACAAAGCAACGACTCGAGTCGCAGTTTGATGTGGCGCTACGTCAGGAGACTATA  266
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2146  ACCGTCACAAAGCAACGACTCGAGTCGCAGTTTGATGTGGCGCTACGTCAGGAGACTATA  2205

Query  267   GCGCGTCTTCGGCAACTGTTGCCAGCAGCGCGTCAACCACTATATACGAGGCGTGTTCTC  326
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2206  GCGCGTCTTCGGCAACTGTTGCCAGCAGCGCGTCAACCACTATATACGAGGCGTGTTCTC  2265

Query  327   CAGCATATGCACCAGGCGTGGCGCTGCTGGAAGGCCAATATCCCGTGGCACGTACGAGAC  386
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2266  CAGCATATGCACCAGGCGTGGCGCTGCTGGAAGGCCAATATCCCGTGGCACGTACGAGAC  2325

Query  387   ATGCCCCCTGAAATCGACAGACTCGTACGGGATTACGTGCAGCAGCGGGCGCAATGGTGG  446
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2326  ATGCCCCCTGAAATCGACAGACTCGTACGGGATTACGTGCAGCAGCGGGCGCAATGGTGG  2385

Query  447   ATAGAGGGTGCTCGACGCTTCGCGATTGCGTTTCGAAGCGGTCGTTCCATGGATAAGGCA  506
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2386  ATAGAGGGTGCTCGACGCTTCGCGATTGCGTTTCGAAGCGGTCGTTCCATGGATAAGGCA  2445

Query  507   CTTGTTCGCCAGATGTACGGGCGACTTGCTCGTCTTGCCGTTCACCATGAGCAGGCGCAT  566
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2446  CTTGTTCGCCAGATGTACGGGCGACTTGCTCGTCTTGCCGTTCACCATGAGCAGGCGCAT  2505

Query  567   CAGCGGCAGTATTTGGAGCACGGTCCCTTTCTGTTGCCGACCGAAGCAGCATCGATATTG  626
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2506  CAGCGGCAGTATTTGGAGCACGGTCCCTTTCTGTTGCCGACCGAAGCAGCATCGATATTG  2565

Query  627   TACAGATTCGCAACGTACCTGGAGAAAGCAGGTGTTGCCGACTCATGCCCTTGGCAATTG  686
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2566  TACAGATTCGCAACGTACCTGGAGAAAGCAGGTGTTGCCGACTCATGCCCTTGGCAATTG  2625

Query  687   CCTTGGTTCGCCGCAGATAACGCGGATTCCGGGCTAGCTCCCGCATTACTGCAACTCGCG  746
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2626  CCTTGGTTCGCCGCAGATAACGCGGATTCCGGGCTAGCTCCCGCATTACTGCAACTCGCG  2685

Query  747   ATCGAGCGTCTCCGCCCGGAAACGGAGGGCGAGCACTGTCCCCCGGGGCTCGGCGAAC  804
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2686  ATCGAGCGTCTCCGCCCGGAAACGGAGGGCGAGCACTGTCCCCCGGGGCTCGGCGAAC  2743


Query  1     TGGTTCGCCGCAGATAACGCGGATTCCGGGCTAGCTCCCGCATTACTGCAACTCGCGATC  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2629  TGGTTCGCCGCAGATAACGCGGATTCCGGGCTAGCTCCCGCATTACTGCAACTCGCGATC  2688
```

```
Query  61    GAGCGTCTCCGCCCGGAAACGGAGGGCGAGCACTGTCCCCCGGGGCTCGGCGAACTTGTC  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2689  GAGCGTCTCCGCCCGGAAACGGAGGGCGAGCACTGTCCCCCGGGGCTCGGCGAACTTGTC  2748

Query  121   ACAGAGGCAGAGGCAGCGCCTGCAGCGATGCTGTATAGAATCCGGCAGCGACTGCAGGCG  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2749  ACAGAGGCAGAGGCAGCGCCTGCAGCGATGCTGTATAGAATCCGGCAGCGACTGCAGGCG  2808

Query  181   GTGCAAATGCGTCACCAGGTGGAACTCCGCTTCTACGATGACGCGGACCTGACGCCTGTG  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2809  GTGCAAATGCGTCACCAGGTGGAACTCCGCTTCTACGATGACGCGGACCTGACGCCTGTG  2868

Query  241   TATCGGATCGATTCCTTTGAGCGACTGGTGGACGCTTTCCTCGATCAATGGCTCTGGTAC  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2869  TATCGGATCGATTCCTTTGAGCGACTGGTGGACGCTTTCCTCGATCAATGGCTCTGGTAC  2928

Query  301   AAGGCCACGGAGACGCGCCTTTTTCCGGCCCATGTCCAGCCATGCGACGATGAGTTGCCG  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2929  AAGGCCACGGAGACGCGCCTTTTTCCGGCCCATGTCCAGCCATGCGACGATGAGTTGCCG  2988

Query  361   CCCGTTCACGTCCTCCGTTTCGTCGAACGCTTGGATGCTATACCTCATCTGTGGACACTC  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2989  CCCGTTCACGTCCTCCGTTTCGTCGAACGCTTGGATGCTATACCTCATCTGTGGACACTC  3048

Query  421   GGAACAGCACCAAACAAGAGCTTCCTGGTCCTCATACAAACGCCATTGCCAGGGCTATTT  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3049  GGAACAGCACCAAACAAGAGCTTCCTGGTCCTCATACAAACGCCATTGCCAGGGCTATTT  3108

Query  481   CAGCGCGCGGATTTGCTGGTTCTGGATCGATTGCTGCGACAGTTGCTCGCACCCGAAATC  540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3109  CAGCGCGCGGATTTGCTGGTTCTGGATCGATTGCTGCGACAGTTGCTCGCACCCGAAATC  3168

Query  541   GTTGATTATCTAATTGCCCGCTGTAATGCAACGATTACGTTCAAAGACATGCGTTATACG  600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3169  GTTGATTATCTAATTGCCCGCTGTAATGCAACGATTACGTTCAAAGACATGCGTTATACG  3228

Query  601   CAGTCGGTTGGCATCCTGCCTGGTTGGGAGTTTTCAGGTTTCCTGCAACAGCTCTATGGT  660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3229  CAGTCGGTTGGCATCCTGCCTGGTTGGGAGTTTTCAGGTTTCCTGCAACAGCTCTATGGT  3288

Query  661   CTGGCGGCGGTAGATCTAGCGTTTCGCGCACCAGATATGGACGCGGACCTGCTGAGCCTG  720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3289  CTGGCGGCGGTAGATCTAGCGTTTCGCGCACCAGATATGGACGCGGACCTGCTGAGCCTG  3348

Query  721   CCGAGCTGGCCGCCTGCCGAACTCGCTGGCGACCATGACGTGCGACGCAACGACGCCCGT  780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3349  CCGAGCTGGCCGCCTGCCGAACTCGCTGGCGACCATGACGTGCGACGCAACGACGCCCGT  3408


Query  1     TTTCCTGCAACAGCTCTATGGTCTGGCGGCGGTAGATCTAGCGTTTCGCGCACCAGATAT  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3267  TTTCCTGCAACAGCTCTATGGTCTGGCGGCGGTAGATCTAGCGTTTCGCGCACCAGATAT  3326

Query  61    GGACGCGGACCTGCTGAGCCTGCCGAGCTGGCCGCCTGCCGAACTCGCTGGCGACCATGA  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3327  GGACGCGGACCTGCTGAGCCTGCCGAGCTGGCCGCCTGCCGAACTCGCTGGCGACCATGA  3386

Query  121   CGTGCGACGCAACGACGCCCGTAGTTGCCCGTCGATGCGGCTGCTTGCCTACGAACGAAT  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3387  CGTGCGACGCAACGACGCCCGTAGTTGCCCGTCGATGCGGCTGCTTGCCTACGAACGAAT  3446

Query  181   TCTCGATCGACTGTATGCGCTGATTCGGGTTCCCGAAGAGACTGCGCGTGTGGCTGTTAC  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3447  TCTCGATCGACTGTATGCGCTGATTCGGGTTCCCGAAGAGACTGCGCGTGTGGCTGTTAC  3506

Query  241   GCGTCTCGAGCAGCGTTACGAACGCCATCCGGAGCGGGAGACACTTCTCGACGCGGCGCT  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3507  GCGTCTCGAGCAGCGTTACGAACGCCATCCGGAGCGGGAGACACTTCTCGACGCGGCGCT  3566

Query  301   GTACCCCAGTCGACGGTGCTGGCCGAGTGCAGTGCGCATGCGCCTGCGCCCCTTCGACTG  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3567  GTACCCCAGTCGACGGTGCTGGCCGAGTGCAGTGCGCATGCGCCTGCGCCCCTTCGACTG  3626

Query  361   TCTGCTGGGTCGAGCCCTCTTTGATGCGATTCGCGATCGCATTGGCCCGGGCGTGAGTGC  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3627  TCTGCTGGGTCGAGCCCTCTTTGATGCGATTCGCGATCGCATTGGCCCGGGCGTGAGTGC  3686

Query  421   GCTCCGACCACTCGTCGAGATGCCCGAATCCCAGACTGCGGTAAGCGTCGTGTCAGGGCC  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3687  GCTCCGACCACTCGTCGAGATGCCCGAATCCCAGACTGCGGTAAGCGTCGTGTCAGGGCC  3746

Query  481   CGAGAATCCCAGTCTTTTCTTCGAAATGTTTGGATTCGAGGTGCGAATGCTCGCAAGCGG  540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3747  CGAGAATCCCAGTCTTTTCTTCGAAATGTTTGGATTCGAGGTGCGAATGCTCGCAAGCGG  3806

Query  541   TTTTGACAGAATCCTGCCTGCTGGCATGGGCAGCGCTCGACCCGCATCAGAAGCAGCAGC  600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3807  TTTTGACAGAATCCTGCCTGCTGGCATGGGCAGCGCTCGACCCGCATCAGAAGCAGCAGC  3866

Query  601   AACTGCAGGGATGCGTTGGTCATTGCCCAGCGGGCGCACCGTGGCGTATGTTCGCGTCTC  660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3867  AACTGCAGGGATGCGTTGGTCATTGCCCAGCGGGCGCACCGTGGCGTATGTTCGCGTCTC  3926
```

```
Query  661   GCCGCACGCGCTGCAGCGATGGGCGCTGGATGTTCAGCGCCTGTTGATGGCGACCATACA  720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3927  GCCGCACGCGCTGCAGCGATGGGCGCTGGATGTTCAGCGCCTGTTGATGGCGACCATACA  3986

Query  721   TGCCCCATTTACGCGCGTGGCAGCCCGGTGGAATGCGTTAGCACTCCACTTTGCCGGCAT  780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3987  TGCCCCATTTACGCGCGTGGCAGCCCGGTGGAATGCGTTAGCACTCCACTTTGCCGGCAT  4046

Query  781   GTATCGCC  788
              ||||||||
Sbjct  4047  GTATCGCC  4054
```

```
Query  1     GTTCGCGTCTCGCCGCACGCGCTGCAGCGATGGGCGCTGGATGTTCAGCGCCTGTTGATG  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3916  GTTCGCGTCTCGCCGCACGCGCTGCAGCGATGGGCGCTGGATGTTCAGCGCCTGTTGATG  3975

Query  61    GCGACCATACATGCCCCATTTACGCGCGTGGCAGCCCGGTGGAATGCGTTAGCACTCCAC  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3976  GCGACCATACATGCCCCATTTACGCGCGTGGCAGCCCGGTGGAATGCGTTAGCACTCCAC  4035

Query  121   TTTGCCGGCATGTATCGCCAAGTCGCAGCGAATGATCCAACCGTACGGCAATTTGTCCAG  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4036  TTTGCCGGCATGTATCGCCAAGTCGCAGCGAATGATCCAACCGTACGGCAATTTGTCCAG  4095

Query  181   CATGCCCAAGAGCGCGTACAGAACAGGATCAAGCTGGGTCTGAACTCGAAGATGCCGGTG  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4096  CATGCCCAAGAGCGCGTACAGAACAGGATCAAGCTGGGTCTGAACTCGAAGATGCCGGTG  4155

Query  241   CGCTTCCCACCGGTGGTGTTCTACGCTCCGCGCAGCCTCGGCGGCCTCGAAATGATGAAC  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4156  CGCTTCCCACCGGTGGTGTTCTACGCTCCGCGCAGCCTCGGCGGCCTCGAAATGATGAAC  4215

Query  301   ATAGGCCACGACCCGGTGCCGTCGCTGTGGTCCTTCATACCAAGCTGGCTCGATGAGATC  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4216  ATAGGCCACGACCCGGTGCCGTCGCTGTGGTCCTTCATACCAAGCTGGCTCGATGAGATC  4275

Query  361   GCTGACGCGGAGTTGCTGGAGCGCGAGCTTCAAGAACGAGCGCAGGCATTTGGGGCACAC  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4276  GCTGACGCGGAGTTGCTGGAGCGCGAGCTTCAAGAACGAGCGCAGGCATTTGGGGCACAC  4335

Query  421   CTCGATCGGCGACTGCTGCCACCAGCATGGCTGCATCGCGGTCTGCCACGTTTGGCTGCA  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4336  CTCGATCGGCGACTGCTGCCACCAGCATGGCTGCATCGCGGTCTGCCACGTTTGGCTGCA  4395

Query  481   CGCTATCACCCGCAAGGAGCGTTGTGGACGCTCGATCACGGCTATCGTGTGCGAAGTCTC  540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4396  CGCTATCACCCGCAAGGAGCGTTGTGGACGCTCGATCACGGCTATCGTGTGCGAAGTCTC  4455

Query  541   TTGCGCGTGCACGTTACCGGGCGGAAGAACGCCCTCTGGTGGCTAGATTTCATGCACGAC  600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4456  TTGCGCGTGCACGTTACCGGGCGGAAGAACGCCCTCTGGTGGCTAGATTTCATGCACGAC  4515

Query  601   GGACGCCTTTGGGAGCTGGACGACTACCGGAGTCAGGTTACGCATGCTTTGGGTGGCGTG  660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4516  GGACGCCTTTGGGAGCTGGACGACTACCGGAGTCAGGTTACGCATGCTTTGGGTGGCGTG  4575

Query  661   CCGGCAATCCTGTCGCATACGCTCTTTGCGGCAACTGGGTATCGAGATTGGCGAGGCATC  720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4576  CCGGCAATCCTGTCGCATACGCTCTTTGCGGCAACTGGGTATCGAGATTGGCGAGGCATC  4635

Query  721   GTCTGGGGCGACCACGGCTTTGAGCACAAACTCGCGGG  758
             |||||||||||||||||||||||||||||||||||||||
Sbjct  4636  GTCTGGGGCGACCACGGCTTTGAGCACAAACTCGCGGG  4673
```

```
Query  1     GCATGCTTTGGGTGGCGTGCCGGCAATCCTGTCGCATACGCTCTTTGCGGCAACTGGGTA  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4557  GCATGCTTTGGGTGGCGTGCCGGCAATCCTGTCGCATACGCTCTTTGCGGCAACTGGGTA  4616

Query  61    TCGAGATTGGCGAGGCATCGTCTGGGGCGACCACGGCTTTGAGCACAAACTCGCGGGCAG  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4617  TCGAGATTGGCGAGGCATCGTCTGGGGCGACCACGGCTTTGAGCACAAACTCGCGGGCAG  4676

Query  121   GCCGTTGACGCGGGCGCAGCGATCTGGACTTGTCCAGATACCCAATCGCCGATTCACGCT  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4677  GCCGTTGACGCGGGCGCAGCGATCTGGACTTGTCCAGATACCCAATCGCCGATTCACGCT  4736

Query  181   CTGGTGGTCGCCGACCATCAATCGCAGCCGGGTGTACATGGGCTTCCGCGCCCAGCTCGA  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4737  CTGGTGGTCGCCGACCATCAATCGCAGCCGGGTGTACATGGGCTTCCGCGCCCAGCTCGA  4796

Query  241   TCTGACGGGGATCTTCATGTACGGCAAGCTTTCGACGCTCAAAATCAGCCTTTTGCAGGT  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4797  TCTGACGGGGATCTTCATGTACGGCAAGCTTTCGACGCTCAAAATCAGCCTTTTGCAGGT  4856

Query  301   GTTCCGAGGCCATCTCTGGCAGCGTATTCATGAAAGTCTGGTACTTGATTTGTGCAAAGC  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4857  GTTCCGAGGCCATCTCTGGCAGCGTATTCATGAAAGTCTGGTACTTGATTTGTGCAAAGC  4916

Query  361   ATTGGATACAGAGCTGGGGGCGCGCTCGCGCGAGGCGCGTGTTGCGGTCACGGTGCAGAA  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Sbjct  4917  ATTGGATACAGAGCTGGGGGCGCGCTCGCGCGAGGCGCGTGTTGCGGTCACGGTGCAGAA  4976

Query  421   GGAACGAATCCATCCGCGCAAGTCTTATCGCATGCACTGGAGCAGCGCGGATATTAGAAT  480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4977  GGAACGAATCCATCCGCGCAAGTCTTATCGCATGCACTGGAGCAGCGCGGATATTAGAAT  5036

Query  481   CGATTTCGAGCAACCGATTCTCGTGTCCGGTGACGCCTTGCCCGTGGACGAGGCCGTCGC  540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5037  CGATTTCGAGCAACCGATTCTCGTGTCCGGTGACGCCTTGCCCGTGGACGAGGCCGTCGC  5096

Query  541   CTTCGAATCCAATGCGGGCGGTCAAGGGCGATCGTCGCTGCGGCCGCCAGGTAGCGACGC  600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5097  CTTCGAATCCAATGCGGGCGGTCAAGGGCGATCGTCGCTGCGGCCGCCAGGTAGCGACGC  5156

Query  601   GGCAGCCAAAATCTATTGGATTGATGTCCAGCTCCGTTGGGGCGACTACGACGACCATGA  660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5157  GGCAGCCAAAATCTATTGGATTGATGTCCAGCTCCGTTGGGGCGACTACGACGACCATGA  5216

Query  661   TGCCGCGCAATACGCAGCGCAAAAATTTCGGGCGTATACAGCACCGGGAGCGTCGAGTCT  720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5217  TGCCGCGCAATACGCAGCGCAAAAATTTCGGGCGTATACAGCACCGGGAGCGTCGAGTCT  5276

Query  721   GTACCCGTCTCACCATGGACTGGTCGTGGTCTTTGATCTAGCCTACGGCGAGTGGTCTGC  780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5277  GTACCCGTCTCACCATGGACTGGTCGTGGTCTTTGATCTAGCCTACGGCGAGTGGTCTGC  5336

Query  781   TTTTGGGCAC  790
              ||||||||||
Sbjct  5337  TTTTGGGCAC  5346
```

```
Query  1     TGATGCCGCGCAATACGCAGCGCAAAAATTTCGGGCGTATACAGCACCGGGAGCGTCGAG  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5214  TGATGCCGCGCAATACGCAGCGCAAAAATTTCGGGCGTATACAGCACCGGGAGCGTCGAG  5273

Query  61    TCTGTACCCGTCTCACCATGGACTGGTCGTGGTCTTTGATCTAGCCTACGGCGAGTGGTC  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5274  TCTGTACCCGTCTCACCATGGACTGGTCGTGGTCTTTGATCTAGCCTACGGCGAGTGGTC  5333

Query  121   TGCTTTTGGGCACGCCTGCGCCATGGGCGTCGTGTCGATCGTGACCGAGGAATCGCGCCG  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5334  TGCTTTTGGGCACGCCTGCGCCATGGGCGTCGTGTCGATCGTGACCGAGGAATCGCGCCG  5393

Query  181   GATGCTTATGCATAACCAGGCACTCGCGCTCCTTCGTGAACGCATTCGCAAAGCGCTTCA  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5394  GATGCTTATGCATAACCAGGCACTCGCGCTCCTTCGTGAACGCATTCGCAAAGCGCTTCA  5453

Query  241   GCTCTACGTCATGGAGACGGTGGAAAGCGAGACGCTGCAGGCGGCGACGACTACGACATC  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5454  GCTCTACGTCATGGAGACGGTGGAAAGCGAGACGCTGCAGGCGGCGACGACTACGACATC  5513

Query  301   GGCTTCCGATACGATACCTCTGCTCGTGGGGTGCGGCGGCGACTTGTGGCGGCAACGCCT  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5514  GGCTTCCGATACGATACCTCTGCTCGTGGGGTGCGGCGGCGACTTGTGGCGGCAACGCCT  5573

Query  361   CTGGATCGTGGATGATCGCACTGCGTACAGGCCACATGCGAACGGTGTTATCTGGATATG  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5574  CTGGATCGTGGATGATCGCACTGCGTACAGGCCACATGCGAACGGTGTTATCTGGATATG  5633

Query  421   GGAGACGTCGACAGGGCGACTGTTTGTGAAGATTGTCCATCGGACTACGTGGGCTGGCCA  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5634  GGAGACGTCGACAGGGCGACTGTTTGTGAAGATTGTCCATCGGACTACGTGGGCTGGCCA  5693

Query  481   AACCCGGCGAGCGCAACTCGCCAAGTGGAAATGCGCTGAGCACGTTTTAACCATGCTCCG  540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5694  AACCCGGCGAGCGCAACTCGCCAAGTGGAAATGCGCTGAGCACGTTTTAACCATGCTCCG  5753

Query  541   TTCACAGCCAACTGAAGAGCTACCGCGGGGCATCGTGCTCGCACAAACCGCATCCATGGA  600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5754  TTCACAGCCAACTGAAGAGCTACCGCGGGGCATCGTGCTCGCACAAACCGCATCCATGGA  5813

Query  601   CCCGTTGAAGACGCTCTTGGCAGGCACCGAGTATGCCAAAATTCCTGTGCGTGCCGGTGC  660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5814  CCCGTTGAAGACGCTCTTGGCAGGCACCGAGTATGCCAAAATTCCTGTGCGTGCCGGTGC  5873

Query  661   GGCGGCCATGCCGCTGCAGGCGCTAATGGCGCTGCCGGAGATCCGCGACCGTACTCAGAC  720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5874  GGCGGCCATGCCGCTGCAGGCGCTAATGGCGCTGCCGGAGATCCGCGACCGTACTCAGAC  5933

Query  721   TGCGCGCTCCAGCGAGCTTTCGATCTGGAGTGGCTACGCGGATTGGCTTGAGCATGTGCC  780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5934  TGCGCGCTCCAGCGAGCTTTCGATCTGGAGTGGCTACGCGGATTGGCTTGAGCATGTGCC  5993

Query  781   CGTG  784
             ||||
Sbjct  5994  CGTG  5997
```

```
Query  18    TGTGCGTGCCGGTGCGGCGGCCATGCCGCTGCAGGCGCTAATGGCGCTGCCGGAGATCCG  77
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5859  TGTGCGTGCCGGTGCGGCGGCCATGCCGCTGCAGGCGCTAATGGCGCTGCCGGAGATCCG  5918
```

```
Query  78    CGACCGTACTCAGACTGCGCGCTCCAGCGAGCTTTCGATCTGGAGTGGCTACGCGGATTG  137
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5919  CGACCGTACTCAGACTGCGCGCTCCAGCGAGCTTTCGATCTGGAGTGGCTACGCGGATTG  5978

Query  138   GCTTGAGCATGTGCCCGTGTGGATCGCGTCGGCGCGCTTCCTGCTCCTGCTCCACGCCTT  197
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5979  GCTTGAGCATGTGCCCGTGTGGATCGCGTCGGCGCGCTTCCTGCTCCTGCTCCACGCCTT  6038

Query  198   GGACCGGGCGCCAGAGCGTGTCCTGCAGCTGGTGTGGCCTCAGCGGTCGGCGGACGAGGA  257
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6039  GGACCGGGCGCCAGAGCGTGTCCTGCAGCTGGTGTGGCCTCAGCGGTCGGCGGACGAGGA  6098

Query  258   GAGCGCGGGCTCCGCGACACCTTGGCTGTGGCCCGCGCTTCCCGAGACTGACTGGCGCCG  317
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6099  GAGCGCGGGCTCCGCGACACCTTGGCTGTGGCCCGCGCTTCCCGAGACTGACTGGCGCCG  6158

Query  318   TCTGGAACTAGAGCTCCAGTCGCTGGTGCCCGTGCGTCTACGCCCTGCCCATGTCGCTGG  377
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6159  TCTGGAACTAGAGCTCCAGTCGCTGGTGCCCGTGCGTCTACGCCCTGCCCATGTCGCTGG  6218

Query  378   TGATCAGCCAGGTGGCAGGGATGACGACGGAACGGAGCAGGTGCTGGAGCACACACGACC  437
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6219  TGATCAGCCAGGTGGCAGGGATGACGACGGAACGGAGCAGGTGCTGGAGCACACACGACC  6278

Query  438   GAAGACGGTTGCCGCATTCGACCGGTATGGCAATGTGATTTCGGTGGAGACCACCACGCC  497
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6279  GAAGACGGTTGCCGCATTCGACCGGTATGGCAATGTGATTTCGGTGGAGACCACCACGCC  6338

Query  498   ATTTGAGCGGCAGGAGTACCGCACGTCGTCGGTCACGGTAACGAATGCAGAACAGCAACG  557
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6339  ATTTGAGCGGCAGGAGTACCGCACGTCGTCGGTCACGGTAACGAATGCAGAACAGCAACG  6398

Query  558   GCTACTGTCCTTGTACCGACGACTCCCAGACGTTCTCGAAAACCTGCACGTTGTCGAGCC  617
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6399  GCTACTGTCCTTGTACCGACGACTCCCAGACGTTCTCGAAAACCTGCACGTTGTCGAGCC  6458

Query  618   TGGACACACCAGTGGCAGTCTGTACGGAGGCGAGCGCCCGGAACAGGCGACGATCTCGGC  677
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6459  TGGACACACCAGTGGCAGTCTGTACGGAGGCGAGCGCCCGGAACAGGCGACGATCTCGGC  6518

Query  678   GATGCCGCATCTCGCCAAGCGATTCGAGCTCCATCTACCGAGAGGCCTCGTCCGCGGTCT  737
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6519  GATGCCGCATCTCGCCAAGCGATTCGAGCTCCATCTACCGAGAGGCCTCGTCCGCGGTCT  6578

Query  738   TTTGCTCGCGGGTATGACTTGTGGCCAACTCTGGGGCACGCGCACCGCCAACGCGGATAC  797
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6579  TTTGCTCGCGGGTATGACTTGTGGCCAACTCTGGGGCACGCGCACCGCCAACGCGGATAC  6638

Query  798   AGTGGTTT  805
             ||||||||
Sbjct  6639  AGTGGTTT  6646
```

Prp8

```
Query  1     TCGTCCGCGGTCTTTTGCTCGCGGGTATGACTTGTGGCCAACTCTGGGGCACGCGCACCG  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6566  TCGTCCGCGGTCTTTTGCTCGCGGGTATGACTTGTGGCCAACTCTGGGGCACGCGCACCG  6625

Query  61    CCAACGCGGATACAGTGGTTTGGATCGCCTGGGCGCTGCGTCATCATGACGCCGAGGCAT  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6626  CCAACGCGGATACAGTGGTTTGGATCGCCTGGGCGCTGCGTCATCATGACGCCGAGGCAT  6685

Query  121   CGACACCGTTGCCGAGCATGGTTGAGGCAGCGGCCTGCTTTCGCCCTGGCTCGGTGCAAG  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6686  CGACACCGTTGCCGAGCATGGTTGAGGCAGCGGCCTGCTTTCGCCCTGGCTCGGTGCAAG  6745

Query  181   CGATCGGCTGGATTCAAGCGTGCCTCGATGTGGGTCGGCCACCAGCGCCAGGACCGGGGG  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6746  CGATCGGCTGGATTCAAGCGTGCCTCGATGTGGGTCGGCCACCAGCGCCAGGACCGGGGG  6805

Query  241   TGCATGCGCCCCACGATGTGCAGGTCATACTCTATCTCCAGGGCGAAGAACGCCAAATCG  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6806  TGCATGCGCCCCACGATGTGCAGGTCATACTCTATCTCCAGGGCGAAGAACGCCAAATCG  6865

Query  301   ATTTGGCATCGCCGTCGCTCGATCGCCACGCAGCCGAAGCAGTGCAGACACACGATGTGC  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6866  ATTTGGCATCGCCGTCGCTCGATCGCCACGCAGCCGAAGCAGTGCAGACACACGATGTGC  6925

Query  361   ACGCATCGCGGGCACCGAGTGCGCACGCATTCGGACACGTCATCGGGAGCACGCATGCGA  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6926  ACGCATCGCGGGCACCGAGTGCGCACGCATTCGGACACGTCATCGGGAGCACGCATGCGA  6985

Query  421   TGTCACGCGCCCTGGGTGTCCTGCAGGCGTGGACTGGTGGCGGCTCCGGCGACGTGAACG  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6986  TGTCACGCGCCCTGGGTGTCCTGCAGGCGTGGACTGGTGGCGGCTCCGGCGACGTGAACG  7045

Query  481   ATCCTCAATCCGTGCATGTTCGCGTTCGACTGCGTGATGATTTGAGTGCTGTTTTCCTCT  540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  7046  ATCCTCAATCCGTGCATGTTCGCGTTCGACTGCGTGATGATTTGAGTGCTGTTTTCCTCT  7105

Query  541   ACGCAAAAGATGGTCGGCTGATTGCGCGCCCGCCGGGACGTTTATTCGAAACGATACCGG  600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  7106  ACGCAAAAGATGGTCGGCTGATTGCGCGCCCGCCGGGACGTTTATTCGAAACGATACCGG  7165

Query  601   CACCGATCGAAGAGGGAACTTAG  623
```

```
            ||||||||||||||||||||||
Sbjct  7166  CACCGATCGAAGAGGGAACTTAG  7188
```

Prp8 gene inserted into pMCSG2 (The LIC sequence is presented in bold; pSR797)

```
Query  57    TACTTCCAATCCCAATGCA 76
             ||||||||||||||||||||
Sbjct        --------------------

Query  77    CCCAAACGTGCGTTTTTCGGACGCGACGAAGAGGCAGATCAGACGCTGGACCTGAAGCGA  136
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4     CCCAAACGTGCGTTTTTCGGACGCGACGAAGAGGCAGATCAGACGCTGGACCTGAAGCGA  63

Query  137   AAGCGGCCACGCCGTGCTTTCGGAGCGGATCAGCCATTTTTCAAACCGTATACCTCCCCC  196
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  64    AAGCGGCCACGCCGTGCTTTCGGAGCGGATCAGCCATTTTTCAAACCGTATACCTCCCCC  123

Query  197   GATATTTTCCGCAAACTAGCGAGTCTTGCACGGATCGAGTTGGAGCGAAGCGAAAATGAC  256
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  124   GATATTTTCCGCAAACTAGCGAGTCTTGCACGGATCGAGTTGGAGCGAAGCGAAAATGAC  183

Query  257   CTGGAAAACAGTGCCCACCGTGGTAACCGAGACCGACATATCTCCCGGAACACAGATGCA  316
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  184   CTGGAAAACAGTGCCCACCGTGGTAACCGAGACCGACATATCTCCCGGAACACAGATGCA  243

Query  317   CTGACACTGAACGCGCTCCGCTACCTTCCGCACGCGGTGTACAAATGGCTGGAACACATG  376
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  244   CTGACACTGAACGCGCTCCGCTACCTTCCGCACGCGGTGTACAAATGGCTGGAACACATG  303

Query  377   CCGGCCCCTTGGGAGCCAACGCGGTTTGTACCCGTGATTTTCCATCACACTGGTGCGTTG  436
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  304   CCGGCCCCTTGGGAGCCAACGCGGTTTGTACCCGTGATTTTCCATCACACTGGTGCGTTG  363

Query  437   GCTTTTATCGAGGGTTCACGTCGAGTCCCCGAGGTCGTTCATCGGGCCCAATGGGCGCGT  496
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  364   GCTTTTATCGAGGGTTCACGTCGAGTCCCCGAGGTCGTTCATCGGGCCCAATGGGCGCGT  423

Query  497   TGGTGCGCCTACCTCGACCGTCGTCGTCATGAGGCACACGTCGCCTCGACAAGCAGCGGA  556
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  424   TGGTGCGCCTACCTCGACCGTCGTCGTCATGAGGCACACGTCGCCTCGACAAGCAGCGGA  483

Query  557   AAGCGGCAGCGAGATCTGACGCGTTCCTTCGTACGTTTGCAAGTGCCGGCGTTCGACGAC  616
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  484   AAGCGGCAGCGAGATCTGACGCGTTCCTTCGTACGTTTGCAAGTGCCGGCGTTCGACGAC  543

Query  617   GATGAGCCGTCACCGAGCTTCCGCGAGTTTGTCTACCATCGAACGCTGCCGACACCATTT  676
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  544   GATGAGCCGTCACCGAGCTTCCGCGAGTTTGTCTACCATCGAACGCTGCCGACACCATTT  603

Query  677   GCTAACGATGCAGACGCGGCAGCTGAATACGAGCGAATCCAACAGCTCGCCGCGAATTTG  736
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  604   GCTAACGATGCAGACGCGGCAGCTGAATACGAGCGAATCCAACAGCTCGCCGCGAATTTG  663


Query  25     TTGCACACACCGGAGGCGCGTGCTGCGCTCTACCGGATGAGTTTGCCGTGGGCGCCACAC  84
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  661    TTGCACACACCGGAGGCGCGTGCTGCGCTCTACCGGATGAGTTTGCCGTGGGCGCCACAC  720

Query  85     CCGAACGATCTGGATCCAAGTCAGTTCTATCTGCGAAATCGCTTGACATTGCGACGAGTG  144
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  721    CCGAACGATCTGGATCCAAGTCAGTTCTATCTGCGAAATCGCTTGACATTGCGACGAGTG  780

Query  145    CATCGACTCGAGCATGAGCGTGCAAACCTTCGACGTAAGAGCCTACACGCCGGCTCTTTG  204
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  781    CATCGACTCGAGCATGAGCGTGCAAACCTTCGACGTAAGAGCCTACACGCCGGCTCTTTG  840

Query  205    CACAGCGAGGGCGCGCATCTGCTCCAGGAGGACTGGGACGATTTCACGGAGCTCGGTATG  264
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  841    CACAGCGAGGGCGCGCATCTGCTCCAGGAGGACTGGGACGATTTCACGGAGCTCGGTATG  900

Query  265    CATTCCATCCTGCCTGGTCAACGTCGCGCTCCAGAACTGCACCAGGCTTATCCGTACCTG  324
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  901    CATTCCATCCTGCCTGGTCAACGTCGCGCTCCAGAACTGCACCAGGCTTATCCGTACCTG  960

Query  325    TACGACGCATTCGATGAACGCACCCCGAACCCAGAGCAACCGCTGTGGCACCACCATGCG  384
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  961    TACGACGCATTCGATGAACGCACCCCGAACCCAGAGCAACCGCTGTGGCACCACCATGCG  1020

Query  385    CCGCGGCCGCTCATGTGGCCGCCAACACCACCACCGCACCCGCTCTGGCCGAGTGCGTGG  444
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1021   CCGCGGCCGCTCATGTGGCCGCCAACACCACCACCGCACCCGCTCTGGCCGAGTGCGTGG  1080

Query  445    AGCTGGCAGCCTACGCTAGCACCGCTCCTGCACAGCAGTGCACAAGGCGCTGGCGAGCGC  504
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1081   AGCTGGCAGCCTACGCTAGCACCGCTCCTGCACAGCAGTGCACAAGGCGCTGGCGAGCGC  1140

Query  505    ATCATCCCAGACACCAGCGAAGGCGCCAGCGCTCAGCGTACGCGCTACCGCGCTCATCCG  564
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1141   ATCATCCCAGACACCAGCGAAGGCGCCAGCGCTCAGCGTACGCGCTACCGCGCTCATCCG  1200
```

```
Query  565   GCGCTGGGCGATTTCGATCTGGACGCGTGCGGCGAACGCGTCGGAGCCTGGCTCGATTTG  624
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1201  GCGCTGGGCGATTTCGATCTGGACGCGTGCGGCGAACGCGTCGGAGCCTGGCTCGATTTG  1260

Query  625   CTCTATGCACCGCGTCCATATAAGGGCAGATGTGTTCGAAAACGGAAGCGGATGCAGGAT  684
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1261  CTCTATGCACCGCGTCCATATAAGGGCAGATGTGTTCGAAAACGGAAGCGGATGCAGGAT  1320

Query  685   ATCGACATGGAACGCGAGTTGTACTGGCAACTGGAGCGCGTCCATTACATGCGCCCGAAA  744
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1321  ATCGACATGGAACGCGAGTTGTACTGGCAACTGGAGCGCGTCCATTACATGCGCCCGAAA  1380

Query  745   AGTACTGCCGCAGTCCACAGCTTGTTGAAACGACGCTGTCGACAAGCAGAAGCACAGC  802
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1381  AGTACTGCCGCAGTCCACAGCTTGTTGAAACGACGCTGTCGACAAGCAGAAGCACAGC  1438


Query  27    ACGGAAGCGGATGCAGGATATCGACATGGAACGCGAGTTGTACTGGCAACTGGAGCGCGT  86
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1302  ACGGAAGCGGATGCAGGATATCGACATGGAACGCGAGTTGTACTGGCAACTGGAGCGCGT  1361

Query  87    CCATTACATGCGCCCGAAAAGTACTGCCGCAGTCCACAGCTTGTTGAAACGACGCTGTCG  146
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1362  CCATTACATGCGCCCGAAAAGTACTGCCGCAGTCCACAGCTTGTTGAAACGACGCTGTCG  1421

Query  147   ACAAGCAGAAGCACAGCGTCTTCGCTACCTGCGAAAAGAACGTGGTCGCGGGCGCAAGCA  206
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1422  ACAAGCAGAAGCACAGCGTCTTCGCTACCTGCGAAAAGAACGTGGTCGCGGGCGCAAGCA  1481

Query  207   ATCGACTTTGGCGATCCAGTCGGCCTCTGGAGACGCTGGCGATGAAGCTGAGAAGCAGCA  266
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1482  ATCGACTTTGGCGATCCAGTCGGCCTCTGGAGACGCTGGCGATGAAGCTGAGAAGCAGCA  1541

Query  267   CCGACGGCCGGATCTGCTGAGGACGTTGCGGCAATCGGGAttttttATCGCACCAGTAT  326
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1542  CCGACGGCCGGATCTGCTGAGGACGTTGCGGCAATCGGGATTTTTTTATCGCACCAGTAT  1601

Query  327   GGATTGGTTGGAGGTCGGCATCTGGCTGTGCGACGCTGCACGATCGATGTTTTTGCTACT  386
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1602  GGATTGGTTGGAGGTCGGCATCTGGCTGTGCGACGCTGCACGATCGATGTTTTTGCTACT  1661

Query  387   TTTGAGACGAAAACGATTCTTCTTTCTGAGCATGGACTACAATTTTCAGATCGTGCCGTT  446
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1662  TTTGAGACGAAAACGATTCTTCTTTCTGAGCATGGACTACAATTTTCAGATCGTGCCGTT  1721

Query  447   GCGTACGCTGACGACCAAAGAACGAAAGCAGTCTCGCTTCGGGAATGCGTTTCATTTGAT  506
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1722  GCGTACGCTGACGACCAAAGAACGAAAGCAGTCTCGCTTCGGGAATGCGTTTCATTTGAT  1781

Query  507   GCGCGAATGGATGCGTATGGTGAAGCTGATTGTGGATGTGCATCTGCGGTATCGGGCGGG  566
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1782  GCGCGAATGGATGCGTATGGTGAAGCTGATTGTGGATGTGCATCTGCGGTATCGGGCGGG  1841

Query  567   TCTCGGCGCGGATGCCATTCAACTCGCCGACAGCATCACGTACATCGAATCACATATCGG  626
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1842  TCTCGGCGCGGATGCCATTCAACTCGCCGACAGCATCACGTACATCGAATCACATATCGG  1901

Query  627   TGAACTGACAGGTCTCTATCGTTACAAGTACCGGGTTATGCGCCAAATCCATGCGACCAA  686
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1902  TGAACTGACAGGTCTCTATCGTTACAAGTACCGGGTTATGCGCCAAATCCATGCGACCAA  1961

Query  687   GGACCTCAAGCATCTTATGTATCAGCGTTTCGGATGGTCGGGTCCCGGCAAGGCCTGCTG  746
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1962  GGACCTCAAGCATCTTATGTATCAGCGTTTCGGATGGTCGGGTCCCGGCAAGGCCTGCTG  2021


Query  20    GACCTCAAGCATCTTATGTATCAGCGTTTCGGATGGTCGGGTCCCGGCAAGGCCTGCTGG  79
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1963  GACCTCAAGCATCTTATGTATCAGCGTTTCGGATGGTCGGGTCCCGGCAAGGCCTGCTGG  2022

Query  80    CAACCCCTCTGGCGGCAATGGGTGCACCTACTTCGAGGCCTGATGCCCTTGCTCGAACAG  139
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2023  CAACCCCTCTGGCGGCAATGGGTGCACCTACTTCGAGGCCTGATGCCCTTGCTCGAACAG  2082

Query  140   TGGTTGGGTAACCTGCTGAATCGCCATTTTGAAGGACGCGAGCCGCTGCGCATGGCGCAA  199
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2083  TGGTTGGGTAACCTGCTGAATCGCCATTTTGAAGGACGCGAGCCGCTGCGCATGGCGCAA  2142

Query  200   CGCACCGTCACAAAGCAACGACTCGAGTCGCAGTTTGATGTGGCGCTACGTCAGGAGACT  259
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2143  CGCACCGTCACAAAGCAACGACTCGAGTCGCAGTTTGATGTGGCGCTACGTCAGGAGACT  2202

Query  260   ATAGCGCGTCTTCGGCAACTGTTGCCAGCAGCGCGTCAACCACTATATACGAGGCGTGTT  319
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2203  ATAGCGCGTCTTCGGCAACTGTTGCCAGCAGCGCGTCAACCACTATATACGAGGCGTGTT  2262

Query  320   CTCCAGCATATGCACCAGGCGTGGCGCTGCTGGAAGGCCAATATCCCGTGGCACGTACGA  379
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2263  CTCCAGCATATGCACCAGGCGTGGCGCTGCTGGAAGGCCAATATCCCGTGGCACGTACGA  2322

Query  380   GACATGCCCCCTGAAATCGACAGACTCGTACGGGATTACGTGCAGCAGCGGGCGCAATGG  439
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2323  GACATGCCCCCTGAAATCGACAGACTCGTACGGGATTACGTGCAGCAGCGGGCGCAATGG  2382
```

```
Query  440   TGGATAGAGGGTGCTCGACGCTTCGCGATTGCGTTTCGAAGCGGTCGTTCCATGGATAAG  499
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2383  TGGATAGAGGGTGCTCGACGCTTCGCGATTGCGTTTCGAAGCGGTCGTTCCATGGATAAG  2442

Query  500   GCACTTGTTCGCCAGATGTACGGGCGACTTGCTCGTCTTGCCGTTCACCATGAGCAGGCG  559
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2443  GCACTTGTTCGCCAGATGTACGGGCGACTTGCTCGTCTTGCCGTTCACCATGAGCAGGCG  2502

Query  560   CATCAGCGGCAGTATTTGGAGCACGGTCCCTTTCTGTTGCCGACCGAAGCAGCATCGATA  619
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2503  CATCAGCGGCAGTATTTGGAGCACGGTCCCTTTCTGTTGCCGACCGAAGCAGCATCGATA  2562

Query  620   TTGTACAGATTCGCAACGTACCTGGAGAAAGCAGGTGTTGCCGACTCATGCCCTTGGCAA  679
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2563  TTGTACAGATTCGCAACGTACCTGGAGAAAGCAGGTGTTGCCGACTCATGCCCTTGGCAA  2622

Query  680   TTGCCTTGGTTCGCCGCAGATAACGCGGATTCCGGGCTAGCTCCCGCATTACTGCAACTC  739
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2623  TTGCCTTGGTTCGCCGCAGATAACGCGGATTCCGGGCTAGCTCCCGCATTACTGCAACTC  2682

Query  740   GCGATCGAGCGTCTCCGCCCGGAAACGGAGGGCGAGCACTGTCCCCCGGGGCTCGGCGAA  799
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2683  GCGATCGAGCGTCTCCGCCCGGAAACGGAGGGCGAGCACTGTCCCCCGGGGCTCGGCGAA  2742

Query  800   CTTGTCACAGAGGCAGAGGCAGC  822
             |||||||||||||||||||||||
Sbjct  2743  CTTGTCACAGAGGCAGAGGCAGC  2765


Query  16    GACTCATGCCCTTGGCAATTGCCTTGGTTCGCCGCAGATAACGCGGATTCCGGGCTAGCT  75
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2605  GACTCATGCCCTTGGCAATTGCCTTGGTTCGCCGCAGATAACGCGGATTCCGGGCTAGCT  2664

Query  76    CCCGCATTACTGCAACTCGCGATCGAGCGTCTCCGCCCGGAAACGGAGGGCGAGCACTGT  135
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2665  CCCGCATTACTGCAACTCGCGATCGAGCGTCTCCGCCCGGAAACGGAGGGCGAGCACTGT  2724

Query  136   CCCCCGGGGCTCGGCGAACTTGTCACAGAGGCAGAGGCAGCGCCTGCAGCGATGCTGTAT  195
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2725  CCCCCGGGGCTCGGCGAACTTGTCACAGAGGCAGAGGCAGCGCCTGCAGCGATGCTGTAT  2784

Query  196   AGAATCCGGCAGCGACTGCAGGCGGTGCAAATGCGTCACCAGGTGGAACTCCGCTTCTAC  255
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2785  AGAATCCGGCAGCGACTGCAGGCGGTGCAAATGCGTCACCAGGTGGAACTCCGCTTCTAC  2844

Query  256   GATGACGCGGACCTGACGCCTGTGTATCGGATCGATTCCTTTGAGCGACTGGTGGACGCT  315
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2845  GATGACGCGGACCTGACGCCTGTGTATCGGATCGATTCCTTTGAGCGACTGGTGGACGCT  2904

Query  316   TTCCTCGATCAATGGCTCTGGTACAAGGCCACGGAGACGCGCCTTTTTCCGGCCCATGTC  375
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2905  TTCCTCGATCAATGGCTCTGGTACAAGGCCACGGAGACGCGCCTTTTTCCGGCCCATGTC  2964

Query  376   CAGCCATGCGACGATGAGTTGCCGCCCGTTCACGTCCTCCGTTTCGTCGAACGCTTGGAT  435
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2965  CAGCCATGCGACGATGAGTTGCCGCCCGTTCACGTCCTCCGTTTCGTCGAACGCTTGGAT  3024

Query  436   GCTATACCTCATCTGTGGACACTCGGAACAGCACCAAACAAGAGCTTCCTGGTCCTCATA  495
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3025  GCTATACCTCATCTGTGGACACTCGGAACAGCACCAAACAAGAGCTTCCTGGTCCTCATA  3084

Query  496   CAAACGCCATTGCCAGGGCTATTTCAGCGCGCGGATTTGCTGGTTCTGGATCGATTGCTG  555
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3085  CAAACGCCATTGCCAGGGCTATTTCAGCGCGCGGATTTGCTGGTTCTGGATCGATTGCTG  3144

Query  556   CGACAGTTGCTCGCACCCGAAATCGTTGATTATCTAATTGCCCGCTGTAATGCAACGATT  615
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3145  CGACAGTTGCTCGCACCCGAAATCGTTGATTATCTAATTGCCCGCTGTAATGCAACGATT  3204

Query  616   ACGTTCAAAGACATGCGTTATACGCAGTCGGTTGGCATCCTGCCTGGTTGGGAGTTTTCA  675
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3205  ACGTTCAAAGACATGCGTTATACGCAGTCGGTTGGCATCCTGCCTGGTTGGGAGTTTTCA  3264


Query  18    TGGGAGTTTTCAGGTTTCCTGCAACAGCTCTATGGTCTGGCGGCGGTAGATCTAGCGTTT  77
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3253  TGGGAGTTTTCAGGTTTCCTGCAACAGCTCTATGGTCTGGCGGCGGTAGATCTAGCGTTT  3312

Query  78    CGCGCACCAGATATGGACGCGGACCTGCTGAGCCTGCCGAGCTGGCCGCCTGCCGAACTC  137
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3313  CGCGCACCAGATATGGACGCGGACCTGCTGAGCCTGCCGAGCTGGCCGCCTGCCGAACTC  3372

Query  138   GCTGGCGACCATGACGTGCGACGCAACGACGCCCGTAGTTGCCCGTCGATGCGGCTGCTT  197
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3373  GCTGGCGACCATGACGTGCGACGCAACGACGCCCGTAGTTGCCCGTCGATGCGGCTGCTT  3432

Query  198   GCCTACGAACGAATTCTCGATCGACTGTATGCGCTGATTCGGGTTCCCGAAGAGACTGCG  257
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3433  GCCTACGAACGAATTCTCGATCGACTGTATGCGCTGATTCGGGTTCCCGAAGAGACTGCG  3492

Query  258   CGTGTGGCTGTTACGCGTCTCGAGCAGCGTTACGAACGCCATCCGGAGCGGGAGACACTT  317
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3493  CGTGTGGCTGTTACGCGTCTCGAGCAGCGTTACGAACGCCATCCGGAGCGGGAGACACTT  3552
```

```
Query  318   CTCGACGCGGCGCTGTACCCCAGTCGACGGTGCTGGCCGAGTGCAGTGCGCATGCGCCTG  377
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3553  CTCGACGCGGCGCTGTACCCCAGTCGACGGTGCTGGCCGAGTGCAGTGCGCATGCGCCTG  3612

Query  378   CGCCCCTTCGACTGTCTGCTGGGTCGAGCCCTCTTTGATGCGATTCGCGATCGCATTGGC  437
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3613  CGCCCCTTCGACTGTCTGCTGGGTCGAGCCCTCTTTGATGCGATTCGCGATCGCATTGGC  3672

Query  438   CCGGGCGTGAGTGCGCTCCGACCACTCGTCGAGATGCCCGAATCCCAGACTGCGGTAAGC  497
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3673  CCGGGCGTGAGTGCGCTCCGACCACTCGTCGAGATGCCCGAATCCCAGACTGCGGTAAGC  3732

Query  498   GTCGTGTCAGGGCCCGAGAATCCCAGTCTTTTCTTCGAAATGTTTGGATTCGAGGTGCGA  557
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3733  GTCGTGTCAGGGCCCGAGAATCCCAGTCTTTTCTTCGAAATGTTTGGATTCGAGGTGCGA  3792

Query  558   ATGCTCGCAAGCGGTTTTGACAGAATCCTGCCTGCTGGCATGGGCAGCGCTCGACCCGCA  617
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3793  ATGCTCGCAAGCGGTTTTGACAGAATCCTGCCTGCTGGCATGGGCAGCGCTCGACCCGCA  3852

Query  618   TCAGAAGCAGCAGCAACTGCAGGGATGCGTTGGTCATTGCCCAGCGGGCGCACCGTGGCG  677
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3853  TCAGAAGCAGCAGCAACTGCAGGGATGCGTTGGTCATTGCCCAGCGGGCGCACCGTGGCG  3912

Query  678   TATGTTCGCGTCTCGCCGCACGCGCTGCAGCGATGGGCGCTGGATGTTCAGCGCCTGTTG  737
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3913  TATGTTCGCGTCTCGCCGCACGCGCTGCAGCGATGGGCGCTGGATGTTCAGCGCCTGTTG  3972

Query  738   ATGGCGACCATACATGCCCCATTTACGCGCGTGGCAGCCCGGTGGAATGCGTTAGCACTC  797
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3973  ATGGCGACCATACATGCCCCATTTACGCGCGTGGCAGCCCGGTGGAATGCGTTAGCACTC  4032

Query  798   CACTTTGCCGGCATGTATCGCCAAGTCGCA  827
             ||||||||||||||||||||||||||||||
Sbjct  4033  CACTTTGCCGGCATGTATCGCCAAGTCGCA  4062


Query  16    GCGTATGTTCGCGTCTCGCCGCACGCGCTGCAGCGATGGGCGCTGGATGTTCAGCGCCTG  75
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3910  GCGTATGTTCGCGTCTCGCCGCACGCGCTGCAGCGATGGGCGCTGGATGTTCAGCGCCTG  3969

Query  76    TTGATGGCGACCATACATGCCCCATTTACGCGCGTGGCAGCCCGGTGGAATGCGTTAGCA  135
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3970  TTGATGGCGACCATACATGCCCCATTTACGCGCGTGGCAGCCCGGTGGAATGCGTTAGCA  4029

Query  136   CTCCACTTTGCCGGCATGTATCGCCAAGTCGCAGCGAATGATCCAACCGTACGGCAATTT  195
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4030  CTCCACTTTGCCGGCATGTATCGCCAAGTCGCAGCGAATGATCCAACCGTACGGCAATTT  4089

Query  196   GTCCAGCATGCCCAAGAGCGCGTACAGAACAGGATCAAGCTGGGTCTGAACTCGAAGATG  255
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4090  GTCCAGCATGCCCAAGAGCGCGTACAGAACAGGATCAAGCTGGGTCTGAACTCGAAGATG  4149

Query  256   CCGGTGCGCTTCCCACCGGTGGTGTTCTACGCTCCGCGCAGCCTCGGCGGCCTCGAAATG  315
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4150  CCGGTGCGCTTCCCACCGGTGGTGTTCTACGCTCCGCGCAGCCTCGGCGGCCTCGAAATG  4209

Query  316   ATGAACATAGGCCACGACCCGGTGCCGTCGCTGTGGTCCTTCATACCAAGCTGGCTCGAT  375
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4210  ATGAACATAGGCCACGACCCGGTGCCGTCGCTGTGGTCCTTCATACCAAGCTGGCTCGAT  4269

Query  376   GAGATCGCTGACGCGGAGTTGCTGGAGCGCGAGCTTCAAGAACGAGCGCAGGCATTTGGG  435
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4270  GAGATCGCTGACGCGGAGTTGCTGGAGCGCGAGCTTCAAGAACGAGCGCAGGCATTTGGG  4329

Query  436   GCACACCTCGATCGGCGACTGCTGCCACCAGCATGGCTGCATCGCGGTCTGCCACGTTTG  495
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4330  GCACACCTCGATCGGCGACTGCTGCCACCAGCATGGCTGCATCGCGGTCTGCCACGTTTG  4389

Query  496   GCTGCACGCTATCACCCGCAAGGAGCGTTGTGGACGCTCGATCACGGCTATCGTGTGCGA  555
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4390  GCTGCACGCTATCACCCGCAAGGAGCGTTGTGGACGCTCGATCACGGCTATCGTGTGCGA  4449

Query  556   AGTCTCTTGCGCGTGCACGTTACCGGGCGGAAGAACGCCCTCTGGTGGCTAGATTTCATG  615
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4450  AGTCTCTTGCGCGTGCACGTTACCGGGCGGAAGAACGCCCTCTGGTGGCTAGATTTCATG  4509

Query  616   CACGACGGACGCCTTTGGGAGCTGGACGACTACCGGAGTCAGGTTACGCATGCTTTGGGT  675
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4510  CACGACGGACGCCTTTGGGAGCTGGACGACTACCGGAGTCAGGTTACGCATGCTTTGGGT  4569

Query  676   GGCGTGCCGGCAATCCTGTCGCATACGCTCTTTGCGGCAACTGGGTATCGAGATTGGCGA  735
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4570  GGCGTGCCGGCAATCCTGTCGCATACGCTCTTTGCGGCAACTGGGTATCGAGATTGGCGA  4629

Query  736   GGCATCGTCTGGGGCGACCACGGCTTTGAGCACAAACTCGCGGGCAGGCCGTTGACGCGG  795
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4630  GGCATCGTCTGGGGCGACCACGGCTTTGAGCACAAACTCGCGGGCAGGCCGTTGACGCGG  4689


Query  1     GTTACGCATGCTTTGGGTGGCGTGCCGGCAATCCTGTCGCATACGCTCTTTGCGGCAACT  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4552  GTTACGCATGCTTTGGGTGGCGTGCCGGCAATCCTGTCGCATACGCTCTTTGCGGCAACT  4611
```

```
Query  61    GGGTATCGAGATTGGCGAGGCATCGTCTGGGGCGACCACGGCTTTGAGCACAAACTCGCG  120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4612  GGGTATCGAGATTGGCGAGGCATCGTCTGGGGCGACCACGGCTTTGAGCACAAACTCGCG  4671

Query  121   GGCAGGCCGTTGACGCGGGCGCAGCGATCTGGACTTGTCCAGATACCCAATCGCCGATTC  180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4672  GGCAGGCCGTTGACGCGGGCGCAGCGATCTGGACTTGTCCAGATACCCAATCGCCGATTC  4731

Query  181   ACGCTCTGGTGGTCGCCGACCATCAATCGCAGCCGGGTGTACATGGGCTTCCGCGCCCAG  240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4732  ACGCTCTGGTGGTCGCCGACCATCAATCGCAGCCGGGTGTACATGGGCTTCCGCGCCCAG  4791

Query  241   CTCGATCTGACGGGGATCTTCATGTACGGCAAGCTTTCGACGCTCAAAATCAGCCTTTTG  300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4792  CTCGATCTGACGGGGATCTTCATGTACGGCAAGCTTTCGACGCTCAAAATCAGCCTTTTG  4851

Query  301   CAGGTGTTCCGAGGCCATCTCTGGCAGCGTATTCATGAAAGTCTGGTACTTGATTTGTGC  360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4852  CAGGTGTTCCGAGGCCATCTCTGGCAGCGTATTCATGAAAGTCTGGTACTTGATTTGTGC  4911

Query  361   AAAGCATTGGATACAGAGCTGGGGGCGCGCTCGCGCGAGGCGCGTGTTGCGGTCACGGTG  420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4912  AAAGCATTGGATACAGAGCTGGGGGCGCGCTCGCGCGAGGCGCGTGTTGCGGTCACGGTG  4971

Query  421   CAGAAGGAACGAATCCATCCGCGCAAGTCTTATCGCATGCACTGGAGCAGCGCGGATATT  480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4972  CAGAAGGAACGAATCCATCCGCGCAAGTCTTATCGCATGCACTGGAGCAGCGCGGATATT  5031

Query  481   AGAATCGATTTCGAGCAACCGATTCTCGTGTCCGGTGACGCCTTGCCCGTGGACGAGGCC  540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5032  AGAATCGATTTCGAGCAACCGATTCTCGTGTCCGGTGACGCCTTGCCCGTGGACGAGGCC  5091

Query  541   GTCGCCTTCGAATCCAATGCGGGCGGTCAAGGGCGATCGTCGCTGCGGCCGCCAGGTAGC  600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5092  GTCGCCTTCGAATCCAATGCGGGCGGTCAAGGGCGATCGTCGCTGCGGCCGCCAGGTAGC  5151

Query  601   GACGCGGCAGCCAAAATCTATTGGATTGATGTCCAGCTCCGTTGGGGCGACTACGACGAC  660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5152  GACGCGGCAGCCAAAATCTATTGGATTGATGTCCAGCTCCGTTGGGGCGACTACGACGAC  5211

Query  661   CATGATGCCGCGCAATACGCAGCGCAAAAATTTCGGGCGTATACAGCACCGGGAGCGTCG  720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5212  CATGATGCCGCGCAATACGCAGCGCAAAAATTTCGGGCGTATACAGCACCGGGAGCGTCG  5271

Query  721   AGTCTGTACCCGTCTCACCATGGACTGGTCGTGGTCTTTGATCTA  765
              |||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5272  AGTCTGTACCCGTCTCACCATGGACTGGTCGTGGTCTTTGATCTA  5316


Query  25    CATGATGCCGCGCAATACGCAGCGCAAAAATTTCGGGCGTATACAGCACCGGGAGCGTCG  84
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5212  CATGATGCCGCGCAATACGCAGCGCAAAAATTTCGGGCGTATACAGCACCGGGAGCGTCG  5271

Query  85    AGTCTGTACCCGTCTCACCATGGACTGGTCGTGGTCTTTGATCTAGCCTACGGCGAGTGG  144
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5272  AGTCTGTACCCGTCTCACCATGGACTGGTCGTGGTCTTTGATCTAGCCTACGGCGAGTGG  5331

Query  145   TCTGCTTTTGGGCACGCCTGCGCCATGGGCGTCGTGTCGATCGTGACCGAGGAATCGCGC  204
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5332  TCTGCTTTTGGGCACGCCTGCGCCATGGGCGTCGTGTCGATCGTGACCGAGGAATCGCGC  5391

Query  205   CGGATGCTTATGCATAACCAGGCACTCGCGCTCCTTCGTGAACGCATTCGCAAAGCGCTT  264
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5392  CGGATGCTTATGCATAACCAGGCACTCGCGCTCCTTCGTGAACGCATTCGCAAAGCGCTT  5451

Query  265   CAGCTCTACGTCATGGAGACGGTGGAAAGCGAGACGCTGCAGGCGGCGACGACTACGACA  324
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5452  CAGCTCTACGTCATGGAGACGGTGGAAAGCGAGACGCTGCAGGCGGCGACGACTACGACA  5511

Query  325   TCGGCTTCCGATACGATACCTCTGCTCGTGGGGTGCGGCGGCGACTTGTGGCGGCAACGC  384
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5512  TCGGCTTCCGATACGATACCTCTGCTCGTGGGGTGCGGCGGCGACTTGTGGCGGCAACGC  5571

Query  385   CTCTGGATCGTGGATGATCGCACTGCGTACAGGCCACATGCGAACGGTGTTATCTGGATA  444
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5572  CTCTGGATCGTGGATGATCGCACTGCGTACAGGCCACATGCGAACGGTGTTATCTGGATA  5631

Query  445   TGGGAGACGTCGACAGGGCGACTGTTTGTGAAGATTGTCCATCGGACTACGTGGGCTGGC  504
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5632  TGGGAGACGTCGACAGGGCGACTGTTTGTGAAGATTGTCCATCGGACTACGTGGGCTGGC  5691

Query  505   CAAACCCGGCGAGCGCAACTCGCCAAGTGGAAATGCGCTGAGCACGTTTTAACCATGCTC  564
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5692  CAAACCCGGCGAGCGCAACTCGCCAAGTGGAAATGCGCTGAGCACGTTTTAACCATGCTC  5751

Query  565   CGTTCACAGCCAACTGAAGAGCTACCGCGGGGCATCGTGCTCGCACAAACCGCATCCATG  624
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5752  CGTTCACAGCCAACTGAAGAGCTACCGCGGGGCATCGTGCTCGCACAAACCGCATCCATG  5811

Query  625   GACCCGTTGAAGACGCTCTTGGCAGGCACCGAGTATGCCAAAATTCCTGTGCGTGCCGGT  684
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5812  GACCCGTTGAAGACGCTCTTGGCAGGCACCGAGTATGCCAAAATTCCTGTGCGTGCCGGT  5871

Query  685   GCGGCGGCCATGCCG  699
              |||||||||||||||
```

```
Sbjct  5872  GCGGCGGCCATGCCG  5886


Query  60    AGCCAACTGAAGAGCTACCGCGGGGCATCGTGCTCGCACAAACCGCATCCATGGACCCGT  119
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5759  AGCCAACTGAAGAGCTACCGCGGGGCATCGTGCTCGCACAAACCGCATCCATGGACCCGT  5818

Query  120   TGAAGACGCTCTTGGCAGGCACCGAGTATGCCAAAATTCCTGTGCGTGCCGGTGCGGCGG  179
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5819  TGAAGACGCTCTTGGCAGGCACCGAGTATGCCAAAATTCCTGTGCGTGCCGGTGCGGCGG  5878

Query  180   CCATGCCGCTGCAGGCGCTAATGGCGCTGCCGGAGATCCGCGACCGTACTCAGACTGCGC  239
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5879  CCATGCCGCTGCAGGCGCTAATGGCGCTGCCGGAGATCCGCGACCGTACTCAGACTGCGC  5938

Query  240   GCTCCAGCGAGCTTTCGATCTGGAGTGGCTACGCGGATTGGCTTGAGCATGTGCCCGTGT  299
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5939  GCTCCAGCGAGCTTTCGATCTGGAGTGGCTACGCGGATTGGCTTGAGCATGTGCCCGTGT  5998

Query  300   GGATCGCGTCGGCGCGCTTCCTGCTCCTGCTCCACGCCTTGGACCGGGCGCCAGAGCGTG  359
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5999  GGATCGCGTCGGCGCGCTTCCTGCTCCTGCTCCACGCCTTGGACCGGGCGCCAGAGCGTG  6058

Query  360   TCCTGCAGCTGGTGTGGCCTCAGCGGTCGGCGGACGAGGAGAGCGCGGGCTCCGCGACAC  419
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6059  TCCTGCAGCTGGTGTGGCCTCAGCGGTCGGCGGACGAGGAGAGCGCGGGCTCCGCGACAC  6118

Query  420   CTTGGCTGTGGCCCGCGCTTCCCGAGACTGACTGGCGCCGTCTGGAACTAGAGCTCCAGT  479
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6119  CTTGGCTGTGGCCCGCGCTTCCCGAGACTGACTGGCGCCGTCTGGAACTAGAGCTCCAGT  6178

Query  480   CGCTGGTGCCCGTGCGTCTACGCCCTGCCCATGTCGCTGGTGATCAGCCAGGTGGCAGGG  539
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6179  CGCTGGTGCCCGTGCGTCTACGCCCTGCCCATGTCGCTGGTGATCAGCCAGGTGGCAGGG  6238

Query  540   ATGACGACGGAACGGAGCAGGTGCTGGAGCACACACGACCGAAGACGGTTGCCGCATTCG  599
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6239  ATGACGACGGAACGGAGCAGGTGCTGGAGCACACACGACCGAAGACGGTTGCCGCATTCG  6298

Query  600   ACCGGTATGGCAATGTGATTTCGGTGGAGACCACCACGCCATTTGAGCGGCAGGAGTACC  659
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6299  ACCGGTATGGCAATGTGATTTCGGTGGAGACCACCACGCCATTTGAGCGGCAGGAGTACC  6358

Query  660   GCACGTCGTCGGTCACGGTAACGAATGCAGAACAGCAACGGCTACTGTCCTTGTACCGAC  719
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6359  GCACGTCGTCGGTCACGGTAACGAATGCAGAACAGCAACGGCTACTGTCCTTGTACCGAC  6418

Query  720   GACTCCCAGACGTTCTCGAAAACCTGCACGTTGTCGAGCCTGGACACACCAGTGGCAGTC  779
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6419  GACTCCCAGACGTTCTCGAAAACCTGCACGTTGTCGAGCCTGGACACACCAGTGGCAGTC  6478

Query  780   TGTACGGAGGCGAGCGCCCGGAAC  803
             |||||||||||||||||||||||||
Sbjct  6479  TGTACGGAGGCGAGCGCCCGGAAC  6502


Query  1     CTCGAAAACCTGCACGTTGTCGAGCCTGGACACACCAGTGGCAGTCTGTACGGAGGCGAG  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6433  CTCGAAAACCTGCACGTTGTCGAGCCTGGACACACCAGTGGCAGTCTGTACGGAGGCGAG  6492

Query  61    CGCCCGGAACAGGCGACGATCTCGGCGATGCCGCATCTCGCCAAGCGATTCGAGCTCCAT  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6493  CGCCCGGAACAGGCGACGATCTCGGCGATGCCGCATCTCGCCAAGCGATTCGAGCTCCAT  6552

Query  121   CTACCGAGAGGCCTCGTCCGCGGTCTTTTGCTCGCGGGTATGACTTGTGGCCAACTCTGG  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6553  CTACCGAGAGGCCTCGTCCGCGGTCTTTTGCTCGCGGGTATGACTTGTGGCCAACTCTGG  6612

Query  181   GGCACGCGCACCGCCAACGCGGATACAGTGGTTTGGATCGCCTGGGCGCTGCGTCATCAT  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6613  GGCACGCGCACCGCCAACGCGGATACAGTGGTTTGGATCGCCTGGGCGCTGCGTCATCAT  6672

Query  241   GACGCCGAGGCATCGACACCGTTGCCGAGCATGGTTGAGGCAGCGGCCTGCTTTCGCCCT  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6673  GACGCCGAGGCATCGACACCGTTGCCGAGCATGGTTGAGGCAGCGGCCTGCTTTCGCCCT  6732

Query  301   GGCTCGGTGCAAGCGATCGGCTGGATTCAAGCGTGCCTCGATGTGGGTCGGCCACCAGCG  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6733  GGCTCGGTGCAAGCGATCGGCTGGATTCAAGCGTGCCTCGATGTGGGTCGGCCACCAGCG  6792

Query  361   CCAGGACCGGGGGGTGCATGCGCCCCACGATGTGCAGGTCATACTCTATCTCCAGGGCGAA  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6793  CCAGGACCGGGGGTGCATGCGCCCCACGATGTGCAGGTCATACTCTATCTCCAGGGCGAA  6852

Query  421   GAACGCCAAATCGATTTGGCATCGCCGTCGCTCGATCGCCACGCAGCCGAAGCAGTGCAG  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6853  GAACGCCAAATCGATTTGGCATCGCCGTCGCTCGATCGCCACGCAGCCGAAGCAGTGCAG  6912

Query  481   ACACACGATGTGCACGCATCGCGGGCACCGAGTGCGCACGCATTCGGACACGTCATCGGG  540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Sbjct  6913  ACACACGATGTGCACGCATCGCGGGCACCGAGTGCGCACGCATTCGGACACGTCATCGGG  6972

Query  541   AGCACGCATGCGATGTCACGCGCCCTGGGTGTCCTGCAGGCGTGGACTGGTGGCGGCTCC  600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6973  AGCACGCATGCGATGTCACGCGCCCTGGGTGTCCTGCAGGCGTGGACTGGTGGCGGCTCC  7032

Query  601   GGCGACGTGAACGATCCTCAATCCGTGCATGTTCGCGTTCGACTGCGTGATGATTTGAGT  660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  7033  GGCGACGTGAACGATCCTCAATCCGTGCATGTTCGCGTTCGACTGCGTGATGATTTGAGT  7092

Query  661   GCTGTTTTCCTCTACGCAAAAGATGGTCGGCTGATTGCGCGCCCGCCGGGACGTTTATTC  720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  7093  GCTGTTTTCCTCTACGCAAAAGATGGTCGGCTGATTGCGCGCCCGCCGGGACGTTTATTC  7152

Query  721   GAAACGATACCGGCACCGATCGAAGAGGGAACTTAG  756
              ||||||||||||||||||||||||||||||||||||
Sbjct  7153  GAAACGATACCGGCACCGATCGAAGAGGGAACTTAG  7188
```

Brr2 gene inserted into pQLinkN (The RBS sequence is in blue and the LIC sequence is in bold; pSR656)

```
Query  30   TACTTCCAATCCCACGAGGAGAAATTAACT 60
            ||||||||||||||||||||||||||||||
Sbjct        ------------------------------

Query  61   ATGCCTCAGGAACCTGAACTAGCAAGTATCGAGGTGTCGCCCCCGACTCCCGGCGAAAGC  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGCCTCAGGAACCTGAACTAGCAAGTATCGAGGTGTCGCCCCCGACTCCCGGCGAAAGC  60

Query  121  TCACATTTCATGGAGGCTATACTCCAAGACACTCCAGAATACCGCGTTATATGTGCTGCA  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   TCACATTTCATGGAGGCTATACTCCAAGACACTCCAGAATACCGCGTTATATGTGCTGCA  120

Query  181  GCCGTGCGACGAAAGGCGGCTTTGCCTGTTGTTAGCGCCGACTTGTTCAGCGGTACATGC  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  GCCGTGCGACGAAAGGCGGCTTTGCCTGTTGTTAGCGCCGACTTGTTCAGCGGTACATGC  180

Query  241  TCTGAACAGTCTGCGGCGCGGAGCGCTCTTGGTCAGGTCGGTGACGCTTCCGAGTGGAGC  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  TCTGAACAGTCTGCGGCGCGGAGCGCTCTTGGTCAGGTCGGTGACGCTTCCGAGTGGAGC  240

Query  301  GAAATGATCATTCAAAGGCCGGAGACAACTGAAAGTGCGCTATTGAATGATCGTTCATTT  360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GAAATGATCATTCAAAGGCCGGAGACAACTGAAAGTGCGCTATTGAATGATCGTTCATTT  300

Query  361  GCTGTCGATGATCTAGCGGGTTACATGAAACCTGCCTTTCAGCATATAGCAAGCCTGAAT  420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  GCTGTCGATGATCTAGCGGGTTACATGAAACCTGCCTTTCAGCATATAGCAAGCCTGAAT  360

Query  421  CCTGTCCAGTCCAGCGTTCTTGCTCGGGCGTTACGGATGCATGGAAACGTTCTGGTCTGC  480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CCTGTCCAGTCCAGCGTTCTTGCTCGGGCGTTACGGATGCATGGAAACGTTCTGGTCTGC  420

Query  481  GCACCAACAGGGTCCGGCAAGACGGATATTGCCGTTGCCCTGATTCTTCGCACCCTATTC  540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GCACCAACAGGGTCCGGCAAGACGGATATTGCCGTTGCCCTGATTCTTCGCACCCTATTC  480

Query  541  GAGGAATGCGGTGGCGAGCTGCAAGAGTTCAAATGTGTTTATATTGCTCCAATGCGGGCG  600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  481  GAGGAATGCGGTGGCGAGCTGCAAGAGTTCAAATGTGTTTATATTGCTCCAATGCGGGCG  540

Query  601  CTCGTGGGGGAGCTTCAGCGTTCGCTGAGCGCTCGCCTGCGAACCTATGGAATCTTGGTA  660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  541  CTCGTGGGGGAGCTTCAGCGTTCGCTGAGCGCTCGCCTGCGAACCTATGGAATCTTGGTA  600

Query  661  ACGGAGTGTACCGGGGAGCAGCGTCTGAGTCATCGCGACTTGTGGCGCTCCCATATCCTG  720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  601  ACGGAGTGTACCGGGGAGCAGCGTCTGAGTCATCGCGACTTGTGGCGCTCCCATATCCTG  660

Query  721  GTGACAACTCCTGAAAAATGGGATGTTCTCACTAGAAGAGCCAACGAACGTCCTTTGCTA  780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  661  GTGACAACTCCTGAAAAATGGGATGTTCTCACTAGAAGAGCCAACGAACGTCCTTTGCTA  720
```

```
Query  1    CCTTTGCTACGTTTTCTGAAGATTGTCATCATTGATGAGATCCACGTTCTCGATGATCCA  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  712  CCTTTGCTACGTTTTCTGAAGATTGTCATCATTGATGAGATCCACGTTCTCGATGATCCA  771

Query  61   AAGAGAGGGCCCGTCCTTGAGAGATGCGTAGCTCGTCTACACCATGAAACAGCGTTATTC  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  772  AAGAGAGGGCCCGTCCTTGAGAGATGCGTAGCTCGTCTACACCATGAAACAGCGTTATTC  831
```

```
Query  121   GGCTATCGAGTACGGCTTGTTGGCCTGAGCGCAACGCTTCCAAACTACGATGACGTGGCG  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  832   GGCTATCGAGTACGGCTTGTTGGCCTGAGCGCAACGCTTCCAAACTACGATGACGTGGCG  891

Query  181   GTCTTCATTCGCGCCAGGCTTCACGAGGGCGTGTTCGTCTTCTCTGAGAGAGAACGTCCA  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  892   GTCTTCATTCGCGCCAGGCTTCACGAGGGCGTGTTCGTCTTCTCTGAGAGAGAACGTCCA  951

Query  241   TGCCCCCTGGAGCTGCGCCTGGTGGCACTTCGCCATCGCGTTGGCCTTCTAAGTAAGACC  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  952   TGCCCCCTGGAGCTGCGCCTGGTGGCACTTCGCCATCGCGTTGGCCTTCTAAGTAAGACC  1011

Query  301   CGGTTCGATCAACTTATGAACCATGCTTTGTGGTTCCAGGTCAAAAAATTTGCCGTACAG  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1012  CGGTTCGATCAACTTATGAACCATGCTTTGTGGTTCCAGGTCAAAAAATTTGCCGTACAG  1071

Query  361   CAGATGGAGCAGGTTTTGGTTTTCGTTCACGAGCGAGCAGATACGCTGAAGACCGCGCAC  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1072  CAGATGGAGCAGGTTTTGGTTTTCGTTCACGAGCGAGCAGATACGCTGAAGACCGCGCAC  1131

Query  421   TGGTTGCTTAACGCTGCAGACGGCGAAAACGCTCCCATCCTACAAAGGCGGGAAACGGAA  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1132  TGGTTGCTTAACGCTGCAGACGGCGAAAACGCTCCCATCCTACAAAGGCGGGAAACGGAA  1191

Query  481   TGGCTTTCGCCTTCAATCCGTGAGCATTTCAAAAACTTTGGTGAGATTCTCGCCACGGTG  540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1192  TGGCTTTCGCCTTCAATCCGTGAGCATTTCAAAAACTTTGGTGAGATTCTCGCCACGGTG  1251

Query  541   AAGAAAGGCATTGGAGTCCATCACGCCGGTTTACCGCGAGAAATCCGCCATCTCATGGAG  600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1252  AAGAAAGGCATTGGAGTCCATCACGCCGGTTTACCGCGAGAAATCCGCCATCTCATGGAG  1311

Query  601   CAATTGTTTCTGCAGGGCAGCCTCCGTATTATGATATCTACCGCAACGCTAGCATGGGGA  660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1312  CAATTGTTTCTGCAGGGCAGCCTCCGTATTATGATATCTACCGCAACGCTAGCATGGGGA  1371

Query  661   GTCAATCTGCCGGCGAACGTGGTCGTCATCAAAGGCACGCAGTACTACGACAGTGAGGAG  720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1372  GTCAATCTGCCGGCGAACGTGGTCGTCATCAAAGGCACGCAGTACTACGACAGTGAGGAG  1431

Query  721   GGTCAGACTGTCC  733
             |||||||||||||
Sbjct  1432  GGTCAGACTGTCC  1444
```

```
Query  1     GTTTCTGCAGGGCAGCCTCCGTATTATGATATCTACCGCAACGCTAGCATGGGGAGTCAA  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1317  GTTTCTGCAGGGCAGCCTCCGTATTATGATATCTACCGCAACGCTAGCATGGGGAGTCAA  1376

Query  61    TCTGCCGGCGAACGTGGTCGTCATCAAAGGCACGCAGTACTACGACAGTGAGGAGGGTCA  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1377  TCTGCCGGCGAACGTGGTCGTCATCAAAGGCACGCAGTACTACGACAGTGAGGAGGGTCA  1436

Query  121   GACTGTCCAACTCGCCCCGCTGCACGTCCTGCAGATGCTCGGCAGAGCGGGCCGCTACCC  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1437  GACTGTCCAACTCGCCCCGCTGCACGTCCTGCAGATGCTCGGCAGAGCGGGCCGCTACCC  1496

Query  181   CTTCCACCAGCGTGGTGTAGGCGTGATCATCACAACCGAACCGGAAGCGCCGCTGTATGC  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1497  CTTCCACCAGCGTGGTGTAGGCGTGATCATCACAACCGAACCGGAAGCGCCGCTGTATGC  1556

Query  241   TGCTGTCGTAGCGCACAAGGCCCCCATAGAATCGCATCTCATACCGCAGTTGGCAGATAG  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1557  TGCTGTCGTAGCGCACAAGGCCCCCATAGAATCGCATCTCATACCGCAGTTGGCAGATAG  1616

Query  301   CTTGTTGGCCGAAGTTGCGGGCGGCTCCCTCAGCACCGTCGAAGAGGCAGCAGAGTGGCT  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1617  CTTGTTGGCCGAAGTTGCGGGCGGCTCCCTCAGCACCGTCGAAGAGGCAGCAGAGTGGCT  1676

Query  361   CAAATACACATTCCTCTTCGTTCGAATGCTGCGGAATCCATCTCTCGAGTGGATGCCGCA  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1677  CAAATACACATTCCTCTTCGTTCGAATGCTGCGGAATCCATCTCTCGAGTGGATGCCGCA  1736

Query  421   TTTCGCGAGCAAGCGGCCTGCCGATGGAGAAGATGTTTCCCTCTGGGGCGTTCGGCTCCG  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1737  TTTCGCGAGCAAGCGGCCTGCCGATGGAGAAGATGTTTCCCTCTGGGGCGTTCGGCTCCG  1796

Query  481   ACTTTGCCATTCGGTGGCGAAAGAGCTCGCCCGGAATGAGCTGTTGCGCTACGGCGAAAA  540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1797  ACTTTGCCATTCGGTGGCGAAAGAGCTCGCCCGGAATGAGCTGTTGCGCTACGGCGAAAA  1856

Query  541   TCTAGAGATGACTGTAACAGCTCGCGGTAACGTGGCTTCTGCGTTTATGCTTCCCTACGA  600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1857  TCTAGAGATGACTGTAACAGCTCGCGGTAACGTGGCTTCTGCGTTTATGCTTCCCTACGA  1916

Query  601   CACGCTGCGGACAATAGAGGCATACCTTCACCCAACAGGCGCTCTCCCAGAGTTGATCCA  660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1917  CACGCTGCGGACAATAGAGGCATACCTTCACCCAACAGGCGCTCTCCCAGAGTTGATCCA  1976

Query  661   TCTGCTCGCAGTTGCCTCACCTGCACTTCGGAACCTCAGTCTACCACGGGAAAGCGAAAG  720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1977  TCTGCTCGCAGTTGCCTCACCTGCACTTCGGAACCTCAGTCTACCACGGGAAAGCGAAAG  2036
```

```
Query  721   CCGAGAGTTGCGCA  734
             ||||||||||||||
Sbjct  2037  CCGAGAGTTGCGCA  2050


Query  1     CCATCTGCTCGCAGTTGCCTCACCTGCACTTCGGAACCTCAGTCTACCACGGGAAAGCGA  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1974  CCATCTGCTCGCAGTTGCCTCACCTGCACTTCGGAACCTCAGTCTACCACGGGAAAGCGA  2033

Query  61    AAGCCGAGAGTTGCGCAGGTTCTCTTGGCGCCTACTGATACCGCTTTGGGATCCTGACAG  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2034  AAGCCGAGAGTTGCGCAGGTTCTCTTGGCGCCTACTGATACCGCTTTGGGATCCTGACAG  2093

Query  121   AGACCTCAAAATCTCGGCATTGTTGCAGGCCGCAGTCCTGCCAACGGGCCTAAGCTCTTC  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2094  AGACCTCAAAATCTCGGCATTGTTGCAGGCCGCAGTCCTGCCAACGGGCCTAAGCTCTTC  2153

Query  181   GCCCGCATTATCGCAGGACTGCTCTTCGATACTGGAAGAGTCTCAGCGGATGCTGCGTGC  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2154  GCCCGCATTATCGCAGGACTGCTCTTCGATACTGGAAGAGTCTCAGCGGATGCTGCGTGC  2213

Query  241   GGTGCATGCGCTTTCTGCAACCTTGGGAATGGCTGTACCCATGCGTTTGAGCTTGCTACT  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2214  GGTGCATGCGCTTTCTGCAACCTTGGGAATGGCTGTACCCATGCGTTTGAGCTTGCTACT  2273

Query  301   CGCCAAGAAACTTGAGCATCAACAAACGCGTATTCTGCAAGGCGCCCTCAGGGGCGGAAG  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2274  CGCCAAGAAACTTGAGCATCAACAAACGCGTATTCTGCAAGGCGCCCTCAGGGGCGGAAG  2333

Query  361   CAGCAACGATCAGAGCGGTAGCACAAGCTCCGGCATGCAACCCCGTAAGACGCACAACAG  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2334  CAGCAACGATCAGAGCGGTAGCACAAGCTCCGGCATGCAACCCCGTAAGACGCACAACAG  2393

Query  421   CAAAGGCCAGCCAGAGCAACACAGCGAAACCGAGGCGCTGCGTTGCTTTTCGCCTTCAGT  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2394  CAAAGGCCAGCCAGAGCAACACAGCGAAACCGAGGCGCTGCGTTGCTTTTCGCCTTCAGT  2453

Query  481   CGGCGACTGGATGCCTCCCGCGGTGTTATGCAGCGCTCTGGGGCGTCAGCAGTGTGCCGT  540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2454  CGGCGACTGGATGCCTCCCGCGGTGTTATGCAGCGCTCTGGGGCGTCAGCAGTGTGCCGT  2513

Query  541   TCGTTCCTGGATAGCCTTCGAGAGCGCGCTTCTACCCGTTTCAGCAGACACGTTGCGTTT  600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2514  TCGTTCCTGGATAGCCTTCGAGAGCGCGCTTCTACCCGTTTCAGCAGACACGTTGCGTTT  2573

Query  601   CGAATGTCGCATCGCGCTGCAGGATAGCTCCTGTATCGACAGTACGGAAGCGCTCTGGGT  660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2574  CGAATGTCGCATCGCGCTGCAGGATAGCTCCTGTATCGACAGTACGGAAGCGCTCTGGGT  2633

Query  661   ATCCCTCGAGGANGCCACTGGTGAAAGANCCTTTTTCGCTGCTCGCCTGGNCCTCCGTGA  720
             |||||||||||| |||||||||||||| |||||||||||||||||||| |||||||||
Sbjct  2634  ATCCCTCGAGGATGCCACTGGTGAAAGAGCCTTTTTCGCTGCTCGCCTGGTCCTCCGTGA  2693

Query  721   AGCGAAATATCACCTTGTCGGATGCTGTTCGGTTCCACCGTGGCAGCGAACCTCAGTCGC  780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2694  AGCGAAATATCACCTTGTCGGATGCTGTTCGGTTCCACCGTGGCAGCGAACCTCAGTCGC  2753


Query  1     CTTTTTCGCTGCTCGCCTGGTCCTCCGTGAAGCGAAATATCACCTTGTCGGATGCTGTTC  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2664  CTTTTTCGCTGCTCGCCTGGTCCTCCGTGAAGCGAAATATCACCTTGTCGGATGCTGTTC  2723

Query  61    GGTTCCACCGTGGCAGCGAACCTCAGTCGCTTTCTGGAGAATTGCCGCCGAAGAGCATCT  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2724  GGTTCCACCGTGGCAGCGAACCTCAGTCGCTTTCTGGAGAATTGCCGCCGAAGAGCATCT  2783

Query  121   GGTTGACGACTTTCTGCAAGCTGAGTGTTTGCGCGACCTCCCGTGGCCAGTTGACGATAA  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2784  GGTTGACGACTTTCTGCAAGCTGAGTGTTTGCGCGACCTCCCGTGGCCAGTTGACGATAA  2843

Query  181   CCGAATTGCCCCCGAGATCGGCCGCTTTGTATTTCACGACTTTAGTCGTCTAGCGGCGGC  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2844  CCGAATTGCCCCCGAGATCGGCCGCTTTGTATTTCACGACTTTAGTCGTCTAGCGGCGGC  2903

Query  241   CTGTGAGACCCTGATTCCAGGCTATCTCAAAATGGTGATGGACTGCGCTCCGGGACAGCG  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2904  CTGTGAGACCCTGATTCCAGGCTATCTCAAAATGGTGATGGACTGCGCTCCGGGACAGCG  2963

Query  301   ACCTGGGGTATTGTTTGCGATACCGTTCCGGCCCGAACGGATCCGGTGCCTCGCGGCTTG  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2964  ACCTGGGGTATTGTTTGCGATACCGTTCCGGCCCGAACGGATCCGGTGCCTCGCGGCTTG  3023

Query  361   CCTCACGGTAGCGCGTGAAAAACGCCTCGTGTGCTATGTGTCTCCACATCGCTCGCATCG  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3024  CCTCACGGTAGCGCGTGAAAAACGCCTCGTGTGCTATGTGTCTCCACATCGCTCGCATCG  3083

Query  421   TGAAGTTTTTCGCCAACTCGTACAAAGCTGCAAGTTGCGCTGTGCGGAGCTAAGCATTGA  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3084  TGAAGTTTTTCGCCAACTCGTACAAAGCTGCAAGTTGCGCTGTGCGGAGCTAAGCATTGA  3143

Query  481   TCCGCAGTGGCATAGAAAACTTTCATGCGCTGACGACGATGAGCACGAAGCGCGCTTCCT  540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3144  TCCGCAGTGGCATAGAAAACTTTCATGCGCTGACGACGATGAGCACGAAGCGCGCTTCCT  3203
```

```
Query  541   GATCGGTGACCCCAAACACTTTGCCTCGCTCGTGCATAGCATGCAGTCGGAACAGCTCCG  600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3204  GATCGGTGACCCCAAACACTTTGCCTCGCTCGTGCATAGCATGCAGTCGGAACAGCTCCG  3263

Query  601   TGTGCAAGCACCACTGTGGCTTCTGGACGACCTCGAGTCTGTATCATGGGATCCAAGCTA  660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3264  TGTGCAAGCACCACTGTGGCTTCTGGACGACCTCGAGTCTGTATCATGGGATCCAAGCTA  3323

Query  661   CGAGTTTTTGCTCATGAATCTGGCAAGAGCTGAGCAGGTTCTTTTTTCGGAGCATTTCCA  720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3324  CGAGTTTTTGCTCATGAATCTGGCAAGAGCTGAGCAGGTTCTTTTTTCGGAGCATTTCCA  3383

Query  721   GAATGTCGACCAAGTATTGCGCTGGCTTGGACTGGATCCGTCTCGAATGCTCC  773
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3384  GAATGTCGACCAAGTATTGCGCTGGCTTGGACTGGATCCGTCTCGAATGCTCC  3436


Query  1     GTGGCTTCTGGACGACCTCGAGTCTGTATCATGGGATCCAAGCTACGAGTTTTTGCTCAT  60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3279  GTGGCTTCTGGACGACCTCGAGTCTGTATCATGGGATCCAAGCTACGAGTTTTTGCTCAT  3338

Query  61    GAATCTGGCAAGAGCTGAGCAGGTTCTTTTTTCGGAGCATTTCCAGAATGTCGACCAAGT  120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3339  GAATCTGGCAAGAGCTGAGCAGGTTCTTTTTTCGGAGCATTTCCAGAATGTCGACCAAGT  3398

Query  121   ATTGCGCTGGCTTGGACTGGATCCGTCTCGAATGCTCCAGTTTGAGTGGCAACCTCCGGG  180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3399  ATTGCGCTGGCTTGGACTGGATCCGTCTCGAATGCTCCAGTTTGAGTGGCAACCTCCGGG  3458

Query  181   TGAGGCTTCGCCGCTCATTACCACTGCGGACGAGGCACTTCTGAACCGGCCTCGGCTGGT  240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3459  TGAGGCTTCGCCGCTCATTACCACTGCGGACGAGGCACTTCTGAACCGGCCTCGGCTGGT  3518

Query  241   GGCGAAGCGCTTCTGGCGTGCATTTCGTCGGATAGCGCCTCCTCGAGGACCACAATTCGT  300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3519  GGCGAAGCGCTTCTGGCGTGCATTTCGTCGGATAGCGCCTCCTCGAGGACCACAATTCGT  3578

Query  301   CGCTGTTTGCGCGGTCATGCGCGATGCGCTCTTACTTGGGGCAGAGCTAGCGCGACAGAG  360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3579  CGCTGTTTGCGCGGTCATGCGCGATGCGCTCTTACTTGGGGCAGAGCTAGCGCGACAGAG  3638

Query  361   TCCGTACCGCAGTGAGTGCCCTGCACGTTTGGCAGCGTCGTTAACCGCCGCGGAGCGAGC  420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3639  TCCGTACCGCAGTGAGTGCCCTGCACGTTTGGCAGCGTCGTTAACCGCCGCGGAGCGAGC  3698

Query  421   GCTCCTCGAGCGTGGCATCTGGTTACCCACGAGGAGTACCAGCTACGAGACAGGTGCACC  480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3699  GCTCCTCGAGCGTGGCATCTGGTTACCCACGAGGAGTACCAGCTACGAGACAGGTGCACC  3758

Query  481   CCCGGGGCACACCAATGTCCGCGTCGTAGTCATCGAGATCGGAGAATTACTCCTGCTCCC  540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3759  CCCGGGGCACACCAATGTCCGCGTCGTAGTCATCGAGATCGGAGAATTACTCCTGCTCCC  3818

Query  541   GAGCGATGGCAATGAATACGCCTTCGCATTCATGGTGGGCAACGGCTTTGCGCAGGCATC  600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3819  GAGCGATGGCAATGAATACGCCTTCGCATTCATGGTGGGCAACGGCTTTGCGCAGGCATC  3878

Query  601   GTCGCGCCCACGGGAGAACCTGTTCCGTTGGAGTTGTCGCTACTTTGCCAGGATATGCAG  660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3879  GTCGCGCCCACGGGAGAACCTGTTCCGTTGGAGTTGTCGCTACTTTGCCAGGATATGCAG  3938

Query  661   AGGTCCGGTTTCGGTCCTAGTGCCACCGCATCTTGAGGGACTGATGCAGCTGCCGGCGAA  720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3939  AGGTCCGGTTTCGGTCCTAGTGCCACCGCATCTTGAGGGACTGATGCAGCTGCCGGCGAA  3998

Query  721   TTGGCCTCCA  730
              ||||||||||
Sbjct  3999  TTGGCCTCCA  4008


Query  1     GTTGGAGTTGTCGCTACTTTGCCAGGATATGCAGAGGTCCGGTTTCGGTCCTAGTGCCAC  60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3905  GTTGGAGTTGTCGCTACTTTGCCAGGATATGCAGAGGTCCGGTTTCGGTCCTAGTGCCAC  3964

Query  61    CGCATCTTGAGGGACTGATGCAGCTGCCGGCGAATTGGCCTCCAGTCATAGACTCTGTCG  120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3965  CGCATCTTGAGGGACTGATGCAGCTGCCGGCGAATTGGCCTCCAGTCATAGACTCTGTCG  4024

Query  121   CTGATCTATTCTGGGAGTTATTTATTGTTCATCAGGTAGCACGGGGCACACTGCAGACTG  180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4025  CTGATCTATTCTGGGAGTTATTTATTGTTCATCAGGTAGCACGGGGCACACTGCAGACTG  4084

Query  181   CCGAGCAGGTTGCTGAGCTCATGAGCCGCACTCTCTACCGCTACCGTTACCTGCAGCATA  240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4085  CCGAGCAGGTTGCTGAGCTCATGAGCCGCACTCTCTACCGCTACCGTTACCTGCAGCATA  4144

Query  241   CGGCTGCGAGGCCGCTGTGCGAAGTGATATCGCTGGGCAGCGAGCTCGACAGAATAGCTC  300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4145  CGGCTGCGAGGCCGCTGTGCGAAGTGATATCGCTGGGCAGCGAGCTCGACAGAATAGCTC  4204

Query  301   TAATGACCAGACTCGTCACAGAAAAACTTATCAGCGAATGCAACTTGCTCCGAATTCGCA  360
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4205 TAATGACCAGACTCGTCACAGAAAAACTTATCAGCGAATGCAACTTGCTCCGAATTCGCA  4264

Query  361  ATGACGGGCGCTTTGATTGTCGCGATGTCGTCGCTCCCCGGCTGGCATTCTGTTTCGGTA  420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4265 ATGACGGGCGCTTTGATTGTCGCGATGTCGTCGCTCCCCGGCTGGCATTCTGTTTCGGTA  4324

Query  421  TCGAAGTGCAGACAGCGCGTACGATTTTAGATCGGTTTTCAGGTAGGGTGATAACCAGGA  480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4325 TCGAAGTGCAGACAGCGCGTACGATTTTAGATCGGTTTTCAGGTAGGGTGATAACCAGGA  4384

Query  481  GTGCGCTTTTGGACATGTTCGCTGAAACACTGGTTAGCGAACTCGGTCCTGGTTGGTTGC  540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4385 GTGCGCTTTTGGACATGTTCGCTGAAACACTGGTTAGCGAACTCGGTCCTGGTTGGTTGC  4444

Query  541  CGCGCGACCTCGCATCGAGCGATGAAGCTGGACGCCCTCATTCGCTGTGGCCCATTCGAG  600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4445 CGCGCGACCTCGCATCGAGCGATGAAGCTGGACGCCCTCATTCGCTGTGGCCCATTCGAG  4504

Query  601  GACCAAAAGCGCGTCCTAAGAAAGTGCAGCCCGTTCTCTTTGTACGCACACTGCTGATAT  660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4505 GACCAAAAGCGCGTCCTAAGAAAGTGCAGCCCGTTCTCTTTGTACGCACACTGCTGATAT  4564

Query  661  GGTTGCTGAGATCGCCACCGACGTCACAGTTTTGCTGGCGGAGGCTTTTGGAGGAGCTCA  720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4565 GGTTGCTGAGATCGCCACCGACGTCACAGTTTTGCTGGCGGAGGCTTTTGGAGGAGCTCA  4624

Query  721  GTGGCGCTTG  730
            ||||||||||
Sbjct  4625 GTGGCGCTTG  4634


Query  1    TGGTTGCTGAGATCGCCACCGACGTCACAGTTTTGCTGGCGGAGGCTTTTGGAGGAGCTC  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4564 TGGTTGCTGAGATCGCCACCGACGTCACAGTTTTGCTGGCGGAGGCTTTTGGAGGAGCTC  4623

Query  61   AGTGGCGCTTGGGAGCGACTACTACCGGCACTGGTGTTTACGGGTCTCGCCCGTCTTTGT  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4624 AGTGGCGCTTGGGAGCGACTACTACCGGCACTGGTGTTTACGGGTCTCGCCCGTCTTTGT  4683

Query  121  TGGAGCGCTTTCCGCCAGTTCCACAAGATCACTCGAACGCTTTCGTTCGAGCTTATGGTT  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4684 TGGAGCGCTTTCCGCCAGTTCCACAAGATCACTCGAACGCTTTCGTTCGAGCTTATGGTT  4743

Query  181  AGTGGCTCGCCGTTGCGATCACGAGAGCAGAGCCACCTCACTAAGCCTGCAGACGATTTG  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4744 AGTGGCTCGCCGTTGCGATCACGAGAGCAGAGCCACCTCACTAAGCCTGCAGACGATTTG  4803

Query  241  GATGATGCTGTTGGCGAGGAGCCCGATGGCCGCATAGCACAGATTGAGCATCACATGGGT  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4804 GATGATGCTGTTGGCGAGGAGCCCGATGGCCGCATAGCACAGATTGAGCATCACATGGGT  4863

Query  301  GCCGAATACCGGGAGCGCTGTACGAGTTACCTGCATCAGATGGATCACgagcagcatgag  360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4864 GCCGAATACCGGGAGCGCTGTACGAGTTACCTGCATCAGATGGATCACGAGCAGCATGAG  4923

Query  361  cagcatgagcagcatgagcagcatgagcGCCAAGCGGATCTCGGAAACGAGCCGCACCGG  420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4924 CAGCATGAGCAGCATGAGCAGCATGAGCGCCAAGCGGATCTCGGAAACGAGCCGCACCGG  4983

Query  421  ATAGAACGTCGCGCACGTGGACGTCACCTGCAAGCAAAACTGGCGGCAGCGGCGGGGATG  480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4984 ATAGAACGTCGCGCACGTGGACGTCACCTGCAAGCAAAACTGGCGGCAGCGGCGGGGATG  5043

Query  481  GTGGATTCTGGCGGGAGCATTCAATCCAGAAACGATGCTTGCTTTCATCTTCATCCTCGG  540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5044 GTGGATTCTGGCGGGAGCATTCAATCCAGAAACGATGCTTGCTTTCATCTTCATCCTCGG  5103

Query  541  CAAGGAGGTTCCCTTCGGGTCCAGCAAGCAATACAGCAAGGGACAAGACGGAAACGCGTC  600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5104 CAAGGAGGTTCCCTTCGGGTCCAGCAAGCAATACAGCAAGGGACAAGACGGAAACGCGTC  5163

Query  601  ATAGACGTTCACTGTATTGACATAGCATTGGGGCAGCGAACCGAGAAACCCGTGGATGAG  660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5164 ATAGACGTTCACTGTATTGACATAGCATTGGGGCAGCGAACCGAGAAACCCGTGGATGAG  5223

Query  661  CATATGTACAGGAAGCGACATGCTTGGCTCGTC  693
            |||||||||||||||||||||||||||||||||
Sbjct  5224 CATATGTACAGGAAGCGACATGCTTGGCTCGTC  5256


Query  421  ACGCGTCATAGACGTTCACTGTATTGACATAGCATTGGGGCAGCGAACCGAGAAACCCGT  480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5157 ACGCGTCATAGACGTTCACTGTATTGACATAGCATTGGGGCAGCGAACCGAGAAACCCGT  5216

Query  481  GGATGAGCATATGTACAGGAAGCGACATGCTTGGCTCGTCGTCCGTGATGCCGGCTCCGA  540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5217 GGATGAGCATATGTACAGGAAGCGACATGCTTGGCTCGTCGTCCGTGATGCCGGCTCCGA  5276

Query  541  TACGGTTGTCTACGCGAACCGCGTTTTCTATCGATTTGGACTGCAAAGATTTTGCTTGAA  600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5277 TACGGTTGTCTACGCGAACCGCGTTTTCTATCGATTTGGACTGCAAAGATTTTGCTTGAA  5336
```

```
Query  601   GATCCCCGAGGTCGCCGCATGCAAGCCAACAAAACTCTCCGTTCATGCTTTCGACGAGTT  660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5337  GATCCCCGAGGTCGCCGCATGCAAGCCAACAAAACTCTCCGTTCATGCTTTCGACGAGTT  5396

Query  661   CAGCACGGACGAGGAAATGGAGCGCTGCGTGTTGGTAACATCTGCACAAACGACCGCGGA  720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5397  CAGCACGGACGAGGAAATGGAGCGCTGCGTGTTGGTAACATCTGCACAAACGACCGCGGA  5456

Query  721   TCCGAGTATCTGA  733
             |||||||||||||
Sbjct  5457  TCCGAGTATCTGA  5469
```

Brr2 gene inserted into pPiczA (Underlined is the KpnI and NotI restriction sites, and highlighted in green is the kosak consensus sequence; pSR855)

```
Query  62    CGGATCGGTACCGCCATGGTGCCTCAGGAACCTGAACTAGAA  92
             ||||||||||||||||||||||||||||||||||||||||||
Sbjct         -------------------------------------------

Query  93    TGCCTCAGGAACCTGAACTAGCAAGTATCGAGGTGTCGCCCCCGACTCCCGGCGAAAGCT  152
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2     TGCCTCAGGAACCTGAACTAGCAAGTATCGAGGTGTCGCCCCCGACTCCCGGCGAAAGCT  61

Query  153   CACATTTCATGGAGGCTATACTCCAAGACACTCCAGAATACCGCGTTATATGTGCTGCAG  212
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  62    CACATTTCATGGAGGCTATACTCCAAGACACTCCAGAATACCGCGTTATATGTGCTGCAG  121

Query  213   CCGTGCGACGAAAGGCGGCTTTGCCTGTTGTTAGCGCCGACTTGTTCAGCGGTACATGCT  272
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  122   CCGTGCGACGAAAGGCGGCTTTGCCTGTTGTTAGCGCCGACTTGTTCAGCGGTACATGCT  181

Query  273   CTGAACAGTCTGCGGCGCGGAGCGCTCTTGGTCAGGTCGGTGACGCTTCCGAGTGGAGCG  332
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  182   CTGAACAGTCTGCGGCGCGGAGCGCTCTTGGTCAGGTCGGTGACGCTTCCGAGTGGAGCG  241

Query  333   AAATGATCATTCAAAGGCCGGAGACAACTGAAAGTGCGCTATTGAATGATCGTTCATTTG  392
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  242   AAATGATCATTCAAAGGCCGGAGACAACTGAAAGTGCGCTATTGAATGATCGTTCATTTG  301

Query  393   CTGTCGATGATCTAGCGGGTTACATGAAACCTGCCTTTCAGCATATAGCAAGCCTGAATC  452
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  302   CTGTCGATGATCTAGCGGGTTACATGAAACCTGCCTTTCAGCATATAGCAAGCCTGAATC  361

Query  453   CTGTCCAGTCCAGCGTTCTTGCTCGGGCGTTACGGATGCATGGAAACGTTCTGGTCTGCG  512
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  362   CTGTCCAGTCCAGCGTTCTTGCTCGGGCGTTACGGATGCATGGAAACGTTCTGGTCTGCG  421

Query  513   CACCAACAGGGTCCGGCAAGACGGATATTGCCGTTGCCCTGATTCTTCGCACCCTATTCG  572
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  422   CACCAACAGGGTCCGGCAAGACGGATATTGCCGTTGCCCTGATTCTTCGCACCCTATTCG  481

Query  573   AGGAATGCGGTGGCGAGCTGCAAGAGTTCAAATGTGTTTATATTGCTCCAATGCGGGCGC  632
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  482   AGGAATGCGGTGGCGAGCTGCAAGAGTTCAAATGTGTTTATATTGCTCCAATGCGGGCGC  541

Query  633   TCGTGGGGGAGCTTCAGCGTTCGCTGAGCGCTCGCCTGCGAACCTATGGAATCTTGGTAA  692
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  542   TCGTGGGGGAGCTTCAGCGTTCGCTGAGCGCTCGCCTGCGAACCTATGGAATCTTGGTAA  601

Query  693   CGGAGTGTACCGGGGAGCAGCGTCTGAGTCATCGCGACTTGTGGCGCTCCCATATCCTGG  750
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  602   CGGAGTGTACCGGGGAGCAGCGTCTGAGTCATCGCGACTTGTGGCGCTCCCATATCCTGG  661

Query  751   TGACAACTCCTGAAAAATGGGATGTTCTCACTAGAAGAGCCAACGAACGTCCTTTGCTAC  809
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  662   TGACAACTCCTGAAAAATGGGATGTTCTCACTAGAAGAGCCAACGAACGTCCTTTGCTAC  721


Query  1     CGTCCTTTGCTACGTTTTCTGAAGATTGTCATCATTGATGAGATCCACGTTCTCGATGAT  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  709   CGTCCTTTGCTACGTTTTCTGAAGATTGTCATCATTGATGAGATCCACGTTCTCGATGAT  768

Query  61    CCAAAGAGAGGGCCCGTCCTTGAGAGATGCGTAGCTCGTCTACACCATGAAACAGCGTTA  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  769   CCAAAGAGAGGGCCCGTCCTTGAGAGATGCGTAGCTCGTCTACACCATGAAACAGCGTTA  828

Query  121   TTCGGCTATCGAGTACGGCTTGTTGGCCTGAGCGCAACGCTTCCAAACTACGATGACGTG  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  829   TTCGGCTATCGAGTACGGCTTGTTGGCCTGAGCGCAACGCTTCCAAACTACGATGACGTG  888

Query  181   GCGGTCTTCATTCGCGCCAGGCTTCACGAGGGCGTGTTCGTCTTCTCTGAGAGAGAACGT  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  889   GCGGTCTTCATTCGCGCCAGGCTTCACGAGGGCGTGTTCGTCTTCTCTGAGAGAGAACGT  948

Query  241   CCATGCCCCCTGGAGCTGCGCCTGGTGGCACTTCGCCATCGCGTTGGCCTTCTAAGTAAG  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Sbjct  949   CCATGCCCCCTGGAGCTGCGCCTGGTGGCACTTCGCCATCGCGTTGGCCTTCTAAGTAAG  1008

Query  301   ACCCGGTTCGATCAACTTATGAACCATGCTTTGTGGTTCCAGGTCAAAAAATTTGCCGTA  360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1009  ACCCGGTTCGATCAACTTATGAACCATGCTTTGTGGTTCCAGGTCAAAAAATTTGCCGTA  1068

Query  361   CAGCAGATGGAGCAGGTTTTGGTTTTCGTTCACGAGCGAGCAGATACGCTGAAGACCGCG  420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1069  CAGCAGATGGAGCAGGTTTTGGTTTTCGTTCACGAGCGAGCAGATACGCTGAAGACCGCG  1128

Query  421   CACTGGTTGCTTAACGCTGCAGACGGCGAAAACGCTCCCATCCTACAAAGGCGGGAAACG  480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1129  CACTGGTTGCTTAACGCTGCAGACGGCGAAAACGCTCCCATCCTACAAAGGCGGGAAACG  1188

Query  481   GAATGGCTTTCGCCTTCAATCCGTGAGCATTTCAAAAACTTTGGTGAGATTCTCGCCACG  540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1189  GAATGGCTTTCGCCTTCAATCCGTGAGCATTTCAAAAACTTTGGTGAGATTCTCGCCACG  1248

Query  541   GTGAAGAAAGGCATTGGAGTCCATCACGCCGGTTTACCGCGAGAAATCCGCCATCTCATG  600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1249  GTGAAGAAAGGCATTGGAGTCCATCACGCCGGTTTACCGCGAGAAATCCGCCATCTCATG  1308

Query  601   GAGCAATTGTTTCTGCAGGGCAGCCTCCGTATTATGATATCTACCGCAACGCTAGCATGG  660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1309  GAGCAATTGTTTCTGCAGGGCAGCCTCCGTATTATGATATCTACCGCAACGCTAGCATGG  1368

Query  661   GGAGTCAATCTGCCGGCGAACGTGGT  686
              |||||||||||||||||||||||||
Sbjct  1369  GGAGTCAATCTGCCGGCGAACGTGGT  1394
```

```
Query  1     CGGCGAACGTGGTCGTCATCAAAGGCACGCAGTACTACGACAGTGAGGAGGGTCAGACTG  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1382  CGGCGAACGTGGTCGTCATCAAAGGCACGCAGTACTACGACAGTGAGGAGGGTCAGACTG  1441

Query  61    TCCAACTCGCCCCGCTGCACGTCCTGCAGATGCTCGGCAGAGCGGGCCGCTACCCCTTCC  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1442  TCCAACTCGCCCCGCTGCACGTCCTGCAGATGCTCGGCAGAGCGGGCCGCTACCCCTTCC  1501

Query  121   ACCAGCGTGGTGTAGGCGTGATCATCACAACCGAACCGGAAGCGCCGCTGTATGCTGCTG  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1502  ACCAGCGTGGTGTAGGCGTGATCATCACAACCGAACCGGAAGCGCCGCTGTATGCTGCTG  1561

Query  181   TCGTAGCGCACAAGGCCCCCATAGAATCGCATCTCATACCGCAGTTGGCAGATAGCTTGT  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1562  TCGTAGCGCACAAGGCCCCCATAGAATCGCATCTCATACCGCAGTTGGCAGATAGCTTGT  1621

Query  241   TGGCCGAAGTTGCGGGCGGCTCCCTCAGCACCGTCGAAGAGGCAGCAGAGTGGCTCAAAT  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1622  TGGCCGAAGTTGCGGGCGGCTCCCTCAGCACCGTCGAAGAGGCAGCAGAGTGGCTCAAAT  1681

Query  301   ACACATTCCTCTTCGTTCGAATGCTGCGGAATCCATCTCTCGAGTGGATGCCGCATTTCG  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1682  ACACATTCCTCTTCGTTCGAATGCTGCGGAATCCATCTCTCGAGTGGATGCCGCATTTCG  1741

Query  361   CGAGCAAGCGGCCTGCCGATGGAGAAGATGTTTCCCTCTGGGGCGTTCGGCTCCGACTTT  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1742  CGAGCAAGCGGCCTGCCGATGGAGAAGATGTTTCCCTCTGGGGCGTTCGGCTCCGACTTT  1801

Query  421   GCCATTCGGTGGCGAAAGAGCTCGCCCGGAATGAGCTGTTGCGCTACGGCGAAAATCTAG  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1802  GCCATTCGGTGGCGAAAGAGCTCGCCCGGAATGAGCTGTTGCGCTACGGCGAAAATCTAG  1861

Query  481   AGATGACTGTAACAGCTCGCGGTAACGTGGCTTCTGCGTTTATGCTTCCCTACGACACGC  540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1862  AGATGACTGTAACAGCTCGCGGTAACGTGGCTTCTGCGTTTATGCTTCCCTACGACACGC  1921

Query  541   TGCGGACAATAGAGGCATACCTTCACCCAACAGGCGCTCTCCCAGAGTTGATCCATCTGC  600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1922  TGCGGACAATAGAGGCATACCTTCACCCAACAGGCGCTCTCCCAGAGTTGATCCATCTGC  1981

Query  601   TCGCAGTTGCCTCACCTGCACTTCGGAACCTCAGTCTACCACGGGAAAGCGAAAGCCGAG  660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1982  TCGCAGTTGCCTCACCTGCACTTCGGAACCTCAGTCTACCACGGGAAAGCGAAAGCCGAG  2041

Query  661   AGTTGCGCAGGTTCTCTTGGCGCCTACTGATACCGCTTTGGGATCCTGACAGA  713
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2042  AGTTGCGCAGGTTCTCTTGGCGCCTACTGATACCGCTTTGGGATCCTGACAGA  2094
```

```
Query  1     CCATCTGCTCGCAGTTGCCTCACCTGCACTTCGGAACCTCAGTCTACCACGGGAAAGCGA  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1974  CCATCTGCTCGCAGTTGCCTCACCTGCACTTCGGAACCTCAGTCTACCACGGGAAAGCGA  2033

Query  61    AAGCCGAGAGTTGCGCAGGTTCTCTTGGCGCCTACTGATACCGCTTTGGGATCCTGACAG  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2034  AAGCCGAGAGTTGCGCAGGTTCTCTTGGCGCCTACTGATACCGCTTTGGGATCCTGACAG  2093

Query  121   AGACCTCAAAATCTCGGCATTGTTGCAGGCCGCAGTCCTGCCAACGGGCCTAAGCTCTTC  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2094  AGACCTCAAAATCTCGGCATTGTTGCAGGCCGCAGTCCTGCCAACGGGCCTAAGCTCTTC  2153
```

```
Query  181   GCCCGCATTATCGCAGGACTGCTCTTCGATACTGGAAGAGTCTCAGCGGATGCTGCGTGC  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2154  GCCCGCATTATCGCAGGACTGCTCTTCGATACTGGAAGAGTCTCAGCGGATGCTGCGTGC  2213

Query  241   GGTGCATGCGCTTTCTGCAACCTTGGGAATGGCTGTACCCATGCGTTTGAGCTTGCTACT  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2214  GGTGCATGCGCTTTCTGCAACCTTGGGAATGGCTGTACCCATGCGTTTGAGCTTGCTACT  2273

Query  301   CGCCAAGAAACTTGAGCATCAACAAACGCGTATTCTGCAAGGCGCCCTCAGGGGCGGAAG  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2274  CGCCAAGAAACTTGAGCATCAACAAACGCGTATTCTGCAAGGCGCCCTCAGGGGCGGAAG  2333

Query  361   CAGCAACGATCAGAGCGGTAGCACAAGCTCCGGCATGCAACCCCGTAAGACGCACAACAG  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2334  CAGCAACGATCAGAGCGGTAGCACAAGCTCCGGCATGCAACCCCGTAAGACGCACAACAG  2393

Query  421   CAAAGGCCAGCCAGAGCAACACAGCGAAACCGAGGCGCTGCGTTGCTTTTCGCCTTCAGT  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2394  CAAAGGCCAGCCAGAGCAACACAGCGAAACCGAGGCGCTGCGTTGCTTTTCGCCTTCAGT  2453

Query  481   CGGCGACTGGATGCCTCCCGCGGTGTTATGCAGCGCTCTGGGGCGTCAGCAGTGTGCCGT  540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2454  CGGCGACTGGATGCCTCCCGCGGTGTTATGCAGCGCTCTGGGGCGTCAGCAGTGTGCCGT  2513

Query  541   TCGTTCCTGGATAGCCTTCGAGAGCGCGCTTCTACCCGTTTCAGCAGACACGTTGCGTTT  600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2514  TCGTTCCTGGATAGCCTTCGAGAGCGCGCTTCTACCCGTTTCAGCAGACACGTTGCGTTT  2573

Query  601   CGAATGTCGCATCGCGCTGCAGGATAGCTCCTGTATCGACAGTACGGAAGCGCTCTGGGT  660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2574  CGAATGTCGCATCGCGCTGCAGGATAGCTCCTGTATCGACAGTACGGAAGCGCTCTGGGT  2633

Query  661   ATCCCTCGAGGATGCCACTGGTGAAAGAGCCTTTTTCGCTGCTCGCCTGGTCCTCCGTGA  720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2634  ATCCCTCGAGGATGCCACTGGTGAAAGAGCCTTTTTCGCTGCTCGCCTGGTCCTCCGTGA  2693

Query  721   AGCGAAATATCACCTTGTCGGATGCTGTTCGGTTCCACCGTGGCAGCGAACCTCA  775
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2694  AGCGAAATATCACCTTGTCGGATGCTGTTCGGTTCCACCGTGGCAGCGAACCTCA  2748
```

```
Query  1     TTTTTCGCTGCTCGCCTGGTCCTCCGTGAAGCGAAATATCACCTTGTCGGATGCTGTTCG  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2665  TTTTTCGCTGCTCGCCTGGTCCTCCGTGAAGCGAAATATCACCTTGTCGGATGCTGTTCG  2724

Query  61    GTTCCACCGTGGCAGCGAACCTCAGTCGCTTTCTGGAGAATTGCCGCCGAAGAGCATCTG  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2725  GTTCCACCGTGGCAGCGAACCTCAGTCGCTTTCTGGAGAATTGCCGCCGAAGAGCATCTG  2784

Query  121   GTTGACGACTTTCTGCAAGCTGAGTGTTTGCGCGACCTCCCGTGGCCAGTTGACGATAAC  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2785  GTTGACGACTTTCTGCAAGCTGAGTGTTTGCGCGACCTCCCGTGGCCAGTTGACGATAAC  2844

Query  181   CGAATTGCCCCCGAGATCGGCCGCTTTGTATTTCACGACTTTAGTCGTCTAGCGGCGGCC  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2845  CGAATTGCCCCCGAGATCGGCCGCTTTGTATTTCACGACTTTAGTCGTCTAGCGGCGGCC  2904

Query  241   TGTGAGACCCTGATTCCAGGCTATCTCAAAATGGTGATGGACTGCGCTCCGGGACAGCGA  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2905  TGTGAGACCCTGATTCCAGGCTATCTCAAAATGGTGATGGACTGCGCTCCGGGACAGCGA  2964

Query  301   CCTGGGGTATTGTTTGCGATACCGTTCCGGCCCGAACGGATCCGGTGCCTCGCGGCTTGC  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2965  CCTGGGGTATTGTTTGCGATACCGTTCCGGCCCGAACGGATCCGGTGCCTCGCGGCTTGC  3024

Query  361   CTCACGGTAGCGCGTGAAAAACGCCTCGTGTGCTATGTGTCTCCACATCGCTCGCATCGT  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3025  CTCACGGTAGCGCGTGAAAAACGCCTCGTGTGCTATGTGTCTCCACATCGCTCGCATCGT  3084

Query  421   GAAGTTTTTCGCCAACTCGTACAAAGCTGCAAGTTGCGCTGTGCGGAGCTAAGCATTGAT  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3085  GAAGTTTTTCGCCAACTCGTACAAAGCTGCAAGTTGCGCTGTGCGGAGCTAAGCATTGAT  3144

Query  481   CCGCAGTGGCATAGAAAACTTTCATGCGCTGACGACGATGAGCACGAAGCGCGCTTCCTG  540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3145  CCGCAGTGGCATAGAAAACTTTCATGCGCTGACGACGATGAGCACGAAGCGCGCTTCCTG  3204

Query  541   ATCGGTGACCCCAAACACTTTGCCTCGCTCGTGCATAGCATGCAGTCGGAACAGCTCCGT  600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3205  ATCGGTGACCCCAAACACTTTGCCTCGCTCGTGCATAGCATGCAGTCGGAACAGCTCCGT  3264

Query  601   GTGCAAGCACCACTGTGGCTTCTGGACGACCTCGAGTCTGTATCATGGGATCCAAGCTAC  660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3265  GTGCAAGCACCACTGTGGCTTCTGGACGACCTCGAGTCTGTATCATGGGATCCAAGCTAC  3324

Query  661   GAGTTTTTGCTCATGAATCTGGCAAGAGCTGAGCAGGTTCTTTTTTC  707
             |||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3325  GAGTTTTTGCTCATGAATCTGGCAAGAGCTGAGCAGGTTCTTTTTTC  3371
```

```
Query  1     GACCTCGAGTCTGTATCATGGGATCCAAGCTACGAGTTTTTGCTCATGAATCTGGCAAGA  60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3292  GACCTCGAGTCTGTATCATGGGATCCAAGCTACGAGTTTTTGCTCATGAATCTGGCAAGA  3351

Query  61    GCTGAGCAGGTTCTTTTTTTCGGAGCATTTCCAGAATGTCGACCAAGTATTGCGCTGGCTT  120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3352  GCTGAGCAGGTTCTTTTTTTCGGAGCATTTCCAGAATGTCGACCAAGTATTGCGCTGGCTT  3411

Query  121   GGACTGGATCCGTCTCGAATGCTCCAGTTTGAGTGGCAACCTCCGGGTGAGGCTTCGCCG  180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3412  GGACTGGATCCGTCTCGAATGCTCCAGTTTGAGTGGCAACCTCCGGGTGAGGCTTCGCCG  3471

Query  181   CTCATTACCACTGCGGACGAGGCACTTCTGAACCGGCCTCGGCTGGTGGCGAAGCGCTTC  240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3472  CTCATTACCACTGCGGACGAGGCACTTCTGAACCGGCCTCGGCTGGTGGCGAAGCGCTTC  3531

Query  241   TGGCGTGCATTTCGTCGGATAGCGCCTCCTCGAGGACCACAATTCGTCGCTGTTTGCGCG  300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3532  TGGCGTGCATTTCGTCGGATAGCGCCTCCTCGAGGACCACAATTCGTCGCTGTTTGCGCG  3591

Query  301   GTCATGCGCGATGCGCTCTTACTTGGGGCAGAGCTAGCGCGACAGAGTCCGTACCGCAGT  360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3592  GTCATGCGCGATGCGCTCTTACTTGGGGCAGAGCTAGCGCGACAGAGTCCGTACCGCAGT  3651

Query  361   GAGTGCCCTGCACGTTTGGCAGCGTCGTTAACCGCCGCGGAGCGAGCGCTCCTCGAGCGT  420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3652  GAGTGCCCTGCACGTTTGGCAGCGTCGTTAACCGCCGCGGAGCGAGCGCTCCTCGAGCGT  3711

Query  421   GGCATCTGGTTACCCACGAGGAGTACCAGCTACGAGACAGGTGCACCCCCGGGGCACACC  480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3712  GGCATCTGGTTACCCACGAGGAGTACCAGCTACGAGACAGGTGCACCCCCGGGGCACACC  3771

Query  481   AATGTCCGCGTCGTAGTCATCGAGATCGGAGAATTACTCCTGCTCCCGAGCGATGGCAAT  540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3772  AATGTCCGCGTCGTAGTCATCGAGATCGGAGAATTACTCCTGCTCCCGAGCGATGGCAAT  3831

Query  541   GAATACGCCTTCGCATTCATGGTGGGCAACGGCTTTGCGCAGGCATCGTCGCGCCCACGG  600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3832  GAATACGCCTTCGCATTCATGGTGGGCAACGGCTTTGCGCAGGCATCGTCGCGCCCACGG  3891

Query  601   GAGAACCTGTTCC 613
              |||||||||||||
Sbjct  3892  GAGAACCTGTTCC 3905


Query  1     GTTGGAGTTGTCGCTACTTTGCCAGGATATGCAGAGGTCCGGTTTCGGTCCTAGTGCCAC  60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3905  GTTGGAGTTGTCGCTACTTTGCCAGGATATGCAGAGGTCCGGTTTCGGTCCTAGTGCCAC  3964

Query  61    CGCATCTTGAGGGACTGATGCAGCTGCCGGCGAATTGGCCTCCAGTCATAGACTCTGTCG  120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3965  CGCATCTTGAGGGACTGATGCAGCTGCCGGCGAATTGGCCTCCAGTCATAGACTCTGTCG  4024

Query  121   CTGATCTATTCTGGGAGTTATTTATTGTTCATCAGGTAGCACGGGGCACACTGCAGACTG  180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4025  CTGATCTATTCTGGGAGTTATTTATTGTTCATCAGGTAGCACGGGGCACACTGCAGACTG  4084

Query  181   CCGAGCAGGTTGCTGAGCTCATGAGCCGCACTCTCTACCGCTACCGTTACCTGCAGCATA  240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4085  CCGAGCAGGTTGCTGAGCTCATGAGCCGCACTCTCTACCGCTACCGTTACCTGCAGCATA  4144

Query  241   CGGCTGCGAGGCCGCTGTGCGAAGTGATATCGCTGGGCAGCGAGCTCGACAGAATAGCTC  300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4145  CGGCTGCGAGGCCGCTGTGCGAAGTGATATCGCTGGGCAGCGAGCTCGACAGAATAGCTC  4204

Query  301   TAATGACCAGACTCGTCACAGAAAAACTTATCAGCGAATGCAACTTGCTCCGAATTCGCA  360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4205  TAATGACCAGACTCGTCACAGAAAAACTTATCAGCGAATGCAACTTGCTCCGAATTCGCA  4264

Query  361   ATGACGGGCGCTTTGATTGTCGCGATGTCGTCGCTCCCCGGCTGGCATTCTGTTTCGGTA  420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4265  ATGACGGGCGCTTTGATTGTCGCGATGTCGTCGCTCCCCGGCTGGCATTCTGTTTCGGTA  4324

Query  421   TCGAAGTGCAGACAGCGCGTACGATTTTAGATCGGTTTTCAGGTAGGGTGATAACCAGGA  480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4325  TCGAAGTGCAGACAGCGCGTACGATTTTAGATCGGTTTTCAGGTAGGGTGATAACCAGGA  4384

Query  481   GTGCGCTTTTGGACATGTTCGCTGAAACACTGGTTAGCGAACTCGGTCCTGGTTGGTTGC  540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4385  GTGCGCTTTTGGACATGTTCGCTGAAACACTGGTTAGCGAACTCGGTCCTGGTTGGTTGC  4444

Query  541   CGCGCGACCTCGCATCGAGCGATGAAGCTGGACGCCCTCATTCGCTGTGGCCCATTCGAG  600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4445  CGCGCGACCTCGCATCGAGCGATGAAGCTGGACGCCCTCATTCGCTGTGGCCCATTCGAG  4504

Query  601   GACCAAAAGCGCGTCCTAAGAAAGTGCAGCCCGTTCTCTTTGTACGCACACTGCTGATAT  660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4505  GACCAAAAGCGCGTCCTAAGAAAGTGCAGCCCGTTCTCTTTGTACGCACACTGCTGATAT  4564

Query  661   GGTTGCTGAGATCGCCACCGACGTCACAGTTTTGCTGGCGGAGGCTTTTGGAGGA  715
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4565  GGTTGCTGAGATCGCCACCGACGTCACAGTTTTGCTGGCGGAGGCTTTTGGAGGA  4619
```

```
Query  1     CGCCACCGACGTCACAGTTTTGCTGGCGGAGGCTTTTGGAGGAGCTCAGTGGCGCTTGGG  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4577  CGCCACCGACGTCACAGTTTTGCTGGCGGAGGCTTTTGGAGGAGCTCAGTGGCGCTTGGG  4636

Query  61    AGCGACTACTACCGGCACTGGTGTTTACGGGTCTCGCCCGTCTTTGTTGGAGCGCTTTCC  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4637  AGCGACTACTACCGGCACTGGTGTTTACGGGTCTCGCCCGTCTTTGTTGGAGCGCTTTCC  4696

Query  121   GCCAGTTCCACAAGATCACTCGAACGCTTTCGTTCGAGCTTATGGTTAGTGGCTCGCCGT  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4697  GCCAGTTCCACAAGATCACTCGAACGCTTTCGTTCGAGCTTATGGTTAGTGGCTCGCCGT  4756

Query  181   TGCGATCACGAGAGCAGAGCCACCTCACTAAGCCTGCAGACGATTTGGATGATGCTGTTG  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4757  TGCGATCACGAGAGCAGAGCCACCTCACTAAGCCTGCAGACGATTTGGATGATGCTGTTG  4816

Query  241   GCGAGGAGCCCGATGGCCGCATAGCACAGATTGAGCATCACATGGGTGCCGAATACCGGG  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4817  GCGAGGAGCCCGATGGCCGCATAGCACAGATTGAGCATCACATGGGTGCCGAATACCGGG  4876

Query  301   AGCGCTGTACGAGTTACCTGCATCAGATGGATCACgagcagcatgagcagcatgagcagc  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4877  AGCGCTGTACGAGTTACCTGCATCAGATGGATCACGAGCAGCATGAGCAGCATGAGCAGC  4936

Query  361   atgagcagcatgagcGCCAAGCGGATCTCGGAAACGAGCCGCACCGGATAGAACGTCGCG  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4937  ATGAGCAGCATGAGCGCCAAGCGGATCTCGGAAACGAGCCGCACCGGATAGAACGTCGCG  4996

Query  421   CACGTGGACGTCACCTGCAAGCAAAACTGGCGGCAGCGGCGGGGATGGTGGATTCTGGCG  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4997  CACGTGGACGTCACCTGCAAGCAAAACTGGCGGCAGCGGCGGGGATGGTGGATTCTGGCG  5056

Query  481   GGAGCATTCAATCCAGAAACGATGCTTTCTTTCATCTTCATCCTCGGCAAGGAGGTTCCC  540
             |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Sbjct  5057  GGAGCATTCAATCCAGAAACGATGCTTGCTTTCATCTTCATCCTCGGCAAGGAGGTTCCC  5116

Query  541   TTCGGGTCCAGCAAGCAATACAGCAAGGGACAAGACGGAAACGCGTCATAGACGTTCACT  600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5117  TTCGGGTCCAGCAAGCAATACAGCAAGGGACAAGACGGAAACGCGTCATAGACGTTCACT  5176

Query  601   GTATTGACATAGCATTGGGGCAGCGAACCGAGAAACCCGTGGATGAGCATATGTACAGGA  660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5177  GTATTGACATAGCATTGGGGCAGCGAACCGAGAAACCCGTGGATGAGCATATGTACAGGA  5236

Query  661   AGCGACATGCTTGGCTCGTCGTCCGTGATGCCGGCTCCGATACGGTTGTCTACGCGAACC  720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5237  AGCGACATGCTTGGCTCGTCGTCCGTGATGCCGGCTCCGATACGGTTGTCTACGCGAACC  5296

Query  721   GCGTTTTCTATCGATTTGGACTGCAAAGATTTTGCTTGAAGATCCC  766
             |||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5297  GCGTTTTCTATCGATTTGGACTGCAAAGATTTTGCTTGAAGATCCC  5342


Query  421   AATACAGCAAGGGACAAGACGGAAACGCGTCATAGACGTTCACTGTATTGACATAGCATT  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5133  AATACAGCAAGGGACAAGACGGAAACGCGTCATAGACGTTCACTGTATTGACATAGCATT  5192

Query  481   GGGGCAGCGAACCGAGAAACCCGTGGATGAGCATATGTACAGGAAGCGACATGCTTGGCT  540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5193  GGGGCAGCGAACCGAGAAACCCGTGGATGAGCATATGTACAGGAAGCGACATGCTTGGCT  5252

Query  541   CGTCGTCCGTGATGCCGGCTCCGATACGGTTGTCTACGCGAACCGCGTTTTCTATCGATT  600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5253  CGTCGTCCGTGATGCCGGCTCCGATACGGTTGTCTACGCGAACCGCGTTTTCTATCGATT  5312

Query  601   TGGACTGCAAAGATTTTGCTTGAAGATCCCCGAGGTCGCCGCATGCAAGCCAACAAAACT  660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5313  TGGACTGCAAAGATTTTGCTTGAAGATCCCCGAGGTCGCCGCATGCAAGCCAACAAAACT  5372

Query  661   CTCCGTTCATGCTTTCGACGAGTTCAGCACGGACGAGGAAATGGAGCGCTGCGTGTTGGT  720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  5373  CTCCGTTCATGCTTTCGACGAGTTCAGCACGGACGAGGAAATGGAGCGCTGCGTGTTGGT  5432

Query  721   AACATCTGCACAAACGACCGCGGATCCGAGTATC  754
             |||||||||||||||||||||||||||||||||
Sbjct  5433  AACATCTGCACAAACGACCGCGGATCCGAGTATC  5466
```

Snu114 gene inserted into pQLinkN (The RBS sequence is in blue and the LIC sequence is in bold; pSR647)

```
Query  38   TACTTCCAATCCCACGAGGAGAAATTAACT 68
            ||||||||||||||||||||||||||||||
```

```
Sbjct       ------------------------------

Query  69   ATGAGTTCAGCGTTTCGTGGTGGCGAAACTGATGAGGTCGGCAGTATCCTGGTTCATGGT  128
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGTTCAGCGTTTCGTGGTGGCGAAACTGATGAGGTCGGCAGTATCCTGGTTCATGGT  60

Query  129  GGCGCACGACACGGACGTTCTGACGCTCTGGAGGCCGTAGCTTCCGACGAATGTGTGCCA  188
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GGCGCACGACACGGACGTTCTGACGCTCTGGAGGCCGTAGCTTCCGACGAATGTGTGCCA  120

Query  189  GCTATCGAGACGCTTACAGCAACCACACCCGGCTTACCGCTAGCGATTCCGCGTCGAAGG  248
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  GCTATCGAGACGCTTACAGCAACCACACCCGGCTTACCGCTAGCGATTCCGCGTCGAAGG  180

Query  249  TTGCGCAAACGTCACCGCATCCAAGAGACACAAACCCCTGAACCAATCCCTGCACTCACT  308
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  TTGCGCAAACGTCACCGCATCCAAGAGACACAAACCCCTGAACCAATCCCTGCACTCACT  240

Query  309  CGCGTACGCACGCGAGCACCAAAGCGACATCAGGCACCGAGCGACGCTGGGTTCTATGTC  368
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  CGCGTACGCACGCGAGCACCAAAGCGACATCAGGCACCGAGCGACGCTGGGTTCTATGTC  300

Query  369  CAGGCGCGACCACTGCGCTTCAAGGTGTCGCGAAGGTACCTTTTGCATCTTGCGAAACAC  428
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  CAGGCGCGACCACTGCGCTTCAAGGTGTCGCGAAGGTACCTTTTGCATCTTGCGAAACAC  360

Query  429  GCCGGTCCCGAGCGTATCTGGAACATTCTGGTTGCGGGTCACTACCATCATGGAAAAACA  488
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  GCCGGTCCCGAGCGTATCTGGAACATTCTGGTTGCGGGTCACTACCATCATGGAAAAACA  420

Query  489  AGCCTCATCGACTTGTTGGTAAGTCACCAGCTGCATCCGGCTGCAGCAACCCGTTACATG  548
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  AGCCTCATCGACTTGTTGGTAAGTCACCAGCTGCATCCGGCTGCAGCAACCCGTTACATG  480

Query  549  ATAGGCCCGCGGCGACACCAGCAACAACCGCGCTGGACGGATACGCGTCGGGACGAACTC  608
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  481  ATAGGCCCGCGGCGACACCAGCAACAACCGCGCTGGACGGATACGCGTCGGGACGAACTC  540

Query  609  TCCCGAGGAATGTCGCTTCAGCTTGCTTTCATGCCGCTCTGGGTACCAGACGAGCACGGT  668
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  541  TCCCGAGGAATGTCGCTTCAGCTTGCTTTCATGCCGCTCTGGGTACCAGACGAGCACGGT  600

Query  669  GTATCTCAACTGGTGACGCTGATGGATGCTCCCGGACATGCAGACTTCTTCGATCAGGTT  727
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  601  GTATCTCAACTGGTGACGCTGATGGATGCTCCCGGACATGCAGACTTCTTCGATCAGGTT  660

Query  728  GTGGTGGGTGCAACGCTCGCAGATGCAGTGCTTCTGGTGGTCGACAGTGCCGA  780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  661  GTGGTGGGTGCAACGCTCGCAGATGCAGTGCTTCTGGTGGTCGACAGTGCCGA  713


Query  1     CTCGCAGATGCAGTGCTTCTGGTGGTCGACAGTGCCGAAGGCGTCCTGCTAGGTACAGAG  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  676   CTCGCAGATGCAGTGCTTCTGGTGGTCGACAGTGCCGAAGGCGTCCTGCTAGGTACAGAG  735

Query  61    CGAGTCGTCGCCTTGGCGCTGGAGATGTCGCTTCCGCTGATCCTGGTGTTGACGCAGCTC  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  736   CGAGTCGTCGCCTTGGCGCTGGAGATGTCGCTTCCGCTGATCCTGGTGTTGACGCAGCTC  795

Query  121   GACCGCTTGATACTTGAGTTGCGGTATCCGCCTGATGCCGTGTACCTTAAGTTGAAAGGC  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  796   GACCGCTTGATACTTGAGTTGCGGTATCCGCCTGATGCCGTGTACCTTAAGTTGAAAGGC  855

Query  181   ATCATCGATGCGCTCAACAGCCTTATTGAACGTTTGCAGCCTCAGAAGGTCCCGTACTTT  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  856   ATCATCGATGCGCTCAACAGCCTTATTGAACGTTTGCAGCCTCAGAAGGTCCCGTACTTT  915

Query  241   GACCCAAGGCCACCACATGCCAATGTCCTATTCACGTCTGCAAAGCTGAACATAGCGTTC  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  916   GACCCAAGGCCACCACATGCCAATGTCCTATTCACGTCTGCAAAGCTGAACATAGCGTTC  975

Query  301   AGCCTCATGGACATTGTCGAGAAGTGGTATGGTTCGGCTTTGGAGGTGCAGCGTGAACGT  360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  976   AGCCTCATGGACATTGTCGAGAAGTGGTATGGTTCGGCTTTGGAGGTGCAGCGTGAACGT  1035

Query  361   TCGTGGACCGGTGGCGTTCAGGGCGCGCTCCGGAAGCGCCGAGCGAGGAAACGAACACTC  420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1036  TCGTGGACCGGTGGCGTTCAGGGCGCGCTCCGGAAGCGCCGAGCGAGGAAACGAACACTC  1095

Query  421   GCTGCCTCGCTCTGGGGCGACCGGGTCTACGACAAGCCCACCGGCTTGTTTCTACGCAAA  480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1096  GCTGCCTCGCTCTGGGGCGACCGGGTCTACGACAAGCCCACCGGCTTGTTTCTACGCAAA  1155

Query  481   GCCGCTGTCGCCTGGACGCTGGGCACGGCAAAGCGATCATTTGTTGAGTTTGTTCTGGAG  540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1156  GCCGCTGTCGCCTGGACGCTGGGCACGGCAAAGCGATCATTTGTTGAGTTTGTTCTGGAG  1215

Query  541   CCTCTCTACAAGACGATTGCACTATGTGCGACCCATGAGTCTGGGAGCGGATCGCTGGAG  600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1216  CCTCTCTACAAGACGATTGCACTATGTGCGACCCATGAGTCTGGGAGCGGATCGCTGGAG  1275

Query  601   CAGCTGCTGGTCAATGTGCAAGCTCTGGGAACTCCGCGTGCTTTCGAGCCTCAAGCAGCA  660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Sbjct  1276  CAGCTGCTGGTCAATGTGCAAGCTCTGGGAACTCCGCGTGCTTTCGAGCCTCAAGCAGCA  1335

Query  661   TCGAGTCAGCGAGAGGCGCCTACGGATCTGGAGATGCTCCAGCAGGCGGCAGACTCAGGC  720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1336  TCGAGTCAGCGAGAGGCGCCTACGGATCTGGAGATGCTCCAGCAGGCGGCAGACTCAGGC  1395

Query  721   CCATGGGCTCTTCTCCGTGTTGTCTTCGACATGGTACTGGGCTCGCCAACAGCGTTGTTT  780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1396  CCATGGGCTCTTCTCCGTGTTGTCTTCGACATGGTACTGGGCTCGCCAACAGCGTTGTTT  1455

Query  781   GGAACGCTGACCCGTCGTCGTCGCCACTCGAAATGGTTCCCGGAGCACTGCGCTGAGACG  840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1456  GGAACGCTGACCCGTCGTCGTCGCCACTCGAAATGGTTCCCGGAGCACTGCGCTGAGACG  1515

Query  841   GAGA  844
              ||||
Sbjct  1516  GAGA  1519
```

```
Query  1     AGTCAGCGAGAGGCGCCTACGGATCTGGAGATGCTCCAGCAGGCGGCAGACTCAGGCCCA  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1339  AGTCAGCGAGAGGCGCCTACGGATCTGGAGATGCTCCAGCAGGCGGCAGACTCAGGCCCA  1398

Query  61    TGGGCTCTTCTCCGTGTTGTCTTCGACATGGTACTGGGCTCGCCAACAGCGTTGTTTGGA  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1399  TGGGCTCTTCTCCGTGTTGTCTTCGACATGGTACTGGGCTCGCCAACAGCGTTGTTTGGA  1458

Query  121   ACGCTGACCCGTCGTCGTCGCCACTCGAAATGGTTCCCGGAGCACTGCGCTGAGACGGAG  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1459  ACGCTGACCCGTCGTCGTCGCCACTCGAAATGGTTCCCGGAGCACTGCGCTGAGACGGAG  1518

Query  181   ACCACTGTGCTGGTAGCGACACACTGGCGTTCGCTGGACGGCACCGATACCGTTGCTGTG  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1519  ACCACTGTGCTGGTAGCGACACACTGGCGTTCGCTGGACGGCACCGATACCGTTGCTGTG  1578

Query  241   GGCCGTGTTTGCGGTGCACGACCGCTACGCGCTGATCAGCTGTTTGCAGTTCAACCGGGC  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1579  GGCCGTGTTTGCGGTGCACGACCGCTACGCGCTGATCAGCTGTTTGCAGTTCAACCGGGC  1638

Query  301   GCTGACCGATGCAAGCCGTCGTCGGTaacaacaacaacaacaacaacaacaTGATCACGCAA  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1639  GCTGACCGATGCAAGCCGTCGTCGGTAACAACAACAACAACAACAACAATGATCACGCAA  1698

Query  361   TTGCAGGTGGCTTTCGGACGCGTCTGGTACGACGTTGAGCAAGTCGCCGCCGGCGGTTTA  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1699  TTGCAGGTGGCTTTCGGACGCGTCTGGTACGACGTTGAGCAAGTCGCCGCCGGCGGTTTA  1758

Query  421   GTGCTTCTGCCCAGGGTTTATCCGCCATCACGCGGAACCTACTGCTTTGGCCTCCGTTGT  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1759  GTGCTTCTGCCCAGGGTTTATCCGCCATCACGCGGAACCTACTGCTTTGGCCTCCGTTGT  1818

Query  481   TCCGATGGACCGTACTCCAAGACGCTGGACGCGTGCGTTTATCGCTTGCACAGAGCGCTG  540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1819  TCCGATGGACCGTACTCCAAGACGCTGGACGCGTGCGTTTATCGCTTGCACAGAGCGCTG  1878

Query  541   GGCTATCACGCAATGCCTATCTACGGTATAGCGCTTGCACCGTGTTTGCCGGAGGTGCAC  600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1879  GGCTATCACGCAATGCCTATCTACGGTATAGCGCTTGCACCGTGTTTGCCGGAGGTGCAC  1938

Query  601   AGAACTGCAGAGAGCTTGACGGCAGATTTGGTTGCGCAGTACAGCACCAGCACAGCGAGC  660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1939  AGAACTGCAGAGAGCTTGACGGCAGATTTGGTTGCGCAGTACAGCACCAGCACAGCGAGC  1998

Query  661   GCTGAGCAGCTTCGCCAAGCTTTGGCAATTATTGTTCGAACACATCCCGCCACAGGATTT  720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1999  GCTGAGCAGCTTCGCCAAGCTTTGGCAATTATTGTTCGAACACATCCCGCCACAGGATTT  2058

Query  721   GATGGTAACGGTAGCGTTCATGGTGATGCAGCACACGGTTTTCGAACAGCTTCAACTGCT  780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2059  GATGGTAACGGTAGCGTTCATGGTGATGCAGCACACGGTTTTCGAACAGCTTCAACTGCT  2118

Query  781   GGTGTTGTTTACGGCCCAGGGGAACTCTATCTGGATGTTATACTTCATGAGCTTC  835
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2119  GGTGTTGTTTACGGCCCAGGGGAACTCTATCTGGATGTTATACTTCATGAGCTTC  2173
```

```
Query  1     TTTGGTTGCGCAGTACAGCACCAGCACAGCGAGCGCTGAGCAGCTTCGCCAAGCTTTGGC  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1965  TTTGGTTGCGCAGTACAGCACCAGCACAGCGAGCGCTGAGCAGCTTCGCCAAGCTTTGGC  2024

Query  61    AATTATTGTTCGAACACATCCCGCCACAGGATTTGATGGTAACGGTAGCGTTCATGGTGA  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2025  AATTATTGTTCGAACACATCCCGCCACAGGATTTGATGGTAACGGTAGCGTTCATGGTGA  2084

Query  121   TGCAGCACACGGTTTTCGAACAGCTTCAACTGCTGGTGTTGTTTACGGCCCAGGGGAACT  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2085  TGCAGCACACGGTTTTCGAACAGCTTCAACTGCTGGTGTTGTTTACGGCCCAGGGGAACT  2144

Query  181   CTATCTGGATGTTATACTTCATGAGCTTCGTCGATATTTGGTGCAACGGCATTTGAGTCT  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2145  CTATCTGGATGTTATACTTCATGAGCTTCGTCGATATTTGGTGCAACGGCATTTGAGTCT  2204
```

```
Query  241   GGTATGGCGTTGCCTGCGTACAAATCCGGTACCCTTTGTGGATGCGTTACGGGAAACGAT  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2205  GGTATGGCGTTGCCTGCGTACAAATCCGGTACCCTTTGTGGATGCGTTACGGGAAACGAT  2264

Query  301   CCAAGCCGGTGCAGCAAAAGTGCAAGTAGGCTTGCGAGGCCGTGATGGACGCACGCGGAG  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2265  CCAAGCCGGTGCAGCAAAAGTGCAAGTAGGCTTGCGAGGCCGTGATGGACGCACGCGGAG  2324

Query  361   CGACCAGACTGCGTTGTGGTCGTCGCCAAAGTCATTCTTGGTTTCGGAAGACGAATTCTT  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2325  CGACCAGACTGCGTTGTGGTCGTCGCCAAAGTCATTCTTGGTTTCGGAAGACGAATTCTT  2384

Query  421   CTCTCCCGAGCCCAACGAATGGAGCGACAACGAGGACGCGGCATTTCATCAGCCGCGAGT  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2385  CTCTCCCGAGCCCAACGAATGGAGCGACAACGAGGACGCGGCATTTCATCAGCCGCGAGT  2444

Query  481   CGTGCTTTACGTGGAACCGGTGCACGCTTCGCGCCGGGCTCCAGCTCAGCAGGCAGGAGA  540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2445  CGTGCTTTACGTGGAACCGGTGCACGCTTCGCGCCGGGCTCCAGCTCAGCAGGCAGGAGA  2504

Query  541   GCCGAACGCGGGGGCCTCACCCGGTAGTTGTGCATGCCCTGCATCCACTGCCGAATAGCGT  600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2505  GCCGAACGCGGGGGCCTCACCCGGTAGTTGTGCATGCCCTGCATCCACTGCCGAATAGCGT  2564

Query  601   TCGTTTTGAGCGACGGAGTCTACCTGCTGAGGCCGCTGTGCAGCCAGAAGCGCTCGAAGT  660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2565  TCGTTTTGAGCGACGGAGTCTACCTGCTGAGGCCGCTGTGCAGCCAGAAGCGCTCGAAGT  2624

Query  661   GGAATTTGCAGCCGATCTGGATGGGACAACAGCGCCGTCTGGGCTGCCGGTGACACTCCA  720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2625  GGAATTTGCAGCCGATCTGGATGGGACAACAGCGCCGTCTGGGCTGCCGGTGACACTCCA  2684

Query  721   GGCCCTCTGGGAAGGTCTTCGGTTGGCGAGTCGGCGGGGTCCGCTTCTTCAAGGTCCTGT  780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2685  GGCCCTCTGGGAAGGTCTTCGGTTGGCGAGTCGGCGGGGTCCGCTTCTTCAAGGTCCTGT  2744

Query  781   CGTTGGGGTTCGTTACCACGTGCGTGCTCTC  811
             ||||||||||||||||||||||||||||||||
Sbjct  2745  CGTTGGGGTTCGTTACCACGTGCGTGCTCTC  2775


Query  114   CAGCCAGAAGCGCTCGAAGTGGAATTTGCAGCCGATCTGGATGGGACAACAGCGCCGTCT  173
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2605  CAGCCAGAAGCGCTCGAAGTGGAATTTGCAGCCGATCTGGATGGGACAACAGCGCCGTCT  2664

Query  174   GGGCTGCCGGTGACACTCCAGGCCCTCTGGGAAGGTCTTCGGTTGGCGAGTCGGCGGGGT  233
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2665  GGGCTGCCGGTGACACTCCAGGCCCTCTGGGAAGGTCTTCGGTTGGCGAGTCGGCGGGGT  2724

Query  234   CCGCTTCTTCAAGGTCCTGTCGTTGGGGTTCGTTACCACGTGCGTGCTCTCGAGTGCATT  293
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2725  CCGCTTCTTCAAGGTCCTGTCGTTGGGGTTCGTTACCACGTGCGTGCTCTCGAGTGCATT  2784

Query  294   TGCGGGAGCTCTGCCTGGGAAGCACCACCTCCATGTTGGTCCGCTTGGCATCGGATGCGA  353
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2785  TGCGGGAGCTCTGCCTGGGAAGCACCACCTCCATGTTGGTCCGCTTGGCATCGGATGCGA  2844

Query  354   ACTCGCCTGGTGCTCCTAGCTCGACAGGCAGCGCATAGGGCGCTTCTTGATGCCAAGATG  413
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2845  ACTCGCCTGGTGCTCCTAGCTCGACAGGCAGCGCATAGGGCGCTTCTTGATGCCAAGATG  2904

Query  414   CAGATTCTAGAGCCTTGCTTCCGTTTGCAAGCCGTGGTGCGCGCCGAAAAAGCCGAGCTC  473
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2905  CAGATTCTAGAGCCTTGCTTCCGTTTGCAAGCCGTGGTGCGCGCCGAAAAAGCCGAGCTC  2964

Query  474   ATTTGCCGTCGCTTGCGCAAGGCTTCTGAACTTGTCGAAATCCGGCAGCAGTGGCCCATT  533
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2965  ATTTGCCGTCGCTTGCGCAAGGCTTCTGAACTTGTCGAAATCCGGCAGCAGTGGCCCATT  3024

Query  534   CCAGGCACCTGTTTCGTGATCGTTGACAGTGATGTTCCGGCGCGGGTGCTTGTTCCCGGC  593
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3025  CCAGGCACCTGTTTCGTGATCGTTGACAGTGATGTTCCGGCGCGGGTGCTTGTTCCCGGC  3084

Query  594   CTCGAAGTGATGTTGCGATTTCAGAGCCACGGGCAGGCAAGTGTGCAAGCAACTGTCGAT  653
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3085  CTCGAAGTGATGTTGCGATTTCAGAGCCACGGGCAGGCAAGTGTGCAAGCAACTGTCGAT  3144

Query  654   CCAGCTGTCATGCCGACGTGCAGTGCCCGCTGGATTCCAGTTCCAGGTGATGCGGACAGT  713
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3145  CCAGCTGTCATGCCGACGTGCAGTGCCCGCTGGATTCCAGTTCCAGGTGATGCGGACAGT  3204

Query  714   GTGGAATGCCCACCTCTAGAGGCAGTTGTTGCCTCGGACGACAGCACCGTAACCGAAAAC  773
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3205  GTGGAATGCCCACCTCTAGAGGCAGTTGTTGCCTCGGACGACAGCACCGTAACCGAAAAC  3264

Query  774   ACCCTTGCACCCTGGCTGGTGCGCATCGTACGAATGCGACGGGGACTCGGGACCGACCTC  833
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3265  ACCCTTGCACCCTGGCTGGTGCGCATCGTACGAATGCGACGGGGACTCGGGACCGACCTC  3324

Query  834   TGA  836
             |||
Sbjct  3325  TGA  3327
```

Dib1 gene inserted into pQLinkN (The RBS sequence is in blue and the LIC sequence is in bold; pSR634)

```
Query  49   TACTTCCAATCCCACGAGGAGAAATTAACT 79
            ||||||||||||||||||||||||||||||
Sbjct        ------------------------------

Query  80   ATGGACAGTGCACCGTTGGTGCCGGTACTGGGCTCCATGGCGGCGATTCAACAAGCACTG  139
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGGACAGTGCACCGTTGGTGCCGGTACTGGGCTCCATGGCGGCGATTCAACAAGCACTG  60

Query  140  GCTGAGGAAACCGAGCGTCTGGTTGCCCTGCGCTTTAGTAGCGATCCAGCAGCTGTAGAT  199
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GCTGAGGAAACCGAGCGTCTGGTTGCCCTGCGCTTTAGTAGCGATCCAGCAGCTGTAGAT  120

Query  200  TGTGTCTTCATGGACGAAATCCTCGCAAGATCAGCAGCGCGCGTGCGGAGGTTCGCAGTG  259
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  TGTGTCTTCATGGACGAAATCCTCGCAAGATCAGCAGCGCGCGTGCGGAGGTTCGCAGTG  180

Query  260  GTGTACGGTGTGGACTTGCGGCAGGTTCCGCAGGCTGCGCGGCGCTTCGGCGTTGAGGCG  319
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GTGTACGGTGTGGACTTGCGGCAGGTTCCGCAGGCTGCGCGGCGCTTCGGCGTTGAGGCG  240

Query  320  TGGCGACCCCTGTCGCTCCAGTTCTATTATCGAAAGCGCCTCATCAAGGTGGACTGTGGT  379
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  TGGCGACCCCTGTCGCTCCAGTTCTATTATCGAAAGCGCCTCATCAAGGTGGACTGTGGT  300

Query  380  ACTGGAGACACGGCGCGTCTGACCCGTCCGGTGCCGAGCGTGCAGCAGCTGGTGGACCTC  439
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  ACTGGAGACACGGCGCGTCTGACCCGTCCGGTGCCGAGCGTGCAGCAGCTGGTGGACCTC  360

Query  440  TTCGAAGTCGTCTATCGACAGGCGTTGCGCGGAAAGGGTCTCGCGATGGCGCCGTTCCGA  499
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  TTCGAAGTCGTCTATCGACAGGCGTTGCGCGGAAAGGGTCTCGCGATGGCGCCGTTCCGA  420

Query  500  CTCTAG  505
            ||||||
Sbjct  421  CTCTAG  426
```

SmB gene inserted into pQLinkN (The RBS sequence is in blue and the LIC sequence is in bold; pSR714)

```
Query  33   TACTTCCAATCCCACGAGGAGAAATTAACT 63
            ||||||||||||||||||||||||||||||
Sbjct        ------------------------------

Query  64   ATGGATCTTCTGCCTGTGCTGCGATCCCAGGTTCACGTTCAAACGACCGACGGGCGCCTC  123
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGGATCTTCTGCCTGTGCTGCGATCCCAGGTTCACGTTCAAACGACCGACGGGCGCCTC  60

Query  124  CTAGCGGGCAAGCTGTTAGCGTTCGACGCTCATAGCAATTTATTACTCAGCCACTGTACA  183
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   CTAGCGGGCAAGCTGTTAGCGTTCGACGCTCATAGCAATTTATTACTCAGCCACTGTACA  120

Query  184  GAACGTCGCGGGGAATCAGCGAAACGCTACTTGGGCATGGTGCTGGTGCGCGGGGAGCAT  243
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  GAACGTCGCGGGGAATCAGCGAAACGCTACTTGGGCATGGTGCTGGTGCGCGGGGAGCAT  180

Query  244  GTGCTCGCAGTTATCACGCCCAGAATCACGGAAACTGAACAGAAAACTGCCGCATCTGAA  303
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GTGCTCGCAGTTATCACGCCCAGAATCACGGAAACTGAACAGAAAACTGCCGCATCTGAA  240
```

SmD3 gene inserted into pQLinkN (The RBS sequence is in blue and the LIC sequence is in bold; pSR715)

```
Query  34   TACTTCCAATCCCACGAGGAGAAATTAACT 64
            ||||||||||||||||||||||||||||||
Sbjct        ------------------------------
```

```
Query  65   ATGAGCGGGTATCGACCCGCTGCGTTCGATCTCCCTCGAGCGCTCCTACGCGAAGCAAAG  124
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCGGGTATCGACCCGCTGCGTTCGATCTCCCTCGAGCGCTCCTACGCGAAGCAAAG  60

Query  125  AACCAAATTGTATCGGTAGAGACCAAAAATGGAATGGAGTACCGGGGGCGCCTGGACAAC  184
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   AACCAAATTGTATCGGTAGAGACCAAAAATGGAATGGAGTACCGGGGGCGCCTGGACAAC  120

Query  185  GTGAGCTCGCGGATGAACCTGGTGCTCAGCGCGGTGACGGTATTGAACGCGACTGGCGAG  244
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  GTGAGCTCGCGGATGAACCTGGTGCTCAGCGCGGTGACGGTATTGAACGCGACTGGCGAG  180

Query  245  CGCACCCaaaaaaaTCGTGTTCTCGTCCGTGGTGACAGTATCGTGCTTGTAGTGCTCCCG  304
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  CGCACCCAAAAAAATCGTGTTCTCGTCCGTGGTGACAGTATCGTGCTTGTAGTGCTCCCG  240

Query  305  GAAGCACTAGAAGACGCACCACAGCTGGATGTCCTATTACAGGTAAAGCAGGCCCGGAAG  364
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GAAGCACTAGAAGACGCACCACAGCTGGATGTCCTATTACAGGTAAAGCAGGCCCGGAAG  300

Query  365  GCGGCGATGCACGTGAACAACACTGACCGCAAGTCACGTGGAGCGCGTCGTTCCGAGGCA  424
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  GCGGCGATGCACGTGAACAACACTGACCGCAAGTCACGTGGAGCGCGTCGTTCCGAGGCA  360

Query  425  GACGTACACGAGCGTTCAGGTGCCAGCACGCTGCCACTACCCCAGAGCGAGTCGCAGCCG  484
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  GACGTACACGAGCGTTCAGGTGCCAGCACGCTGCCACTACCCCAGAGCGAGTCGCAGCCG  420

Query  485  CAACTCAAGCGAACTCGAGTTTTCTTGAGTGGTAATGCGGAGACCGTTCAGCGAACCAAA  544
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  CAACTCAAGCGAACTCGAGTTTTCTTGAGTGGTAATGCGGAGACCGTTCAGCGAACCAAA  480

Query  545  GAAGGAGGCGACTCGAACCGGCGGAACGTG  574
            ||||||||||||||||||||||||||||||
Sbjct  481  GAAGGAGGCGACTCGAACCGGCGGAACGTG  510
```

SmD2 gene inserted into pQLinkN (The RBS sequence is in blue and the LIC sequence is in bold; pSR716)

```
Query  29   TACTTCCAATCCCACGAGGAGAAATTAACT 59
            ||||||||||||||||||||||||||||||
Sbjct       ------------------------------

Query  60   ATGCCTCCAGTTGATCAGCCCACTGCTTTAGAAGCGGGGGCTGTAGCGGGACTGACGGTG  119
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGCCTCCAGTTGATCAGCCCACTGCTTTAGAAGCGGGGGCTGTAGCGGGACTGACGGTG  60

Query  120  GCGCAGCTCCGTCGGGAGCTTGCCGCGCGAGAAGCTCCAACCAGCGGTAGAAAGGCTGAA  179
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GCGCAGCTCCGTCGGGAGCTTGCCGCGCGAGAAGCTCCAACCAGCGGTAGAAAGGCTGAA  120

Query  180  CTCCAAAAACGTTTGCTGGACCTGTTAGGTGTGAAACTCGAACAAGAGGCTCGCGATGAG  239
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  CTCCAAAAACGTTTGCTGGACCTGTTAGGTGTGAAACTCGAACAAGAGGCTCGCGATGAG  180

Query  240  GACTCTAGCGTCGCGCCTGGAGCCACACAGGGAGAAGCTGGTCGGGCTACCAACCTTGGA  299
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GACTCTAGCGTCGCGCCTGGAGCCACACAGGGAGAAGCTGGTCGGGCTACCAACCTTGGA  240

Query  300  GACGCAACGACGACATCGTCCGCgcagcagcaggagcagcagcaggagcagcagcaggag  359
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GACGCAACGACGACATCGTCCGCGCAGCAGCAGGAGCAGCAGCAGGAGCAGCAGCAGGAG  300

Query  360  cagcagcaggagcagcagcaggagcagcagcaggagcagcagcaggagcagcagcaggag  419
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  CAGCAGCAGGAGCAGCAGCAGGAGCAGCAGCAGGAGCAGCAGCAGGAGCAGCAGCAGGAG  360

Query  420  cagAAGTTGGCTCAAACCCTGGACCCTGCTGCTCTATCGCCGTCACCGATCCAGTCTAGC  479
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CAGAAGTTGGCTCAAACCCTGGACCCTGCTGCTCTATCGCCGTCACCGATCCAGTCTAGC  420

Query  480  GCGTATCCACAGAGCACGACCACCACGCAACGACGAAAACGACGCTGGGCGGAGCCGGCC  539
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GCGTATCCACAGAGCACGACCACCACGCAACGACGAAAACGACGCTGGGCGGAGCCGGCC  480

Query  540  AGCGCCCCGCCTACCGCGCCCCGGAAACGAAGACCCCTTGATGCACACGACACGCACTTG  599
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  481  AGCGCCCCGCCTACCGCGCCCCGGAAACGAAGACCCCTTGATGCACACGACACGCACTTG  540

Query  600  GATCAAGCTGGGGCAACGCCTGCCGCATCAGAGCTCAGCGCTGCAGCAGAAGCTTCGACA  659
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  541  GATCAAGCTGGGGCAACGCCTGCCGCATCAGAGCTCAGCGCTGCAGCAGAAGCTTCGACA  600

Query  660  TCCTACCAAACGCTAATCGCAGCGACGACCCCAGCAACGACGCAG  704
            |||||||||||||||||||||||||||||||||||||||||||||
Sbjct  601  TCCTACCAAACGCTAATCGCAGCGACGACCCCAGCAACGACGCAG  645
```

```
Query  241  ACCCCAGCAACGACGCAGAGCATTCCGAATAGCAGCGAAAGCGCTGCGTCAGCTCTGAAG  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  628  ACCCCAGCAACGACGCAGAGCATTCCGAATAGCAGCGAAAGCGCTGCGTCAGCTCTGAAG  687

Query  301  CCAGCAGTGCATGCCGCAAACGGCTCGCCGCGAACACCGTTCACGCTGCTCGACCGGTGC  360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  688  CCAGCAGTGCATGCCGCAAACGGCTCGCCGCGAACACCGTTCACGCTGCTCGACCGGTGC  747

Query  361  ATCACCGATCGAGTGCCGTGTCTCGTGAGCTGTCGTCATAATAAAAAGCTCTACGGCACG  420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  748  ATCACCGATCGAGTGCCGTGTCTCGTGAGCTGTCGTCATAATAAAAAGCTCTACGGCACG  807

Query  421  CTGCGCGCCTATGATAAGCACTTTAACCTCATTATGGAGCATGTACGGGAAATCTGGCAG  480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  808  CTGCGCGCCTATGATAAGCACTTTAACCTCATTATGGAGCATGTACGGGAAATCTGGCAG  867

Query  481  GAGTCACAACCCGATCGGCCTCCAGACTTGCGCGAGCGATTCATCTCGCGCCTGTTTGTG  540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  868  GAGTCACAACCCGATCGGCCTCCAGACTTGCGCGAGCGATTCATCTCGCGCCTGTTTGTG  927

Query  541  CGCGGTGACGGCGTGATTTTTATCGTTCGACCNTGCGTATCTGCAACGAGTACAGCGCGC  600
            |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Sbjct  928  CGCGGTGACGGCGTGATTTTTATCGTTCGACCCTGCGTATCTGCAACGAGTACAGCGCGC  987

Query  601  GCACAGCCG  609
            |||||||||
Sbjct  988  GCACAGCCG  996
```

SmE gene inserted into pQLinkN (The RBS sequence is in blue and the LIC sequence is in bold; pSR719)

```
Query  60   TACTTCCAATCCCACGAGGAGAAATTAACT 90
            |||||||||||||||||||||||||||||
Sbjct       ------------------------------

Query  91   ATGCCGAAGGACGCTCTGGACAGACGGATAGTTCCAGAGCAGTTGTTAGCAACGCTGGCG  150
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGCCGAAGGACGCTCTGGACAGACGGATAGTTCCAGAGCAGTTGTTAGCAACGCTGGCG  60

Query  151  CGCCAACAAGCCCGCGTTGAGGTCTGGTTATTCGAAAACACCAGATACTCTCTGGAAGGC  210
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   CGCCAACAAGCCCGCGTTGAGGTCTGGTTATTCGAAAACACCAGATACTCTCTGGAAGGC  120

Query  211  ACCTTGCGCGGCTTCGACGAACACACCAATCTAGTTCTGGTCGACACCGTGGAGCAGTGG  270
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ACCTTGCGCGGCTTCGACGAACACACCAATCTAGTTCTGGTCGACACCGTGGAGCAGTGG  180

Query  271  GGAAGTACTGCAAAGCATAAGCGGCGGACGGTTGCTCTAGGGACGATCCTCCTCAAAGGC  330
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GGAAGTACTGCAAAGCATAAGCGGCGGACGGTTGCTCTAGGGACGATCCTCCTCAAAGGC  240

Query  331  GAAAACGTCGTCCTCGTTCGGTCGCTGGGGATGCCAACCCAGCGAAAAGAGGTCACGCAC  390
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GAAAACGTCGTCCTCGTTCGGTCGCTGGGGATGCCAACCCAGCGAAAAGAGGTCACGCAC  300

Query  391  AGCGCGACTCGGGAG  405
            |||||||||||||||
Sbjct  301  AGCGCGACTCGGGAG  315
```

SmF gene inserted into pQLinkN (The RBS sequence is in blue and the LIC sequence is in bold; pSR720)

```
Query  37   TACTTCCAATCCCACGAGGAGAAATTAACT 87
            |||||||||||||||||||||||||||||
Sbjct       ------------------------------

Query  88   ATGACTGCGACTGGTTTCGCAGAGGCAGTGAAGCCCACAAACCTTCTGAGCGCGCTCCAG  147
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGACTGCGACTGGTTTCGCAGAGGCAGTGAAGCCCACAAACCTTCTGAGCGCGCTCCAG  60
```

```
Query  148  GGAAACAGGGTGTCCGTGCGCCTCAAATGGGACCTGGAGTACACCGGCCTCCTCGCATCG  207
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GGAAACAGGGTGTCCGTGCGCCTCAAATGGGACCTGGAGTACACCGGCCTCCTCGCATCG  120

Query  208  TATGACTCGTACTTCAACCTGGAGCTGGAGCATGCGGAGGAGCTTCAGCCGGACGGCTCA  267
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  TATGACTCGTACTTCAACCTGGAGCTGGAGCATGCGGAGGAGCTTCAGCCGGACGGCTCA  180

Query  268  AGCCTTCCGCTAGGCGACATGATCATTCGCTGTAATAACGTTCTTTATATCCGCGACCTT  327
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  AGCCTTCCGCTAGGCGACATGATCATTCGCTGTAATAACGTTCTTTATATCCGCGACCTT  240

Query  328  CGATCCACAGTGCCGGTCCCGCCTCTATCT  357
            |||||||||||||||||||||||||||||
Sbjct  241  CGATCCACAGTGCCGGTCCCGCCTCTATCT  270
```

SmG gene inserted into pQLinkN (The RBS sequence is in blue and the LIC sequence is in bold; pSR721)

```
Query  67   TACTTCCAATCCCACGAGGAGAAATTAACT 97
            ||||||||||||||||||||||||||||||
Sbjct       ------------------------------

Query  98   ATGGCAAAAGACGAGGTCGATACTGCGGAACTCGAAGCGTTGCTGTTTCATTCCGTCCAA  157
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGGCAAAAGACGAGGTCGATACTGCGGAACTCGAAGCGTTGCTGTTTCATTCCGTCCAA  60

Query  158  GTGTACCTGAACGCGAACAGGTGCGTGCGCGGAAAACTCAGCGGTTTTGATCACTACGCG  217
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGTACCTGAACGCGAACAGGTGCGTGCGCGGAAAACTCAGCGGTTTTGATCACTACGCG  120

Query  218  AACCTGGTGCTGTCGGATGCTCTAGACTGCCGAACGGGTGCGCAACTCGGTCAGGTTTGG  277
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  AACCTGGTGCTGTCGGATGCTCTAGACTGCCGAACGGGTGCGCAACTCGGTCAGGTTTGG  180

Query  278  ATCCGAGGCAACAGTGTCGTTTCAGTGGACCTGCTTCGGGATGTGAACGCAGACCGCACG  337
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  ATCCGAGGCAACAGTGTCGTTTCAGTGGACCTGCTTCGGGATGTGAACGCAGACCGCACG  240

Query  338  GAGCCACCGACCGGCACCGGCTCTGTAGCCGATGACCCCGTGGGTTCTTCGCTTAGCAGC  397
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GAGCCACCGACCGGCACCGGCTCTGTAGCCGATGACCCCGTGGGTTCTTCGCTTAGCAGC  300
```

SmD1 gene inserted into pQLinkN (The RBS sequence is in blue and the LIC sequence is in bold; pSR717)

```
Query  39   TACTTCCAATCCCACGAGGAGAAATTAACT 89
            ||||||||||||||||||||||||||||||
Sbjct       ------------------------------

Query  88   ATGACCCCCTTGCTTTATTTCCTAACTCGCCTTCGAGGTGCCACTGTTACTGTTGAGCTG  147
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGACCCCCTTGCTTTATTTCCTAACTCGCCTTCGAGGTGCCACTGTTACTGTTGAGCTG  60

Query  148  AAAGATGGGACGAAGGCCACGGGAACTGTACAGCGAGTGGATAATGAGATGAACGTTTAC  207
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   AAAGATGGGACGAAGGCCACGGGAACTGTACAGCGAGTGGATAATGAGATGAACGTTTAC  120

Query  208  CTGCTGAACGCTTCCGTTACTGGAAAACCTCCAGCCGAGCTCCCCTCTGCTTCTCTGGAG  267
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  CTGCTGAACGCTTCCGTTACTGGAAAACCTCCAGCCGAGCTCCCCTCTGCTTCTCTGGAG  180

Query  268  ACGCACGCGGCCCAGGTCGTCGCCCCTTGGACCGAGCGATTCAGTGAACCGGATGCCTCA  327
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  ACGCACGCGGCCCAGGTCGTCGCCCCTTGGACCGAGCGATTCAGTGAACCGGATGCCTCA  240

Query  328  GCTATGAGTCGTCGGAATCAACCTCAGCAAAAGGCGAGAGAATATCGAATCCGGGGATCT  387
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCTATGAGTCGTCGGAATCAACCTCAGCAAAAGGCGAGAGAATATCGAATCCGGGGATCT  300

Query  388  ACGGTTCGATATATCATTCTGCCCGAGTCATTGAACCTGGAGAGCGCTTTGAAAGAAACG  447
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  ACGGTTCGATATATCATTCTGCCCGAGTCATTGAACCTGGAGAGCGCTTTGAAAGAAACG  360
```

```
Query  448  CGCAAGTTCAGCCCCAGAACACGATATCAGAAAGAGAGACAC  489
            |||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CGCAAGTTCAGCCCCAGAACACGATATCAGAAAGAGAGACAC  402
```

SmE gene inserted into pQLINKH (The RBS sequence is in blue, the start codon in pink, the seven histidines is in green, the TEV site is in red, and in **bold** is the PmlI restriction site sequence; pSR723)

```
Query  165  GAATTCAGGAGAAATTAACTATGAAACATCACCATCACCATCACCATGAGAATCTGTACTTCCAATCCCACG 128
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct        ---------------------------------------------------------------------------

Query  129  CCGAAGGACGCTCTGGACAGACGGATAGTTCCAGAGCAGTTGTTAGCAACGCTGGCGCGC  188
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4    CCGAAGGACGCTCTGGACAGACGGATAGTTCCAGAGCAGTTGTTAGCAACGCTGGCGCGC  63

Query  189  CAACAAGCCCGCGTTGAGGTCTGGTTATTCGAAAACACCAGATACTCTCTGGAAGGCACC  248
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  64   CAACAAGCCCGCGTTGAGGTCTGGTTATTCGAAAACACCAGATACTCTCTGGAAGGCACC  123

Query  249  TTGCGCGGCTTCGACGAACACACCAATCTAGTTCTGGTCGACACCGTGGAGCAGTGGGGA  308
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  124  TTGCGCGGCTTCGACGAACACACCAATCTAGTTCTGGTCGACACCGTGGAGCAGTGGGGA  183

Query  309  AGTACTGCAAAGCATAAGCGGCGGACGGTTGCTCTAGGGACGATCCTCCTCAAAGGCGAA  368
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  184  AGTACTGCAAAGCATAAGCGGCGGACGGTTGCTCTAGGGACGATCCTCCTCAAAGGCGAA  243

Query  369  AACGTCGTCCTCGTTCGGTCGCTGGGGATGCCAACCCAGCGAAAAGAGGTCACGCACAGC  428
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  244  AACGTCGTCCTCGTTCGGTCGCTGGGGATGCCAACCCAGCGAAAAGAGGTCACGCACAGC  303

Query  429  GCGACTCGGGAG  440
            ||||||||||||
Sbjct  304  GCGACTCGGGAG  315
```

Appendix 2

Table 24 Summary of the expression of the proteins using different constructs: After insertion of the genes, each vector was given a name (here called pSR). The Snu114 gene was inserted into pMCSG2 by another lab member. (*) represents the insertion of the gene into pMCSG2. (**) represents the insertion of the gene into pPICZA. (***) represents the insertion of the gene into pQLinkH.

Gene-containing construct	pSR#	Sequenced	Expression
Dib1	634	Yes	Successful
Snu114	647	Yes	Failed
Prp8	655	Yes	Failed
Prp8*	797	Yes	Failed
Brr2	656	Yes	Failed
Brr2**	855	Yes	Failed
Snu114*	767	Yes	Failed
Sm B	715	Yes	Not tested
Sm D3	716	Yes	Not tested
Sm D2	717	Yes	Not tested
Sm D1	719	Yes	Not tested
Sm E	720	Yes	Not tested
Sm F	721	Yes	Not tested
Sm G	723	Yes	Not tested
Sm E***	724	Yes	Not tested
Sm D1/D2	733	No	Not tested
Sm D3/B	735	No	Not tested
Sm E/G	734	No	Not tested
Sm E-HIS/G	736	No	Not tested
Sm D3/B/D1/D2	739	No	Not tested
Sm D3/B/D1/D2/F	743	No	Not tested
Sm F-HIS/ E/G	744	No	Not tested
Sm E-HIS/G/ D3/B/D1/D2/F	752	No	Successful
Sm F-HIS/ E/G/ D3/B/D1/D2	751	No	Successful
SmE-HIS/G/ D3/B/D1/D2/F/U5	755	No	Not tested
SmF-HIS/E/G/ D3/B/D1/D2/U5	753	No	Not tested
Sm Prp8/Dib1	708	No	Failed
Sm Brr2/ Prp8/Dib1	712	No	Failed
SmFHIS/E/G/D3/B/D1/D2/U5/Brr2/Prp8/Dib1	829	No	Failed
SmFHIS/E/G/D3/B/D1/D2/U5/Brr2/Prp8/Dib1/Snu114	762	No	Failed