National Institute for Medical Research

Division of Molecular Structure

In vitro and *in vivo* studies of RNA recognition by Rna15

A thesis submitted by

Laura Elizabeth Robertson

In partial fulfilment of the requirements of

University College London

for the degree of Doctor of Philosophy in Structural and Molecular Biology

July 2014

Declaration

I, Laura Elizabeth Robertson, declare that the work presented in this thesis is my own. This work was carried out in the laboratory of Dr. Ian A. Taylor in the Division of Molecular Structure at the MRC National Institute for Medical Research. Where information has been derived from other sources, I confirm this has been indicated in the text.

Acknowledgements

I would firstly like to thank my supervisor Dr. Ian Taylor for giving me the opportunity to work on this project. I have learnt a lot and know that I am very lucky to have worked in such a brilliant lab. Also, I would like it put in writing that I beat both Dr. Ian Taylor and Dr. Steve Smerdon at pool, that was probably the easiest part of the PhD by far…

I am also very grateful to current and past members of the Taylor lab. To Dave Goldstone who taught me a lot and wound me up even more but it made for such a fun time and I am very grateful to him. To Joseph Hedden, for his help with getting the yeast side of my project off the ground and providing A LOT of laughs, some actually with him. I would like to thank Valerie, who was always there for advice and became my "London mum" over the last 5 years, I will miss you a lot. I would like to thank Laurence Arnold for his advice and for the vast collection of extremely work appropriate jokes. I would also like to thank David Schwefel and Neil Ball for their help and advice.

To Dr. Steve Martin for providing so much help with the fluorescence, I hope it wasn't too painful! To Prof. Andres Ramos, for his help and guidance throughout this PhD. To super cool Kat for her help with the NMR spectroscopy. To Dr. Peter Thorpe, for his advice on all things yeast. To Simon Pennell, who deserves a gigantic bottle of whisky for the critical reading of this thesis. I am extremely grateful to have worked in such a fantastic department with people who have been so kind and generous with their help and support. I would especially like to thank, Aylin, Manuela, Lizzi, Jan, Matt and Tom for their support and putting up with my moaning/homesickness!

I would like to thank my mum and dad for their unconditional love and encouragement. I would also like to thank my family for their love and support. I'd like to thank my very close friends, Stephanie Laing and Cara Howie for their love, support and offerings of copious amounts of tea and pizza. Finally, I'd like to thank Stuart, who I am so lucky to have met and without whom I never would have managed this PhD. I know it has meant that we were apart for 4 years but hopefully it'll be worth it!!! I dedicate this thesis to you!

Abstract

For successful gene expression in eukaryotes, mRNA transcripts are processed in the 5' and 3' untranslated regions (UTRs) and non-coding mRNA is spliced out. These processes are crucial in determining the fate of the mRNA transcript. Rna14 and Rna15 are subunits of Cleavage factor 1A (CF1A) and required for 3' end processing in *S. cerevisiae*. Structural and biophysical data have determined a number of residues within the RNA recognition motif (RRM) of Rna15 that interact directly with RNA. However, although the crystal structure of the RRM revealed a mechanism for the preferential recognition of G/U nucleotides by Rna15 the sequence specificity for Rna15 is still extensively debated. This thesis applies combination of *in vivo* and *in vitro* techniques aimed to characterise Rna15-RNA binding, Rna15 interaction with Rna14 and examine consequences for RNA processing and yeast viability *in vivo*. An *in vitro* mutational/biophysical analysis is presented that reveals the residues essential for the Rna15-RNA interaction and application of NMR-quantified Scaffold Independent Analysis (SIA) demonstrates a clear GU-bias in the *in vitro* consensus sequence. However, given these strong effects, surprisingly only extensive mutation of the RRM produces growth defects in *S. cerevisiae* and qRT-PCR experiments employing a small subset of genes show only slight effects on polyA site selection. By contrast, an RNA-sequencing (RNA-Seq) global analysis of expression and transcriptional readthrough reveals that expression of over 100 *S. cerevisiae* genes is severely affected when the RRM of Rna15 is deleted and in addition the 3'-UTR of the mRNA of a sample set of 40 genes is significantly different to wild type. These results indicate that only severe reduction of Rna15-RNA interactions result in defects in transcriptional and 3' end processing, hypothesized to be due in part to functional redundancy. Nevertheless, the global changes observed upon deletion of the Rna15 RRM are striking and reinforce the link between 3'-end processing, transcriptional regulation and gene expression.

Table of Contents

Declaration 2

Acknowledgements 3

Abstracts 4

Table of Contents 5

List of figures 9

List of tables 11

List of abbreviations 12

1. Introduction 13

1.1 Eukaryotic mRNA processing 13
1.1.1 5’ end capping 13
1.1.2 mRNA splicing 18
1.1.3 mRNA export 20

1.2 mRNA processing at the 3’ end 22
1.2.1 Protein factors involved in polyadenylation in *S. cerevisiae* 23
1.2.1.1 CFI 24
1.2.1.1.1 Pcf11 26
1.2.1.1.2 Clp1 29
1.2.1.1.3 Rna14 33
1.2.1.1.4 Rna15 39
1.2.1.1.5 Hrp1 48
1.2.1.2 CPF 54
1.2.1.2.1 Glc7 and Ssu72 56
1.2.1.2.2 Cft1 56
1.2.1.2.3 Mpe1 57
1.2.1.2.4 Yth1 57
1.2.1.2.5 Cft2 58
1.2.1.2.6 Ysh1 59
1.2.1.2.7 Ydh1 60
1.2.1.3 Poly(A) polymerase 61
1.2.1.4 Poly(A) binding protein 62

1.3 Sequences that direct polyadenylation 63

1.4 Model for polyadenylation in *S. cerevisiae* 66

1.5 Coupling processing and transcription 67
1.5.1 5’ end capping and 3’ end processing 67
1.5.2 Splicing and 3’ end processing 68
1.5.3 Export and 3’ end processing 68
1.5.4 Transcription and 3’ end processing 70
1.5.4.1 The importance of the poly(A) site in regulation of transcription 70
1.5.4.2 The role of Ssu72 and Pta1 in transcription regulation 71
1.5.4.3 Gene looping and transcription regulation 72
1.5.4.4 Transcription termination – The allosteric and torpedo models 73

1.6 Mammalian homologues **77**

1.7 RNA sequencing **79**

1.8 Research objectives **82**

2. Materials and Methods 84

2.1 Bioinformatics **84**
2.1.1 DNA and protein information 84
2.1.2 RNA sequencing analysis 84
2.1.2.1 Global analysis of expression 84
2.1.2.2 Analysis of 3' end processing 85

2.2 Molecular Biology **86**
2.2.1 Bacterial Strains 86
2.2.2 Yeast Strains 86
2.2.3 Plasmid constructs employed in *in vitro* studies 87
2.2.4 Mutant constructs employed in *in vitro* studies 87
2.2.5 Plasmid constructs employed in *in vivo* studies 88
2.2.6 Mutant constructs employed in *in vivo* studies 88
2.2.7 Polymerase chain reaction (PCR) 90
2.2.8 Agarose gel electrophoresis 91
2.2.9 Determination of DNA concentration 91
2.2.10 Restriction enzyme digest 91
2.2.11 Ligation 92
2.2.12 Ligation independent cloning 93
2.2.13 Transformation of bacterial strains 94
2.2.14 Site-directed mutagenesis 94

2.3 Protein purification **95**
2.3.1 Protein expression 95
2.3.2 Cell lysis and sample preparation 96
2.3.3 Strep-Tactin® affinity chromatography 96
2.3.4 Nickel-affinity chromatography 97
2.3.5 Size exclusion chromatography 97
2.3.6 SDS-PAGE 98
2.3.7 Protein concentration, storage and dialysis 98
2.3.8 Determination of protein concentration 99

2.4 *In vitro* analysis of Rna15-RNA binding **99**
2.4.1 Fluorescence spectroscopy of Rna15 constructs 99
2.4.2 Fluorescence spectroscopy of the Rna14-Rna15 complex with RNA 100
2.4.3 Competition assays 101
2.4.4 Scaffold independent analysis (SIA) 102
2.4.5 Nuclear magnetic resonance (NMR) 103
2.4.6 Circular Dichroism 103
2.4.7 Fluorescence anisotropy experiments 103

2.5 Yeast methods **104**
2.5.1 Yeast transformation 104
2.5.2 5-FOA plasmid shuffle 105
2.5.3 Measuring *S. cerevisiae* cell density 106
2.5.4 Growth rate assays 107
2.5.5 Viable cell count 107
2.5.6 Western blotting 108
2.5.7 Preparation of total RNA 110
2.5.8 Analysis of RNA quality 110
2.5.9 Quantitative real time PCR 111
2.5.10 RNA sequencing – library preparation and sequencing 115

3. *In vitro* analysis of Rna15-RNA binding 116

3.1 Introduction and Overview 116

3.2 Analysis of Rna15, Rna15 (2-94) and Rna15 (2-103) RNA binding *in vitro* 117
3.2.1 Cloning Rna15 *in vitro* constructs 117
3.2.2 Expression and purification of Rna15 *in vitro* constructs 117
3.2.3 Principles of fluorescence spectroscopy 120
3.2.4 Fluorescence spectroscopy of Rna15, Rna15 (2-94) and Rna15 (2-103) 122

3.3 Rna15-RNA interactions – Investigations of residues that direct RNA binding 123
3.3.1 Cloning, expression and purification of Rna15 (2-103) point mutations 124
3.3.2 Circular dichroism and mass spectroscopy 126
3.3.3 Analysis of residues in the RRM that direct RNA binding 129
3.3.3.1 Mutational analysis of residues involved in site I & II RNA binding 130
3.3.3.2 Mutational analysis of residues involved in site III RNA binding 135
3.3.3.3 Mutational analysis of surface lysine residues 137

3.4 Summary 142

4. Analysis of Rna15-RNA binding *in vivo* 144

4.1 Introduction and Overview 144

4.2 Establishing the Rna14 and Rna15 haploid strains 144

4.3 Analysis of Rna15-RRM mutations on growth phenotype of *S. cerevisiae* 149

4.4 Deletion of the RRM is lethal on 5-FOA at 37°C 159

4.5 Analysis of transcription in Rna15 mutants 166

4.6 Preparation of libraries for RNA sequencing 168

4.7 Analysing RNA sequencing data 172
4.7.1 Analysis of expression of genes using RNA sequencing 172
4.7.2 Analysis of the 3' UTR of genes using RNA sequencing 180

4.8 Analysis of processing events using RNA sequencing 183
4.8.1 Analysis of processing in genes involved in metabolism and translation 183
4.8.2 Analysis of processing in genes involved in cell cycle and cell wall maintenance 186
4.8.3 Analysis of poly(A) site selection in ACT1, TDH2, ADH1 and YPT1 189
4.8.4 RNA sequencing conclusions 193

4.9 Summary 194

5. Characterising specificity and tetramer-RNA binding 196

5.1 Introduction and Overview 196

5.2 Generating a Rna15 consensus sequence 197

5.3 Identifying the orientation of the RNA oligonucleotide 204

5.4 The effect of the Rna14-Rna15 tetramer on RNA binding 211
5.4.1 Expression of Rna14 and Rna15 211
5.4.2 Fluorescence titration of the Rna14-Rna15 tetramer with RNA 212

5.5 Deletion of the Rna15 binding interaction domain is lethal in *S. cerevisiae* 218

5.6 Summary **221**

6. Discussion 222

6.1 The importance of the C-terminal tail of Rna15 **222**

6.2 The importance of residues within the RRM of Rna15 **223**

6.3 Mutational analysis of the RRM *in vivo* **226**

6.4 Nucleotide specificity of Rna15 **228**

6.5 The influence of Rna14 on Rna15-RNA binding **231**

6.6 Concluding remarks **232**

7. Appendix 234

7.1 Oligonucleotide primers for PCR **234**

7.1.1 LIC primers for cloning *in vitro* constructs 234
7.1.2 Primers for cloning *in vivo* constructs 234
7.1.3 Primers for mutagenesis 235

7.2 Circular dichroism of Rna15 2-103 constructs **237**

7.3 Mass spectrometry of Rna15 and mutant constructs **238**
7.3.1 Representative mass spectrum deconvolution report 238
7.3.2 Mass spectrometry results for Rna15 and mutant constructs 239

7.4 Analysis of RNA quality **240**

7.5 Primers and probes employed in qRT-PCR experiments **241**

7.6 C_T and quantity data from qRT-PCR experiments **242**
7.6.1 Rna15 Wild type 242
7.6.2 Rna15 Δ16-94 242

8.0 Bibliography 243

List of Figures

Figure 1.1	5’ end capping in *S. cerevisiae*	17
Figure 1.2	Schematic representation of the CF1A complex	25
Figure 1.3	Structure of the Pcf11-CID bound to a RNAPII CTD peptide	28
Figure 1.4	The Clp1 and ATP, Mg^{2+}-Pcf11 complex	32
Figure 1.5	Structural analysis of the Rna14-Rna15 binding interface	37
Figure 1.6	The crystal structure of residues 16-103 of Rna15	41
Figure 1.7	NMR structural analysis of Hrp1 and Rna15 mediated RNA binding	43
Figure 1.8	RNA binding at site I within the RRM of Rna15	45
Figure 1.9	RNA binding at site II within the RRM of Rna15	47
Figure 1.10	Hrp1-RNA structures	50
Figure 1.11	Schematic representation of CFI bound to RNA	53
Figure 1.12	The Allosteric and Torpedo models of transcription termination	76
Figure 2.1	Principle of qRT-PCR	112
Figure 3.1	Purification of Rna15 and Rna15 (2-94) protein after overexpression in *E. coli*	119
Figure 3.2	Jablonski diagram	121
Figure 3.3	Purification of Rna15 (2-103) mutant constructs	125
Figure 3.4	Circular dichroism of Rna15 (2-103) and mutants Y27A and K90E	128
Figure 3.5	Measured emission spectra of mutant S24T	129
Figure 3.6	Determination of K_d values for Rna15 2-103, Y27F,Y27A, R87K and R87A RNA binding using fluorescence spectroscopy	133
Figure 3.7	Determination of the K_d of Rna15 (2-103), K90E, K90R and K90A RNA binding	140
Figure 4.1	Schematic representation of the 5-FOA selection assay	146
Figure 4.2	Testing the Rna14 and Rna15 *S. cerevisiae* strains	148
Figure 4.3	Western analyses of Rna15 wild type and mutant constructs	149
Figure 4.4	5-FOA selection assay of point mutations within the Rna15 RRM	152
Figure 4.5	5-FOA selection assay demonstrating the effect of point mutations in the RRM of Rna15	153
Figure 4.6	Growth of *S. cerevisiae* Rna15 RRM point mutations on minimal media contaning 1.5% formamide	155
Figure 4.7	Growth assay to determine the effect of temperature and 1.5% formamide on growth of Rna15 RRM site I mutants.	156
Figure 4.8	Growth of site I mutants on 1.5% formamide at 37ºC	158
Figure 4.9	5-FOA assay of Rna15 double, triple and deletion mutants	161
Figure 4.10	Analysis of deletion of the RRM on growth phenotype of *S. cerevisiae*	162
Figure 4.11	Growth of the *S. cerevisiae* mutants	165
Figure 4.12	Transcription of a subset of genes in the *rna15 Δ16-94* mutant	167
Figure 4.13	ScriptSeq™ v2 RNA-Seq Library Preparation Kit Protocol	169

Figure 4.14	Formation of cDNA colonies on a flow cell	170
Figure 4.15	Illumina sequening of cDNA colonies	171
Figure 4.16	Expression of genes involved in metabolism	175
Figure 4.17	Expression of genes involved in maintaining the cell wall (A) and in transport of molecules (B).	176
Figure 4.18	Expression of genes involved in modifying RNA molecules (A) and in mediating protein modification (B)	177
Figure 4.19	Expression of genes induced by stress factors (A) and required for numerous cellular processes (B)	178
Figure 4.20	Expression of genes of unknown function	179
Figure 4.21	Representative aligned RNA sequencing data	181
Figure 4.22	Analysis of the 3' UTR	182
Figure 4.23	mRNA amounts relative to wild type at three distinct sites within the 3'UTR of genes	185
Figure 4.24	Relative amounts of mRNA for genes involved in maintenance of the cell wall and cell cycle	188
Figure 4.25	Poly(A) site selection in the ACT1 gene (A) and the TDH2 gene (B)	191
Figure 4.26	Poly(A) site selection in the ADH1 gene (A) and the YPT1 gene (B)	192
Figure 5.1	Scaffold independent analysis	200
Figure 5.2	SIA generated consensus sequence	202
Figure 5.3	Assigned 2D-NMR spectra of Rna15 (2-103)	205
Figure 5.4	Orientation of the RNA oligonucleotide bound to the RRM of Rna15	207
Figure 5.5	Analysis of the effect of the TET dye on RNA binding	210
Figure 5.6	Size exclusion purification of Rna14 and Rna15 protein after overexpression in *E. coli*.	211
Figure 5.7	The Rna14-Rna15 tetramer	213
Figure 5.8	Hypothetical tetramer-RNA binding	215
Figure 5.9	Rna14-Rna15-RNA binding curves	217
Figure 5.10	Deletion of the Rna15 interaction domain	219
Appendix 7.2	Circular dichroism of Rna15 2-103 mutant constructs	237
Appendix 7.3.1	Representative deconvoluted mass spectrum report	238
Appendix 7.4	Analysis of RNA quality	240

List of tables

Table 1.1	3' end processing machinery in *S. cerevisiae* and their mammalian counterparts	55
Table 2.1	Genotypes of *E. coli* strains	86
Table 2.2	Genotypes of *S. cerevisiae* strains	87
Table 2.3	Wild type constructs used in this study	89
Table 2.4	Typical PCR reaction conditions	90
Table 2.5	Typical cycle for PCR reactions	91
Table 2.6	Typical restriction digest conditions	92
Table 2.7	Ligation reaction	93
Table 2.8	Minimal media used for expression of ^{15}N labelled Rna15 2-103	96
Table 2.9	Typical yeast transformation mix	105
Table 2.10	Typical reaction setup for a qRT-PCR experiment	114
Table 2.11	Thermal cycling parameters used in a typical qRT-PCR reaction	114
Table 3.1	Fluorescence spectroscopy analysis of Rna15 deletion constructs	123
Table 3.2	Fluorescence spectroscopy analysis of Rna15 (2-103) site I and site II mutants	132
Table 3.3	Fluorescence spectroscopy analysis of Rna15 (2-103) mutants in site III	137
Table 3.4	Fluorescence spectroscopy analysis of lysine residues	139
Table 4.1	List of RNA15 RRM point mutations	150
Table 4.2	Summary of Rna15 wild type and mutant RNA binding affinites, viability and growth rate assays in *S. cerevisiae*	164
Table 4.3	Genes involved in metabolism	184
Table 4.4	Genes involved in translation	184
Table 4.5	Genes involved in maintaining the composition of the cell wall	187
Table 4.6	Genes involved in regulation and progression of the cell cycle	187
Table 5.1	RNA pools employed in SIA experiments	197
Table 5.2	SIA results showing a clear consensus for G/U rich nucleotides	201
Table 5.3	Binding constants of Rna15 (2-103) with SIA generated RNA oligonucleotides	203
Table 5.4	Fluorescence spectroscopy of the Rna15-Rna15 tetramer with RNA	216
Appendix 7.1.1	LIC primers for cloning *in vitro* constructs	234
Appendix 7.1.2	Primers for cloning *in vivo* constructs	234
Appendix 7.1.3	Primers for mutagenesis	235
Appendix 7.3.2	Mass spectrometry results for Rna15 and mutant constructs	239
Appendix 7.5	Primers and probes employed in qRT-PCR experiments	241
Appendix 7.6.1	C_T and quantity data from qRT-PCR experiments – Rna15 Wild type	242
Appendix 7.6.2	C_T and quantity data from qRT-PCR experiments – Rna15 Δ16-94	242

List of abbreviations

Abbr.	Definition	Abbr.	Definition
5-FOA	5-fluoroorotic acid	mRNA	Messenger RNA
ATP	Adenosine triphosphate	mRNP	Messenger ribonucleoprotein
AU	Absorbance units	N.I.M.R.	National Institute for Medical Research
CD	Circular dichroism	NMR	Nuclear magnetic resonance
cDNA	Complementary DNA	OD	Optical density
CF I	Cleavage factor I	PAGE	Polyacrylamide gel electrophoresis
CF II	Cleavage factor II	PAS	Polyadenylation signal
CF1A/B	Cleavage factor 1A/B	PCR	Polymerase chain reaction
CID	C-terminal interacting domain	PDB	Protein Data Bank
CPF	Cleavage and polyadenylation factor	PE	Positioning element
CPSF	Cleavage and polyadenylation specificity factor	pI	Isoelectric point
CstF	Cleavage stimulation factor	qPCR	Quantitative real time PCR
C_T	Threshold cycle	qRT-PCR	Quantitative reverse transcription PCR
CTD	C-terminal domain	RACE	Rapid amplification of cDNA ends
Da	Dalton	RNA	Ribonucleic acid
DNA	Deoxyribonucleic acid	RNAPII	RNA polymerase II
DSE	Downstream element	RNA-seq	RNA sequencing
DTT	Dithiothreitol	RRM	RNA recognition motif
EDTA	Ethylenediaminetetraacetic acid	SDS	Sodium dodecyl sulphate
EE	Efficiency element	SEC	Size exclusion chromatography
EM	Electron microscopy	SELEX	Systematic evolution of ligands by exponential enrichment
EST	Expressed sequence tag	SIA	Scaffold independent analysis
HA	Human influenza haemaglutinin	snoRNA	small nucleolar RNA
IPTG	Isopropyl β-D-1-thiogalactopyranoside	snRNA	small nuclear RNA
kDa	Kilodalton	TET	Tetrachlorofluoroscein
LIC	Ligation independent cloning	UTR	Untranslated region
M_r	Molecular mass	YPD	Yeast peptone-dextrose medium

1. Introduction

1.1 Eukaryotic mRNA processing

The central dogma of molecular biology, DNA-RNA-Protein, is an essential feature of cellular life and describes the flow of information from DNA to the resulting proteins. Messenger RNA (mRNA) is central to gene expression as it carries the message encoded in the DNA from the nucleus to the cytoplasm for translation into protein. Processing allows entry into the cytoplasm where the translation machinery resides and additionally affords the message protection from degradation to ensure it is efficiently translated. Processing of mRNA comprises addition of a 5' cap, synthesis of a 3' poly(A) tail and deletion of non-coding intron sequences. The work presented in this thesis focuses on the process of polyadenylation which will be introduced in detail. However, it is pertinent to briefly describe the other processing events which immature mRNAs undergo.

1.1.1 5' end capping

Capping of mRNA at the 5' end is essential for maturation of eukaryotic mRNA transcripts and important for the maintenance of mRNA integrity in preventing degradation by 5'-3' exonucleases present in the nucleus and cytoplasm (1,2). In addition, the 5’ cap is required for successful translation initiation as it is recognised by the translation machinery and this interaction promotes mRNA direction toward ribosomes. Binding by the translation initiation factor, eIF4F, results in mRNA circularisation further protecting the mRNA and aiding translation (3,4). Capping is also tightly coupled with transcription and occurs within elongation of the first 25-30 nucleotides. Capping, transcription initiation and elongation are co-dependent processes involving proteins with functions in more than one process (5).

The 5' cap consists of a 7-methylguanosine linked to the end of the transcript by a 5'-5' triphosphate bridge (6). In yeast, three classes of enzymes are responsible for formation of the cap structure, a RNA triphosphatase (Cet1), a RNA guanylyltransferase (Ceg1) and a RNA methyl transferase (Abd1) (7). The capping machinery localises to the 5' end of the mRNA transcript as it emerges from RNA polymerase II (RNAPII) during transcription (8). Recruitment is dependent on the phosphorylation state of the C-terminal domain (CTD) of RNAPII that is composed of a number of heptad repeats $YSPTSPS_n$. Phosphorylation of serine 5 and serine 7 (S5, S7) within the heptad repeats results in recruitment of Ceg1 and Abd1 to the 5’ end of the nascent transcript (9,10). The Kin28 subunit of transcription factor TFIIH is responsible for phosphorylation at this site (9). Cet1 does not associate with the phosphorylated CTD of RNAPII and is recruited to the transcript via interaction with Ceg1 (11–14).

The CTD of RNAPII provides a platform for the capping enzymes and acts to stabilise them and modulate their enzymatic activities. Phosphorylation of RNAPII therefore has a major influence on the capping process (14,15). During transcription initiation the CTD of RNAPII is steadily dephosphorylated leading to dissociation of the Ceg1-Cet1 complex prior to transcription elongation. However, association of Abd1 continues into the coding region (9).

In *S. cerevisiae*, the process of capping is essential to cell growth and mutation of either the guanylyltransferase or methyltransferase that result in loss of catalytic activity are lethal *in vivo* (16–20). Successful addition of a 7-methylguanosine to the end of the newly synthesised transcript first requires Cet1 hydrolysis of the 5' triphosphate to a diphosphate (6). Following this phosphatase reaction, the 5' end is then capped with guanine mono-phosphate (GMP) by Ceg1 (Figure 1.1) (6). Ceg1 contains a

number of sequence and structural motifs which are conserved in their order and spacing across capping enzymes from DNA viruses to higher eukaryotes (21). GTP is bound to motifs I, III, IIIa, IV, V and VI and the crystal structure of the *Chlorella* Ceg1 demonstrated that binding of GTP to the motifs induced a conformational change in the enzyme allowing hydrolysis to occur (22). The enzyme is composed of two domains with a deep cleft between them which shifts from an open to closed state when GTP is bound. The closed state brings motif VI into contact with the γ and β phosphates of GTP resulting in their reorientation so that they are exposed to attack by the charged amino group of a conserved lysine residue found in motif I within a KXDG sequence characteristic of nucleotidyl transferases (22–24). This lysine residue is also conserved across diverse species (16,18,21,25–28). The evolutionary importance of these conformational changes is demonstrated by analysis of the guanylyltransferase enzyme in *S. cerevisiae.* The same residues responsible for maintaining contacts with GTP in the *Chlorella* crystal structure were mutated in the yeast homologue and resulted in loss of capping at the 5' end *in vivo* (20,28).

Regulation of the capping pathway is performed by the capping enzymes themselves. The 33-kDa subunit of the vaccinia virus capping enzyme is able to stimulate methyltransferase activity 50- to 100- fold by association with the catalytic subunit (22-23). In *S. cerevisiae* similar regulation is observed where the guanylyltransferase activity of Ceg1 is regulated by Cet1 (15). GTP hydrolysis by Ceg1 increased ~13-fold when in complex with Cet1 and an increase in GTP affinity was also observed (15). In addition, the CTD of RNAPII and Cet1 have been shown to allosterically regulate Ceg1 activity (14). Association of Ceg1 with the CTD alone inhibits guanylyltransferase activity but this inhibition is relieved upon binding of Cet1 to form a Ceg1/Cet1/CTD heterotrimer (14,15). In higher eukaryotes, a single protein

performs both the phosphatase and guanyltransferase roles (8). The mouse guanylyl transferase domain is able to functionally complement deletion of yeast, Ceg1 and the full length protein is able to complement deletion of Cet1 *in vivo* (13,21,31,32).

Following addition of GMP by Ceg1, the cap is then methylated by a methyltransferase (Figure 1.1). In *S. cerevisiae* this function is performed by Abd1 (6). In addition to recruitment of the capping machinery, phosphorylation of the CTD also mediates stimulation from transcription initiation to elongation (33). In addition to its methyltransferase activity in 5' capping, Abd1 has been shown to exert other effects on transcription (34). Temperature sensitive mutations in Abd1 have been isolated that result in a reduction in occupancy of RNAPII at the 5' end of genes PGK1, ENO2, GAL1 and GAL10 (34). In genes TEF1 and ACT1, RNAPII was shown to be distributed abnormally throughout the gene compared to wild type, most likely caused by aberrant transition from transcription initiation to elongation (34). Indeed, Abd1 has been demonstrated to associate with the CTD of RNAPII during transcription elongation (9). Together, these results demonstrate a role for Abd1 in transcription elongation efficiency, possibly as a checkpoint to ensure transcription proceeds after successful addition of the 5' cap (34).

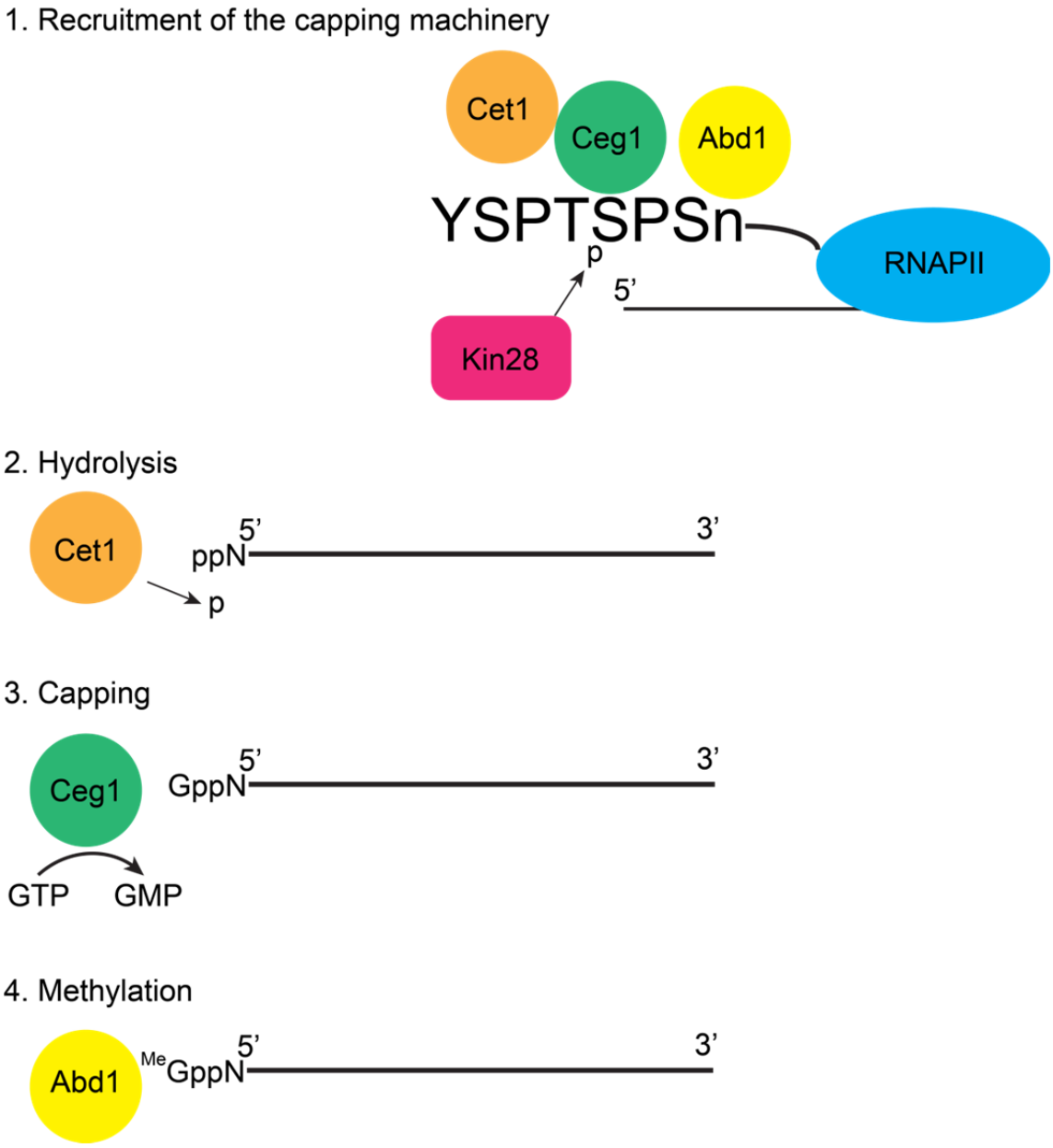


Figure 1.1 5' end capping in *S. cerevisiae*. 1. The capping machinery is recruited to the nascent transcript by association with the RNAPII CTD. **2.** Cet1 completes hydrolysis of the of the 5' triphosphate. **3.** Ceg1 then caps the 5' end with GMP. **4.** Abd1 completes formation of the 5' cap following addition of a methyl group.

1.1.2 mRNA Splicing

Following transcription, the resulting RNA product is a precursor or pre-mRNA which contains non-coding introns as well as coding exons. The introns are removed by the spliceosome, a large protein RNA complex, guided by short consensus sequences present at the intron-exon boundary (35–37). These consist of the 5' splice site, the branch point sequence (present in the intron) and the 3' splice site (38). The spliceosome is required to fold the mRNA transcript bringing these sequences together so that the intron in question can be excised (38). Different splice sites can be recognised and paired within a single mRNA transcript leading to alternative splicing. This means that a single mRNA transcript can encode different protein isoforms resulting in diverse and large proteomes. Sequencing of mRNA products reveals that over 95% of the human genome is alternatively spliced, with the products determined predominantly by stage of development, tissue type and external stimulus (39). In *S. cervisiae*, the spliceosome consists of five small nuclear RNAs; U1, U2, U4, U5 and U6 snRNAs, and over 100 different polypeptides (40). The spliceosome in humans is more complex with the major spliceosome complex consisting of the same snRNAs but over 300 different proteins. Humans also have a second spliceosome unit called the minor spliceosome which consists of U11, U12, U4atac, U5 and U6atac snRNAs (41). Although there are some major differences between the yeast and human systems, a high degree of conservation is observed whereby 85% of the yeast spliceosome has a human homologue (40). The yeast spliceosome could be described as an "evolutionary conserved core" around which the human system has evolved (40). Adaptation and evolution of the human spliceosome means that the spliceosome machinery is able to exert an increasingly complex role in alternative splicing. The spliceosome proteins bind to the snRNAs to form ribonucleoprotein complexes and function to direct splicing. Other proteins are required to regulate the timing of the process and confirm the final product has been

correctly spliced (35–37).

Due to the number of proteins and the highly complex nature of splicing, it has taken many years to elucidate the mechanism (35–37). The current understanding of the process is as follows. In the human spliceosome, U1 begins splicing by binding to the 5' exon splice site along with other proteins of the splicing machinery which bind to the branchpoint of the intron. The "A complex" is then formed by displacement of splicing proteins by U2 and subsequent binding of U2 to the intron to be excised. The "B complex" is then formed by addition of the U4/U6.U5 tri-snRNP. A catalytically active complex, termed the B* complex, carries out excision at the 5' intron requiring dissociation of U1 and U4, addition of the Prp19 complex, destabilisation of proteins bound to U2 and association of the pre-mRNA, U2 and U6 snRNAs (40,42–44). The B* complex is present on pre-mRNA allowing catalytic access to the consensus sequences determining both the 5' and 3' splice sites and the branch point present within the intron (38). The phosphodiester bond bridging the 5' exon with the intron is cleaved by attack of a 2' hydroxyl present in an adenosine residue within the branch point consensus sequence. This complex, where the intron has dissociated from the 5' exon, is called the C1 complex. The 5' exon is then free to attack the 3' splice site resulting in the C2 complex and exon ligation. The spliceosome is then recycled to begin intron exclusion at a new site (42,43).

Splicing, like capping, is tightly coupled to transcription (45,46). The 65 kDa component of the U2 snRNP mammalian splicing factor has been demonstrated to bind phosphorylated S2 in the CTD of RNAPII (47,48). In mammals, truncation of the CTD of RNAPII results in inefficient splicing and loss of co-localisation of splicing factors (49,50). Moreover, reconstituted splicing experiments *in vitro* using HeLa cell extracts

demonstrate that phosphorylation of the CTD of RNAPII result in the recruitment of the spliceosome, hyper-phosphorylation was shown to stimulate splicing whilst hypo-phosphorylation inhibited splicing activity (51).

1.1.3 mRNA Export

Following transcription and RNA processing, the transcript is exported from the nucleus through nuclear pores (5,52,53). This requires the function of many proteins which bind to and stabilise the nascent mRNA during transcription to form a messenger ribonucleoparticle (mRNP) (5,52,53). Binding of proteins which signal mRNA export from the nucleus also work to stop RNA:DNA hybridisation (54). Such pairing can act to stall RNA polymerase II along the DNA template resulting in decreased transcriptional efficiency (54). Furthermore, movement of DNA during hybridisation can result in DNA damage which can lead to genome instability. In *S. cerevisiae*, this is observed in cells lacking the export factor THO (55).

THO is a multisubunit complex which is required to direct export of mRNA from the nucleus. The core complex consists of Tho2, Hpr1, Mft1 and Thp2 (56). Other proteins such as Sub2, Yra1 and Mex-67 have also been observed to be associated with the complex (57,58). THO associates with chromatin during transcription and recruits Sub2 to the emerging mRNA (59,60). Sub2 then recruits Yra1 which acts as an adaptor protein, linking Mex-67 with the mRNA. Interaction with Mex-67 displaces Sub2 from Yra1 as interaction with both proteins occurs through the same domain (61–64). Export is facilitated through interaction of Mex-67 with nucleoporins lining the nuclear pore (5,52,53). The formation of the THO-Sub2-Yra1 complex has been termed the TREX complex due to its ability to link transcription and export (65). The THO complex is an

essential shuttling pathway with conserved protein counterparts in higher eukaryotes (5,52,53).

Mex-67 relies on adaptor proteins like Yra1 due to its low affinity for pre-mRNA (52). It has been demonstrated that multiple export pathways exist in *S. cerevisiae* that do not require the function of Yra1 for recruitment of Mex-67 (66,67). Indeed, Yra1 does not associate with all yeast transcripts (68). In addition to Yra1 mediated association, THO can also transport Mex-67 to the mRNP by association with the Hpr1 component (69). This interaction is dependent on the ubiquitin ligase, Rsp5 that polyubquitinates Hpr1 creating a substrate for the ubiquitin-associated (UBA) domain of Mex-67 or the proteasome (70). Following association of Mex-67 with the nucleoporin, Hpr1 is free to disassociate allowing entry of the pre-mRNA into the cytoplasm (71). This rearrangement of the pre-mRNA leads to proteasome directed degradation of polyubiquitinated Hpr1, possibly resulting in rapid displacement of other processing factors coating the pre-mRNA (72).

Along with the TREX complex, a TREX2 complex exists that directly links transcription and export. TREX-2 complex consists of Sac3, Thp1 and the SAGA histone acetylase complex (52). It has been suggested that TREX2 acts at different stages of mRNA export than the THO complex (53). THO associates with the chromatin early in export to prevent RNA-DNA hybridisation and mediates export through the formation of TREX (59–64). TREX-2 facilitates export immediately prior to mRNA entry to the cytoplasm by interacting both with the mRNA and the nuclear pore (73–75). Both Thp1 and Mex-67 interact with Sac3 (75). Sac3 associates with the nuclear pore complexes to allow export of the mRNP and along with Thp1 associates

with the SAGA complex through interaction with Sus1, a SAGA complex component (75,76).

1.2 mRNA processing at the 3' end

Polyadenylation is the name given to processing of mRNA at the 3' end and it is this process that this thesis will focus on. Apart from histone mRNAs, all mRNAs are polyadenylated in yeast. Polyadenylation consists of a cleavage reaction in the 3' untranslated region (UTR) of mRNA followed by addition of between 70 and 90 adenosine nucleotides. Polyadenylation is a highly complex process and unlike 5' capping requires the coordination of a large number of protein factors into complexes that are highly conserved throughout evolution (5,8,77–79). Many of these protein factors are recruited to the site of polyadenylation by interaction with the CTD of RNAPII. Phosphorylation of S2 and Y1 in the heptad repeats signals for recruitment of the 3’ end processing machiney during transcription elongation (9,80). Polyadenylation is directed through the presence of sequences in the 3' UTR. The cleavage reaction occurs at a point in the UTR determined by specific protein factors and sequences surrounding the cleavage site. In *S. cerevisiae*, cleavage normally occurs 10-30 nucleotides from an A-U rich polyadenylation signal (5,77).

Accurate processing of mRNA at the 3' end is of the utmost importance to maintain cell viability and growth (77). Normal 3' end processing facilitates export of mRNA to the cytoplasm (81). Disruption of polyadenylation leads to a 10 fold decrease in the ratio of cytoplasmic to nuclear mRNA which results in a decrease in protein expression (82). The poly(A) tail also enhances stability of the transcript acting to prevent degradation from the 3' end (83). Poly(A) binding protein (Pab1) also associates with the poly(A) tail to provide additional protection by 3'-5' exonucleases (83,84). In

addition, the poly(A) tail also enhances translation of mRNA into protein through indirect interaction with the 5' cap (3,83,85).

The process of polyadenylation is tightly coupled to transcription with components of the 3' end processing complex interacting with the CTD of RNAPII and transcription factors to mediate successful transcription (77,84,85). Alternative polyadenylation can occur where multiple polyadenylation sites exist within a single gene. This normally occurs in the 3' UTR and results in regulation of gene expression at the post-transcriptional stage (88).

Abnormal polyadenylation results in thalassemias (89,90) and lysosomal storage disorder in humans (91). Incorrect processing can affect the expression of oncogenes linking aberrant 3' end processing with cancer development (92). Viral proteins also target host cell processing machinery so that viral protein synthesis is favoured over host protein synthesis (93–96). Despite the importance of 3' end processing, relatively little mechanistic information is known and the full extent of the protein-protein interactions which contribute to successful processing is unclear. However, recent advances in technology such as RNA sequencing have meant that knowledge of proteins involved in 3' end processing is growing.

1.2.1 Protein factors involved in polyadenylation in *S. cerevisiae*

A number of distinct protein assemblies are responsible for 3' end mRNA processing. These were initially identified as four large protein complexes named CF I, CF II, PF I and Pap1 (97). CF I and CF II together were sufficient to carry out cleavage of mRNA at the 3' end without the aid of PF I and Pap1. However, optimal addition of the poly(A) tail required the presence of all processing factors (97). PF I was originally

thought to be required to link CF I and CF II with Pap1 since it was not involved in the cleavage step (97). It is now known that PF I, CF II and Pap1 work together to carry out their roles in polyadenylation in a large complex known as the Cleavage and Polyadenylation Factor complex (CPF) (98,99). Once the poly(A) tail has been added the PAN complex, consisting of the Pan2 nuclease and the Pan3 regulatory protein, trims the poly(A) tail to the correct length (100,101). Prior to the introduction of genome wide RNA sequencing analysis, polyadenylation experiments were performed on a number of highly transcribed housekeeping genes, such as GAL7, ACT1 and CYC1. A selection of these experiments are described in the next sections illustrate the role of the individual components of the 3' end processing machinery.

1.2.1.1 CF I

CF I was identified using *in vitro* reconstruction whereby precursor RNA molecules containing both the CYC1 and GAL7 gene poly(A) sites were monitored for cleavage and subsequent addition of the poly(A) tail with partially purified yeast extracts (97). These assays revealed that both cleavage and polyadenylation required CF I (97). Further purification of CF I revealed that it was made up of two complexes; CF1A and CF1B (102). CF1A is made up of Rna15, Rna14, Clp1 and Pcf11 (103–105) while CF1B is made up of just one protein, Hrp1 (106). Within CF1A, Rna14 and Rna15 form a 2:2 tetramer (107) accompanied by Pcf11 and Clp1 monomers to form a 6 component complex (108). A schematic representation of the components of CF1A is shown in Figure 1.2.

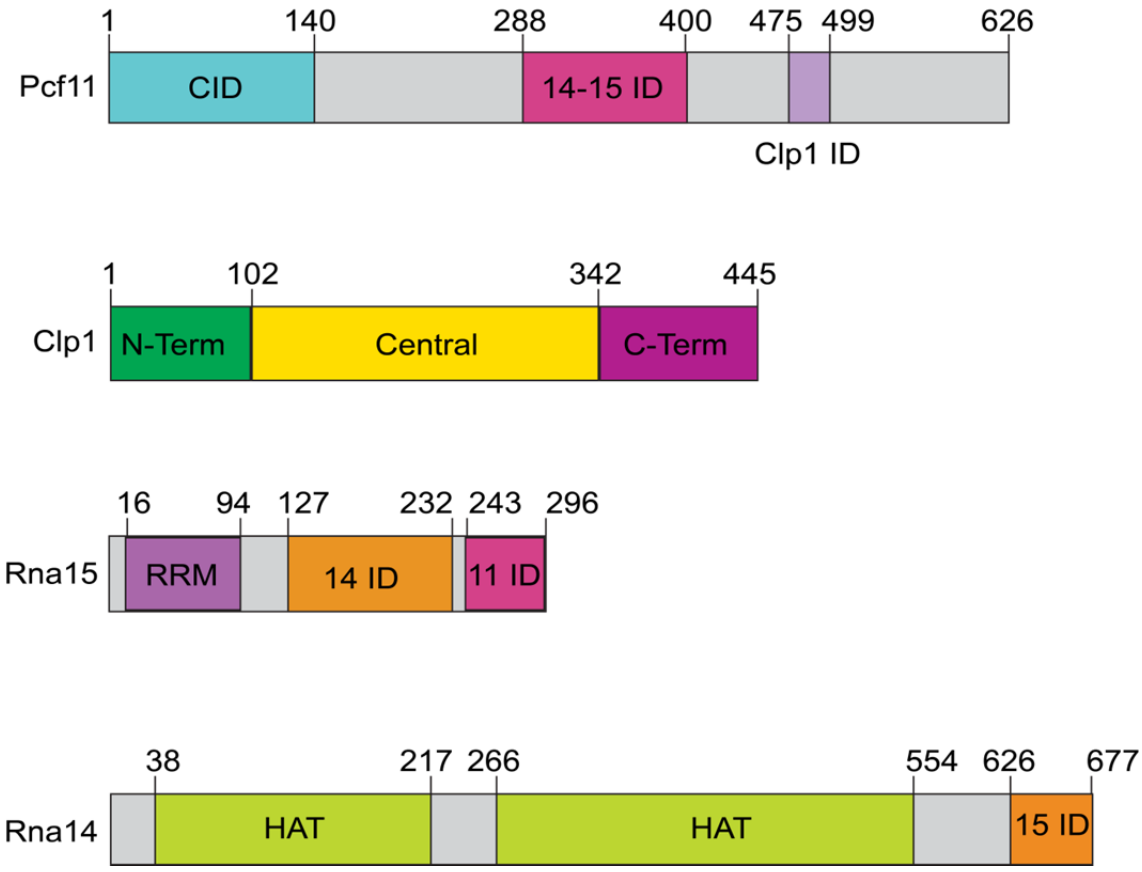


Figure 1.2 Schematic representation of the CF1A complex. Protein domains and interaction domains (ID) are shaded.

1.2.1.1.1 Pcf11

Pcf11 is the 71 kDa subunit of cleavage factor 1A (CF1A) in *S. cerevisiae* and was discovered as part of the polyadenylation machinery by yeast 2-hybrid experiments involving already established components Rna14 and Rna15 (103,104,109). Temperature sensitive point mutations in Pcf11 proved lethal at a permissive temperature when combined with temperature sensitive point mutations in Rna14 and Rna15 (104). Genomic deletion of Pcf11 results in lethality whilst temperature sensitive mutants demonstrated a shortening of poly(A) tail length and an apparent decrease in the concentration of ACT1 mRNA *in vivo* (104). In addition, *in vitro* experiments with mutant Pcf11 resulted in a failure of pre-mRNA cleavage and polyadenylation steps as had been observed for Rna14 and Rna15 (103,104). Pcf11 co-purifies with Rna15 under non-denaturing conditions allowing purification of functional CF I (97,104). Pcf11 acts as a scaffold protein which forms extensive interactions with all other proteins in CF1A (110–113). It also mediates coupling of transcription with polyadenylation through interaction with the CTD of RNAPII (112,114,115).

Pcf11 makes RNAPII CTD contacts through a C-terminal interacting domain (CID) present in the N-terminus of the protein (Figure 1.3) (116,117). X-ray structures show that the CID folds into 8 α helices arranged in a righthanded superhelical conformation (116,117). Pcf11 binds to phosphorylated S2 within the conserved heptad repeats of the CTD of RNAPII (48). A structure of the CID in complex with a derived peptide possessing a phosphorylated serine demonstrates that binding is mediated by a groove formed by helices 2, 4 and 7 (Figure 1.3). Sequence alignments show conservation of residues present on the surface of this groove from yeast to human suggesting that this interaction is universal (116). The backbone of the RNAPII derived peptide forms 7 hydrogen bonds with the side chains of residues present within the CID

(116). Residues Y_1 and P_3 of the heptad repeats present in the CTD of the peptide are buried in the hydrophobic groove of the CID with Y_1 hydrogen bonded to D68 in the CID of Pcf11 (116). The phosphorylated S2 residue present in the peptide does not form direct contacts with Pcf11 suggesting that Pcf11 can bind both phosphorylated and un-phosphorylated residues (115). However, phosphorylation of the S2 in the heptad repeats greatly increases binding affinity (112,115).

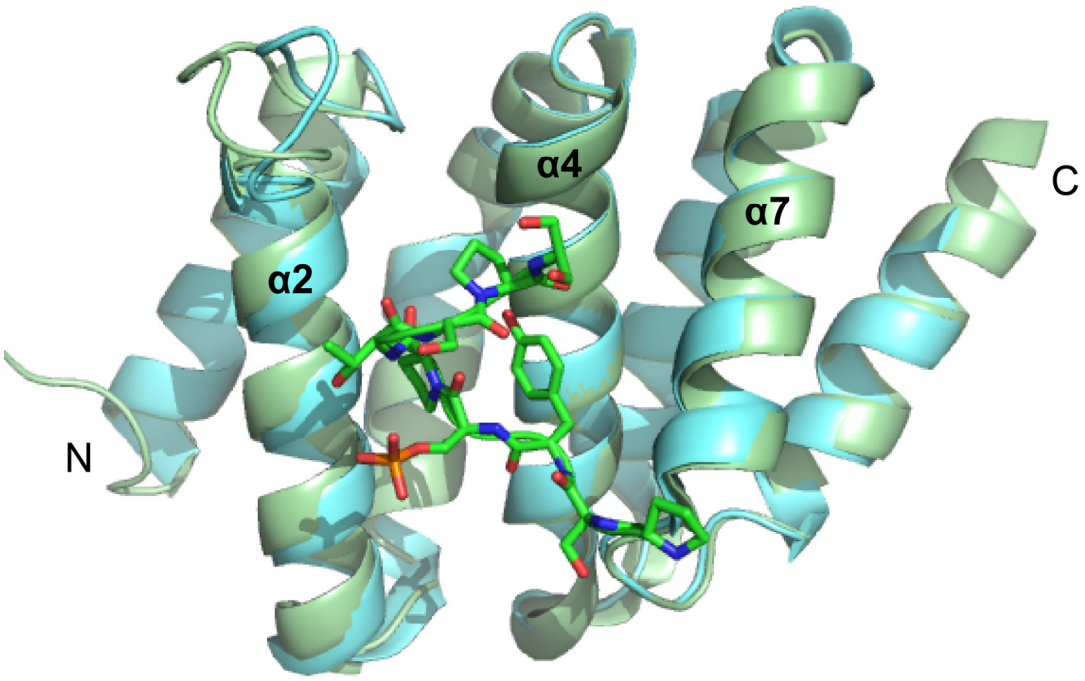


Figure 1.3 Structure of the Pcf11-CID bound to a RNAPII CTD peptide. (PDB ID 1SZA, 116) (PDB ID 2BFO, 117). A structural analysis of the solved structure is shown. Nobel et al (blue) and Meinhart and Cramer (cyan).

Pcf11 binds Clp1 through a short stretch of 25 amino acids (residues 475-499) present between two zinc finger motifs and overlaps a previously identified Clp1 binding site (112,118). Three conserved residues are present in the 25 amino acid stretch, R480, W482 and W489 and they form hydrophobic or hydrogen bonding interactions with Clp1(118). Although the CF1A complex does not have a direct homologue in mammalian processing systems, the protein factors which make up CF1A do have homologues. The residues which mediate Clp1 binding in Pcf11 are conserved from yeast to human suggesting this method of interaction is conserved (118). Mutation of these conserved residues has been demonstrated to increase poly(A) site readthrough by up to 20-fold (119).

As well as interaction with Clp1, Pcf11 makes contact with protein factors that are part of the CPF complex. GST pulldown experiments have identified interactions of Pcf11 with CPF components Cft1, Cft2, Ysh1, Pta1 and Ssu72 (120–124).

1.2.1.1.2 Clp1

Clp1 is a 50kDa protein and was identified by purification of CF1A and subsequent peptide microsequencing (105). It is known to interact with Cft1, Cft2 and Pta1 of CPF allowing cross talk between the polyadenylation machineries (120,121). Of all the components of CF1A, least is known about the exact role of Clp1 other than its requirement for the cleavage and polyadenylation steps (123). Despite this it has been well characterised in terms of structure. The crystal structure of Clp1 reveals three domains, a large central domain flanked by two smaller domains present at the N and C terminus (118).

The central domain is composed of a 7 stranded beta sheet surrounded by 7 α-helices and is a member of the SIMBI class of NTPases (Figure 1.4, A) (118,125). A P-loop or Walker A motif is observed in the loop connecting the β1 sheet and α2 helix. This feature is synonymous of NTPases and often allows binding and hydrolysis of nucleotides. High performanice liquid chromatography (HPLC) revealed the ability of Clp1 to bind ATP. Residues within the P-loop and the N-terminal domain of Clp1 act to secure the ATP nucleotide in place through extensive hydrogen bonding (Figure 1.4, B) (118). The γ-phosphate of the base forms hydrogen bonds with the side chains of Q133 and K136 in the P-loop as well as the ε amino group of K321. The β phosphate forms hydrogen bonds with the backbone and side chains of aromatic residue Y137. Further interactions are observed between the β phosphate group and backbone amides of Y134, G135 and K136. The α phosphate makes fewer contacts to the protein than the β phosphate, only forming hydrogen bonds with the main chains of Y137 and S138 and the side chain hydroxyl of S138. The N terminal domain mediates interaction with the ribose sugar by forming hydrogen bonds between the hydroxyl group present on carbon 2 of the ribose and the backbone carbonyl of K72. The ε amino group of K72 indirectly supports ATP binding by interaction with A315 within the C terminus of the P-loop. This interaction acts to stabilise conformation of the P-loop and in turn aids the van der Waals interactions between the P-loop and the bound ATP (118). Two switch regions characteristic of ATPases are also present within the central domain. Both contain conserved motifs which, with the addition of a magnesium ion, aid in binding of ATP (118).

The N-terminal domain comprises of a 4 stranded β sheet backed by a 3 stranded β sheet in a sandwich conformation. Long loops are present between the 7 β sheets structure and the degree of curvature means the sandwich structure is quite globular

(118). The C terminal domain is larger than the N terminal domain and consists of a 7 stranded β sheet combined with 3 short α helices (118) (Figure 1.4, A).

The ATP binding site of Clp1 also displays a degree of binding specificity with no bound GTP observed during HPLC analysis (118). The crystal structure of the ATP binding site revealed that D33, present in the N-terminal domain of Clp1, formed a hydrogen bond with the exocyclic N6 amino group of the adenine base (Figure 1.4, B). The presence of an O6 carbonyl oxygen in the corresponding position of the guanine base would not be able to form the same hydrogen bond thus inhibiting binding (118).

The structure of the Clp1:ATP complex demonstrates an apparent capability to support ATP hydrolysis. However, no such reaction was observed during *in vitro* experiments involving both Clp1 and Clp1 bound to other factors of the polyadenylation machinery (118). Indeed, in higher eukaryotes Clp1 homologues are kinases that phosphorylate 5' hydroxyl ends of RNA (126,127). Furthermore, a point mutation, G135R, in the P-loop of Clp1 is not lethal until combined with point mutations in the part of the protein that mediates Pcf11 binding (128). Crystal structure of the Clp1 G135R mutant reveals a conformational change in the P-loop of Clp1 which prevents ATP binding (129). However, Clp1 mutants were still viable suggesting that neither binding of ATP nor ATP hydrolysis is required by Clp1 during cleavage and polyadenylation (128,130).

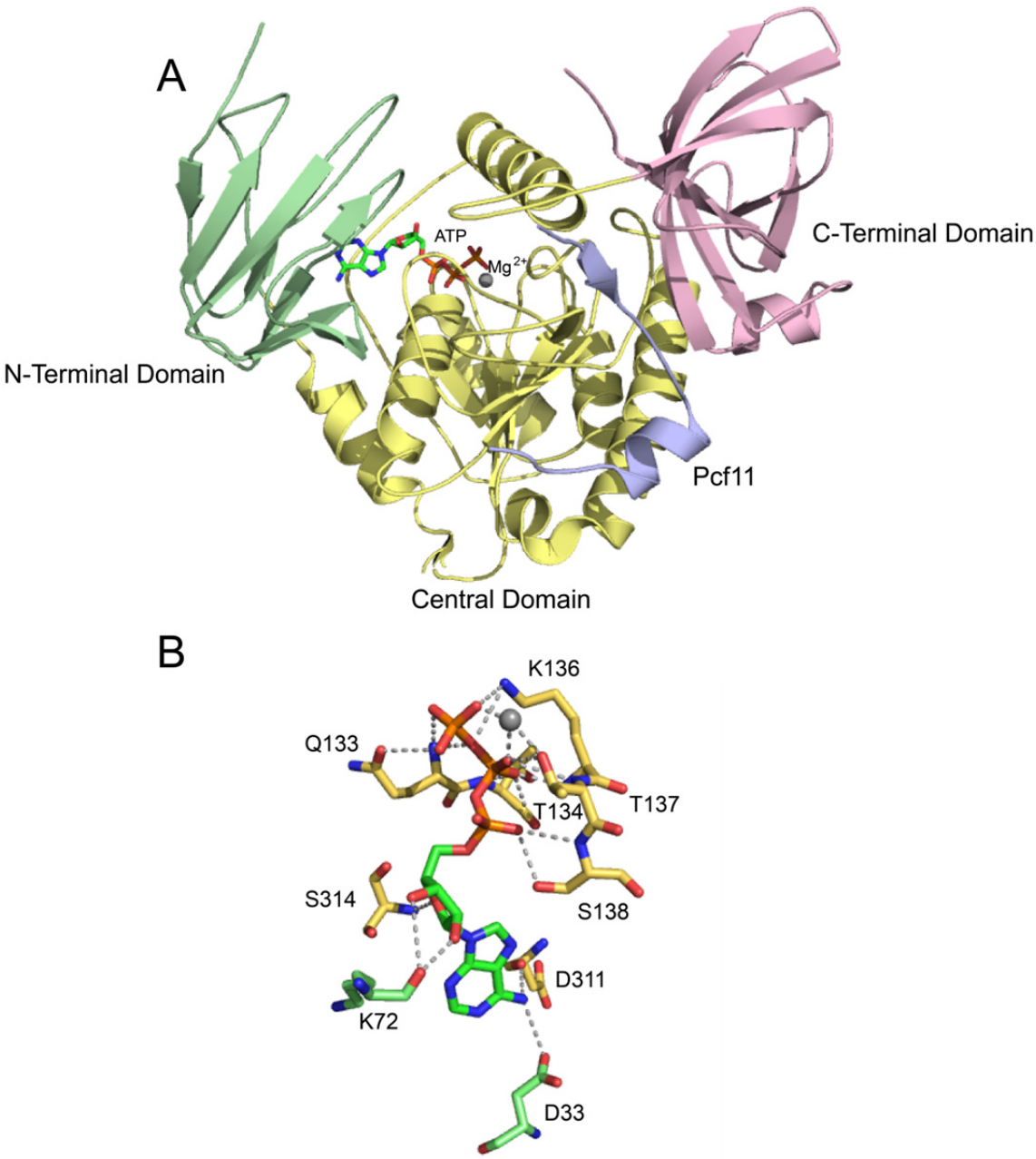


Figure 1.4 The Clp1 and ATP, Mg^{2+}-Pcf11 complex. (PDB ID 2NPI, 118). A. Clp1-ATP-Mg^{2+}-Pcf11 binding. The structure of the N-, central and C-terminal domains of Clp1 bound to Mg^{2+} and ATP. Pcf11 (blue) is also shown bound to the central domain. The N-terminal domain forms a β-sandwich whilst the C-terminal structure consists of a 7 stranded βsheet with 2αhelices. The central domain is formed of a 7 stranded βsheet surrounded by 7αhelices. The P-loop and residues in the N-terminal domain mediate interaction with ATP and Mg^{2+}. Pcf11 interacts with the central domain of Clp1. **B. ATP and Mg^{2+} binding in Clp1.** The residues within the N-terminal and P-loop domain involved in mediating ATP and Mg^{2+} binding are shown. Interactions are highlighted by the presence of dashed lines.

Structural analysis revealed a conserved Pcf11 interaction region within the central domain of Clp1 (Figure 1.4, A) forming part of a hydrophobic cleft which is able to interact with Pcf11 (118). Combination of the mutations G135R, H267Q and L341S generate a Clp1 temperature sensitive mutant which results in loss of the Clp1-Pcf11 interaction (128). *In vitro* experiments employing a synthetic CYC1 pre-mRNA transcript revealed that the *clp1* mutant was unable to carry out the cleavage reaction required during processing (128). These results suggest that association of Pcf11 and Clp1 is essential to maintain wild type levels of processing.

1.2.1.1.3 Rna14

Mutations in genes RNA14 and RNA15 were found to be responsible for temperature sensitive phenotypes characterised by a decrease in the total amount of mRNA at the restrictive temperature, 37ºC (109). Pulse chase experiments revealed no obvious decrease in transcriptional activity and it was therefore suggested that the proteins had a role in mRNA stability (109) but the function remained unclear. Further analysis of temperature sensitive mutations in both RNA14 and RNA15 demonstrated deficient cleavage and poly(A) tail addition (103). Localisation studies of Rna14 found it was localised to both the nucleus and the cytoplasm (131). Subcellular fractionation experiments revealed that the cytoplasmic fraction of Rna14 localised to the mitochondria implicating a role for Rna14 in mitochondrial metabolism (132). Fraction complementation assays and purification of CF1A together revealed both Rna14 and Rna15 form part of the CF1A complex (102,103). Rna14 is a 76 kDa subunit of CF1A and shares 25% sequence identity to mammalian polyadenylation factor CstF-77. The homologous regions include the highly conserved HAT (Half-a-TPR) repeats (133,134). These repeats are so named because they lack the conserved glycine and alanine residues found in the conserved tetratrico-peptide (TPR) motifs; WX_2LGX_2Y and

AX_3FX_2A (133). The motifs form two helices named helixA and helixB and together make 1 TPR/HAT repeat (133). The number of these repeats vary across species but are observed in all Rna14 homologues (133).

Rna14 is able to interact to form a homodimer (107). Structure analysis of CstF-77 from both *Mus musculus* and *Encephalitozoon cuniculi* revealed Cst-F77 forms these homodimers through interaction of the conserved HAT repeats (135,136). In the murine structure, the HAT repeats were shown to encompass residues 1-550 of CstF-77 (135). A C- terminal collection of HAT repeats was identified between residues 242-549 that formed a stable domain on their own. The domain included HAT repeats 6-12 which formed 7 pairs of anti-parallel α helices (135). The helices form a curved structure with helix 6 almost at a right angle to helix 12. The first 5 HAT repeats, present in an N terminal domain, form the same structure of 5 pairs of α helices. However, HAT repeats 1-5 do not curve and sit at an acute angle relative to the C-terminal HAT repeats (135). Association into a CstF-77 homodimer is achieved through interaction between HAT repeats 6-12 in both CstF-77 monomers (135). The α helices are anti-parallel to one another so that HAT repeat 12 and the immediate C terminal residues of one monomer are in contact with repeats 7-10 of the other (135). Comparison of mouse and *E. cuniculi* structures reveal high conservation in the structural basis of both CstF-77 homologues (136). Furthermore, the structure of the *E. cuniculi* and *M. musculus* CstF-77 homologue homodimer reveals that the association results in a slight bend or kink with an angle of 110º (135,136). Rna14 interacts with Rna15 within the CF1A complex demonstrated by yeast two hybrid analysis (106) and coimmunoprecipitation (102). The association of the Rna14 homodimer results in the formation of an Rna14-Rna15 tetramer with a stoichiometry of 2:2 demonstrated by electron microscopy (107). Electron micrographs revealed tetramers formed as rod

shaped particles possessing a kink/bend of 110º which is in agreement with the solved CstF-77 homodimer structures (107). Sedimentation equilibrium results revealed the tetrameric association constant of Rna14 and Rna15 was $K_a = 6 \times 10^5$. Further, the orientation of associated heterodimers varied greatly between complexes demonstrating that although the interaction is relatively weak it is also flexible (107).

Using the murine CstF-77 homodimer as a model, two conserved residues involved in dimer binding were mutated in Rna14 in order to characterise the mechanism of Rna14 homodimerisation. Mutations R562E and Y563S resulted in loss of Rna14 dimerisation demonstrating the mode of binding is conserved through Rna14 homologues (108). Although such disruption in binding was shown not to affect Rna14-Rna15 binding it did affect Rna15-RNA binding (108). Rna15 tethers the CF1A complex to the RNA via a RNA recognition motif (109). An electromobility gel shift assay using a labelled GAL7 3'UTR mRNA template revealed disruption of the Rna14 homodimer resulted in a decreased affinity for the RNA which was rescued upon increasing concentration of *rna14* and wild type Rna15 protein (108). By increasing concentrations of *rna14* and wild type Rna15, the template RNA acts as a bridge where two available Rna15-Rna14 heterodimers are able to bind simulating the role of the tetramer (108). These result suggest that two Rna15 RRM domains are required to efficiently bind the mRNA and such binding is dependent on formation of the Rna14-Rna15 tetramer through association of the Rna14 homodimer (108). Wild type Clp1 and Pcf11 have also demonstrated the ability to rescue the defective tetramer (108). Cross-linking of wild type Pcf11, Clp1, Rna15 and *rna14* revealed a molecular weight close to wild type (108).

Association analysis of a series of truncated versions of both Rna14 and Rna15 revealed the Rna14-Rna15 binding interface was located within residues 626-677 and 127-232 respectively (137). Structural studies using NMR highlighted the extensive nature of the binding interface (Figure 1.5). Rna15 forms a central core of 4 α helices that is surounded by two Rna14 α helices joined by a loop. Additional helices at the N- and C- terminal region of Rna15 act to further secure the interaction. A clasp region present between α0 and α1within the N-terminus of Rna15 lacks secondary structure but acts to hold α1' of Rna14 in place (137).

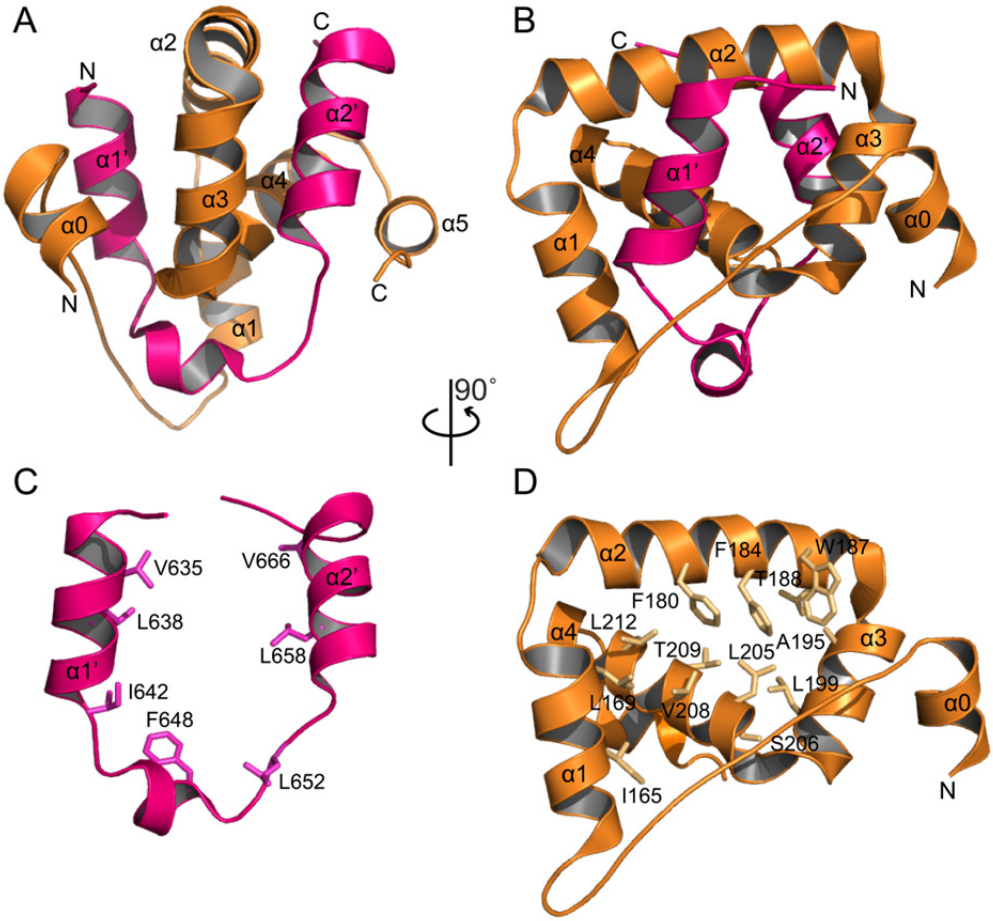


Figure 1.5 Structural analysis of the Rna14-Rna15 binding interface. (PDB ID 2L9B, 137). A. The Rna14-Rna15 binding interface. Rna14 is depicted in pink whilst Rna15 is shown in orange. Rna14 forms a monkey-tail like structure where α-helices α1' and α2' wrap around a helical core formed by Rna15. A clasp region is present in Rna15 present between α-helix 0 and α-helix 1. **B.** A 90° orientation of the binding interface. Four α-helices interact to form a tight helical core held in place by extensive side chain interactions. **C.** The residues highlighted on α-helices α1' and α2' represent those of Rna14 that interact with Rna15 to hold the monkey-tail like structure in place around the Rna15 core. **D.** The extensive nature of the side chain interactions are represented that act to stabilise the central core. Also, highlighted are residues T188, S206 and T209 that are found buried away from the interaction interface.

The region of Rna14 involved in mediating contact with Rna15 forms a so-called monkey-tail structure (Figure 1.5, C). The monkeytail structure forms a tight association with the Rna15 central core through extensive side chain interactions. A combination of hydrophobic and aromatic residues including V666, L658, F648, V635 and L638 present on α 1' , α2' and the loop connecting the two helices act to stabilise the interaction (137).

Both the crystal structure of the Rna14-Rna15 complex from *Kluyveromyces lactis* and small angle X-Ray scattering analysis suggest that the 2 Rna15 monomers sit near the centre of the Rna14:Rna14 dimer (108,138). In both the mouse and *E. cuniculi* Cst-F77 dimer structures the C-terminal domain was truncated but the most C-terminal part of the protein did face the centre of the CstF-77 dimer (135,136).

Rna14 is referred to as a scaffolding protein as it binds both Rna15 and CF1B component, Hrp1. By forming interactions with both Rna15 and Hrp1 it assists these proteins in their individual roles and interaction with both Rna15 and Hrp1 increases efficiency of polyadenylation (107,139). Its association with factors within the CF II complex (99) infers that Rna14 may maintain a role in crosstalk between CF I and CF II. Binding of CF1A to the RNA may cause Rna14 to signal the initiation of the next polyadenylation step as it associates with Pfs2 and Fip1, both components of the CPF complex (99,139). However, localisation of Rna14 in the cytoplasm and in mitochondria suggests a role for Rna14 in addition to mRNA polyadenylation (131). However, mutational analysis of Rna14 did not produce any obvious impairment in replication or expression of mitochondrial DNA (132).

1.2.1.1.4 Rna15

Rna15 has been shown to associate with Pcf11 as well as Rna14 within the CF1A complex and with Sub1 of the CPF complex (140). Rna15 possesses an N-terminal RRM domain present between residues 16-94 that binds to the pre-mRNA transcript to tether the CF1A complex to the 3' UTR (141). As in other RRM containing RNA binding proteins, the conserved RNA binding motifs, RNP-1 and RNP-2, are present within Rna15 (141). Rna15 interacts with Rna14 within the CF1A complex through a C-terminal binding interface present between residues 127-232 that form a helical structure with a hinge domain in order to lock onto Rna14 (Figure 1.5) (137). The extensive interface of this interaction results in extremely tight binding as measured by AUC sedimentation equilibrium experiments (107). Rna15 interacts with Rna14 through a central core formed by 4 α-helices. This core structure is maintained through interactions between side chains within the core helices of Rna15 (Figure 1.5, D). Aromatic residues, F180, F184 and Y187, present on helix 2, act to stabilise the structure in addition to the aliphatic side chains, I165, L169, A195, L199, L205, V208 and L212. Polar residues, T188, S206 and T209, are buried away from the central core and are inaccessible from the surface (137). The appearance of buried polar residues is unusual and in other Rna15 homologues they are replaced by hydrophobic residues (137). Mutation T209A and S206A don't disrupt interaction but do affect thermal stability of the protein and it has been suggested that the presence of these polar residues may act in a coupled binding and folding reaction (137). This model is supported by experiments employing 2M urea that demonstrated local unfolding between residues 142 and 150 in Rna15 suggesting a scenario whereby under the correct conditions, the clasp region could unfold to allow dissociation of the two factors *in vivo* (137).

The primary role of Rna15 in polyadenylation appears to be poly(A) site selection. The crystal structure of the RRM of Rna15 reveals a 4 stranded β-sheet backed by two α-helices as is observed in most RRM folds (Figure 1.6) (141). However, there is much debate over the particular site at which Rna15 binds to direct polyadenylation. It has been suggested that Rna15 itself shows no specificity for RNA sequence and requires Hrp1 and Rna14 to help mediate RNA binding to the adenosine rich (A-rich) positioning element (PE) (139). Mutational analysis of conserved sites present near the polyadenylation signal in GAL7 3' UTR mRNA show that Rna15 interaction is essential for polyadenylation *in vivo* (139). NMR spectroscopy using the GAL7 anchoring RNA containing conserved UA rich efficiency element (EE) and the A rich positioning element (PE) revealed that Rna15 bound RNA very weakly with a dissociation constant in the high micromolar range with very little specificity (142). Crosslinking experiments involving GAL7 3' UTR mRNA demonstrated an ability of Rna15 to recognize and bind A rich RNA sequences in the presence of both Hrp1 and Rna14 (139). In addition, such interaction was shown to be dependent on both Hrp1 and Rna14 as Rna15 alone did not bind the A-rich sequence (139).

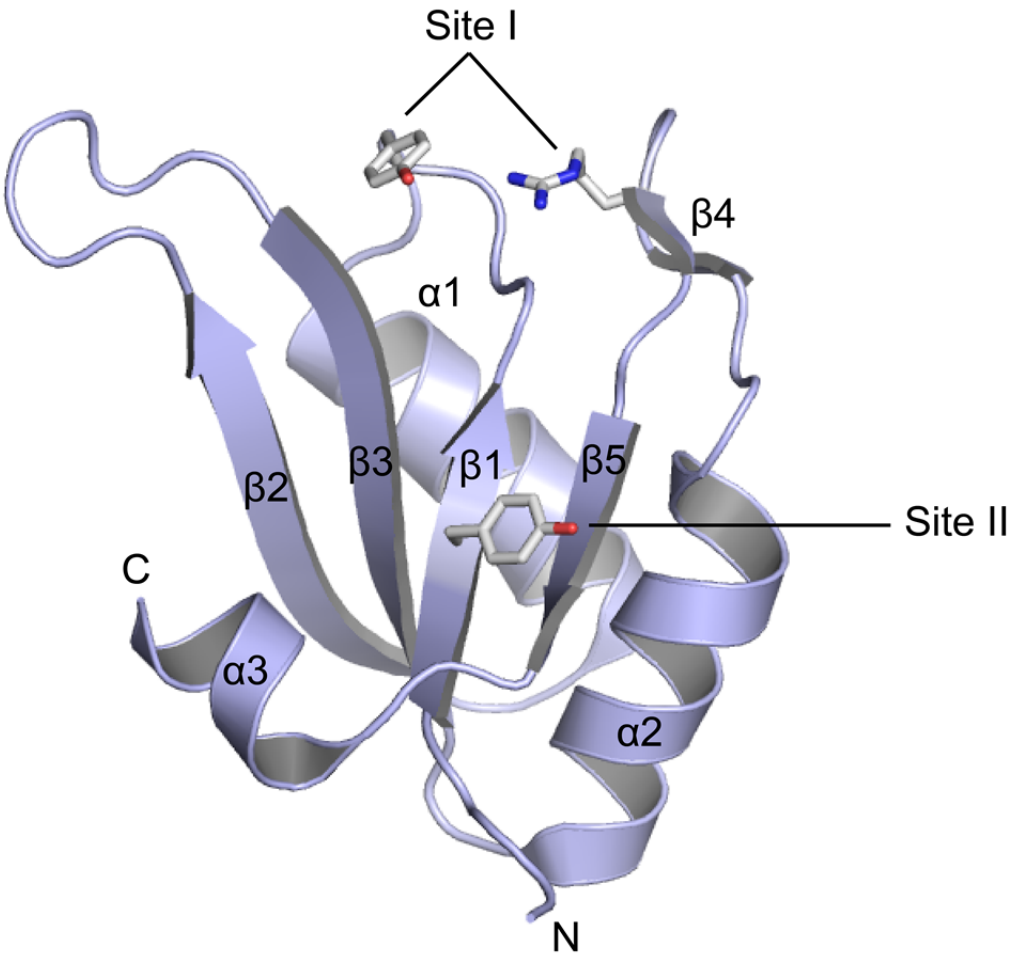


Figure 1.6. The crystal structure of residues 16-103 of Rna15. (PDB ID 2X1B, 141). Residues 16-94 of Rna15 make up the RRM of Rna15. The RRM forms a four stranded β-sheet backed by two α-helices. Also shown are residues involved in mediating RNA binding in site I and site II.

NMR spectroscopy experiments revealed that the RRMs of Rna15 and Hrp1 form a horseshoe shape when bound to the PE and EE of GAL7 respectively (Figure 1.7, A) (142). Each RRM is present at a 90° angle relative to one another with the β sheet of each facing the centre of the horseshoe shape. The structure provided a possible mechanism as to why Hrp1 influences Rna15 binding and site selection. A protein-protein contact was seen between Hrp1 residue D193 and Rna15 residue R87 (Figure 1.7, B). Experiments employing *in vitro* cleavage and polyadenylation reactions revealed that mutation R87D in Rna15 results in defects in the cleavage reaction. This result suggests association with Hrp1 is required for efficient Rna15 mediated site selection and processing. NMR spectroscopy experiments employing a Rna15 R87A mutant and the GAL7 3' UTR mRNA sequence revealed the Rna15 mutant bound RNA more weakly than wild type. This decrease in Rna15-RNA binding allowed binding of Hrp1, effectively competing the mRNA transcript away from Rna15, a feature not observed in wild type. These results suggest that Hrp1 directs Rna15 to the A rich PE sequence through protein-protein interactions and further bridging by Rna14 is able to hold the complex in place (142).

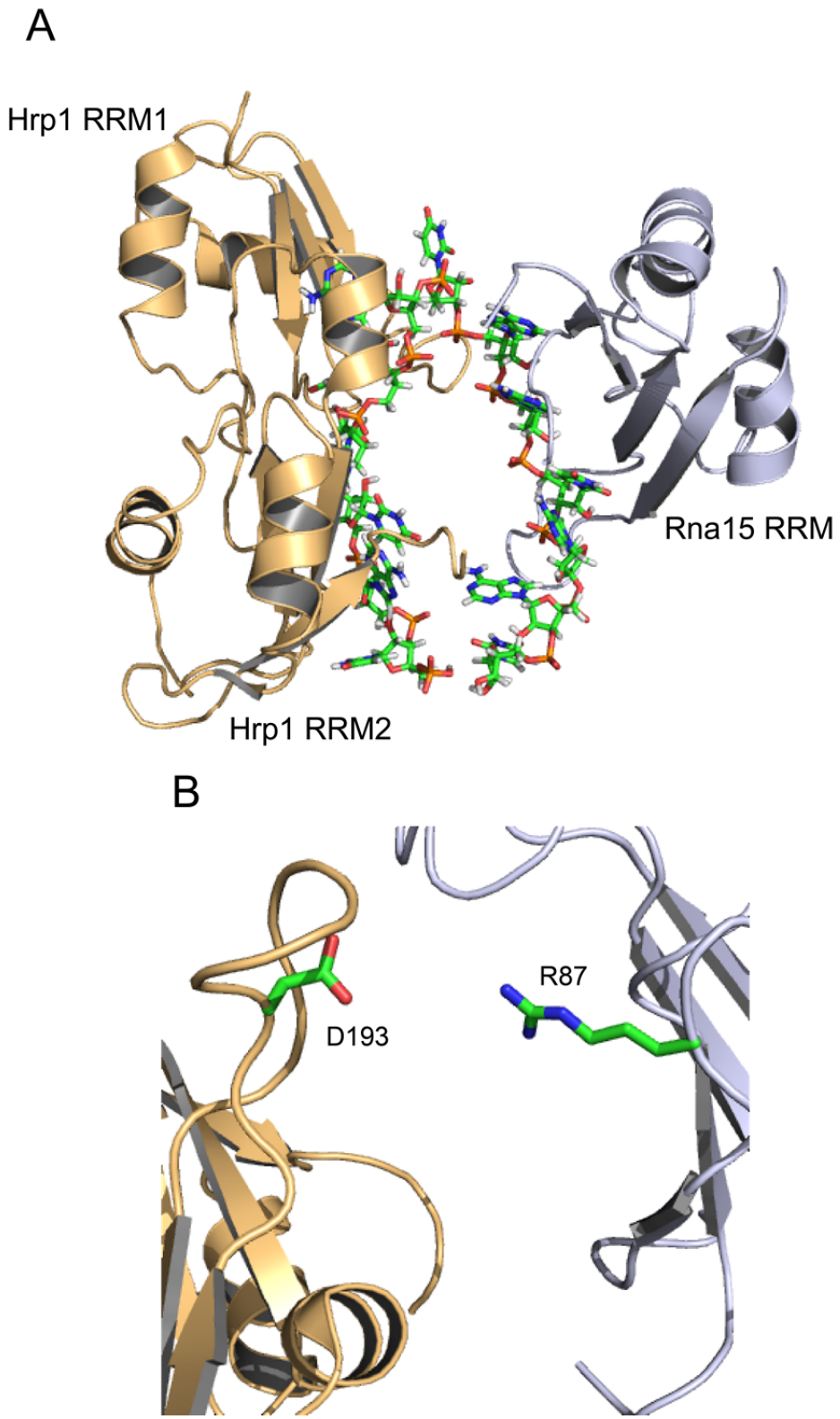


Figure 1.7 NMR structural analysis of Hrp1 and Rna15 mediated RNA binding. (PDB ID 2KM8, 142). A. The RRMs of both Hrp1 and Rna15 form a horseshoe shape which results in curvature of the RNA oligonucleotide. RNA binding occurs over the face of the RRM domains where the oligonucleotide interacts with residues present on the 4-stranded β-sheets of each RRM. **B.** Contacts between D193 of Hrp1 and R87 of Rna15 were suggested to maintain Rna15 contact with the positioning element.

In contrast, the X-ray crystal structure of the RRM of Rna15 bound to RNA convincingly displays a clear specificity of the protein for G/U nucleotides (141) with two binding sites on the face of the central 4 stranded β sheet (Figure 1.4). The first site is present on the loop connecting the β1 and α1. Here the side chains of residues R87 and Y27 form the walls of a binding pocket (Figure 1.8). Two structures demonstrate both G and U nucleotides are able to stack against the aromatic group and planar side chain. Upon RNA binding, R87 undergoes a conformational change which allows the nucleotide to stack against the side chain. This new re-orientation of R87 is stabilised by hydrogen bonding between the side chain and the main chain carbonyl and side chain hydroxyl of S24. Watson-Crick like hydrogen bonding between the base and the backbone of residues in the protein acts to further stabilise the interaction. Uracil bases hydrogen bond to the backbone amide of Y27 through the C-4 carbonyl and the imino proton of N3 is shared with the backbone carbonyl of I25. For guanine, the C-6 carbonyl hydrogen bonds to the backbone amide of Y27 and the imino proton of N1 is shared with the carbonyl of I25. These hydrogen bonds are observed in standard Watson-Crick AU and GC base pairing and these interactions form the basis of G/U selectivity as in both A and C nucleotides an exocyclic amino group replaces the interacting carbonyl found in G and U nucleotides and there is no imino proton present. The structure of the RRM and the site I binding pocket in particular is conserved in the human homologue of Rna15, CstF-64. CstF-64 is known to bind G/U rich sequences known as downstream sequence elements (143,144). Together these data imply that binding of G/U rich sequences driven by specificity in the site I binding pocket is a conserved feature.

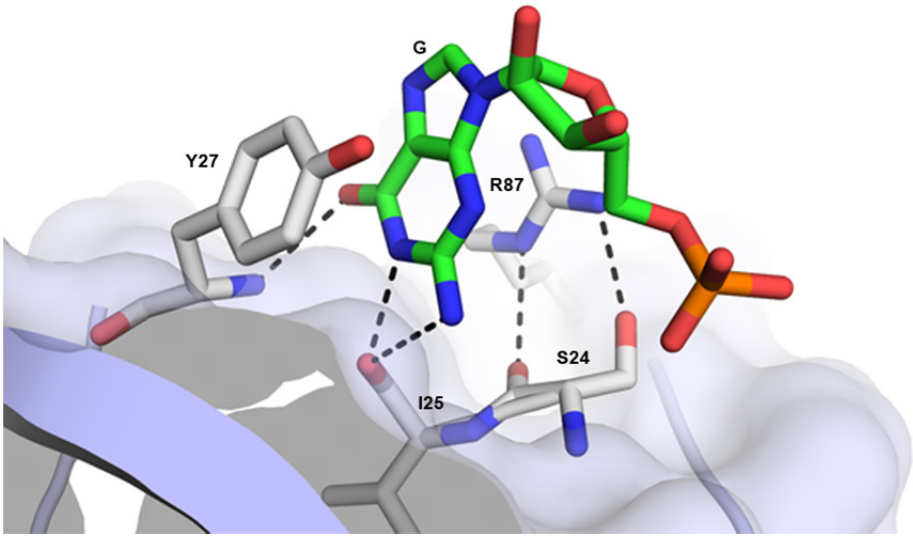


Figure 1.8 RNA binding at site I within the RRM of Rna15. (PDB ID 2X1A, 141). Y27 and R87 form the walls of a binding pocket that the base of the nucleotide is able to stack against. R87 forms hydrogen bonds with the backbone and side chain of S24 to maintain the correct orientation to form the binding pocket. Watson-Crick hydrogen bonding interactions between the base of the nucleotide and the backbone of I25 and Y27 act to stabilise the nucleotide within the binding pocket.

A second binding site is present on the face of the RRM on β1 of the central sheet (Figure 1.9) (141). Y21 is a conserved residue present in the RNP-2 RNA binding motif and stacks against a guanine nucleotide. Unlike site I, Site II binding relies mainly on this stacking interaction with only a single hydrogen bond between the exocyclic amino group on the N2 of guanine and the backbone carbonyl of Y93 arguing against a role for site 2 in G/U selectivity. NMR spectroscopy employing UG rich RNA oligonucleotides highlighted the presence of a third site present on the face of the RRM on the β-1, β-2 and β-3 of the central sheet (141). Residues Y61 and F63 form this potential third site and in addition to this site other residues present on the surface of the RRM were also disturbed by RNA titration. Taken together these data show that the RNA binding surface extends over the face of the 4 stranded β sheet.

Fluorescence spectroscopy employing both A/C and G/U rich RNA oligonucleotides highlights a clear preference of Rna15 for G/U sequences (141). A range of RNA oligonucleotides were employed with increasing G/U nucleotide content (141). Strongest binding was observed with a G/U nucleotide; K_A $2.1x10^5$, weakest binding was observed with an A/C rich oligonucleotide with a K_A of $<1.0x10^4$. Indeed, GU rich sequences are present within the 3' UTR in *S. cerevisiae* (145,146) and mutation of such sites in the 3' UTR of genes CYC1 and ADH1 results in deficient cleavage and polyadenylation (147,148).

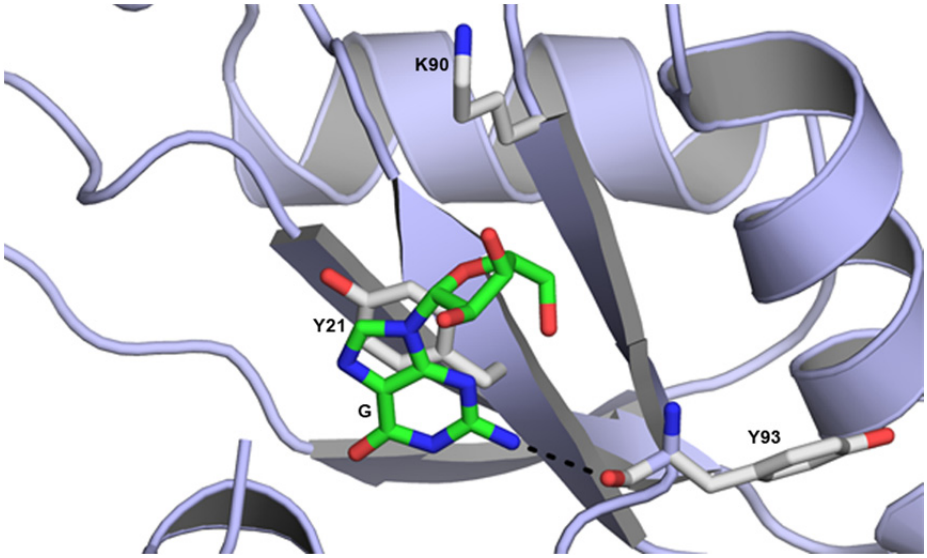


Figure 1.9 RNA binding at site II within the RRM of Rna15. (PDB ID 2X1F, 141). Site II is predominantly mediated by a stacking interaction between the nucleotide and the aromatic of Y21. This interaction is held in place by a single hydrogen bond between the base of the nucleotide and the backbone of Y93.

Immediately C-terminal to the RRM in Rna15 is a short flexible region termed the tail of the RRM and is present between residues 95 and 103 (141). In the crystal structure of the free RRM this so-called tail region forms a short α-helix that covers β1-β3 of the RRM. When RNA is bound, the tail region adopts secondary structure with an extended β conformation. In CstF-64, the tail region has been implicated in regulation of RNA binding. This region also forms an α-helix which packs against the central β-sheet of the RRM. It has been suggested that upon RNA binding the helix moves giving access to residues F19 and F61 allowing stacking against a uracil base. NMR spectroscopy data show that the tail region in Rna15 also forms an interaction with the central β-sheet and as observed in CstF-64 is displaced upon RNA binding (141). Following the tail is a linker region which is present between residues 103 and 126 which acts to bridge the RRM to the C-terminal Rna14, Pcf11 and Sub1 interacting domains.

1.2.1.1.5 Hrp1

CF1B consists of a single 73 kDa protein, Hrp1 that binds the immature mRNA 3' UTR during processing (106). Binding occurs via 2 RRMs that are present between residues 158-233 and 244-318 (149). SELEX experiments (Systematic evolution of ligands by exponential enrichment) revealed that Hrp1 bound UA_{4-8} rich sequences, termed efficiency elements (EE), with a high affinity (150). Deletion of the UA rich sequence results in diminished binding of Hrp1 measured by an inability to crosslink the protein to mRNA (150). NMR titration using $[UA]_n$ RNA oligonucleotides aimed to characterise the mechanism of RNA binding (149). Upon introduction of the RNA, a large number of chemical shift movements were observed relative to free protein with a slow exchange rate suggesting a tight interaction. Furthermore, 2 Hrp1 molecules were

observed bound to the longer UA sequences demonstrating the possibility of multiple Hrp1 binding *in vivo* where more than one efficiency element is available (149).

Structural analysis using NMR determined the RRMs of Hrp1 behave as independent entities with no NOEs observed between the two domains in the absence of RNA (142). Comparison of two Hrp1 structures bound to RNA reveals both RRMs of Hrp1 fold into βαββαβ structure as seen in the Rna15 RRM with the β strands folded into an antiparallel β-sheet and the α-helices packed alongside it in both RRM domains (Figure 1.10) (142,149). The two RRM domains are joined by a flexible linker that undergoes a conformational change upon RNA binding to form an α-helix (142). In the bound structure, RNA buries itself into a V-shaped cleft formed by the 2 RRM domains so that the phosphates are accessible on the surface of the interface (142). Residues in each RRM direct binding specificity. W168 in RRM1 is conserved across several fungal Hrp1-like proteins and stacks against an adenine base and appears critical for RNA binding since a conservative mutation to phenylalanine results in a decrease in binding affinity (142,149). RRM1 mediates binding of 3’ adenine and uracil residues by direct contact with residues present in the RNP1 and RNP2 motifs. Residues, G249 and G285, in RRM2 make direct contact with a uracil base present at position 1 of the RNA oligonucleotide (142). It has been suggested that Hrp1 binding of the AU rich efficiency element is required to enhance the cleavage reaction but is not essential. A more detailed account of the sequences that direct 3' end processing is presented in section 1.3.

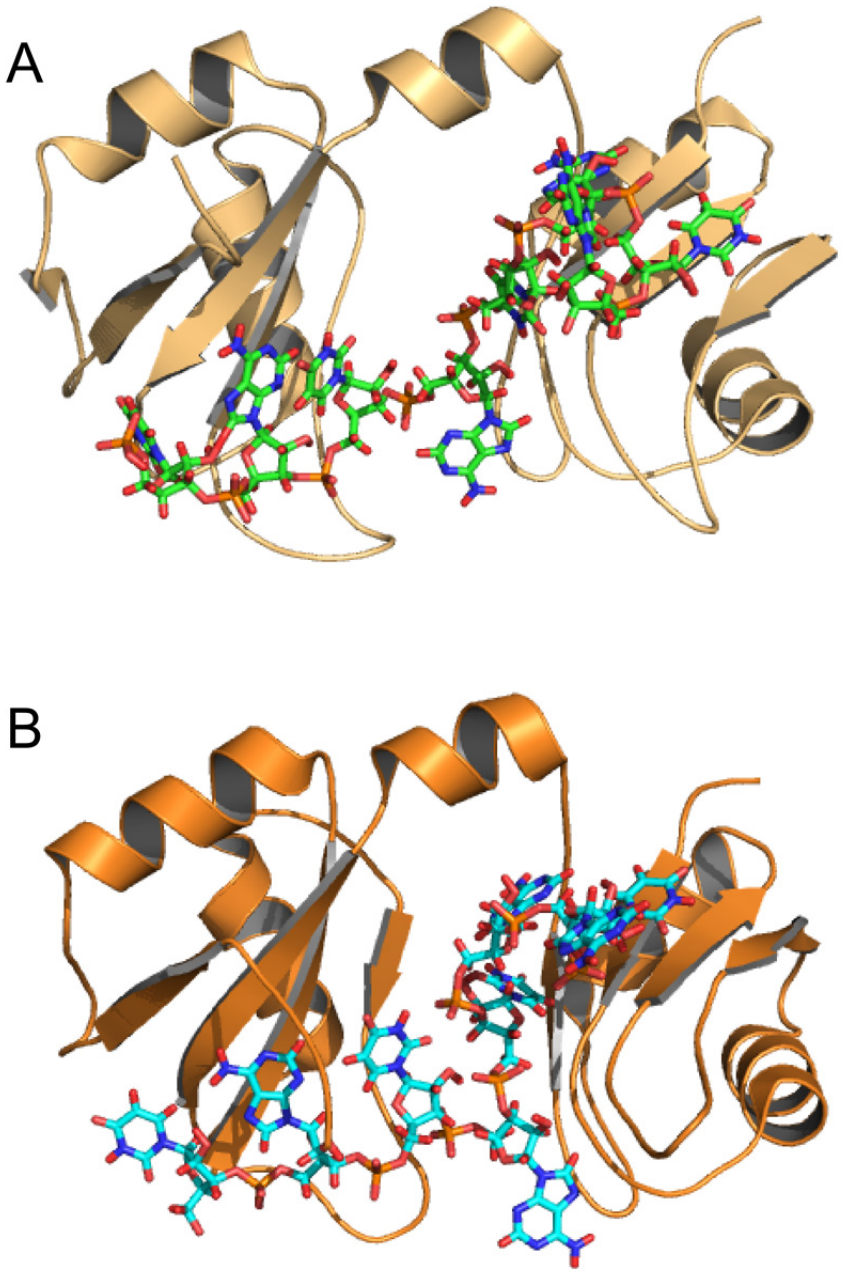


Figure 1.10 Hrp1-RNA structures. (PDB ID 2KM8, 142) (PDB ID 2CJK, 149). The RRMs of Hrp1 form a cleft in which the RNA oligonucleotide is able to bind. The four stranded β-sheet of each Hrp1 RRM faces one another to mediate RNA binding.

Recent structural analysis of the Rna15-Rna14-Hrp1 complex has revealed that Hrp1 interacts with Rna14 via both RRMs on the opposite face to that used for RNA binding (151). Mutations in this Hrp1 interaction surface result in defects in polyadenylation *in vitro* and cell lethality at 37ºC (151). Computational modelling based on existing structural data has been used to characterise the structure of the Rna15-Rna14-Hrp1-RNA complex. This method, with the assistance of previously characterised structural data, was employed to generate a structural model of the complex. It is predicted that the complex forms a U-shape with Rna14 at the bottom and Hrp1 and Rna15 forming two arms either side (151).

Hrp1 has been shown to repress a temperature sensitive phenotype in yeast strains carrying a mutation in the NPL3 allele (66). Npl3 was also found to associate with polyadenylated RNAs *in vivo* (152). Temperature sensitive *npl3* mutants were defective for localisation of some nuclear proteins and nuclear export of mRNA with poly(A) tails (152,153). Both Npl3 and Hrp1 are known to shuttle between the nucleus and cytoplasm and appear to display some functional redundancy implicating Hrp1 in both polyadenylation and mRNA export (154,155). This is supported by additional structural similarities as each protein contains an Arginine-Glycine rich box (RGG box) (156). It has been demonstrated that methylation of the RGG boxes in these proteins by Hmt1p is critical for export from the nucleus (66,157). However, methylation of Hrp1 does not affect RNA binding of the efficiency element and it has been hypothesised that methylation of Hrp1 is only required for binding of RNA in export of the polyadenylated mRNA transcript (150).

The abundance of mutational, biophysical and structural data has led to the hypothetical modelling of the structure of CF1 that may form *in vivo* (108). A schematic

figure representing a 2:2:1:1 CF1A complex with Hrp1 and RNA bound is shown (Figure 1.11). In the model, the Rna14 homodimer interacts with 2 Rna15 monomers to form a tetramer within the complex. Rna15 also interacts with Pcf11 although which Rna15 monomer is responsible for this interaction is still unclear. Rna14 forms a scaffold interaction between Rna15 and Hrp1 to secure these proteins onto the mRNA transcript. Pcf11 interacts with Rna14, Rna15 and Clp1.

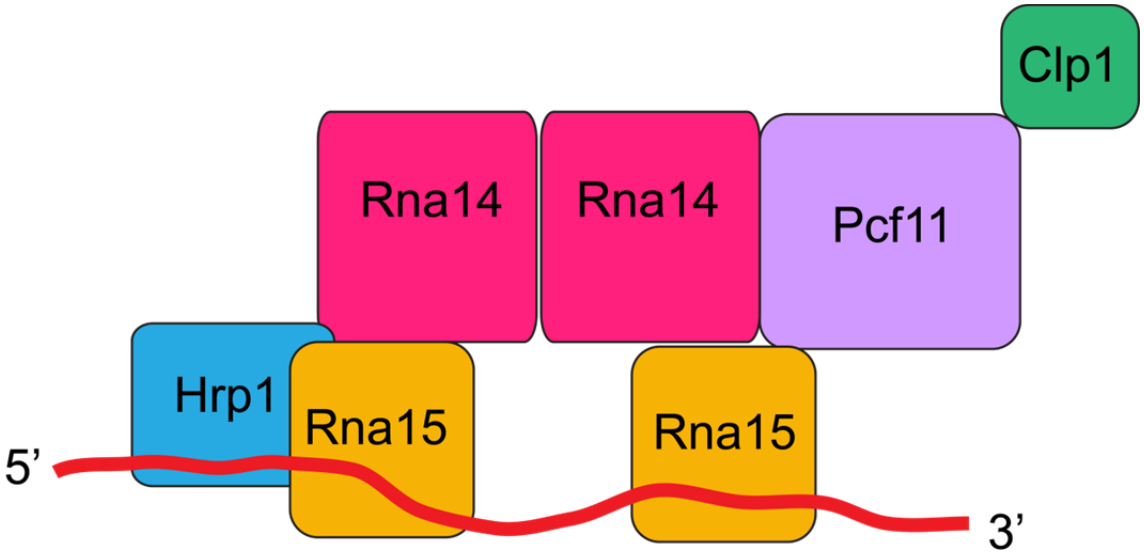


Figure 1.11 Schematic representation of CFI bound to RNA. Biophysical and structural data have demonstrated that CFI consists of an Rna14-Rna15 tetramer, a Pcf11 monomer, a Clp1 monomer and Hrp1 tethered to CF1A through interactions with Rna14. Pcf11 associates with all components of CF1A whilst Clp1 associates only with Pcf11.

1.2.1.2 CPF

Three protein complexes, CF I, CF II, PF I and protein factor Poly(A) polymerase 1 (PAP1) were identified as being responsible for polyadenylation in *S. cerevisiae* requiring both CFI and CFII for the cleavage reaction and CF II, PF I and PAP1 for the polyadenylation step (97). Further purification studies highlighted the presence of a CF II-PF I complex (98,99) which was termed the cleavage and polyadenylation factor (CPF) (99) in which Pap1 protein interacts with the Fip1 subunit of PF I to successfully synthesise the poly(A) tail (158).

In addition to the CPF complex, mutations in Ref2 resulted in inefficient polyadenylation at poly(A) sites suggesting a role in poly(A) site selection (159). Along with Ref2, another five proteins (Pti1, Swd2, Glc7, Ssu72 and Syc1) have been identified known collectively as the APT1 (associated with Pta1) complex. The APT1 complex associates with core CPF component, Pta1. The core CPF unit comprises CF II and PF I whilst addition of the APT1 complex generates a holo-CPF complex required for cleavage and polyadenylation (160). Components of the CPF complex associate with factors involved in transcription initiation and influence phosphorylation state of the CTD of RNAPII providing a direct link between processing and transcription (161–163). A list of the proteins required for successful polyadenylation and their mammalian counterparts is shown in Table 1.1.

Protein Factor (Yeast)	Complex (Yeast)	Protein Factor (Mammalian)	Complex (Mammalian)	Function
Pcf11	CF1A	hPcf11	$CFII_m$	Interacts with the CTD of RNAPII
Rna14	CF1A	CstF-77	CstF	Contains HAT repreats and bridges CF1A to CPF
Clp1	CF1A	hClp1	$CFII_m$	Walker-A motif binds ATP
Rna15	CF1A	CstF-64	CstF	RNA binding through a N-terminal RRM
Hrp1	CF1B	-	-	Binds RNA via 2 RRMs
Cft1	CPF	CPSF160	CPSF	Binds RNA possibly via β-propellar repeats
Cft2	CPF	CPSF100	CPSF	Member β-CASP superfamily and associates with RNA
Ysh1	CPF	CPSF73	CPSF	Contains metallo-β-lactamase domain and member of β-CASP superfamily. Is required for the cleavage reaction.
Yth1	CPF	CSPF30	CPSF	Binds RNA via zinc finger structure
Fip1	CPF	hFip1	CPSF	Recruits PAP
Pta1	CPF	Symplekin	-	CPF factor recruitment
Pfs2	CPF	CstF-50	CstF	Contains WD-40 repeats and bridges CPF to CF1A
Pti1	CPF	CstF-64	CstF	Bnds RNA via RRM
Mpe1	CPF	-	-	Binds RNA via a Zinc knuckle
Ref2	CPF	-	-	Enhances processing efficiency
Swd2	CPF	-	-	Contains WD-40 repreats and recruits CPF
Syc1	CPF	-	-	Regulates processing reaction
Ssu72	CPF	hSsu72	-	Protein phosphatase: dephosphorylates CTD of RNAPII and regulates transcription
Glc7	CPF	-	-	Protein phosphatase: dephosphorylates CTD of RNAPII
Pap1	-	Pap	-	Required for addition of the poly(A) tail

Table 1.1 3' end processing machinery in *S. cerevisiae* and their mammalian counterparts

1.2.1.2.1 Glc7 and Ssu72

CPF components have been shown to direct transcription and processing during early transcription. Regulation of transcription initiation and elongation is directed by Ssu72 as it binds transcription initiation factor TFIIB and is able to suppress mutations in the *TFIIB* gene (161,162). Both Ssu72 and Glc7 have been shown to interact with the CTD of RNAPII and have been implicated in the regulation of transcription termination by dephosphorylation (163). RNAPII is phosphorylated upon association with particular promoter sequences (164). Systematic phosphorylation of the CTD has been demonstrated as being responsible for recruitment of processing factors required for mRNA capping and transcription termination at the precise point in the elongation process that they are required (9,165). Phosphorylation of Y1 of the heptad repeats is required to inhibit the association of transcription termination factors before they are required (80). Glc7 and Ssu72 dephosphorylate Y1 and S5 respectively during elongation promoting association of termination factors (163,166–168). In addition to regulation of the transcription process, other components of CPF have been shown to influence poly(A) site selection by binding to the mRNA transcript. These components are discussed in the following sections.

1.2.1.2.2 Cft1

Cft1 is the largest subunit of CPF and mutations in the CPF subunit Cft1 result in defects in both cleavage and polyadenlyation *in vitro* (169). Filter binding experiments using RNA purified from *S. cerevisiae* demonstrate Cft1 binds both CYC1 and GAL7 RNAs with a K_d of 5 and 15 nM respectively (169). RNA binding by residues 500-750 of Cft1 are predicted to form five β-propeller repeats (169). Cft1 has been shown to mediate site selection within poorly defined regions surrounding cleavage site (169). Northern blot analysis of the 3' UTR of the ACT1 mRNA revealed

4 sites utilised by alternative polyadenylation. In wild type *S. cerevisiae* the most proximal site is favoured whilst, in Cft1 mutants, the distal sites are utilised (169). However, global analysis of poly(A) tails and their length have shown that Cft1 does not exert an effect on all mRNAs but only a subset (169).

1.2.1.2.3 Mpe1

Effects on poly(A) site selection have been observed with other components of CPF. Northern blotting analysis of ACT1 poly(A) tails revealed that mutations in CPF component, Mpe1 resulted in higher usage of distal poly(A) sites rather than proximal site utilised by wild type (170). It has been suggested Mpe1 interacts with RNA through use of its zinc knuckle domain supported by the fact that the zinc knuckle domain of mammalian polyadenylation factor, CPSF-30 is known to bind mRNA (171). Mpe1 is essential for cell viability and *in vitro* assays have revealed mutation in Mpe1 results in deficient cleavage and polyadenylation. Antibodies raised against Mpe1 inhibit polyadenylation in wild type cell extracts. Furthermore, mutations in Mpe1 have been shown to suppress mutations in Pcf11 demonstrating the interdependence of the two complexes to carry out efficient processing (170).

1.2.1.2.4 Yth1

Yth1 is the yeast homolog of the 30kDa subunit of mammalian CPSF-30 and has been implicated in both cleavage and addition of the poly(A) tail (171). CPSF-30 binds RNA and favours U rich sequences (171). Mutation in Yth1 leads to detrimental effects on both cleavage and polyadenylation *in vitro*. Furthermore, Yth1 has been shown to bind RNA through use of its second zinc finger (ZF2) present within a five zinc finger structure. Yth1 binding within the 3' UTR has been mapped using RNase H protection assays employing overlapping oligonucleotides complimentary to the 3' UTR of CYC1.

The experiments revealed a region near the cleavage site free from degradation and an additional small amount of protection was observed near the efficiency element (172).

Mutation in ZF2 also leads to a decrease in RNA polyadenylation efficiency implicating Yth1 RNA binding as an important feature for RNA cleavage. The binding of Yth1 near the cleavage site suggests that it is required to stabilise the CPF complex whilst cleavage takes place. In addition, ZF2 has been implicated in binding other components of CPF such as Cft1 and Cft2. Mutations in ZF2 may weaken their interaction and thus the interaction of CPF at the cleavage site (172).

The homologue of Yth1 in *Drosophila melanogaster*, Clipper, was demonstrated as having endonucleolytic cleavage activity specific for RNA hairpins (173), although no such activity has been reported for Yth1 to date. It is possible that endonucleolytic cleavage requires the other subunits of CPF to induce a conformational change allowing the cleavage reaction to proceed (172).

1.2.1.2.5 Cft2

Cft2 is the 96 kDa subunit of CPF, a homologue of mammalian counterpart CPSF-100, and is a member of the β-CASP superfamily (77). Like Cft1 and Yth1, Cft2 also interacts with RNA near the polyadenylation site (147,169,172). Cft2 is essential for cell viability (98) and is required for cleavage and polyadenylation *in vitro* and poly(A) site recognition of ACT1 mRNA *in vivo* (121). Cft2 has also been shown to bind U-rich elements within the 3' UTR of mRNA around the poly(A) site (147). Mutations within Cft2 have detrimental effects on cleavage of RNA *in vitro* (121). Therefore, it has been suggested that like Cft1, Cft2 is required to facilitate poly(A) site selection (121). The ability of core CPF proteins (Cft1, Cft2, Yth1 and Mpe1) to bind

RNA and act to facilitate poly(A) site selection provides additional support for recognition and tethering of the polyadenylation machinery to the poly(A) site (147).

Both Cft1 and Cft2 interact with the phosphorylated CTD of RNAPII and have similar functions in RNA binding and RNAPII binding (121,169). However, run-on experiments employing probes designed to hybridise to the CYC1 mRNA termination site demonstrate that mutation of Cft1 severely affects transcription termination in the CYC1 gene (169) whereas mutations in Cft2 does not (121). The apparent lack of Cft2 role in transcription termination suggests that interaction with RNAPII occurs predominantly to direct the timing of polyadenylation machinery recruitment (121).

1.2.1.2.6 Ysh1

The above results highlight the dynamic nature of the CPF complex in site selection and regulation of transcription elongation and termination. Once associated with the pre-mRNA, the CPF complex is responsible for cleavage at the poly(A) site. CPF factor, Ysh1 is required for cell viability (174). Both Ysh1 and its mammalian homologue, CPSF-73 contain a highly conserved metallo-β-lactamase domain at the N-terminus (175). The metallo-β-lacatamase family are a group of zinc dependant hydrolases implicating CPSF-73 and Ysh1 in the cleavage reaction of polyadenylation (176,177). In addition, both Ysh1 and its mammalian counterpart, CPSF-73, contain a β-CASP domain which is involved in coordinating two metal ions. Loss of metal ion binding leads to failure of the cleavage reaction during polyadenylation which is rescued upon addition of $ZnCl_2$. Furthermore, the cleavage reaction is inhibited in nuclear extracts upon introduction of zinc specific chelators and EDTA (178).

The crystal structure of human CPSF-73 shows the structural basis of nuclease activity (179). Residues 1-208 form a canonical metallo-β-lactamase structure with a four layered αβ/βα shape. Residues 395-460 form an additional part of the metallo-β-lactamase region and add to the central β sheet structure. The β-CASP domain is present between residue 209-394 and forms a 6 stranded β-sheet that is surrounded by 6 α-helices on both sides. The coordination of metal ions is mediated by 5 conserved motifs, 4 of which are present in the metallo-β-lactamase region and the fifth present within the β-CASP domain. Zinc was shown to bind the protein with an extremely high affinity and the presence of the metal ion suggests the protein is capable of nuclease activity. Residues present in the 5 motifs act to stabilize the zinc ion within the active site and are conserved with other metallo-β-lactamase containing proteins (179). These conserved motifs are present in *S. cereviase* and mutation of one of the conserved residues within the motif results in lethality (180). A conserved histidine residue is utilised as a general acid for the nuclease reaction which is activated by an Asp/Glu residue. Mutation of the histidine residue result in lethality in *S. cerevisiae* (180).

1.2.1.2.7 Ydh1

Ydh1 also contains a metallo-β-lactamase and β-CASP domains as observed in CPF subunit, Ysh1 (175). However, the zinc binding residues are not conserved in the mammalian counterpart of Ydh1, CPSF-100 (177,181). The crystal structure of yeast Ydh1 reveals a similar domain arrangement with no zinc ions present (179). The loss of zinc ions both in mammalian CPSF-100 and yeast Ydh1 suggests that they are not catalytically active.

1.2.1.3 Poly(A) polymerase 1

Once the cleavage reaction has been completed, the next step is addition of the poly(A) tail. Pap1 is required to synthesise the poly(A) tail and is required for both cleavage and polyadenylation in mammals but is only required for the polyadenylation step in yeast (87). *S. cerevisiae* Pap1 has a molecular weight of 63 kDa and is able to elongate RNA primers in isolation although this requires the addition of other protein factors in order to achieve specificity of the polyadenylation site and regulation of poly(A) tail length (182).

Pap1 is directed to the polyadenylation machinery by Fip1 (183). Fip1 is a component of the CPF complex and interacts directly with Pap1 (158) through a linker region (184). Indeed, *in vitro* experiments lacking the Fip1 linker resulted in severe defects in addition of the polyadenylate tail (184). Fip1 has low conservation across species and structural prediction suggests that a large amount of the protein is disordered (184,185). However, these predictions also reveal the presence of conserved structured domains within Fip1 that have been found to mediate interaction with the C-terminus of Yth1 (172,184). Mutation of the C-terminal region of Yth1 results in defects in polyadenylation *in vitro* due to weakening of the interaction with Fip1 which leads to a decrease in Fip1-Pap1 recruitment to the CPF complex (172). The recruitment of Pap1 to the CPF complex facilitates efficient Pap1 function, as Pap1 alone polyadenylates any substrate in a non-specific and distributive manner (98,186). Therefore, the Yth1-Fip1-Pap association mediates efficient accurate addition of the poly(A) tail (172).

The structure of Pap1 has been solved and shows three globular domains. The active site requires the coordination of two Mg^{2+} ions in order to carry out catalysis (187). The first is required to act as a cosubstrate and binds as $MgATP^{2-}$. The second is

required to stabilise the negative charge and orient the newly cleaved 3' end and the ATP molecule. A recent ternary structure of Pap1, MgATP and a 5 nucleotide poly(A) revealed that the N and C terminal domains of Pap1 form a cleft type structure (188). A conformational change is induced in this cleft when the 3' mRNA end and MgATP bind resulting in a closed conformation that enables catalysis to proceed (188). Pap1 shows a preference for polyriboadenylate and kinetic studies support the notion that an induced fit model may drive nucleotide specificity (189).

1.2.1.4 Poly(A) binding protein 1

Regulation of the poly(A) tail length has been ascribed to poly(A) binding protein (Pab1) (105,190,191) and depletion of Pab1 results in lengthened poly(A) tails *in vivo* (191). Pab1 associates with the tail when it is around 11-14 nucleotides in length to stimulate initiation of translation and regulate mRNA degradation through the deadenylation pathway (192–195). Binding of Pab1 continues until the poly(A) tail has been completely synthesised and results in an 80-fold increase in polyadenylation efficiency (196,197). Poly(A) tail length is controlled by the 3'-5' exonuclease Pab1-dependent poly(A) nuclease, PAN and Nab2 (79,105,190,198). Nab2 associates with the poly(A) tail through three tandem Cys-Cys-Cys-His zinc fingers. Mutation of *nab2* leads to defects in length of the poly(A) tail resulting in transcripts that are hyperadenylated (198). In addition Nab2 mutants displayed a decrease in poly(A) RNA binding affinity *in vitro* and these mutations resulted in aberrant poly(A) tail length *in vivo* (199).

1.3 Sequences that direct polyadenylation

In addition to the large number of protein factors required to carry out 3’ end mRNA processing, a number of RNA sequence elements are required to direct cleavage and polyadenylation. As has been mentioned previously, these elements act as recognition sequences for the components of the polyadenylation machinery to bind.

Experiments to define these sequences have demonstrated that addition of the 3' UTR of the CYC1 gene transcript led to cleavage and polyadenylation of artificial RNAs *in vitro* (200,201). Cleavage and polyadenylation were found to be dependent upon a 38 bp region of the 3' UTR (202). Deletion of this region in *cyc1-512* mutants resulted in increased elongation of mRNAs but with a 90% reduction in total amount of CYC1 mRNA in *S. cerevisiae* (202). The effect deletion of these regions within the 3' UTR had on transcription termination *in vivo* also demonstrated a clear coupling of the polyadenylation and transcription processes (203).

Further analysis of this *cyc1-512* mutant uncovered sequence elements that act in concert to direct processing at different polyadenylation sites within the CYC1 gene (204,205). The first was initially characterised by an upstream element characterised by either UAUAUA, UAGnnnUAUGUA or UUUUUAUA motifs. These sequence motifs were able to enhance efficiency in the *cyc1-512* mutant and are now known as the efficiency element (204,205). The efficiency element is thought to be positioned around 10-30 nucleotides from the positioning element although this is highly variable from gene to gene (205). The second is a downstream element composed of the sequences UUAAGAAC and AAGAA in the CYC1 gene (204,205). These sequences directed the position of the polyadenylation site and are therefore known as the positioning element (204,205). It was also demonstrated that cleavage within the 3' UTR of the ADH1 gene

was positioned at a sequence composed of a pyrimidine followed by three adenosine residues, $Py(A)_3$ (206).

Insertion of an effeciency element, UAGUAUGUA, present in the cauliflower mosiac virus (CaMV) demonstrated the functionality of this polyadenylation site in *S. cerevisiae* (207). Mutational analysis of the efficiency element revealed that the hexanucleotide UAUAUA was required within the efficiency element to produce optimum polyadenylation *in vivo* (207). Single point mutations in the UAUAUA sequence can have profound effects on polyadenylation. Mutation of the first and last U nucleotides result in a reduction in activity to below 20% and 14 out of 18 single base substitutions within the element result in a decrease in polyadenylation to below 50% (207). However, in comparison with the conserved mammal polyadenylation motif, AAUAAA, where all mutations but one results in a decrease in polyadenylation to below 20%, the yeast motif is fairly degenerate.

The cleavage site, $Py[A]_3$, was first characterised in the ADH1 gene (206). Replacement of any of the three adenosine nucleotides following the pyrimidine base abolished cleavage at the respective site in *in vitro* processing experiments (206). Yeast cDNA sequences revealed that of the 8 different 3' ends arising due to alternative polyadenylation sites within the CBP1 gene, 6 of these ended with the motif $Py[A]_n$ (208). Furthermore, analysis of GAL7 3' ends revealed that the $Py[A]_n$ sequence was utilised as a cleavage site in this gene in both *in vitro* and *in vivo* processing experiments (209). Taken together these data implicate the $Py[A]_n$ as the most common signal for cleavage in *S. cerevisiae.*

Positioning elements were identified in the CYC1 3' UTR of *S cerevisiae* with the motif UUAAGAAC and AAGAA (205). The positioning element is normally placed in the same position from gene to gene, usually 16-27 nucleotides upstream of the polyadenylation site (205). Mutagenesis experiments revealed that the presence of AAAAAA and AAUAAA sequences resulted in optimum processing efficiency (210). However, the positioning element is largely degenerate with only one residue critical for activity (210). The degenerate nature of the poly(A) elements coupled with the functional redundancy demonstrated by the 3 polyadenylation sequence elements makes a true consensus sequence for each difficult to define (210,211).

Recently, a genome wide search of polyadenylation signals has provided a more detailed analysis of the polyadenylation sites used in *S. cerevisiae* (145). An expressed sequence tag (EST) library of 3425 *S. cerevisiae* cDNA sequences was employed to annotate polyadenylation sites within the yeast genome. The findings were largely in agreement with previous data with the UAUAUA sequence most commonly used. Point mutations within this sequence (UACAUA, UAUGUA) were also observed in efficiency elements in a significant amount of genes. The most common positioning element was AAUAAA, the same as that observed in higher eukaryotes, although this varies from gene to gene (145). The poly(A) site/cleavage site was previously identified as being a pyrimidine followed by 3 or 4 nucleotides (206). However, the genome wide analysis revealed U-rich sequences both before and after the cleavage sites not previously annotated (145). Presence of an A or C nucleotide surrounded by the U-rich sequences produces optimum cleavage (145). This suggests that yeast 3' UTRs are not that different to higher eukaryotic 3'UTRs where U-rich sequences are also present (145,212–214).

1.4 Model for polyadenylation in *S. cerevisiae*

From extensive analysis of the protein factors and sequences required for successful polyadenylation, we can start to build a detailed picture of the 3' mRNA processing event. As previously described processing of mRNA occurs cotranscriptionally and thus means that both transcription and processing are tightly coupled events.

Recruitment of processing factors has been demonstrated as early as transcription initiation. Ssu72 associates with TFIIB during transcription initiation and along with Glc7, plays an important role in dephosphorylation of the CTD of RNAPII during transcription elongation. Other factors also associate with the CTD of RNAPII in a phosphorylation dependent manner and it is likely that association with the CTD stimulates recruitment of the remaining processing factors through extensive protein-protein interactions. Once recruited to the nascent mRNA transcript, poly(A) site selection is achieved through interaction of CF1A (Rna15), CF1B and several components of CPF. Endonucleolytic cleavage of the pre-mRNA is then mediated by action of the β-CASP and metallo-β-lactamase domain of Ysh1p. After cleavage, Fip1 interacts with Pap1 to recruit the polymerase to the polyadenylation machinery. Recruitment is facilitated through contacts between Fip1 and Yth1. Pap1 then binds the newly cleaved 3' end and generates the poly(A) tail.

Pap1 can function independently of the processing machinery to synthesise the poly(A) tail, however, it does require interaction of other proteins to regulate tail length. Poly(A) binding protein (Pab1) is required to regulate the length of the poly(A) tail. It binds adenosine residues preferentially through use of RRMs. Further, it has been suggested that interaction with Rna15 recruits the protein to the nascent mRNA

transcript. Association of Pab1 adjacent to PAP results in a 80-fold increase in efficiency of polyadenylation. Pab1 also associates with poly(A) tails to stimulate translation initiatiation (194).

1.5 Coupling processing and transcription

Successful gene expression requires transcription, mRNA processing, nuclear export and translation. These processes are tightly coupled as is demonstrated by the association of protein factors with roles in different aspects of gene expression. Close coupling and cross talk between these processes is required as each is reliant on the successful outcome of both upstream and downstream events (86).

1.5.1 5’ end capping and 3' end processing

Capping and polyadenylation together provide a framework for subsequent translation of the mRNA transcript. Translation initiation factors are able to bind both the mRNA 5' cap and Pab1 which covers the poly(A) tail. In *S. cerevsiae,* translation initiation factor eIF4E is able to bind the 5' cap and in association with eIF4G and eIF3, is able to recruit the 40S ribosomal unit (215). These translation factors are also able to interact with Pab1 (3). These interactions induce circularisation of the transcript (85). Mutations which disrupt interaction between eIF4E and EIF4G or association of Pab1 and eIF4G inhibits translation initiation *in vitro* (194,216–218). Circularisation is thought to promote ribosome recycling enhancing translation efficiency. It also ensures that only intact mRNAs are translated (219).

1.5.2 Splicing and 3’ end processing

Protein factors involved in splicing are also thought to be involved in stimulating cleavage and polyadenylation through interaction with the 3' end processing machinery (220,221). The polyadenylation machinery has also been implicated in regulation of splicing of the most 3' intron (222). Recent findings have implicated U1 snRNP in alternative polyadenylation. In the bovine papilloma virus, the late polyadenylation signal is positioned near a 5' splice site which U1 snRNP binds to. Once bound the U1 70K subunit of the U1 snRNP interacts with Pap to inhibit polyadenylation (223). A subsequent whole genome study in U1 snRNP knockout HeLa cells revealed instances where cleavage and polyadenylation occurred at cryptic polyadenylation sites found within introns (224). Further whole genome analysis in *D. melanogaster* revealed cryptic polyadenylation sites within 30% of all introns. Two mechanisms were proposed to allow progression of the transcriptional machinery past polyadenylation sites. The first implied competition between the splicing machinery and the polyadenylation machinery with the splicing machinery able to splice out the polyadenylation signal containing intron before polyadenylation can take place. The second is a kinetic model whereby RNAPII reaches the splice site before recruitment of the polyadenylation machinery to the polyadenylation signal (225). These studies were performed in higher eukaryotes and have not been conducted in *S. cerevisiae.* Although 95% of *S. cerevisiae* genes lack introns (226), the presence of some cryptic polyadenylation sites within introns has been demonstrated in the yeast genome (227).

1.5.3 Export and 3' end processing

Addition of the poly(A) tail is essential for mRNA export (82,155,228). Replacement of signals known to direct polyadenylation with a cis-acting ribozyme within reporter constructs resulted in accumulation of the corresponding mRNA in the

nucleus (82,155). Coupling of 3' end processing and mRNA export is demonstrated in temperature sensitive CFI or poly(A) polymerase alleles where defects in export are observed (228). Temperature sensitive mutations in Rna14 and Rna15 result in mRNA retention in the nucleus around 8 minutes following a shift to the restrictive temperature (226). This is also observed in some mutations within Pap1, Hrp1 and Pcf11. In addition, processing activity was shown to be inhibited at this restrictive temperature but not at 23ºC. Mutations within components of the export machinery affect polyadenylation of mRNA transcripts. Strains with mutations in export factors Mex67 and Rat1 display transcripts with hyperadenylated poly(A) tails (229).

Hrp1 has been implicated in mRNA nuclear export as it is able to shuttle to the cytoplasm (226). The association of Hrp1 and the export pathway demonstrates a direct link between polyadenylation and mRNA export (226). Hrp1 was first identified as a suppresor of the temperature sensitive export defect phenotype in *npl3* mutants (66). Npl3 binds to the poly(A) tail of mRNAs and is able to shuttle rapidly between the cytoplasm and nucleus (230,231). Npl3 has also been implicated in polyadenylation site selection by competing with the polyadenylation machinery for RNA binding (232). It has been shown that high concentrations of recombinant Npl3 inhibit the cleavage and polyadenylation reaction *in vitro.* It was therefore suggested that Npl3 directs the cleavage and polyadenylation machinery to genuine poly(A) sites inhibiting the use of cryptic sites by competitively binding the RNA (232).

A hypothetical model suggests that Hrp1 binds during polyadenylation to package the mRNA transcript for export and stays until it has been exported from the nucleus (106). Once in the cytoplasm, Hrp1 is released from the mRNP package and reintroduced into the nucleus (106). It has been shown that Hrp1 is able to be

reintroduced to the nucleus through use of the nuclear import receptor Kap104p in yeast (233).

1.5.4 Transcription and 3' end processing

Coupling of transcription and processing is essential to maintain the efficiency of gene expression, exemplified by the observation that processing of mRNAs injected into oocytes is less efficient than co-transcriptional processing *in vivo* (234). The poly(A) site is cleaved ~30 seconds after it is transcribed (235). When uncoupled from transcription cleavage and polyadenylation has been shown to take over 20 minutes (78). The CTD of RNAPII is known to regulate aspects of mRNA processing through the recruitment of protein factors (236–238). The CTD is key to tethering these factors at the nascent mRNA transcript. It provides a scaffold for their assembly and greatly increases their local concentration (78).

1.5.4.1 The importance of the poly(A) site in regulation of transcription

Conversely, elements of the 3' end processing machinery have been implicated in maintaining and driving successful transcription (239). Experiments employing a tetracyclin inducible β-globin gene possessing a SV40 poly(A) signal was used to generate HEK293 cell lines. Mutation of this poly(A) site resulted in a decline in mRNA levels. ChIP analysis revealed association of RNAPII was increased downstream of the poly(A) site in the mutant upon comparison with the wild type (239). By contrast, association of RNAPII at the promoter and poly(A) site was decreased 3 fold and 2 fold respectively and wild type mRNA transcript levels were 5-fold higher than those observed in the mutant (239). The ChIP experiments revealed that relative to wild type, association of transcription factor TBP with the promoter decreased 4 fold in the mutant (239). This reduction correlates with an increase in association of TBP with

the coding sequence and terminator regions (239). In addition, a reduced amount of transcription factor TFIIB coupled to the promoter and an increased amount in the coding sequence is also observed. Both mutant and wild type have similar levels of TFIIB associated with the terminator region of the transcript but an increase in readthrough of the proximal poly(A) site utilised in wild type is apparent in the mutant (239). It has been suggested that the poly(A) site mediates recycling of transcription factors and RNAPII back to the promoter for transcription initiation (239). Indeed, hyperphosphorylation of S2 in the CTD of RNAPII correlates with an increase in association of human S2 kinase CDK9 at the distal poly(A) site and is likely responsible for the aberrant association of transcription factors within the coding region (239). Similarly, in *S. cerevisiae*, depletion of the yeast CDK9 homologue, Ctk1, results in irregular association of transcription factors in the elongation complex during transcription (240). Similar effects on TBP and TFIIB are observed when Pcf11 is depleted by RNA interference suggesting a general linkage between transcription rates and polyadenylation (239).

1.5.4.2 The role of Ssu72 and Pta1 in transcription regulation

Ssu72 is recruited at the 5' end and maintains a dual function in both 3' end processing and transcription initiation, roles that are independent of one another (166). Ssu72 was identified as an interaction partner of transcription factor TFIIB impliying a function in transcription initiation/elongation (161,162). It contains a CX_5RS motif characteristic of tyrosine phosphatases and has been shown to desphosphorylate S5 on the CTD (166,241,242). Depletion of Ssu72 leads to an increase in levels of S5 phosphorylation and apparent inhibition of transcription *in vitro*. It has been suggested that Ssu72 dependent dephosphorylation of RNAPII at S5 generates a

hypophosphorylated form of RNAPII allowing recycling to the promoter where it is able to participate in transcription initiation (166).

Pta1 associates with Ssu72 and the transcription factor, Sub1, which interacts with the transcription machinery at the promoter (122). Sub1 also makes contacts with transcription factor TFIIB although mutational analysis suggests that Sub1 and Ssu72 proteins cannot bind simultaneously (162,243). Interaction of Pta1 with Sub1 and Ssu72 is also mutually exclusive and a hypothetical model has been proposed whereby interactions between these proteins act to regulate the association of protein factors and the process of transcription (122). At transcription initiation Ssu72 associates with TFIIB at the promoter to recruit other components of CPF. ChIP analysis of proteins associated at the promoter have identified the presence of CPF components (114). Upon promoter clearance, the phosphorylated CTD associates with the CPF complex, Ssu72 dissociates from TFIIB and associates with Pta1 in the CPF complex. TFIIB is then free to bind Sub1. During elongation, Sub1 associates with Rna15 and upon selection of a poly(A) site by Rna15, transcription termination is signalled. Transcription termination is facilitated by the dissociation of Sub1 from Rna15 and its association with Pta1 subsequently stimulating release of Ssu72 and resulting in cleavage within the 3’ UTR (122).

1.5.4.3 Gene looping and transcription termination

Interaction of components of the 3' end processing machinery have been shown to facilitate gene looping (244). Gene looping is the association of the 5' and 3' ends of DNA undergoing transcription and it is thought to function in transcription regulation (245,246). Loop formation is driven by transcription factor TFIIB which is present at both ends of the transcript (247,248). In addition to its interaction with Ssu72, TFIIB

can also bind to Rna15 (161,162,247,248). A complex of TFIIB, Rna14, Rna15, Pcf11 and Pap1 was found to be present exclusively within looping competent strains of *S. cerevisiae*. In looping defective strains, this complex was not formed and transcription activation was inhibited (244). Gene looping has been proposed to facilitate transcription by promoting recycling of RNAPII from the terminator sequence to the promoter (249) resulting in enhanced transcription as the rate limiting recruitment step is minimised (250). Mutations in Clp1 resulted in decreased recruitment of CF1A to the 3' end of transcripts which led to defects in transcription termination (250). As well as these effects, loss of association of RNAPII at the promoter was also observed (250). Gene looping has also been demonstrated as dependent on Ssu72 and Pta1 (249). Mutation of Ssu72 or depletion of Pta1 resulted in inhibition of loop formation (249). Another model suggests that gene looping enables Pta1 and Ssu72 to associate with TFIIB at the promoter to facilitate transfer of the polymerase and allow Ssu72 directed dephosphorylation of S5 in the CTD (249).

1.5.4.4 Transcription termination – The allosteric and torpedo models

Components of CF1A have been implicated in transcription termination of the CYC1 gene in *S. cerevisiae* (251). Mutation in Rna14, Rna15 and Pcf11 resulted in defects in transcription termination at the 3' end of the CYC1 gene (251). However, mutation in Pap1, Fip1 and Yth1 had little effect on transcription termination implicating only those protein factors involved in site selection and cleavage (251). Indeed, it has been observed that recognition and cleavage of the poly(A) site signals termination of transcription (252). Two models named the torpedo (253) and allosteric model (254–256), have been proposed that may explain how transcription termination occurs in *S. cerevisiae* (Figure 1.12) (257). Both models rely on the poly(A) site for recognition and induction of transcription termination but differ in the mechanism

behind which transcription termination is directed. In the allosteric model, recognition of the poly(A) site results in either destabilisation or conformational change in RNAPII and the elongating complex. This signals for recruitment of termination factors which facilitate in dissociation of RNAPII from the DNA template. The torpedo model relies on endonucleolytic cleavage at the poly(A) site which then allows entry of 5'-3' exonuclease Rat1 which degrades the uncapped cleaved transcript. The resulting short mRNA induces arrest of RNAPII and the elongation complex and promotes termination (257). Evidence in support of both models has been demonstrated. For the allosteric model, plasmid driven transcription in *Xenopus laevis* oocyte nuclei was visualised by electron microscopy (EM) and demonstrated that cleavage was not essential to transcript termination (258). Further EM analysis of over 100 genes in *Drosophila melanogaster* showed that in majority of cases the transcript was released before cleavage at the poly(A) site (259). Furthermore, it was demonstrated that whilst mutation of the poly(A) signal has detrimental effects on the cleavage and polyadenylation reactions it does not affect transcription termination in the FBP1 gene. This suggests that recruitment of the 3' end complex machinery is in itself sufficient to direct transcription termination (260). In support of the torpedo model a mutational analysis of *S. cerevisiae* Rna15 showed no effect on recruitment of the polyadenylation machinery to the poly(A) site but did impair the cleavage reaction leading to detrimental effects in transcription termination (251). In addition, the 5'-3' RNA exonuclease Rat1 was demonstrated as a requirement for transcription termination (261). ChIP experiments revealed that inactivation of Rat1 results in an increase in RNAPII association downstream of the poly(A) site suggesting that Rat1 is required to maintain wild type RNAPII association and to facilitate normal transcription termination (261).

The contradiction between the torpedo and allosteric models can be reconciled by an integration of the two models whereby degradation mediated by Rat1 does occur but does not result in transcription termination (262). Rather, Rat1 functions in termination as it enhances recruitment of 3' end processing factors to the cleavage site (262). Rat1 recruitment to the 3' end is also dependent upon interaction with Pcf11 (262). Pcf11 is implicated in transcription termination and it has been suggested that it directs conformational changes in the CTD of RNAPII through interaction with its CID (112,263). Indeed, Pcf11 is essential for both 3' end processing and transcription termination (112). However, mutations in Pcf11 outside of the CID also induced defects in transcription termination suggesting that the stimulation of conformational changes within the CTD of RNAPII does not occur (262). Therefore, the termination model can be modified such that neither conformational changes nor RNA degradation result in transcription termination but that cleavage is carried out by components of the 3' end machinery bound to RNAPII CTD and Rat1. The cleavage reaction then results in Rat1 mediated degradation of the cleaved RNA and an allosteric change is transmitted to RNAPII causing transcription termination (262). The CPF component, Cft1, is proposed to maintain crosstalk by binding both the mRNA transcript and RNAPII (169). Upon site recognition, this crosstalk is thought to initiate transcript termination which is consistent with the above integrated termination model (169).

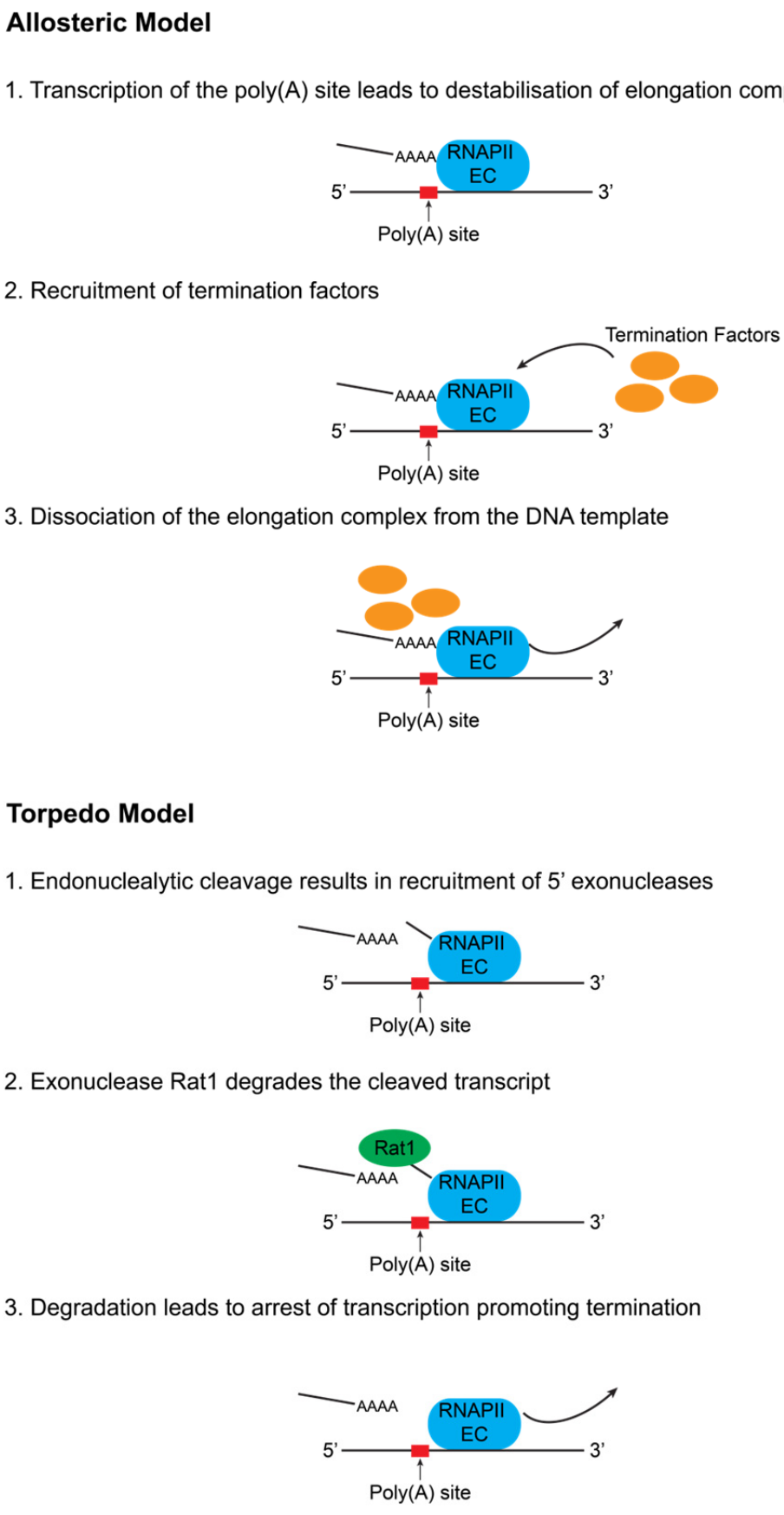


Figure 1.12 The Allosteric and Torpedo models of transcription termination.

1.6 Mammalian homologues

The yeast 3' end processing machinery is very similar to that found in higher eukaryotes and many comparisons can be drawn based upon structure and function (Table 1.1) (8). The mammalian 3' end processing machinery is made up of a number of protein factors that can be grouped into complexes, as observed in yeast, that act in concert to carry out polyadenylation *in vivo* (264–266). These include the Cleavage and Polyadenylation Specificity Factor (CPSF), Cleavage and stimulation Factor (CstF), Cleavage Factor I (CF I_m) and Cleavage Factor II (CF II_m). Other protein factors have been identified that are required for processing. These include Poly(A) Polymerase (PAP), poly(A) binding protein (PABP) and the C-terminal domain of RNA polymerase II. All these protein factors, apart from PABP are required for *in vitro* cleavage and only CPSF, PAP and PABP are required for *in vitro* polyadenylation (87).

The homologues in each system are often arranged into different complexes. For example Yth1 and Fip1 present in the yeast CF II complex of CPF are actually present within the PF I component of CPSF in the mammalian system and are named CPSF-30 and hFip1 respectively. CF1A also contains subunits with homologues present within complexes CstF and CF II_m of the mammalian system. However, neither Hrp1 (yeast) or CstF-50 (human) have homologues in other systems. Also, homologues in either system can have different functions, for example, CPSF-160 recognises the AAUAAA polyadenylation signal whereas its homologue in yeast, Cft1, is mostly associated with the cleavage site (87).

Some factors have homologues that carry out the same function in yeast as in mammals. For example, Rna15 recognition of G/U rich nucleotides is also observed in mammalian counterpart, CstF-64 (141,143,144). CPSF-73 was identified as the factor

required to carry out endonucleolytic cleavage through use of its β-CASP and metallo-β-lactamase domains (176,177). These domains are also present in Ysh1 and have been demonstrated to carry out this function in the yeast system (175).

The sequence elements required to direct polyadenylation in mammals have been identified. They consist of the polyadenylation signal (PAS), the downstream element (DSE), the cleavage site and the auxiliary upstream element. The polyadenylation signal is similar to that observed in yeast. Extensive studies employing 13,000 human and mice expressed sequence tags (ESTs) demonstrated that only 4% of genes did not possess a PAS and that AAUAAA and AUUAAA are the most common PAS sequences present in 70% and 15% of transcripts respectively (267). Mutations in this sequence element result in deficient polyadenylation and increased amounts of unprocessed mRNA in *Xenopus laevis* oocytes (268). Patients with mutations in the final base of the hexamer sequence (A to G) suffer from α and β-thalassaemia (89).

The DSE is less conserved than the PAS and deletion of this sequence element results in a 3-fold reduction in general polyadenylation (269). The DSE can be both G-U rich and U rich but mRNA transcripts can have both, one or neither of these sequences (270–272). Point mutations in this sequence element have small effects on cleavage activity while deletion has a more significant impact (273,274).

The cleavage site is generally positioned between the PAS and the DSE (275). The sequence around the cleavage site is not conserved but in vertebrate pre-mRNAs optimal cleavage was observed where a CA dinucleotide was present and this is observed in 59% of 269 sequences examined (276–278).

The auxiliary upstream element is located upstream of the PAS and does not have a consensus sequence but is often U-rich and promotes enhancement of cleavage and polyadenylation (279). The downstream element is generally G rich but lacks a conserved distance from the cleavage site and can be present more than once within a gene (280–284).

1.7 RNA Sequencing

Recent advances in RNA sequencing technology have provided a powerful tool to observe the function and importance of individual components of the large polyadenylation machinery. The advantage this has over previous methods is that effects of mutations within the processing machinery can be observed on a genomic scale whereas previous experiments demonstrated effects only on a small number of housekeeping genes. This left room for argument over particular findings in that although a particular affect was observed in the expression of one gene it was possible that this was not maintained throughout the whole genome. The observation of alternative polyadenylation, differences in gene expression and bidirectional transcription only complicates this issue further. RNA sequencing eliminates these potential problems and can, for example, elucidate the location of polyadenylation sites within 3' UTRs and the frequency at which they are used whilst only requiring small amounts of experimental material.

Previous genome wide studies were derived from Expressed Sequence Tag (EST) databases but the high error rate and lack of coverage precludes an in depth analysis of polyadenylation sites. This is in part due to low numbers of full length ESTs, chimeric sequences, internal cDNA priming events and low quality sequencing data at the end of EST sequences (285–287). Other techniques such as Rapid Amplification of

cDNA Ends (RACE), RT-PCR and nuclease sensitivity assays are not able to provide the same depth, sensitivity and throughput as that obtained by RNA sequencing. The 3' ends of yeast mRNAs were first analysed by RNA sequencing in 2008 but the resolution did not achieve distinct mapping of individual poly(A) sites (288). However, it did reveal that 74.5% of the unique sequence in the yeast genome is transcribed and globally mapped the 3' UTR boundaries of 5212 genes. Overlapping genes were discovered whereby transcription occurred in both the sense and antisense direction. Of 4646 known genes, 793 contain overlapping 3' UTRs (288). Furthermore, sequencing data analysis revealed regions of the genome which are transcribed that had not been previously identified. The length of the 3' UTR ranges from 0-1461 bases with a median length of 104 bases (288). Sequencing analysis also revealed 540 genes had the ability to utilise more than 1 poly(A) site. An example of this was observed in the ACT1 gene in which at least 2 regions were present where a poly(A) tail was added (288).

A more in depth analysis of 3' end in both human and *S. cerevisiae* transcripts was presented in Ozsolak *et al*. This study obtained 7,036,730 reads by direct RNA sequencing, a method which eliminates the need for reverse transcription, which were aligned to the yeast genome (289). Each read corresponded to a polyadenylation site of an independent transcript (289). The results were in agreement with the previous study but due to higher resolution of the sequencing technique identified additional cleavage sites which are utilised to a lesser degree. The median 3' UTR length of 5759 open reading frames was 166 nucleotides (289). In 72.1% of genes analysed polyadenylation sites were separated by at least 50 nucleotides and in most a longer spacing in between sites is observed (289). This further enforces the complexity of the *S cerevisiae* polyadenylation system. Bidirectional transcription is evident in 14% of yeast genes where open reading frames are orientated in a tail to tail fashion and have overlapping 3'

ends (289). Another 14% contain cleavage sites within the exons or introns of open reading frames which links with splicing effects on alternative polyadenylation (section 1.1.2) (289).

The sequences surrounding cleavage sites were also analysed on a genome wide scale. In humans, a novel UUUUUUUUU sequence was discovered around 21 nucleotides upstream of the polyadenylation site present in both genic and intergenic regions (289). Palindromic sequences were also discovered and coincide with the U rich motifs (289). The role of these elements remains unclear but it is likely that these elements assist in positioning of the poly(A) site or maintaining efficiency of the cleavage reaction. No novel motifs were identified for *S. cerevisiae* but a yeast-like efficiency element was identified in a number of human poly(A) sites immediately upstream of the cleavage site (289). This suggests that yeast-like polyadenylation patterns exist within the human transcriptome/genome. Further RNA sequencing analysis of the fission yeast, *S. pombe*, revealed regulatory sequences required to direct polyadenylation (290). The efficiency element UGUA was found to be most prevalent within polyadenylation sites, a sequence which has previously been described as an upstream element in both *S. cerevisiae* and humans (145,290,291). Also, U rich sequences were observed in 3' UTRs of *S. pombe*, as seen in *S. cerevisiae*, highlighting a potentially conserved mechanism for polyadenylation (290,292).

The extensive characterisation of 3' UTRs in *S. cerevisiae* and higher eukaryotes provides detailed information on the mechanism of polyadenylation. Recent RNA sequencing studies aimed to characterise the importance of proteins involved in polyadenylation. A novel method aimed to characterise the effect on cleavage position when protein factors were mutated. This involved isolating poly(A) tails and the

immediate 5' transcript sequence from wild type and mutant strains (293). The technique was first implemented to characterise the structure of 3' UTRs in *C. elegans* and was then adapted to look at the state of poly(A) tails in *S. cerevisiae* in both wild type and export factor Yra1 depletion mutants (293,294). Comparison of wild type and Yra1 depleted mutants revealed that depletion of Yra1 resulted in effects on poly(A) site choice (294). This was observed by analysing 2.70 million uniquely mapped reads in the wild type and 1.58 million in the mutant sample (294). The ability of this technique to provide this amount of coverage and depth of the genome provides a powerful tool to demonstrate the importance and influence of 3' end processing factors.

1.8 Research Objectives

Rna14 and Rna15 are essential proteins required for successful 3’ end processing in *S. cerevisiae*. Rna15 is known to bind RNA within the 3’ UTR to direct positioning of the processing reaction. The crystal structure of the RRM of Rna15 has been solved and illustrates two sites that direct RNA binding. Further NMR experiments revealed additional residues that are affected by titration of an RNA oligonucleotide. Despite these findings the mechanism by which Rna15 interacts with RNA to direct polyadenylation is poorly understood. In addition, specificity of Rna15 is debated and whether Rna14 association with Rna15 enhances RNA binding affinity is unclear.

The objectives of this thesis are therefore to understand the importance of Rna15-RNA binding for 3’ end processing and transcription activity in the entire transcriptome of *S. cerevisiae*. In order to achieve this a series of *in vitro* experiments are employed including fluorescence spectroscopy and nuclear magnetic resonance (NMR) to identify the residues crucial to maintain wild type RNA binding, characterise further the nucleotide specificity of Rna15 and analyse effect of Rna14 on the Rna15-

RNA interaction. The importance of RNA binding residues demonstrated in the *in vitro* experiments is tested further by *in vivo* experiments whereby growth phenotype of *S. cerevisiae* is assayed. The impact of these mutations on transcription and processing is analysed by RNA sequencing and qRT-PCR.

2. Materials and Methods

2.1 Bioinformatics

2.1.1 DNA and protein information

The 5' UTR and coding sequences of Rna14 and Rna15 were obtained from the *Saccharomyces* genome database (http://www.yeastgenome.org/). Automated DNA sequencing was performed by GATC Biotech (http://www.gatc-biotech.com/en/index.html) with results analysed using SeqMan Pro (v9 1.0) software from the DNASTAR Lasergene 9 sequence analysis suite. Purification and concentration of a recombinant protein requires calculation of its theoretical isoelectric point (pI), molar extinction coefficient at 280 nm (ε_{280}) and molecular weight (M_r). These parameters were calculated using the ExPASy Prot Param tool (http://web.expasy.org/protparam/).

2.1.2 RNA-Sequencing Analysis

2.1.2.1 Global analysis of expression

A global analysis of expression in wild type, Y27A, Y27A/R87A/K90E and Δ16-94 was completed by Dr. Nikolay (N.I.M.R.). Reads were aligned to the yeast genome (ensemble version 75) using Tophat (version 2.0.9). Raw counts were determined using the union method in htseq-count (version 0.5.4p3) and mappings were filtered for a phred score of >10. EdgeR (Version 3.2.4) was used for filtering of lowly expressed features, normalisation and statistical analysis. Statistically significant genes were determined using the exact binomial test (295) and genes with FDR < 0.05 are reported. All mutants were analysed for statistically significant differences in expression upon comparison with wild type.

2.1.2.2 Analysis of 3' end processing

mRNA Sequencing data from Rna15 mutant and wild type *S. cerevisiae* strains were obtained from the high throughput sequencing facility (N.I.M.R). Analysis of the subsequent data was carried out using the program Tophat (http://tophat.cbcb.umd.edu/), Bowtie (http://bowtie-bio.sourceforge.net/index.shtml), the Cufflinks package (including cufflinks, cuffmerge and cuffdiff) and CummeRbund (http://cufflinks.cbcb.umd.edu/). Tophat aligns the short reads generated during the sequencing reaction to a reference genome using bowtie which is a general purpose short read aligner. Cufflinks then assembles transcripts with cuffcompare comparing these transcript assemblies to the annotated reference genome. Cuffmerge is used to merge two or more transcript assemblies. The merging of transcript assemblies provides a uniform basis for calculating expression levels and tests the statistical significance of observed changes. Cuffdiff then acts to distinguish between differentially expressed genes. The method of analysis is detailed in (296).

Aligned reads were viewed using the Integrative Genomics Viewer (IGV) Version 2.3.3.2. The 3' UTR of genes were analysed by measuring read number at three arbitrary sites within the 3' UTR in both wild type and mutant. Read count was normalised against expression of the gene and changes in amount of mRNA in the 3' UTR were expressed relative to wild type amount. Genes were grouped based on their role in cellular processes. Identification of appropriate genes for analysis was achieved using the Saccharomyces Genome Database (http://www.yeastgenome.org/).

In order to determine differences in processing activity between wild type and mutant *S. cerevisiae* strains, the amount of reads at each poly(A) site were measured and normalised against gene expression. The difference in mRNA amount prior to

processing and after was calculated for each site and expressed as a percentage of processing activity. This was carried out for both wild type and mutant and processing in each was compared. Poly(A) site positions were located in the aligned reads using annotated poly(A) sites listed in the Saccharomyces Genome Database (http://www.yeastgenome.org/).

2.2 Molecular Biology

2.2.1 Bacterial Strains

E. coli NovaBlue Singles™ competent cells (Novagen) were used for cloning purposes whilst recombinant proteins were expressed in *E. coli* BL21(DE3, pLysS) Singles™ competent cells (Novagen). The expression strain contains a chromosomal copy of T7 polymerase gene which is under the control of the lac operon promoter. Addition of IPTG induces expression of the T7 polymerase which is required for recombinant protein expression from the vectors employed. The genotypes used in this study are detailed in Table 2.1.

Strain Name	Genotype
NovaBlue	*endA1 hsdR17* (r_{k12}^- m_{k12}^+) *supE44 thi-1 recA1 gyrA96 relA*
BL21 (DE3)	*F⁻ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm* (DE3) pLysS (Cam^R)

Table 2.1. Genotypes of *E. coli* strains.

2.2.2 Yeast Strains

The Δ*rna14* and Δ*rna15* deletion yeast strains were obtained from the European Saccharomyces Cerevisiae Archive for Functional Analysis (EUROSCARF-http://web.uni-frankfurt.de/fb15/mikro/euroscarf/yeast.html). The strains provided were diploid where one chromosomal copy of both the RNA14 and RNA15 gene had been deleted by insertion of a Kanamycin cassette. Sporulation and subsequent selection of

the deletion strain was completed by Dr. J. Hedden (N.I.M.R). Genotypes of *S. cerevisiae* used in this study are shown in Table 2.2.

***S. cerevisiae* Strain** (EUROSCARF accession no.)	**Genotype**
rna14Δ/RNA14 (Y26194)	BY4743; Mat a/a; his3D1/his3D1; leu2D0/leu2D0; lys2D0/LYS2; MET15/met15D0; ura3D0/ura3D0; *rna14*::kanMX4/*RNA14*
rna15Δ/RNA15 (Y24412)	BY4743; Mat a/a; his3D1/his3D1; leu2D0/leu2D0; lys2D0/LYS2; MET15/met15D0; ura3D0/ura3D0; *rna15*::kanMX4/*RNA15*

Table 2.2 Genotypes of *S. cerevisiae* strains.

2.2.3 Plasmid constructs employed in *in vitro* studies

Plasmid constructs employed in the *in vitro* studies are detailed in Table 2.3. A series of Rna15 deletion mutants were cloned by ligation independent cloning in order to investigate the role of regions flanking the RRM in RNA binding. Constructs were cloned into pET52b.

A pETDUET vector containing both full length RNA14 and RNA15 was supplied by Dr. L. Arnold, N.I.M.R. In this construct Rna14 was tagged with a 6x histidine tag allowing purification of both proteins as a heterodimer complex.

2.2.4 Mutant constructs employed in *in vitro* studies

A series of point mutations were also introduced into *in vitro* construct Rna15 (2-103) by site directed mutagenesis following the QuikChange protocol (Stratagene).

2.2.5 Plasmid constructs employed in *in vivo* studies

Constructs employed in *in vivo* studies were generated by PCR amplification of the target sequence from *Saccharomyces cerevisiae* genomic DNA. Forward primers were designed to amplify from 500bp 5' of the start site to include the promoter region of the gene and included a *SacII* restriction site. Reverse primers employed for amplification of RNA15 included either an *Xba1* or *Pst1* site. In the case of RNA14, reverse primers used contained either a *Nhe1* or a *Pst1* restriction site as *Xba1* is present in the open reading frame of RNA14. Restriction digestion of *Nhe1* sites produces DNA ends able to ligate into an *XbaI* cleaved site. Following digestion of amplified DNA, the inserts were then ligated into yeast shuttle vector, pRS315 (provided by Dr. J Hedden, N.I.M.R). The pRS315 vector contained a 444 nucleotide sequence encoding a triple repeat of human influenza haemagglutinin DNA (HA-tag). The HA tag was present immediately 3' of a stop codon flanked by two restriction sites, XbaI (5') and Pst1 (3'). Four constructs were generated exploiting the presence of these two restriction sites. Restriction digest of the *Xba1* site and subsequent ligation resulted in insertion of both genes so that they were untagged. However, digestion of DNA at the Pst1 site and ligation results in an in frame fusion so that the HA tag is present on the C-terminal end of both Rna14 and Rna15.

2.2.6 Mutant constructs employed in *in vivo* studies

RNA14 (Δ626-677) construct was cloned by using wt RNA14-pRS315 HA tagged construct as a template for PCR. Primers were designed so that the 3’ primer annealed to and extended through the vector whereas the 5’ primer annealed adjacent to residue 626 effectively deleting residues 626-677 from the construct. An additional ligation step was added to re-circularise the DNA.

RNA15 RRM single, double and triple mutations were inserted by site directed mutagenesis. The RNA15 RRM deletion mutant Δ16-94 was cloned by PCR where primers annealed to the sequence immediately adjacent to the coding sequences for residues both 16 and 94 and amplified outwards to effectively delete residues 16-94 from the construct. A ligation step was added to re-circularise the DNA.

Construct Name	Vector	Organism	Marker	Function
Rna15-FL	pET52b (novagen)	*E. coli*	AmpR	Expression of full length Rna15
Rna15-2-94	pET52b (novagen)	*E. coli*	AmpR	Expression of residues 2-94 of Rna15
Rna15-2-103	pET52b (novagen)	*E. coli*	AmpR	Expression of residues 2-103 of Rna15
Rna15-16-94	pET52b (novagen)	*E. coli*	AmpR	Expression of residues 16-94 of Rna15
Rna15 WT	pRS315	*S. cerevisiae*	AmpR/LEU2	Complementation of *Rna15Δ* strain
Rna14 WT	pRS315	*S. cerevisiae*	AmpR/LEU2	Complementation of *Rna14Δ* strain
Rna14 WT: Rna15 WT	pET-Duet1 (novagen)	*E. coli*	AmpR	Expression of both full length Rna14 and Rna15

Table 2.3 Wild type constructs used in this study. Mutant constructs are listed in chapter 3.

2.2.7 Polymerase Chain Reaction (PCR)

Template DNA was amplified by PCR using the Kod Hot Start Polymerase kit (Novagen) using a Mastercycler PCR machine (Eppendorf). In general, 50 µL reactions were set up in 200 µL thin-walled tubes (Abgene/Thermoscientific) with either 100 ng genomic DNA or 10 ng of plasmid DNA. The protocol consisted of 20-40 cycles (depending whether template DNA was genomic or plasmid) of strand denaturation, primer annealing and extension. Following PCR, 5 µL of the reaction was taken and resolved on a 1 % agarose gel in order to determine the product size and assess yield. PCR products were purified using the QIAquick PCR Purification kit (Qiagen) according to manufacturer's instructions and eluted in 30 µL of dH_2O. A typical reaction set up and cycling conditions are shown in Tables 2.4 and 2.5.

Component	Volume (µL)	Final concentration
KOD Hot Start Buffer (10X)	5	1X
$MgSO_4$ (25 mM)	3	1.5mM
dNTPs (2mM each)	5	0.2mM (each)
Forward oligonucleotide primer (10 µM)	1.5	0.3µM
Reverse oligonucleotide primer (10 µM)	1.5	0.3µM
Template DNA	X	10 ng / 100 ng
KOD Hot Start	1	1 U/µL
dH_2O	X	-
Final reaction volume	**50**	

Table 2.4 Typical PCR reaction conditions.

Step	Temperature (°C)	Time
1. Polymerase Activation	95	2 mins
2. Strand Denaturing	95	20 s
3. Primer Annealing	55-57	10 s
4. Extension	70	20 s/ kb
5. Repeat Steps 2-4 for 20-40 cycles		
6. Final Extension	70	2 mins
7. Cooling	4	5 mins

Table 2.5 Typical cycle for PCR reactions. Annealing temperature was adapted for optimum amplification of the product.

2.2.8 Agarose gel electrophoresis

Agarose gel electrophoresis was used to determine DNA fragment size by comparison with the DNA size marker 1kb plus DNA ladder (Invitrogen). Gels were prepared by dissolving 1% (w/v) agarose (Bio Rad) in 1X TAE buffer (40mM Tris Acetate pH 8.3, 1mM EDTA). Ethidium Bromide (Bio-Rad) at a final concentration of 0.5 µg/mL was added to allow visualization of DNA bands by ultraviolet (UV) light at 254 nm.

2.2.9 Determination of DNA Concentration

DNA concentrations were determined using UV spectrophotometry by measuring absorbance at 260 nm on a NanoDrop® ND-1000 spectrophotometer (ThermoScientific).

2.2.10 Restriction Enzyme Digest

PCR products and plasmid vectors were digested using enzymes purchased from New England Biolabs. Double digests were carried out using a buffer suitable for both enzymes according to the NEBuffer Activity Chart for Restriction Endonucleases. Reactions were carried out for 1 hour at 37ºC. Where buffers and temperatures could

not be matched sequential digestions were carried out for two hours employing the correct buffers and temperature. Typical reaction conditions are detailed in Table 2.6.

Following digestion of PCR products, the reactions were purified using the QIAQuick PCR purification kit (Qiagen) following the manufacturer's instructions. DNA was eluted in 30 µL of dH_2O. Once digested, vector DNA was treated with Antarctic Phosphatase (NEB) to remove 5'-phosphate groups in order to prevent recircularisation and gel purified (section 2.2.3) using the QIAquick gel extraction kit (Qiagen) following the manufacturer's instructions.

Component	Volume (µL)	Final Concentration
NEB Buffer (10X)	3	1X
BSA (100X)	0.3	1.5mM
DNA (vector 1-3µg) (total PCR Product)	1-3/30	0.2mM (each)
NEB Enzyme 1 (10 U/µl)	1.5	0.3µM
NEB Enzyme 2 (10 U/µl)	1.5	0.3µM
dH_2O to final volume	**30**	

Table 2.6 Typical restriction digest conditions. For sequential digest the reaction was incubated with enzyme 1 in the correct buffer at the correct temperature for 1 hour. Heat inactivation at 65ºC for 10 minutes was performed before incubation with enzyme 2. An additional clean-up step was included before incubation with enzyme 2 using the QIAquick PCR purification kit (Qiagen). The reaction was held for a second hour in the correct buffer at the optimum temperature for enzyme 2.

2.2.11 Ligation

Ligation reactions were carried out using T4 DNA ligase (NEB) and incubated overnight at 16 °C. A typical reaction setup is detailed in Table 2.7.

Reagent	Volume (µL)
Linearized Vector (25 ng/µL)	4
Digested Insert (x ng/µL)	X
Reaction Buffer (10x)	2
T4 DNA ligase (400 U/µL)	1
dH_2O to final volume	**20**

Table 2.7 Ligation reaction. The amount of vector used was always 100 ng. A molar ratio of 3:1 (PCR Insert:vector) was used in all ligation reactions where the exact amount of insert added is dependent on concentration of the PCR product.

2.2.12 Ligation Independent Cloning

Constructs used for *in vitro* studies were cloned using Ligation Independent Cloning (LIC). Primers were designed to include a 21 base pair overhang corresponding with the LIC site in the cloning vector, pET52b. Deletion mutant constructs were cloned using primers that annealed at different sites within the Rna15 gene allowing amplification of the desired sequences only. PCRs were set up using 100 ng of *Saccharomyces cerevisiae* genomic DNA and followed the KOD HotStart Protocol as described in section 2.2.7. Extension time during the PCR cycle was altered depending on construct length. PCR products were purified using the QIAquick PCR Purification kit (Qiagen) according to the manufacturer's instructions and eluted in 30 µL of dH_2O. Purified PCR products together with pre-cut and purified pET52b vector (provided by Vangelis Christodoulou, N.I.M.R) were treated separately with T4 DNA polymerase. Around 0.2 pmole of PCR insert were treated with dATP while 1 pmole of vector was treated with dTTP for 20 minutes at room temperature. The enzyme was heat inactivated through incubation at 75°C for 20 minutes. Following heat inactivation, LIC reactions were carried out by adding 2 µl of insert to 1 µL vector (~50ng/µL) with 5 minutes incubation at room temperature. Following incubation, 1 µL EDTA was added and the reaction was incubated for a further 5 minutes. Finally, 2 µL of the LIC reaction

was transformed into 50 µL of chemically competent *E. coli* NovaBlue singles™ (Novagen) as described in section 2.2.8.

2.2.13 Transformation of Bacterial Strains

Plasmid DNA was transformed into *E. coli* Novablue or BL21 DE3 competent cells for cloning and protein expression purposes respectively. Generally, transformations were carried out using approximately 10 ng/µL of plasmid DNA which was then added to a 50 µL aliquot of the relevant competent cells and incubated on ice for 5-10 minutes. Cells were heat-shocked at 42 °C and then placed on ice for 2 minutes. Around 1mL of Luria-Bertani broth (LB) was added and the cells incubated at 37 °C for 1 hour. A 200 µL sample was plated on LB agar containing the appropriate antibiotic. Plates were incubated overnight at 37 °C.

2.2.14 Site-directed Mutagenesis

Site-directed mutagenesis of both protein expression and yeast shuttle vector constructs were performed using PCR as described in previous section using the KOD Hot Start DNA polymerase kit (Novagen). Mutagenic primers were designed in accordance with the QuikChange protocol (Stratagene) where overlapping primers contained the desired mutation and the KOD HotStart protocol was utilised for amplification. Following successful PCR, parent wild type DNA was digested using *DpnI*. This works on the premise that parent strand DNA is methylated after synthesis in *E. coli*. *DpnI* targets this methylation resulting in complete digestion of parental DNA. Constructs were then transformed into NovaBlue strain *E. coli* (Novagen) (as described in section 2.2.13) and plated on the appropriate selective medium. Plasmids were recovered using Qiaquick mini prep kit (Invitrogen) and successful mutagenesis was verified by DNA sequencing (GATC biotech).

2.3 Protein purification

2.3.1 Protein expression

Plasmids were transformed into BL21 (DE3) cells for expression using the method described in section 2.2.8. Cells were plated on LB-agar supplemented with ampicillin at 100 µg/mL and incubated at 37°C overnight. Starter cultures of 100 mL were inoculated with a single colony the following day and incubated at 37°C overnight with shaking (220 RPM). Around 7.5 mL of starter culture was used to inoculate each of the 12 2 L flasks containing 750 mL of LB + ampicillin media for large scale expression. Cultures were incubated at 37°C until cell density reached 0.6 OD. Protein expression was induced using 1mM IPTG and the temperature was decreased to 25°C and left to grow overnight following induction (~16 hours). Cells were then harvested by centrifugation at 4000 RPM for 20 minutes at 4ºC, supernatant was discarded and cells frozen and stored at -20ºC.

Rna15 (2-103) protein expressed for NMR spectroscopy was grown following the same protocol as above but in minimal media containing ^{15}N ammounium sulphate. The components of minimal media are detailed in Table 2.8.

^{15}N-M9 minimal media	Volume	Final Concentration
Water	800ml	-
$CaCl_2$ (1M)	300µl	0.3 mM
$MgSO_4$ (1M)	1ml	1 mM (each)
Thiamine (1 mg/ml)	1ml	1µg/ml
Biotin (1 mg/ml)	1ml	1µg/ml
Ampicillin (100mg/ml)	1ml	100 µg/ml
^{15}N Ammonium Sulphate	1g	1mg/ml
Glucose (20%)	20ml	0.4%
M9 salts (10X)	100ml	1X
Final Volume	**1L**	

Table 2.8 Minimal media used for expression of ^{15}N labelled Rna15 2-103.

2.3.2 Cell lysis and sample preparation

All protein purifications were carried out at 4ºC. In general cells were re-suspended in lysis buffer (100 mM Tris pH 8.3, 300 mM NaCl, 0.5 mM TCEP) using 10mL of buffer per gram of pellet. One complete Protease Inhibitor cocktail tablet (Roche) per 50 mL of lysis buffer was added to prevent protein degradation, according to manufacturer's instructions. In addition, 1µL of Benzonase nuclease (Novagen) and 1µL rLysozyme (Novagen) were added to reduce viscosity and aid lysis. Samples were sonicated on ice at 40% power, 10-12 times x 30 sec with 30 sec rest between cycles on a Branson Sonifier 450 sonicator. Lysate was clarified by centrifugation at 20000 xg for 1 hour at 4ºC in a Beckman JA-25.50 and the supernatant collected. Samples of whole cell extract and cleared lysate were analysed by SDS-PAGE.

2.3.3 Strep-Tactin® affinity chromatography

Proteins tagged with Strep-Tag® II (iba) was initially purified using Strep-Tactin® affinity chromatography. Cleared lysate was applied to a 5mL column of Strep-Tactin® high performance resin (iba) packed into a polyprep disposable column (BioRad). Once bound, the protein was washed with 5 X 5mL of lysis buffer and then

eluted in 5 X 5 mL aliquots of 10 mM desthiobiotin. Protein was detected within the fractions by spotting 10 µL onto chromatography paper (Whatman) and staining with Coomassie Blue (BioRad). A sample of the wash fraction and protein fraction were taken for analysis by SDS-PAGE. Protein was dialysed into 20 mM Tris-HCl pH 8.3, 150mM NaCl, 0.5mM TCEP at 4°C overnight. The streptavidin tag was cleaved overnight during dialysis using 3C protease (produced in house). Approximately 1 mg of 3C protease was used for every 10 mg of target protein.

2.3.4 Nickel-affinity chromatography

Nickel affinity chromatography was used as an initial purification step for his-tagged Rna14 during purification of both Rna14 and Rna15 expressed from a pET duet vector. Cells were lysed as above in lysis buffer suitable for nickel affinity chromatography (100 mM Tris-HCl pH 8.3, 300 mM Nacl, 0.5 mM TCEP, 10 mM imidazole). Cleared lysate was applied to a 5 mL of Ni-NTA Superflow resin (Qiagen) packed in a XK-16 column (GE Healthcare). Resin was then washed with lysis buffer to remove any non-specifcally bound proteins. The protein was then eluted with a 20-500 mM imidazole gradient over 50 mL. Fractions eluted from the column were analysed using SDS-PAGE and further purified using SDS-PAGE. Imidazole was removed from the sample during size exclusion chromatography.

2.3.5 Size exclusion chromatography

Size exclusion chromatography (SEC) was used as a final polishing step to improve sample homogeneity. Protein samples were concentrated to just under 5 mLs and loaded onto either a superdex-75 or superdex-200 (26/60 or 16/60) columns (GE Healthcare) depending on protein size and sample volume. Columns were equilibrated and sample eluted in 20 mM Tris-HCl pH 8.3, 300 mM NaCl, 0.5 mM TCEP (SEC

buffer) using an AKTA Prime (GE Healthcare). Fractions containing protein were analysed by SDS-PAGE, pooled and concentrated.

2.3.6 SDS-PAGE

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was used to analyse the purity of the protein sample after each step of purification. NuPAGE Bis-Tris precast polyacrylamide gels were used with 1X NuPAGE MES SDS running buffer (Invitrogen). Gels were run in an XCell Sure Lock mini-cell using a Power-Ease 500 power supply (Invitrogen). A 10 µL sample of protein was taken and mixed with 3X SDS loading buffer and 100mM dithiothreitol to a final concentration of 33 mM. Samples were heated at 95 °C for 5 minutes prior to loading. Protein sizes were estimated against SeeBlue® Plus2 molecular marker (Novagen). Gels were developed by staining with Coomassie Blue solution (50 % Methanol, 10 % acetic acid, 0.05 % Coomassie Brilliant Blue R-250) for 5 minutes and destained by washing with destaining solution (10 % Acetic Acid, 10 % Methanol).

2.3.7 Protein concentration, storage and dialysis

Proteins were concentrated by centrifugal ultrafiltration using Vivaspin concentrators (Vivaproducts) of the appropriate molecular weight cutoff to ~10 mg/mL. For buffer exchange during purification proteins were dialysed in 2 L of the appropriate buffer overnight at 4 °C. Samples were injected into a Slide-a-lyzer dialysis cassettes (ThermoScientific) of appropriate size and molecular weight cut off. The sample was then dialysed against 2 L of the appropriate buffer for ~2 hours, at which point the sample was placed in fresh dialysis buffer and left overnight at 4°C. Purified protein was flash frozen in liquid nitrogen and stored at -80°C in 200 µL aliquots.

2.3.8 Determination of protein concentration

Protein concentrations were determined by UV absorption spectroscopy using a spectrophotometer (Cary 50 Bio UV-Visible Spectrophotometer). The protein sample was diluted 1:10 and 200 µL was added to a cuvette. The absorbance was scanned between 220-400 nm and protein concentration (c) was calculated from Beer-Lambert's Law:

Equation 2.1 $A = \varepsilon.c.l$

Where A=absorbance, ε= molar extinction coefficient (M^{-1} cm^{-1}), c=concentration (M) and l = cell path length.

2.4 *In vitro* analysis of Rna15-RNA binding

2.4.1 Fluorescence spectroscopy of Rna15 constructs

Dissociation constants for the interaction of proteins with Tetrachlorofluorescein (TET)-labelled RNA were determined at 20°C in 20 mM Tris-HCl pH 8.3, 150 mM NaCl, 0.5 mM TCEP. Titrations were performed keeping RNA at a constant concentration with increasing protein concentration. Protein was introduced to the RNA in ~2µM steps and changes in fluorescence were monitored. Titrations were performed on a Jasco FP-6300 spectrofluorimeter with excitation at 515 nm and emission at 560 nm. Experiments were repeated three times and the derived equilibrium dissociation constants are an average of the experiments. Dissociation constants were derived from data by using a 1:1 interaction model as follows.

1-1 interaction model - The interaction of a protein (P) with a ligand (L) to form a simple 1-1 complex (PL) is represented by the following scheme:

$$P + L \underset{K_d}{\rightleftharpoons} PL$$

with the dissociation constant, K_d, defined as $K_d = \frac{[P][L]}{[PL]}$

For any mixture of P and L the observed fluorescence signal (F_{OBS}) is given by**:**

Eq 2.2 $F_{OBS} = F_P[P] + F_L[L] + F_{PL}[PL] = F_P[P_0] + F_L[L_0] + (F_{PL}-F_P-F_L)[PL]$

where $[X_0]$, $[X]$, and F_X represent the total concentration, the equilibrium concentration, and the molar fluorescence coefficient of species X (note $F_P=0$). A non-linear least-squares fit to equation **(2.2)** with [PL] calculated using equation **(2.3)** yields the K_d for the interaction and the F_X values.

Eq 2.3 $$[PL] = \frac{(K_d + [P_0] + [L_0]) - \sqrt{(K_d + [P_0] + [L_0])^2 - 4[P_0][L_0]}}{2}$$

2.4.2 Fluorescence spectroscopy of the Rna14-Rna15 complex with RNA

Titrations were performed on a Jasco FP-6300 spectrofluorimeter using an excitation wavelength of 515 nm and emission of 560nm. Experiments were performed at 20°C in 20 mM Tris-HCl pH 8.3, 150 mM and NaCl 0.5 mM TCEP. Titration experiments involved using 3 RNA oligonucleotides each possessing two GUGUGU sequences separated by an increasing number of nucleotides (see section 5). 355 µM of unlabelled RNA oligonucleotide spiked with 5 µM of a labelled version of the same oligonucleotide was titrated and the change in fluorescence was observed. Total protein concentration was kept constant at 30 µM Rna14-Rna15 protein by including the proteins in the oligonucleotide stock. Around 2 µL additions of the stock were made and the change in fluorescence recorded. A binding isotherm was generated by calculating the difference between the titration and a control titration with no protein. Each experiment was repeated three times and the derived equilibrium dissociation was determined using a 1:1 binding model (section 2.4.1).

2.4.3 Competition assays

In order to determine whether the tetrachlorofluoroscein (TET) label present on the 5' end of the RNA oligonucleotide affected the measured dissociation constant of the Rna15 protein for the RNA, a series of competition assays were performed. Competition assays were performed on a Jasco FP-6300 spectrofluroimeter at 20°C in 20 mM Tris-HCl pH 8.3, 150 mM NaCl, 0.5 mM TCEP. Titrations followed the protocol detailed in section 2.4.1.1. However, in this case an excess of unlabelled RNA was included in the cuvette. Data were fitted following the scheme detailed below.

When two ligands (L and N) compete for binding to a single site on a protein (P) the following scheme is appropriate:

$$P + L \underset{K_{d,\,L}}{\rightleftharpoons} PL$$

$$P + N \underset{K_{d,\,N}}{\rightleftharpoons} PN$$

The dissociation constants for the two ligands are defined as:

$$K_{d,L} = \frac{[P][L]}{[PL]} \qquad K_{d,N} = \frac{[P][N]}{[PN]}$$

For any mixture of P, L and N the observed fluorescence signal (F_{OBS}) is given by**:**

Eq 2.4 $$F_{OBS} = F_P[P] + F_L[L] + F_N[N] + F_{PL}[PL] + F_{PN}[PN]$$

A non-linear least squares fit to equation **(2.4)** yields the $K_{d,N}$ value (with $K_{d,L}$ fixed) for the interaction and the F_X values. Since the protein and the displacing ligand (N) are non-fluorescent then $F_N = F_{PN} = F_P = 0$. For this system, the concentrations are calculated in the following way. The free protein concentration is given by the root of the following equation:

$$C_3[P]^3 + C_2[P]^2 + C_1[P] + C_0 = 0$$

where

$$C_3 = 1$$
$$C_2 = -[P_0] + K_{d,L} + K_{d,N} + [L_0] + [N_0]$$
$$C_1 = -[P_0]K_{d,L} - [P_0]K_{d,N} + K_{d,L}K_{d,N} + [L_0]K_{d,N} + [N_0]K_{d,L}$$
$$C_0 = -[P_0]K_{d,L}K_{d,N}$$

[PL] and [PN] can then be calculated from equations **(2.5)** and **(2.6)** and the remaining concentrations from the expressions for the K_ds or from conservation of mass (e.g., [L] = [L_0] - [PL]).

Eq 2.5
$$[PL] = \frac{[L_0][P]}{K_{d,L} + [P]}$$

Eq 2.6
$$[PN] = \frac{[N_0][P]}{K_{d,N} + [P]}$$

2.4.4 Scaffold independent analysis (SIA)

Scaffold Independent Analysis (SIA) was performed in collaboration with Katherine Collins (Ramos Group, N.I.M.R) following the method described in Beuth *et al. 2007*. A pool of 16 RNA oligonucleotides was employed to determine a potential consensus sequence for optimal Rna15 binding. Concentration of Rna15 protein was maintained at 25 μM while increasing amount of each RNA pool was added at the following ratios 1:1, 1:3. Titrations were performed in 20 mM Tris-HCl pH 7.0, 40 mM NaCl, 0.5 mM TCEP at 25°C on a Bruker Avance spectrometer equipped with cryoprobe and operating at 700 MHz ^{1}H frequency. The weighted average chemical shift perturbations at each titration point for 16-21 non-overlapped peaks were calculated using the equation **(2.7)**.

Eq 2.7
$$\Delta\delta_{av} = \sqrt{(\Delta\delta_H)^2 + (\Delta\delta_N/10)^2}$$

The results were visualised using sparky assignment and integration software http://www.cgl.ucsf.edu/home/sparky/.

2.4.5 Nuclear Magnetic Resonance (NMR)

Prior to NMR experiments Rna15 (2-103) was dialysed in 20 mM Tris-HCl pH 7.0, 50 mM NaCl and 0.5 mM TCEP. Titrations were performed on a Bruker 600 NMR MHz NMR spectrophotometer. 25 µM Rna15 (2-103) was titrated against 5' labelled and unlabelled RNA oligonucleotide, GUGUGU (1:1). A free 25 µM Rna15 (2-103) protein sample was also included. Movement and position of peaks were calculated using equation (**2.7)** detailed in section 2.4.4 and visualised using integration software http://www.cgl.ucsf.edu/home/sparky/. Previously assigned NMR spectra were employed to assign peaks to corresponding Rna15 amino acids.

2.4.6 Circular Dichroism

For CD experiments, protein samples were dialysed into 20 mM Tris-HCl pH 8.3 150 mM NaCl 0.5 mM TCEP overnight at 4°C. The sample was concentrated to 0.16 mg/mL and far-UV spectra was recorded on a Jasco-J815 spectropolarimeter using a quartz cuvette with a 1mm pathlength for Rna15 (2-103) and all Rna15 mutants. Spectra were taken with a 200nm/min scan rate, 0.25 s time constant and 1nm spectral bandwidth. CD was measured from 260nm-200nm. Thermal unfolding experiments were carried out on Rna15 (2-103) and Rna15 (2-103) Y27A by recording CD at 204 nm as the temperature was increased from 20°C to 80°C. During thermal experiments, CD was measured with a 8 s time constant and a 1nm spectral bandwidth. Final spectra were the average of 20 scans. Data analysis was carried out with Dr S. Martin (N.I.M.R) using in house software. The CD signal was measured in millidegrees (mdeg) and converted to the mean residue weight CD extinction coefficient ($\Delta\varepsilon_{mrw}$) using equation 2.8.

Eq 2.8 $$\Delta\varepsilon mrw = \frac{S\cdot(mrw)}{(3290)\cdot C\cdot l}$$

Where S is the CD signal in millidegrees, *mrw* is the protein mean residue weight (a value close to 108 Da), C is the protein concentration in mg/mL and *l* is the pathlength of the cuvette. Fitting of protein secondary structure content from far-UV CD spectra was carried out using the CDpro program (with Dr S. Martin, N.I.M.R): the CONTIN, SELCON3 and CDSSTR methods of secondary structure estimation from CD data were used as recommended by Sreerama and Woody (297). The final estimated secondary structure content was calculated by averaging the results provided by each method.

2.4.7 Fluorescence anisotropy experiments

Fluorescence anisotropy was carried out on a Jasco FP-6300 spectrofluorimeter. The excitation wavelength used was 515 nm and the emission was 560 nm. Vertical and horizontal polarisation was measured on 0.2 µM labelled RNA (5'TET-GUGUGU) alone before adding a saturating concentration of Rna15 2-103 protein (20 µM). The polarisation values were correlated to movement of the protein in solution using the physical theory of Brownian motion by Perrin shown in chapter 3.

2.5 Yeast methods

2.5.1 Yeast transformation

The desired *S. cerevisiae* strain to transform was streaked onto YPD agar and incubated at 30°C for 2 days. Following incubation, cells were inoculated into 20 mL of YPD media and incubated overnight at 30°C with shaking (180 RPM). The following morning, 1 mL of the overnight culture was inoculated into 50 mLs of YPD media and left to incubate at 30°C for 5 hours. Cells were then harvested by centrifugation at 3,000 *xg* for 2 minutes and washed with 30 mL of dH_2O and 1 mL of 100 mM lithium acetate, before being resuspended in 1 mL of 100 mM lithium acetate at a ratio of 5:1 (lithium acetate:cells). A transformation mix containing treated cells, the plasmid to be

transformed and a number of transformation reagents was prepared (Table 2.9). The transformation reaction was mixed thoroughly by vortex and incubated at 30°C for 30 minutes and then 42°C for 30 minutes. Following heat-shock, cells were placed on ice for 5 minutes and then harvested by centrifugation at 6200 *xg* for 1 minute. Cells were resusupended in 600 µL of dH_2O and 200 µl was plated onto the appropriate selective medium (in this study minimal meda minus leucine for LEU2 plasmid constructs). Plates were incubated for 2-3 days at 30°C.

Component	Volume (µL)
Competent cells	50
Polyethylene glycol 3350K (PEG)	240
Lithium Acetate (1M)	36
Salmon sperm DNA (1µg/µl)	50
Plasmid DNA to be transformed (100ng/µl)	2
Final volume	**378**

Table 2.9 Typical yeast transformation mix.

2.5.2 5-FOA plasmid shuffle

The 5-FOA plasmid shuffle technique was used to visualise the effects of mutations in target genes, RNA14 and RNA15. Initially, the target, URA and LEU genes are deleted from the haploid genome of *S. cerevisiae*. The target genes are essential and so are introduced on a URA3 plasmid to maintain cell viability. The mutated copy of the target gene can then be transformed on a LEU2 plasmid in order to visualise the effect of the mutation on phenotype. Selection for cells which have been successfully transformed with the mutated copy of the target gene is achieved through use of 5-fluoroorotic acid (5-FOA). This method works on the basis that cells which are URA^+ are sensitive to 5-FOA and are not viable allowing effect of the mutation to be observed.

Sensitivity to 5-FOA is due to the presence of the wild type copy of the target gene on a plasmid which contains the URA gene. This encodes the orotidine 5'-phosphate decarboxylase enzyme (OMP-decarboxylase) which converts 5-FOA to 5-flurouracil. This compound is toxic and results in cell death by inhibition of thymidylate synthase which is involved in the production of thymidine for DNA synthesis and repair. Therefore, cells where the URA plasmid is still present are no longer viable. Populations of cells where the URA plasmid has been lost due to asymmetric plasmid segregation during mitosis maybe viable if they possess the LEU2 plasmid containing the RNA15 gene only if the mutation itself does not result in loss of viability. Both the *rna14Δ* and *rna15Δ* haploid strains supported by wild type RNA14 and RNA15 on the pRS316-URA construct were supplied by Dr. J. Hedden (N.I.M.R). These strains were both transformed with LEU2 plasmids containing the desired RNA14 or RNA15 construct as described in section 2.5.1. An individual colony was taken from the transformant plate and streaked onto minimal media minus leucine and incubated at 30°C for 2-3 days. The 5-FOA assay was completed as follows; cells were taken from the minimal media and diluted into 1mL of dH_2O. A further dilution of 30 µL into 1mL of dH_2O was made to obtain the first point in the dilution series. Three serial dilutions of 1:10 ratio (cells:dH2O) were made from the first point in the series and 5 µL of each dilution was plated. Cells were plated onto minimal media containing 5-FOA (1mg/mL) and YPD. Plates were allowed to dry and incubated at 22°C for 5 days. To look for temperature sensitivity, a duplicate set of plates were incubated at 37°C (or other specified temperatures) for 2-3 days.

2.5.3 Measuring *S. cerevisiae* cell density

For growth curves, RNA purification etc, the amount of yeast cells contained within a culture had to be measured accurately. In order to do this a haemocytometer

was used. Counting cells in this way is only practical if the density of the culture is between $1x10^5$ and $5x10^6$ cells/mL. Therefore, overnight cultures grown to saturation (~$1x10^8$) were diluted 100-fold. The heamocytomoter is made up of a number of grids measuring 1 mm^2. Each grid is made up of 16 squares identical in size. Following addition of 100µl of cell culture, a coverslip was used to cover the cells. Since the distance between the coverslip and the grid is fixed at 100 µm, the volume measured is 0.1 mm^3. Therefore, subsequent counting of cells within the 16 square of the grid can be multiplied by 10^4 to convert the amount to cells/ml.

2.5.4 Growth rate assays

Following transformation and 5-FOA selection of wild type and mutant Rna15 constructs, a small clump of cells was streaked onto YPD agar and incubated for 2 days at 30ºC. A 20 mL culture of YPD was then inoculated and incubated at 30ºC, shaking (180 RPM). The following morning, the cell density of the overnight culture was measured by counting cells on a haemocytometer grid to determine the amount of cells within the culture (cells/mL). A 50 mL culture of fresh YPD was prepared and cells added from the overnight culture to achieve a cell density of 1 X 10^6 cells/mL. Cultures were incubated at 30ºC with shaking (180 RPM) for 10 hours and cell density was determined once every hour.

2.5.5 Viable cell count

Following 5-FOA selection, Rna15 WT, Rna15 Y27A, R87A, K90E and Rna15 Δ16-94 were streaked onto YPD-agar and incubated at 30°C for 2 days. A small amount of cells were picked and inoculated into 20mL of YPD broth + 2% glucose. Cultures were incubated overnight at 30°C with shaking (180 RPM). The following morning, cells were counted. Cells were diluted into 50 mLs of YPD to give a starting density of

$1x10^6$cells/ml and incubated at 30°C with shaking (180 RPM) until each culture reached a density of $1x10^7$cells/ml. At this point, cultures were diluted to give a 3 point 100 fold dilution series of each mutant; $1x10^6$, $1x10^4$ and $1x10^2$. Equal amounts from each point were plated onto minimal agar -Leucine in triplicate. Plates were incubated at 30°C for 2-3 days.

2.5.6 Western blotting

Strains transformed with HA tagged wild type and mutant Rna14 and Rna15 proteins were employed in western blotting. A HA tag was added onto the C-terminal end of the protein to allow detection of the protein within the yeast cell lysate. Western blotting analysis was employed as a control experiment to distinguish whether mutation of either Rna14 and Rna15 resulted in unstable protein within the cell. A LEU construct containing the desired gene was transformed into the appropriate haploid strain. A single colony was picked following transformation and streaked onto -LEU minimal media and incubated at 30ºC for 2-3 days. Cells were inoculated into 20 mL of minimal media and incubated at 30ºC with shaking (180 RPM) overnight. The following morning 2 mL of overnight culture was diluted into 50 mL of minimal media and incubated at 30ºC with shaking (180 RPM) for 5 hours. Cells were then harvested by centrifugation at 3000 *xg* and washed with 1 mL PBS. Cells were then spun down to form a pellet and the supernatant disposed. Cells were then snap frozen in liquid nitrogen and held at -80ºC.

The following morning, cells were resuspended in 500 µL yeast lysis buffer (50 mM Tris-HCl pH 7.4, 250 mM NaCl, 0.5 mM EDTA). Nonidet P-40 was added to a final concentration of 0.2% v/v in order to support lysis and 5µL of 100X Halt Protease Inhibitor Cocktail (Thermo scientific) was added to inhibit protein degradation. To

achieve lysis, chilled 500µm diameter acid washed glass beads (Sigma) were added to each sample. Yeast cells were lyzed using a FastPrep-24 ribolyzer (MP Biomedicals) where cells were subjected to 4-5 10 second bursts with 2 minutes rests on ice in between. Whole cell extract was collected by centrifugation twice at 20, 000 *xg* for 15 minutes at 4ºC. The protein concentration of the cleared lysate was determined using Bio-Rad protein assay following the manufacturers protocol (Bio-Rad), using an Eppendorf BioPhotometer (Eppendorf).

For Western blotting a sample of 2mg/mL lysate, 100 mM DTT and 1X SDS was prepared. Samples were then heated 95ºC for 5 minutes and subjected to SDS-PAGE. After SDS-PAGE, the gel was incubated in Western Transfer buffer (25mM Tris pH 8.3, 194mM glycine, 10% methanol) for around 15 minutes. Proteins were transferred from the gel to an Immobilon-P Transfer membrane (Millipore) in a BioRad Trans-Blot Cell (100 V for 1 hour). Following protein transfer, the membrane was blocked overnight with 10mL 1X Western Blocking Reagent (Roche) at 4ºC with gentle agitation.

The following morning, the membrane was probed with 1:1000 dilution of mouse Anti-HA antibody (clone12CA5, Roche) for either 1 or 5 hours at room temperature. The primary antibody was diluted in 0.5X Western Blocking Reagent (Roche). After 1 hour, the membrane was washed 4 times for 10 minutes with 20 mLs of Tris buffered saline plus 0.1% Tween-20 (TBST) to remove any non-specifically bound antibody. The secondary antibody (peroxidase-conjugated goat anti-mouse antibody, Thermo Scientific) was diluted to a final conentration of 1:5000 in 0.5X Western Blocking Reagent. The membrane was incubated with the secondary antibody for 1 hour at room temperature. The membrane was washed as described previously.

SuperSignal West chemiluminescent substrate was applied in accordance with the manufacturer's protocol (Thermo Scientific) and protein bands were visualised using X-Ray film (Amersham Hyperfilm MP, GE Healthcare). Film was exposed for 2 minutes 30 seconds and the films were developed using a Fujifilm FPM-3800 AD film processor.

2.5.7 Preparation of total RNA

Deletion strains which were successfully transformed with the desired gene were streaked onto YPD agar following the 5-FOA assay. Plates were incubated for 2 days at 30ºC. Cells were inoculated into 20 mL YPD and incubated overnight at 30ºC with shaking (180 RPM). The following morning cell density of the overnight culture was determined using a haemocytometer. Fresh 50 mL cultures were prepared with the appropriate amount of overnight culture to give a cell density of 2 X 10^6 cells/mL. The cultures were grown for 8 hours at 30ºC with shaking (180 RPM). Following incubation, cells were harvested by centrifugation at 3000*xg* and the cell pellet flash frozen in liquid nitrogen and stored at -80ºC overnight. The following day, total RNA was prepared from cells using the RiboPure Yeast Kit (Ambion) in accordance with the manufacturer's protocol. Purified RNA was aliquoted into 10 µL aliquots, flash frozen in liquid nitrogen and stored at -80ºC. The concentration of the RNA samples was measured using a NanoDrop spectrophotometer (ThermoScientific) and the purity of the sample was analysed using the A_{260}:A_{280} ratio (a high purity sample has a A_{260}:A_{280} ratio of 1.8 to 2.1).

2.5.8 Analysis of RNA quality

Purified RNA was analysed for degradation using an Agilent 2100 bioanalyser 384 (Agilent Technologies). This method is an electrophoretic assay based on the

principles of traditional gel electrophoresis but in chip format. Each chip contains 12 wells and an additional well for a molecular size marker. Preparation of the chip involves filling the wells with a sieving polymer and a fluorescent dye. The sieving polymer separates the sample in question by size while the fluorescent dye integrates with the RNA sample so that a fluorescent signal is emitted upon excitation by a laser. Once the wells and channels of the chip are filled and the RNA sample in question has been added, the chip becomes an integrated electrical circuit where pin like electrodes fit into the wells present on the chip. The eukaryote total RNA Nano samples 2100 expert software (Agilent Technologies) plots fluorescence intensity versus molecular size (nt) and produces an electrogram for each sample (appendices). For RNA analysis, intact RNA is demonstrated by two distinct peaks and bands that are representative of 28S and 18S ribosomal RNA. This was demonstrated in wild type and Rna15 mutant RNA samples purified from *S. cerevisiae* (Appendix 7.4).

2.5.9 Quantitative real time PCR

Quantitative real time PCR (qPCR) is a technique that is used to quantify amount of DNA target within a sample. qPCR experiments employing reverse transcribed RNA (qRT-PCR) provide a direct measurement of the amount of a target RNA within a sample allowing analysis of DNA expression. The amount of a target is quantified by monitoring the amplification of the sequence using a fluorescence reporter system. The experiments presented in this study were completed using the 5' nuclease (TaqMan) assay that employs hydrolysis probe technology and utilises the 5'-3' exonuclease activity of *Thermus aquaticus* DNA polymerase. During the annealing step, a dual labelled hydrolysis probe hybridises to its complementary sequence within the target. The probe maintains a fluorescent marker present on the 5' end which is near a non-fluorescent quencher present on the 3' end of the probe. Prior to amplification,

excitation which leads to emission of fluorescence by the fluorescent marker is quenched due to the close proximity of the 3' end quencher. Amplification by the DNA polymerase in the 5'->3' direction from the forward primer results in cleavage of the fluorescent marker from the probe. This frees the probe resulting in emission of fluorescence and allowing detection of the fluorescent signal illustrated in Figure 2.1.

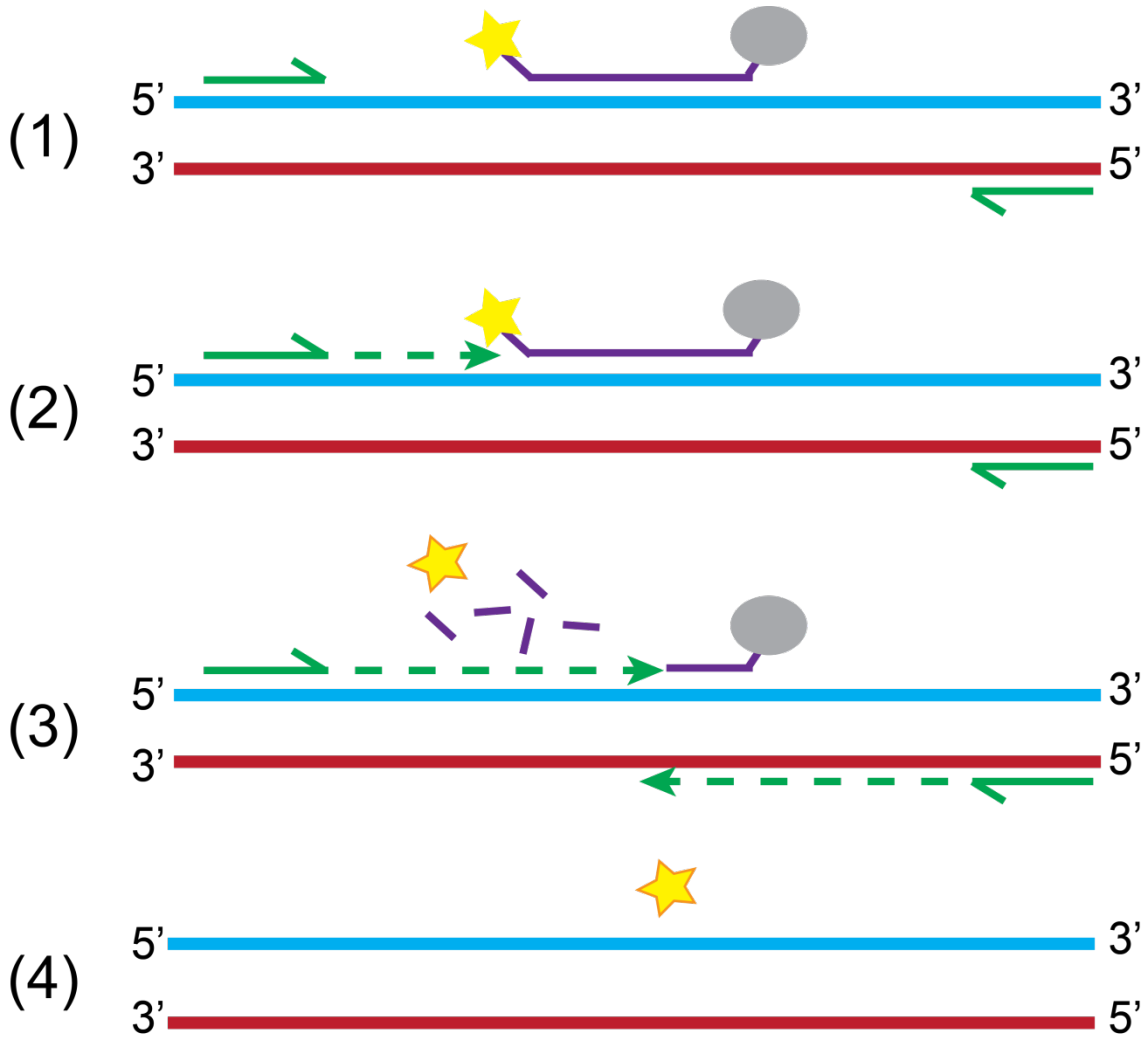


Figure 2.1 Principle of qRT-PCR. **1.** The forward and reverse primers (green) bind to the sense strand (blue) and antisense strand (red) respectively. The TaqMan probe (purple) also hybridises to its target sequence. **2.** The DNA polymerase amplifies both strands. **3.** Amplification of the sense strand results in degradation of the probe due to the 5’->3’ exonuclease activity of the polymerase. This releases the 5’ fluorophore (yellow star) from the quencher (grey sphere) resulting in a fluorescence emission upon excitation. **4.** The thermal cycler can then quantify the fluorescent signal resulting in a direct correlation with the amount of cDNA.

The most useful measurement for quantification of a target is determination of the threshold cycle (C_T). This is the point at which a significant increase in fluorescence signal is detected (in the exponential phase of PCR). There are different methods that can be used to derive absolute or relative quantities of target from C_T values. These are detailed in Applied Biosystems User Bulletin #2) (298). In the experiment presented in this study, data were analysed to generate relative amounts by using the standard curve method. This involves the inclusion of a RNA or DNA dilution series of known quantity to generate a calibration curve allowing a correlation between fluorescence signal and mRNA quantity. Following standard curve analysis, relative quantification was determined by firstly normalising to a data set unbiased by the presence of the mutation (SNR52) and secondly dividing against quantity of a calibrator to generate a final target quantity expressed as an n-fold difference to the calibrator (ie mutant and wild type).

qRT-PCR experiments were carried out on an Applied Biosystems 7500 Fast Real Time PCR system. Primers and probes employed were designed using the Primer Express 2.0 software (Applied Biosystems) and purchased from Applied Biosystems. Primer and probe sequences can be found in appendices. All target amplicons were detected by the 5' nuclease (TaqMan) assay using 5'-6-carboxyfluorescein (6-FAM) 3' minor groove binder (MGB) labelled probes and TaqMan One Step RT PCR Master Mix reagents kit (Applied Biosystems). Reactions were set up on ice in 0.1 mL MicroAmp Fast Optical 96-Well Reaction Plates (Applied Biosystems). The reaction setup used is described in Table 2.10 and the cycling parameters are presented in Table 2.11. Each target sequence was detected in triplicate and each plate was repeated three times (i.e. quantity of each target was measured 9 times). An eight point, 2 fold dilution series of *S. cerevisiae* genomic DNA (Stratagene) ranging from 250ng-1.95ng was used to generate a standard curve required for quantification of cDNA amount in both targets

and controls. The reaction threshold for each target was generated automatically by the 7500v2.0.5 analysis software from Applied Biosystems.

Reagent	Volume (µL)
Master Mix (2X)	12.5
Reverse transcriptase (40X)	0.625
Forward primer (10µM)	2.25
Reverse primer (10µM)	2.25
TaqMan probe (10µM)	0.625
RNA template 20ng/µl *or* Genomic DNA (serial dilution)	5
dH_2O (RNase free)	1.75
Total	**25**

Table 2.10 Typical reaction setup for a qRT-PCR experiment.

Step	Temperature	Time
1	48ºC (Reverse transcription)	15 minutes
2	95ºC (DNA polymerase activation)	10 minutes
3	95ºC (melting)	15 seconds
4	60ºC (annealing/extension)	1 minute
5	Repeat 3-4, 40 cycles	

Table 2.11 Thermal cycling parameters used in a typical qRT-PCR reaction. These conditions are identical to those specified by the Applied Biosystems 7500 Fast Real-Time PCR system software v2.0.5 for a standard one step amplification from an RNA template.

Changes in amount of mRNA and therefore changes in expression were measured from five genes (TDH2, YPT1, ACT1, CYC1 and ADH1) in both wild type and Rna15 RRM mutant *S. cerevisiae* strains. Primers and probes were designed to hybridise and amplify within the open-reading frame or within the 3' UTR prior to the poly(A) site. Data were analysed using the standard curve method and quantities of mutant cDNA were expressed relative to wild type cDNA quantity.

2.5.10 RNA sequencing – library preparation and sequencing

RNA sequencing experiments are discussed in detail in chapter 4. RNA was purified from Rna15 wild type and Y27A, Y27A/R87A/K90E, Δ16-94 mutants. Prior to library preparation, RNA samples were analysed using the bioanalyser following the manufacturers' instructions. Ribosomal RNA (rRNA) was removed from RNA using the RiboZero™ Magnetic Gold kit (Illumina) following the manufacturer's protocol. The RNA sequencing library was generated using the ScriptSeq™ v2 RNA-Seq Library Preparation kit (Illumina) following the manufacturer's protocol. Sequencing was carried out on a Illumina GAIIx as a single read 40 cycles run. The resulting fast-Q files were generated using Illumina CASAVA 1.8.2.

3.1 Introduction and Overview

Rna15 binds RNA through an RNA recognition motif engaging the CF1A complex onto the 3' UTR to direct polyadenylation. The aim of this chapter is to characterise the mechanism of Rna15-RNA binding and identify the residues responsible. Previous NMR experiments revealed an α-helical structure formed by residues 96-101 of Rna15 packs against the RRM to occlude RNA binding sites II and III. Upon RNA binding this α-helical structure is displaced, a feature that is also observed in human Rna15 homologue, CstF-64 (140,141). This observation demonstrates a possible mechanism for regulation of Rna15-RNA binding.

The crystal structure of the free Rna15 RRM forms a four stranded β-sheet backed by two α-helices (141). The crystal structure bound to RNA revealed two sites were able to direct RNA binding. The first, a binding pocket formed by the side groups of Y27 and R87 and a second, formed by the side group of Y21 (141). Fluorescence spectroscopy experiments revealed that mutation of these residues results in a loss in RNA binding affinity. Additional NMR spectroscopy demonstrated that residues over the face of the RRM were also involved in binding RNA including those observed in the crystal structure (141). However, the mechanism by which these residues bind RNA and also the extent to which they are involved in RNA binding remains unknown.

This chapter aims to explore in further detail Rna15-RNA binding properties discussed above. The role of the C-terminal tail in RNA association and the effect of residues present on the surface of the RRM on RNA binding affinity were investigated using fluorescence spectroscopy.

3.2 Analysis of Rna15, Rna15 (2-94) and Rna15 (2-103) RNA binding *in vitro*

3.2.1 Cloning Rna15 *in vitro* constructs

A full length Rna15 construct and two shorter constructs of Rna15 comprising residues 2-94 and residues 2-103 were prepared. The Rna15 (2-94) construct consisted of the disordered N-terminal region and the RRM whilst the 2-103 construct additionally included the C-terminal tail. Rna15, Rna15 (2-94) and Rna15 (2-103) were amplified from *S. cerevisiae* genomic DNA (invitrogen) by PCR employing ligation independent cloning (LIC) adapted primers (see section 2.2.12). Constructs were cloned into vector pET52b so as to include an N-terminal Streptavidin tag, cleavable via a 3C protease site immediately downstream. Constructs were transformed into *E. coli* strain Novablue (Novagen) and clones verified by DNA sequencing.

3.2.2 Expression and purification of Rna15 *in vitro* constructs

Rna15, Rna15 (2-94) and Rna15 (2-103) were expressed as described in section 2.3.1. Initial expression tests determined that optimum expression in all cases was achieved by induction at 25ºC overnight. Initial purification of Rna15 employed streptactin affinity chromatography. The protein successfully bound to the streptactin column through interaction with the N-terminal strep II tag. Due to high expression not all protein was captured and some is detected in the flowthrough (Figure 3.1 A, FT). A sample of the eluted protein was taken for analysis (Figure 3.1 A, E). The remainder of the protein sample was dialysed overnight in dialysis buffer with 3C protease to cleave the strep II tag. The cleaved protein was then applied to a size exclusion chromatography column for final purification (Figure 3.1 A, Cl). Following elution from the column, peak fractions were examined by SDS-PAGE. Peak fractions were

pooled and concentrated using Vivaspin concentrators with a molecular weight cut-off of 20K (Figure 3.1 A, 1-4).

Rna15 (2-94) and (2-103) were purified using the same method. Rna15 (2-94) was purified by exploiting the strep II tag present on the N-terminal end of the protein. However, high level of overexpression meant that not all of the protein was captured and some was detected in the flowthrough (Figure 3.1 B, FT). Rna15 (2-94) protein was eluted from the column and dialysed overnight with 3C protease in order to cleave the strep II tag. Following 3C cleavage, the protein was applied to a size exclusion column for final purification (Figure 3.1 B, Cl). The band is shifted slightly when compared to pre-cleaved Rna15 (2-94) protein (Figure 3.1 B, E), demonstrating the strep II tag has been successfully cleaved. Peak fractions eluted from the size exclusion column were pooled and concentrated using a Vivaspin concentrator with a 3 kDa molecular weight cut off (Figure 3.1 B, 1-4).

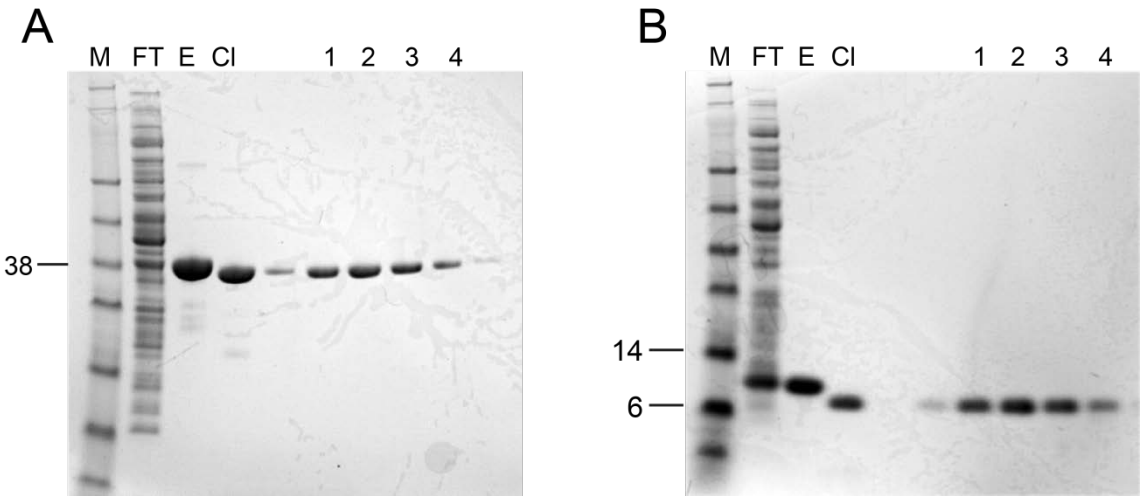


Figure 3.1 Purification of Rna15 and Rna15 (2-94) protein after overexpression in *E. coli*. Each stage of purification was analysed. **A.** SDS-PAGE analysis of Rna15 purification: M, Marker (SeeBlue® Plus 2 by life technologies); FT, Flow-through from Streptactin affinity column; E, Elution faction; Cl, cleaved protein. Peak fractions pooled from size exclusion chromatography are shown in lanes 1-4. **B.** SDS-PAGE analysis of Rna15 (2-94).

3.2.3 Principles of fluorescence spectroscopy

Fluorescence spectroscopy was selected to analyse RNA binding affinity as it is a sensitive technique and therefore does not require a large amount of working material. Fluorescence spectroscopy requires fluorescent markers called fluorophores that become excited when a photon of the appropriate energy is absorbed. Absorption results in an electronic transition from the ground state to a higher energy excited state. This excess energy is then released so that the electrons of the fluorophore return to the lowest vibratrional level of the first excited state. This process is illustrated in the Jablonski diagram (Figure 3.2). This release of energy coincides with the release of a photon in the form of fluorescence. The fluorophore is extremely sensitive to changes in its environment. Experimental parameters such as temperature and solvent polarity can affect the emitted fluorescence signal as they can affect the way in which energy is lost upon excitation. Therefore, fluorescence spectra need to be collected using accurate temperature control and appropriate buffer conditions. This sensitivity of fluorophores to change in their environment is exploited in binding experiments and can be used to determine binding affinities. In the experiments presented in this study, a 6 nucleotide RNA, 5' UGUUGU 3' (ThermoScientific) labelled at the 5' terminus with the fluorophore tetrachlorofluoroscein (TET) was employed. Addition of Rna15 protein results in formation of a TET-RNA-protein complex altering the environment of the fluorophore and affecting the emission spectra recorded following excitation. In the case of fluorescence titrations presented in this study, a red shift was observed upon addition of Rna15 protein. Data were fit to a 1:1 binding model where one oligonucleotide binds to one Rna15 monomer. All titrations were repeated three times and the average and standard deviation of the mean calculated.

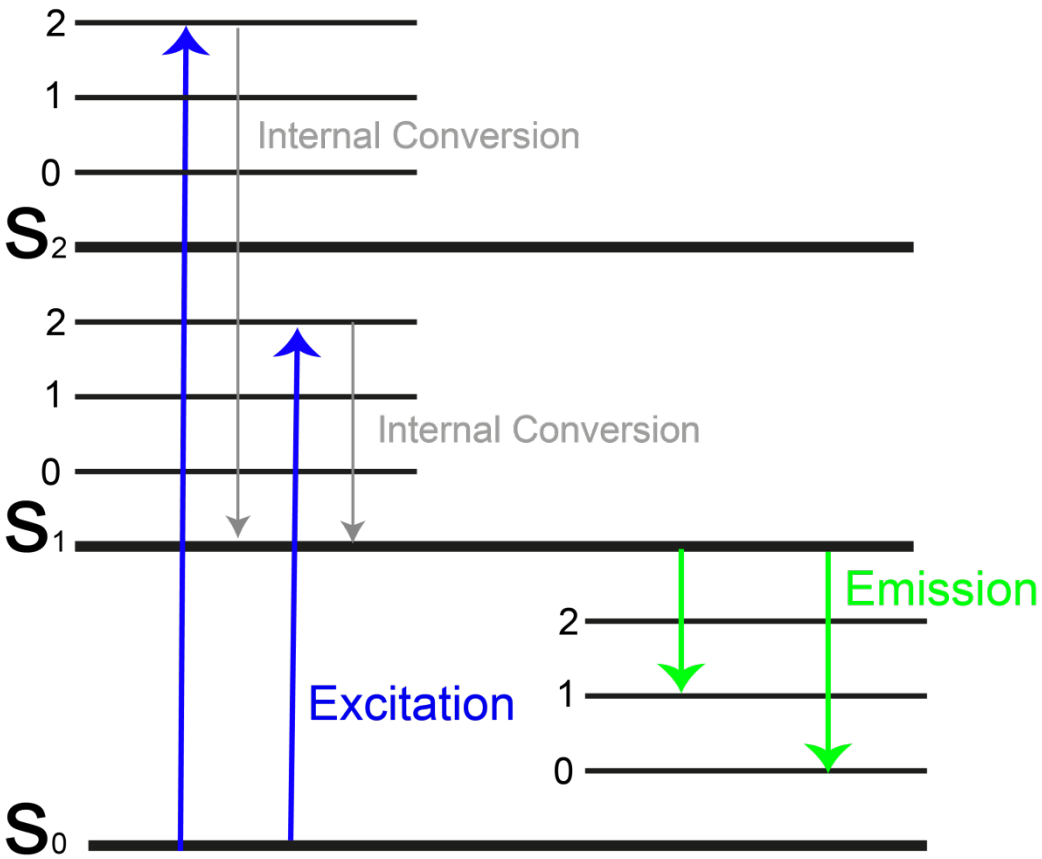


Figure 3.2 Jablonski diagram. Absorption of a photon due to excitation results in an electronic transition from the ground energy level (S_0) to an excited energy state (S_1, S_2). Excess energy is lost via internal conversion until the lowest vibrational state of the excited state is reached (S_1). A photon is them emitted as fluorescence to return the specices to S_0. The species may return to one of multiple vibrational states of S_0.

3.2.4 Fluorescence Spectroscopy of Rna15, Rna15 (2-94) and Rna15 (2-103)

The 5' UGUUGU 3' sequence was chosen as previously published data showed that a Rna15 construct composed of the RRM, tail and part of the linker (residues 16-111) bound this RNA oligonucleotide with a K_d 4.8μM (139). Fluorescence spectroscopy was used to characterise the role of the C-terminal region in RNA binding by determining their effect on RNA binding affinity.

Analysis of fluorescence titrations revealed that association with the oligonucleotide was similar for all constructs. The full length Rna15 construct and the Rna15 (2-103) construct both bound with a K_d of 6.4μM ± 2.1μM and ± 0.5μM respectively (Table 3.1). The similar binding affinity of both these constructs demonstrates RNA binding is mediated wholly by the RRM present between 16-94 but does not rule out the possibility that the N-terminal region and C-terminal tail (residues 2-16 and 94-103 respectively) may affect RNA binding indirectly either by regulating binding or stabilising the RNA-RRM interaction. However, these experiments demonstrate residues present C-terminal to the tail of Rna15 may not exert any influence on RNA binding. In the construct where the tail domain is excluded, construct Rna15 (2-94), the K_d is 7.6μM ± 2.6μM (Table 3.1) which is very similar to the observed K_d when the tail region is present (Rna15 2-103 K_d of 6.4μM ± 0.5μM). This result suggests that the tail does not exert any major regulatory role in binding affinity that results in an effect on Rna15 binding affinity.

Following characterisation of RNA binding, Rna15 (2-103) was chosen as a template for a mutagenic analysis of specific residues within the RRM for RNA binding on the basis that the fluorescence titrations performed yielded the smallest error in RNA binding.

Rna15 Construct	K_d ($x10^{-6}$ M^{-1})
Rna15 2-296	6.4 ± 2.1
Rna15 2-94	7.6 ± 2.6
Rna15 2-103	6.4 ± 0.5

Table 3.1 Fluorescence spectroscopy analysis of Rna15 deletion constructs. Dissociation constants of Rna15 deletion constructs when titrated with RNA 5'-UGUUGU-3'. Measured values are the average of three independent titrations and their standard deviation of the mean calculated.

3.3 Rna15-RNA interactions - Investigation of residues that direct RNA binding

The crystal structure of the RRM of Rna15 bound to a GU dinucleotide revealed key residues involved in RNA binding, Y27, R87, S24 and Y21. These residues have been identified in several studies as critical for RNA binding (139). However, additional residues present on the surface of the RRM were implicated in RNA binding by NMR spectroscopy experiments employing the 5'UGUUGU3' oligonucleotide (141). The ^{15}N-^{1}H HSQC spectrum of Rna15 16-111 was recorded at increasing stoichiometric concentrations of the ribo-oligonucleotide. Chemical shift peaks corresponding to Y61 and F63 were affected by introduction of RNA and have been implicated in RNA binding in other studies (139,141). These neighbouring aromatic residues present a potential third site for binding whereby a nucleotide could stack against the aromatic ring of both Y61 and F63 as observed at site I and site II. The free structure of the RRM also highlighted the presence of lysine residues present on the surface that could potentially interact with the RNA oligonucleotide through charge interactions mediated by the side chain. NMR spectroscopy identified chemical shift perturbations upon introduction of RNA in K59 which is present near the potential third site (141). Additional NMR studies revealed that K90 makes contacts with the phosphate group of a bound nucleotide and K48 was present in close proximity to the RNA (142). As they

are accessible, charged residues present on the surface of the RRM and therefore, potentially important for RNA binding they were included in mutagenesis studies.

3.3.1 Cloning, expression and purification of Rna15 (2-103) point mutations

Using the Rna15 (2-103) construct as a template, point mutations were introduced by site directed mutagenesis at the residues detailed above. The resultant mutant Rna15 (2-103) constructs were then expressed and purified as for the wild type protein. A typical purification is demonstrated in Figure 3.3. Following lysis, a sample of cleared lysate was analysed by SDS-PAGE (Figure 3.3 A, S). An intense band of approximately 10kDa is clearly visible. This intense band is missing from the unbound fraction eluted from the streptavidin column suggesting that this represents the protein of interest (Figure 3.3 A, FT). Samples were removed for analysis by SDS-PAGE following elution and 3C protease treatment to remove the affinity tag (Figure 3.3, lanes E and Cl respectively). The slight increase in protein mobility (Figure 3.3 A, E-Cl) demonstrates that the tag has been successfully cleaved. Proteins were then applied to a size exclusion column for further purification. A sample of the peak fractions were taken for analysis and then pooled and concentrated using Vivaspin concentrators with a 3kDa molecular weight cut-off (Figure 3.3 A, 1-4). An example of a typical size exclusion chromatography experiment is shown in Figure 3.3 B. The second peak corresponds to that of the Rna15 2-103 protein whilst the third peak corresponds to the Strep II tag. Indeed, when a sample of this peak is analysed by SDS-PAGE a faint low molecular band is observed corresponding to the streptavidin tag which is just over 2kDa in this construct. The first peak corresponds to the void volume. A sample of this peak was analysed by SDS-PAGE but no protein species were observed using coomassie staining which suggests that this peak consists of nucleic acids. The

A_{260}/A_{280} ratio of the purified protein was ~0.5 demonstrating no significant nucleic acid contamination. All mutant constructs were expressed and purified in exactly the same way.

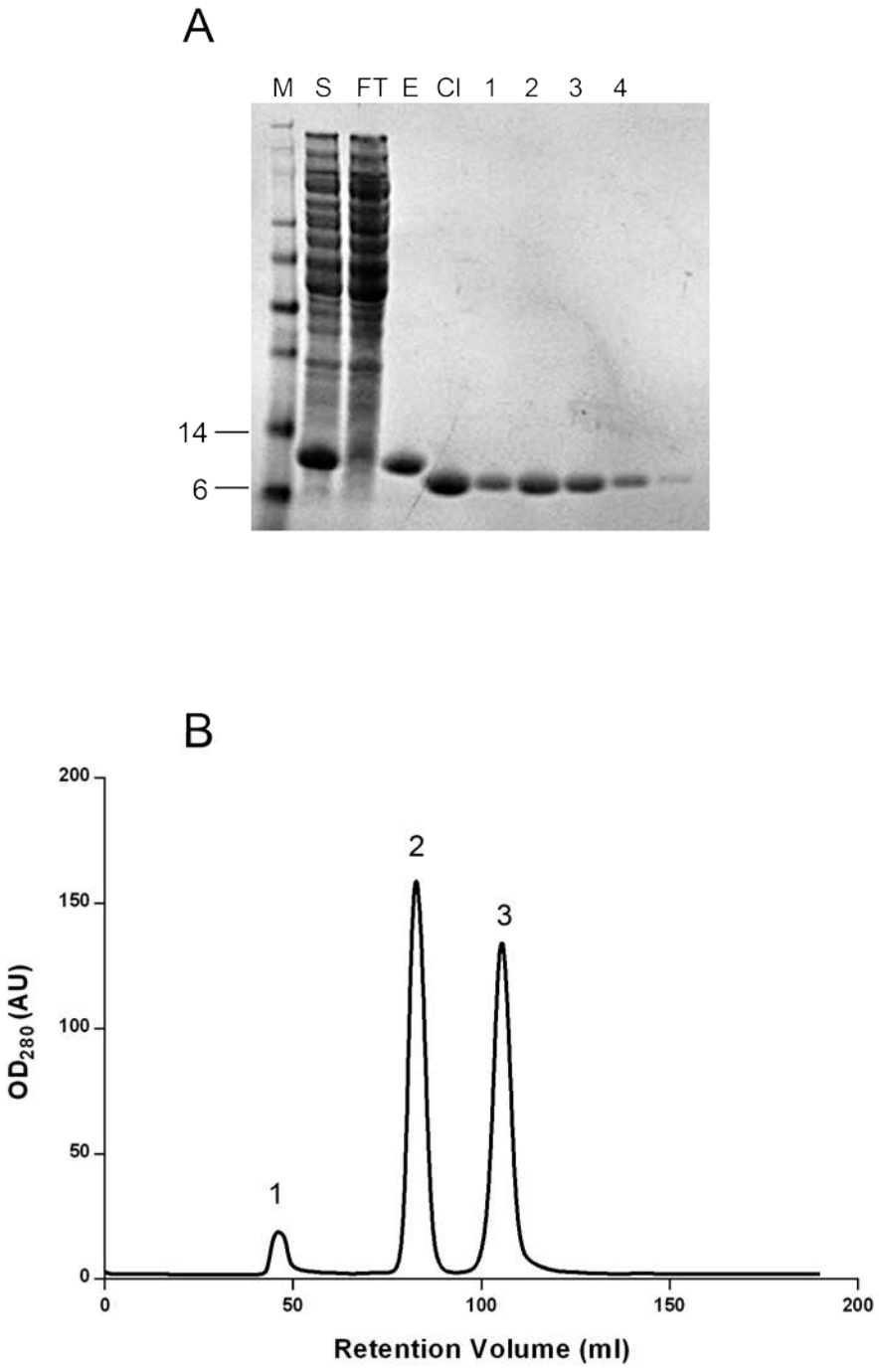


Figure 3.3 Purification of Rna15 (2-103) mutant constructs. A. SDS analysis of Rna15 (2-103) purification. **B.** Size exclusion chromatography profile demonstrating a typical Rna15 (2-103) mutant purification. Peak no. 1 is the void volume, peak no.2 is the target protein and peak no.3 is the cleaved strep II tag.

3.3.2 Circular dichroism and mass spectroscopy

In order to determine whether point mutations made in Rna15 (2-103) affected the secondary structure of the protein, circular dichroism (CD) was performed. This method measures the differentially absorbed left- and right- handed circularly polarised light that is sensitive to the conformation of protein structure. As CD exploits the absorption properties of a protein, the chromophores that contribute to an absorption spectrum are identical to those that contribute to a CD spectrum. CD signal is only detected when a molecule is achiral so proteins are ideal candidates as backbone of the protein provides a plane of symmetry.

The tertiary and sometimes quaternary structure of proteins can be derived from the near-UV CD spectrum as Tryphtophan, Tyrosine and Phenylalanine absorb between wavelengths 310-255 nm. CD absorption in the far-UV (below 250nm) reveals the secondary structure of a protein, for example α-helical, β-sheet, β-turn or disordered. The structural results are low resolution but are extremely sensitive to differences in conformation of the protein. Usually α-helical proteins produce a intense negative signal with two peaks at 208 and 222nm and a strong positive signal between wavelengths 191-193nm. Proteins that are predominantly formed of β-sheets or β-turns exhibit a negative signal between 210-225nm and a strong positive band between 190-200 nm.

Therefore, CD was used to determine whether mutation of residues within the RRM of Rna15 had any effect on the structural integrity of the protein when compared to wild type Rna15 (2-103). Mutations were selected on the basis of their location on the surface of the crystal structure of the RRM so they were not expected to alter the conformation of the four stranded β-sheet backed by two α-helices structure of the RRM. CD spectra were recorded for wild type Rna15 (2-103) and all mutant Rna15 (2-103) proteins (Appendix 7.2). The spectra of all Rna15 (2-103) mutant

proteins were identical to that of the wild type. The proteins had significant secondary structure demonstrated by the absence of a strong negative signal between wavelengths 195-200nm and absence of a weak signal between 215 and 230nm. Rna15 (2-103) is composed both of β-sheets and α-helices with a small amount (residues 2-16) being highly disordered. In proteins with high α-helical and β-sheet structures, the α-helical signal of the protein tends to be dominant over the signal produced by the β-sheet structural component. CD spectra for proteins like Rna15 (2-103) would normally produce spectra with a signal at either 222, 208 or between 190-195nm. Often an additional signal may be shown between 210 and 220nm because of the overlapping α-helical and β-sheet components. As expected, the CD spectra for all were extremely similar and revealed a high negative signal around 210nm demonstrating a predominantly α-helical and β-sheet structure as expected given the crystal structure of the RRM of Rna15 (Figure 3.4). Therefore, we can conclude that mutations presented in this study do not affect the conformation of the RRM structure.

In addition to CD, the molecular weight of all Rna15 (2-103) proteins, including wild type, were measured using mass spectrometry (Appendix 7.3). The expected molecular weight was calculated for all proteins and compared with the derived molecular weights determined by mass spectrometry. These results combined with sequencing of all constructs adequately demonstrate that mutagenesis of residues within the RRM of Rna15 was successful. A thermal shift assay was performed on Rna15 (2-103) and Y27A and demonstrated that both maintained observed secondary structure up to 60°C.

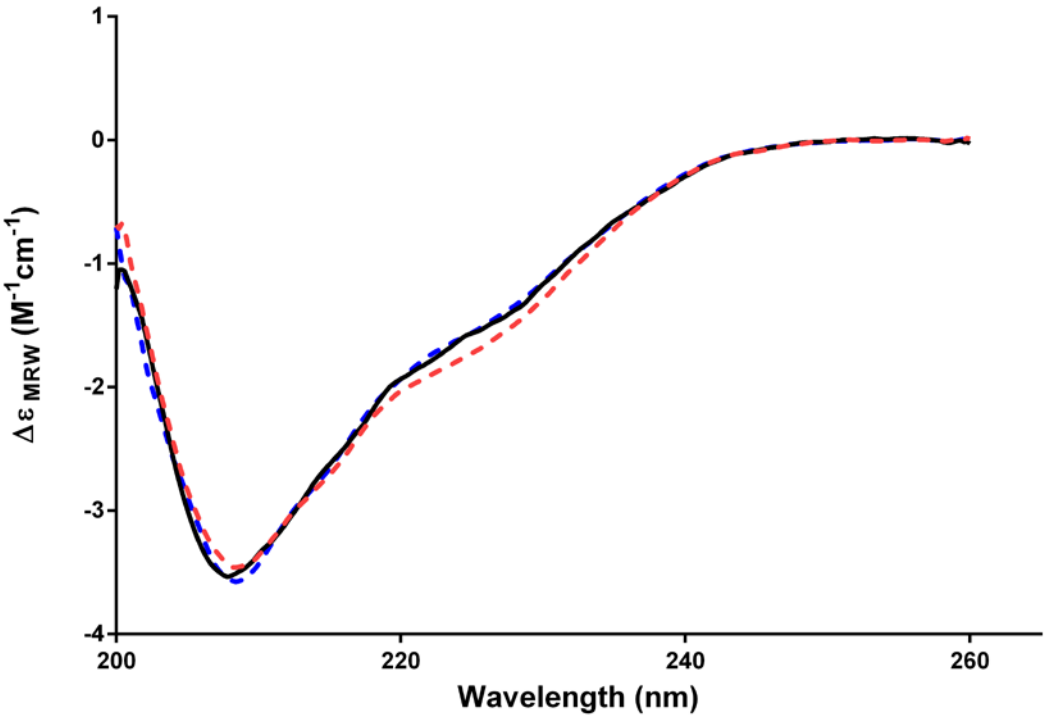


Figure 3.4 Circular Dichroism of Rna15 (2-103) and mutants Y27A and K90E. The above CD spectra demonstrate identical plots for Rna15 2-103 (black), Y27A (red) and K90E (blue). Plots are representative of all mutant constructs. Complete spectra of all constructs can be found in appendices.

3.3.3 Analysis of residues in the RRM that direct RNA binding

Fluorescence titrations were carried out in triplicate for each mutant and the dissociation constant and the standard deviation of the mean calculated. The 5'-TET-UGUUGU-3' ribo-oligonucleotide was used to provide a direct comparison to wild type RNA binding affinity demonstrated in previous experiments. Representative emission spectra from a single S24T fluorescence spectroscopy experiment are shown (Figure 3.5). The mutants prepared can be divided into three groups centred on site I/site II, site III and the surface lysines and will be considered separately.

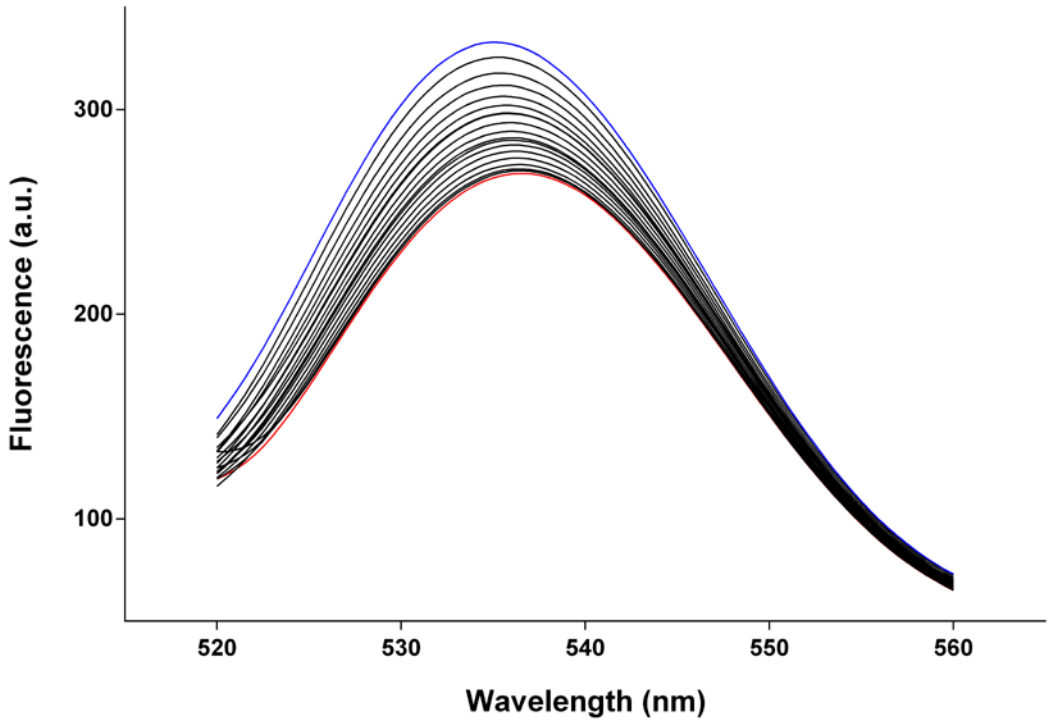


Figure 3.5 Measured emission spectra of mutant S24T. The above spectra are typical of a fluorescence titration. The fluorescence signal decreased upon addition of protein. The initial fluorescence emitted prior to addition of protein is represented by the blue plot. The last scan measured in excess of S24T protein is represented by the red plot.

3.3.3.1 Mutational analysis of residues involved in site I and II RNA binding

Residues R87, Y27, S24 and Y21 were mutated by either conservative or non-conservative changes at each position (Table 3.1). In each case, the non-conservative mutation was to alanine as the CH_3 side chain would be expected to disrupt the polar or charged interactions mediated by the wild type side chains to the RNA oligonucleotide. Furthermore, the side chain of alanine is small and does not actively interact with other residues and should not, therefore, affect binding of RNA mediated by proximal residues. As such any effects on RNA binding can be directly attributed to the residue in question. The conservative mutations were chosen to determine whether addition of a chemically similar amino acid altered binding affinity.

Site I binding is mediated by Y27 and R87 which together form the walls of the binding pocket through the aromatic side chain of Y27 and the planar side chain of R87 which allows stacking of both a G and U nucleotide. Unsurprisingly, mutation R87A results in a large decrease in binding affinity for the oligonucleotide resulting in a dissociation constant of 43 μM ± 10.8 μM (Figure 3.6). The loss of the planar side chain likely leads to disruption of the wall of the binding pocket preventing stacking of the nucleotide. Also, the loss of the stacking interaction may indirectly lead to loss of the Watson-Crick hydrogen bonding also observed at this site since the loss of one side of the pocket wall may mean that the base of the nucleotide cannot make the necessary hydrogen bonds. The conservative mutation R87K also displays a decrease in RNA binding with a dissociation constant of 33.8 μM ± 9.6 μM (Figure 3.6). This infers that the role of R87 is likely structural and loss of planarity in the R87K mutant may impede stacking. This supports the hypothesis that R87 interacts with the RNA predominantly through the stacking interaction with the side chain as observed in the crystal structure.

Mutation of Y27 results in a decrease in RNA binding. Alanine substitution results in a more than 10 fold reduction in binding (K_d 76.7μM ± 16μM), presumably from loss of the stacking interactions, resulting in a overall decrease in stability of the binding pocket (Figure 3.6). Loss of the stacking at one side of the binding pocket in Y27A may also affect the orientation of the bound nucleotide and therefore lead to loss of the hydrogen bonding interactions demonstrated in the crystal structure.

Interestingly, mutation Y27F demonstrates a slight increase in RNA binding affinity resulting in a dissociation constant of 4.3 μM ± 0.2 μM (Figure 3.6). This is likely due to composition of the aromatic side chain of phenylalanine. The delocalised electron arrangement present in the aromatic side chain due to the absence of the tyrosine hydroxyl group may facilitate improved stacking against the base of the nucleotide. S24 was selected for mutation as hydrogen bonding between its side chain hydroxyl group and the side chain of R87 acts to hold R87 in place in order to form the binding pocket. Unexpectedly, the conservative mutation of S24T results in a larger decrease in binding affinity than the non-conservative mutation S24A, 18.1μM ± 2.1μM and 8.8μM ± 0.5μM respectively. It is possible that backbone interactions made by S24A are adequate to hold R87 in the correct orientation to stack against the nucleotide. Indeed, it is not as effective as the wild type demonstrating that hydrogen bonding between the hydroxyl present in the side chain of S24 is required to maintain wild type RNA binding affinity. The decrease in binding affinity observed in S24T is surprising as like serine, theronine is both polar and contains a hydroxyl group. It is possible that the slight increase in overall size of the side chain may affect the interaction with R87 and that the orientation and position of the hydroxyl group present in threonine does not adequately facilitate hydrogen bonding.

Rna15 2-103 construct	K_d ($x10^{-6}$ M^{-1})
Wild type	6.4 ± 0.5
R87K	33.8 ± 9.6
R87A	43.6 ± 10.8
Y27F	4.3 ± 0.2
Y27A	76.7 ± 16
S24T	18.1 ± 2.1
S24A	8.8 ± 0.5
Y21F	3.1 ± 0.9
Y21A	28 ± 2.1

Table 3.2 Fluoresence spectroscopy analysis of Rna15 (2-103) site I and site II mutants. Titrations were repeated three times for each mutant and the average of these along with the standard deviation of the mean is displayed in the above table.

.

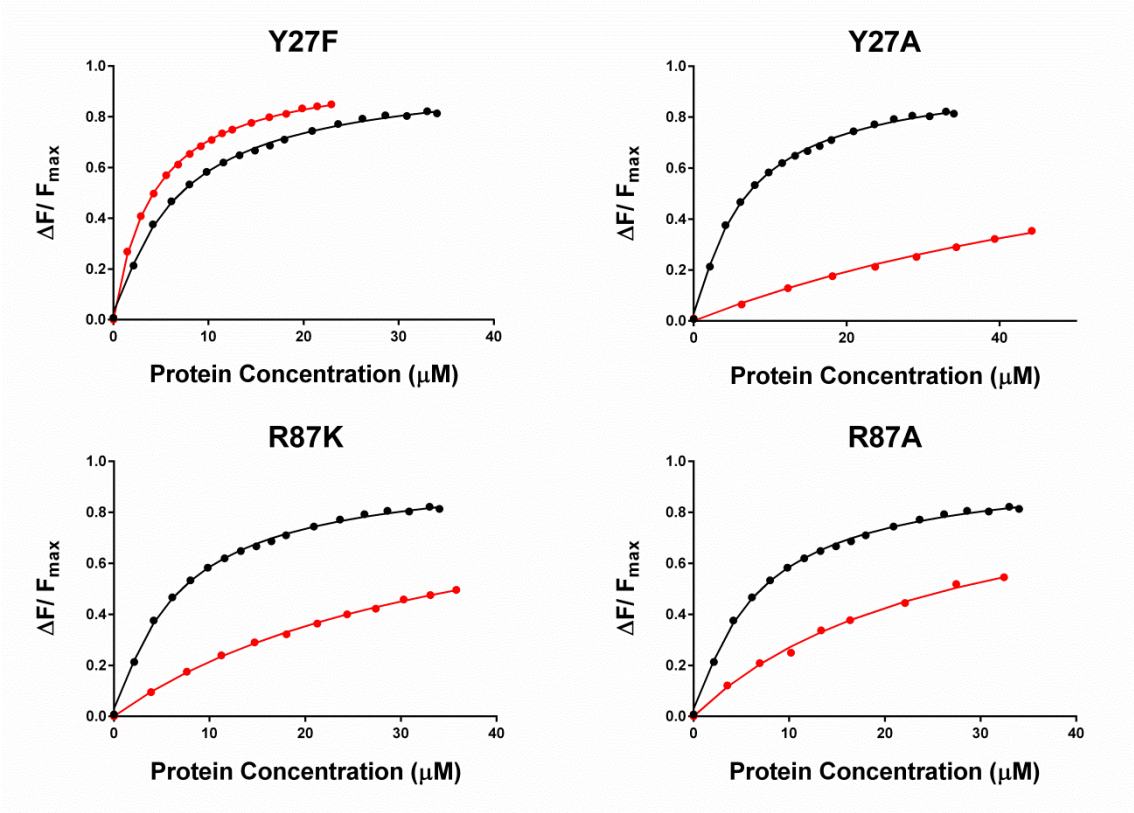


Figure 3.6 Determination of K_d values for Rna15 (2-103), Y27F, Y27A, R87K and R87A RNA binding using fluorescence spectroscopy. Protein was titrated into labelled RNA and the change in signal observed. The symbols represent the experimental data and the solid lines are the fit to the data using a single site binding model. Wild type-RNA binding is represented by black lines and symbols while mutant-RNA binding is represented by red lines and symbols.

In conclusion, mutational analysis of the site I binding pocket reveals that both R87 and Y27 are essential for RNA binding. Mutation of S24 to alanine and threonine results in slight decrease in affinity. The crystal structure revealed that both the backbone and side chain hydroxyl groups interact with R87 to hold it in the correct orientation to stack against the nucleotide. It is apparent from the above experiments that this interaction although important is not essential to maintain RNA binding. This is possibly due to interactions between the backbone hydroxyl of threonine and R87 which may act to hold R87 in the correct orientation in the absence of wild type S24. Therefore, although mutational analysis implies S24 interaction is not essential, mutation S24A and S24T results in a moderate decrease in binding demonstrating its requirement for optimum RNA binding.

From the crystal structure, the site II binding pocket is predominately formed by a stacking interaction between Y21 and the base of a guanine nucleotide. Similar to Y27F in site I, Y21F also produces an increase in binding affinity of the same order of magnitude (K_d 3.1μM ± 0.9μM). Again this is likely due to the altered composition of the aromatic phenylalanine which has a continuous arrangement of delocalised electrons without the presence of a hydroxyl group as observed in tyrosine. This could lead to an increase in stability in the stacking interaction. Mutation Y21A leads to a more than 4-fold decrease in RNA binding resulting in a dissociation constant of 28μM ± 2.1μM. This is more than likely due to loss of the base stacking interaction. These results demonstrate that Y21 is important but not as essential to RNA binding as R87 and Y27, possibly due to the lack of extensive hydrogen bond interactions as seen in site I.

3.3.3.2. Mutational analysis of residues involved in site III RNA binding

In addition to residues in sites I and II other residues are thought to be involved in RNA binding due to their chemical shift movement during NMR RNA titrations. This revealed that residues present on the surface of the entire RRM were affected by introduction of RNA and highlighted a possible third stacking interaction site present at residues Y61 and Y63. Taking into account this evidence and the structure of the RRM, residues were selected for mutational analysis on the basis of position within the RRM and chemical features which may facilitate RNA binding. The positioning of the two aromatic rings of Y61 and F63 on the surface of the RRM could potentially form two sides of a binding pocket as observed in site I. To investigate RNA binding at this position both Y61 and F63 were mutated in point and double mutation constructs (Table 3.3). In the point mutation constructs Y61 was mutated to phenylalanine, alanine and serine. Mutation Y61S did not affect RNA binding in the Rna15 (2-103) construct as the measured K_d was similar to that of wild type, 8.7µM ± 1.2µM. Mutation Y61A resulted in a two-fold decrease in RNA binding with a measured K_d of 11.7 µM ± 0.6 µM. A slight increase in RNA binding affinity is observed in the Y61F mutation, 2.5 µM ± 1 µM the same order of magnitude as the site I Y27F mutation. The increased binding affinity in both the Y27F and Y61F suggest that a stacking interaction also exists at site III. The apparent lack of effect on the RNA binding affinity in the Y61 mutants suggests that although involved, it is not essential for RNA binding. A similar pattern was also observed in the F63 point mutations. Conservative mutation F63Y had no apparent affect in RNA binding (K_d 7.4µM ± 1.5µM) whilst non-conservative mutation to alanine only produced a slight decrease in RNA binding affinity upon comparison with wild type (K_d 15.4µM ± 3.9µM). A more pronounced affect is observed in F63S where the K_d is 21.9µM ± 8.6µM demonstrating a larger decrease in RNA binding affinity. Point mutations at both Y61 and F63 do not display a dramatic reduction in RNA

binding as seen in the mutational analysis of site I. Therefore, both Y61 and F63 mutations were combined in double mutation constructs in order to determine whether mutation of both resulted in significant loss of RNA binding affinity. Mutations Y61F/F63Y and Y61S/F63S displayed similar RNA binding affinities with dissociation constants of 22.3 µM ± 2 µM and 17.6 µM ±1.1 µM respectively. The decrease in binding in the conservative Y61F/F63Y double mutation is surprising as the Y61F and F63Y single mutations had such minimal effect. It is possible that an effect in binding is only observed when both are mutated because both Y61 and F63 may act in a redundant manner to stack against the nucleotide base. The Y61S/F63S RNA binding affinity (K_d 17.6 µM ± 1.1 µM) suggests that the presence of serine at both positions does slightly reduce binding affinity for the RNA upon comparison with wild type RNA binding affinity. However, the presence of the double mutation does not result in a significantly larger decrease in binding affinity than that observed in the Y61S and F63S point mutations where the dissociation constant measured was 8.7 µM ± 1.2 µM and 21.9 µM ± 8.6 µM respectively. Since mutation F63S affects Rna15 2-103 binding affinity it is not surprising that the K_d presented in the double is comparable to that displayed in this mutant. These results seem to suggest that mutation Y61S is tolerated while F63S is not. This is perhaps due to the orientation of the serine side chains at either position where at position 61 they are able to interact in some way with the nucleotide to support RNA binding but at position 63 the side chain of serine obstructs interaction with the RNA. However, the double mutation Y61A/F63A resulted in a 6-fold reduction in RNA binding affinity (K_d of 36.2 µM ± 4.7 µM. The decrease in RNA binding observed is comparable to those seen in single mutations of site I. The decrease in binding reinforces the NMR data suggesting that these residues are also important for RNA binding. In the crystal structure, the side chain of R87 was shown to stack against the base of the nucleotide; moreover, interactions observed at site II were also

predominantly stacking interactions mediated by Y21. Mutation at site III results in reduction in binding affinity comparable to that observed in the R87 and Y21 mutants. Therefore, it is possible that nucleotide binding at site III is mediated by a stacking interaction between the base of the nucleotide and the aromatic rings of Y61 and F63. However, structural evidence is required to provide an adequate argument of the nature of site III mediated RNA binding.

Rna15 2-103 construct	K_d ($x10^{-6}$ M^{-1})
Wild type	6.4 ± 0.5
F63Y	7.4 ± 1.5
F63A	15.4 ± 3.9
F63S	21.9 ± 8.6
Y61F	2.5 ± 1
Y61A	11.7 ± 0.6
Y61S	8.7 ± 1.2
Y61F, F63Y	22.3 ± 2
Y61A, F63A	36.2 ±4.7
Y61S, F63S	17.6 ± 1.1

Table 3.3 Fluorescence spectroscopy analysis of Rna15 (2-103) mutants in site III.

3.3.3.3 Mutational analysis of surface lysine residues

In addition to the three discete RNA binding sites, the effect of RRM surface lysine residues on RNA binding was analysed by mutagenesis. Positively charged surface lysines, K48, K59 and K90, were selected for mutagenesis due to their surface accessibility and proximity to the three mapped binding sites.

Lysines, K90 and K48 were mutated to alanine, arginine and glutamate whilst K59 was mutated to arginine and glutamate (Table 3.4). There is no K59A mutant as site directed mutagenesis failed to generate a positive clone after several attempts. The

aim of mutational analysis was to determine whether these lysine residues formed charge-charge interactions with the RNA. For K48R a slight increase in RNA binding affinity is observed (K_d is 3.7 µM ± 0.61 µM). This is unsurprising given that arginine has two NH_2 groups present within the side chain which may facilitate a stronger electrostatic interaction with the negatively charged RNA backbone. However, K48E and K48A do not result in any effect on RNA binding affinity. The mutation to alanine signifies a complete loss in charge whilst mutation to glutamate represents a charge reversal to a negatively charged amino acid. The resulting dissociation constant for K48E is 7µM ± 0.2µM and for K48A is 6.6µM ± 3.3µM and statistically indistinguishable from the measured wild type dissociation constant. The wild type binding demonstrated in the K48A and K48E mutants demonstrates that K48 does not play a significant role in RNA binding.

Residue K59 was also mutated to arginine and glutamate due to its position near to site III. As with K48R, K59R resulted in a slight increase in RNA binding affinity (K_d 3.3µM ± 0.7µM) and again this is likely due to the increase in positive charge leading to an increased non-specific attraction for the negatively charged RNA. Analysis of K59E revealed that the measured K_d was 28.2 µM ± 7.7 µM which demonstrates a decreased RNA binding affinity compared to wild type. Mutant K59E produces a more pronounced effect on RNA binding affinity than K48E. The measured value implies that K59 maintains some type of contact with the bound RNA oligonucleotide. However, the affinity for the K59E mutant is not sufficiently decreased to represent the loss of a contact critical for RNA binding. The apparent loss in RNA binding in the K59E mutant implies a charge-charge interaction. However, without at

K59A binding affinity it is possible that the measured K59R and K59E binding affinities are due to non-specific binding.

K90 was also mutated due to is position between sites I and II. As observed with the other lysine to arginine mutants, K90R gave tighter binding with an affinity of 2.1µM ± 0.4µM. Further information as to the nature of the K90 interaction was provided by mutation K90A. The observed dissociation constant was 18.9µM ± 3.3µM, demonstrating a weaker RNA binding affinity when compared to wild type. The observed decrease in binding affinity indicates an interaction between K90 and the bound RNA. Moreover, mutation to K90E resulted in a loss of RNA binding. The lack of binding in the K90E mutant suggests that the RNA oligonucleotide is repelled from the face of the RRM. The nature of the K90-RNA interaction is likely based on charge since mutation to an acidic residue completely abolishes RNA binding. Other interactions may be mediated by K90 resulting in the slight decrease in RNA binding in the K90A but K90E shows that the charged group exerts the dominant effect.

Rna15 2-103 Construct	K_d ($x10^{-6}$ M^{-1})
Wild type	6.4 ± 0.5
K48E	7 ± 0.2
K48R	3.7 ± 0.61
K48A	6.6 ± 3.3
K59E	28.2 ± 7.7
K59R	3.3 ± 0.7
K90E	-
K90R	2.1 ± 0.4
K90A	18.9 ± 3.3

Table 3.4 Fluoresence spectroscopy analysis of lysine residues. Experiments were repeated three times and the average shown in the above table.

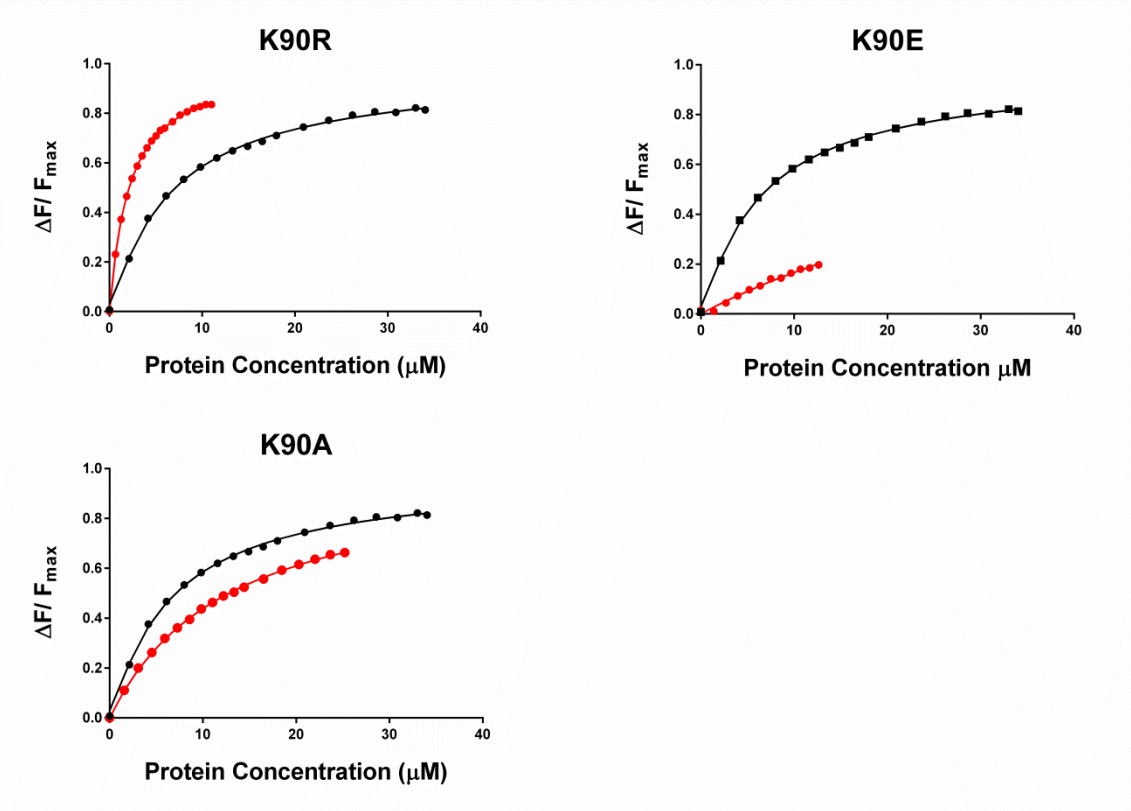


Figure 3.7 Determination of the K_d of Rna15 (2-103), K90E, K90R and K90A RNA binding. The symbols represent the signal observed upon increasing concentration of protein while the solid line represents the fit. Black symbols and lines represent wild type-RNA binding while red lines and symbols represent mutant-RNA binding.

The results presented in this chapter provide insight into the mechanism of RNA binding in Rna15. Site I is clearly essential to maintain binding to the UGUUGU RNA oligonucleotide. This is not unexpected given the extensive network of interactions displayed in the crystal structure. Residue K90 forms an essential charge-charge interaction with the RNA oligonucleotide that is crucial for RNA binding. The mutational analysis also implicated residues within site II and site III in RNA binding but these sites appear less critical to RNA binding affinity. K59 interacts with the RNA oligonucleotide in similarly non-essential fashion. However, these results do not demonstrate whether Rna15 binds one continuous stretch of RNA or more than one RNA oligonucleotide. Previous results suggest that the stoichiometry of the Rna15-RNA interaction is 1:1 (107, 108, 142). In addition, NMR spectroscopy experiments revealed that residues across the entire surface are affected by titration of an RNA oligonucleotide. Therefore, it is highly likely that a single RNA oligonucleotide binds across the surface of the RRM and the residues presented here mediate binding of the RNA. The three sites and charge-charge interactions with K90 appear sufficient to anchor the RNA. However, additional residues are affected by titration of RNA during NMR spectroscopy experiments perhaps with an accessory role in RNA binding. In order to generate a more detailed analysis of the extent of residues involved in RNA binding further mutational analysis of the residues present on the surface of the RRM is required.

The data provide a detailed analysis of the mechanism behind Rna15 RRM RNA binding. However, it was not known whether Rna15 could distinguish between G/U nucleotides, at different positions in the binding sites. Such specificity may account for the differing effects observed in mutational analysis using only the 5' TET-UGUUGU 3' sequence. Therefore, mutation of these residues does not characterise the binding specificity of Rna15 (2-103) in terms of a consensus sequence for binding. Specificity

for both guanine and uracil nucleotides has already been demonstrated in the crystal structure of the bound RRM at site I in the crystal structure but has not been shown for sites II or III.

3.4 Summary

This chapter has characterised Rna15-RNA binding using *in vitro* techniques. Initially, a number of deletion constructs were cloned, expressed and purified. These constructs were designed to facilitate analysis of the role of the regions flanking the RRM of Rna15. Deletion of the tail region (residues 94-103) did not have any effect on Rna15 mediated RNA binding affinity demonstrated by fluorescence titrations. Previous NMR experiments suggested a possible regulatory role of the tail regions which may cover possible RNA recognition sites present on the RRM (139). However, these experiments did not provide significant evidence that the tail region has a role in regulation of RNA binding as not effect in RNA binding affinity was observed. However, this does not mean that the tail does not perform this regulation of RNA binding and further characterisation is required.

Extensive mutational analysis of the RRM reveals that Y27, R87 and K90 are essential for Rna15 (2-103) binding to the UGUUGU RNA oligonucleotide. This is illustrated by the decrease in RNA binding affinity demonstrated by fluorescence titration employing the Rna15 (2-103) mutants. The defect in RNA binding demonstrated by the K90E mutation demonstrates a previously uncharacterised site for RNA binding. Other mutants impose a less severe effect on RNA binding but still affect the binding affinity of Rna15 (2-103) implying that residues present across the surface of the RRM are required to maintain wild type RNA binding. Results presented in this chapter have provided a deeper insight into the mode of RNA binding by the RRM of

Rna15. The importance of the previously annotated binding sites present on the RRM has been demonstrated in terms of their influence on binding affinity.

4.1 Introduction and Overview

This chapter aims to characterise the effect of point mutations within the RRM of Rna15 on the overall health of *S. cerevisiae* by comparing the growth phenotype of wildtype and Rna15 mutants. In order to do this the Rna15 RRM mutations studied in Chapter 3 are introduced into haploid strains of *S. cerevisae* to determine the effect of these mutations on growth. This allows the correlation between RNA binding and overall health through analysis of growth of the cell to be investigated.

The effect of these mutations on 3' end processing activity was assessed by Quantitative Reverse Transcription Polymerase Chain Reaction (qRT-PCR) to determine whether mutation of Rna15 affected expression of 5 housekeeping genes, ACT1, CYC1, ADH1, TDH2 and YPT1. Finally, any effects on transcription and potential readthrough of proximal polyadenylation sites were assayed by RNA sequencing (RNA seq). This technique is extremely sensitive and allows the sequencing and mapping of total mRNA resulting in analysis of transcription within Rna15 strains on a global level.

4.2 Establishing the Rna14 and Rna15 haploid strains

Both RNA14 and RNA15 genes were inserted into yeast shuttle vector, pRS315, as described in Chapter 2. The resulting clones were then transformed into haploid *S. cerevisiae* strains where the genomic copy of the gene had been deleted by insertion of a Kanamycin cassette. Both RNA14 and RNA15 are essential genes so both haploid strains are kept viable by introduction of wild type genes present in pRS316 yeast shuttle vectors. Constructs were transformed into *S. cerevisiae* and selected for using a 5-FOA assay. The 5-FOA selection method exploits the URA^+ marker present on the

pRS316 vector and results in selection of the LEU^+ pRS315 construct. The RNA14 and RNA15 pRS315 constructs were transformed into the appropriate haploid *S. cerevisiae* Δrna14/Δrna15 strain containing a wild type copy of each gene on the pRS316-URA construct. The pRS315-LEU constructs were selected for by incubation on minimal media minus leucine. Cells containing the pRS315-LEU construct and therefore viable on -LEU media were taken for 5-FOA selection. Selection was achieved by incubating yeast on 5-FOA containing media. The URA marker present in the pRS316 vector encodes the enzyme orotidine 5-phosphate decarboxylase which functions in uracil synthesis. This enzyme converts the 5-FOA to the toxic compound, 5-fluorouracil, resulting in cell death. Therefore, cells viable on 5-FOA containing media following transformation must contain only the pRS315 construct (Figure 4.1). Both tagged and untagged RNA14 and RNA15 pRS315 constructs were transformed into their respective Δrna14 and Δrna15 *S. cerevisiae* strains containing the pRS316-URA constructs in order to demonstrate whether presence of the HA tag affected either Rna14 or Rna15 function and thus resulting in inviable cells.

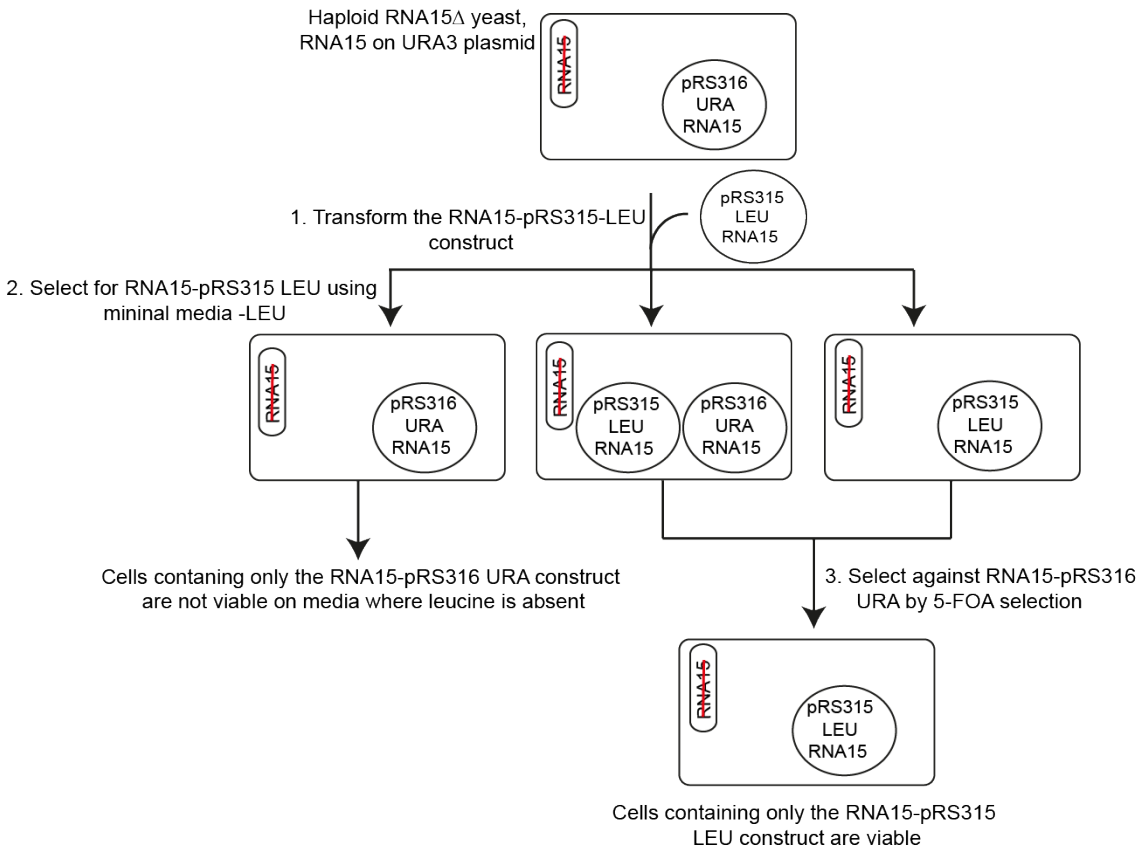


Figure 4.1 Schematic representation of the 5-FOA selection assay.

Transformation of wildtype RNA14 and RNA15 on pRS315-LEU are shown in Figure 4.2. The top two panels show transformation of RNA14. Cells are shown as spots and are present as a dilution series where cells were diluted serially 1:10 four times. The first row of cells represents a control to demonstrate that the 5-FOA assay worked as expected. The *S. cerevisiae* strain used is a haploid Δ*rna14* cell line kept viable through presence of the RNA14-pRS316 yeast shuttle vector. Cells are viable on rich medium at 22ºC but are unable to grow on minimal medium due to the presence of 5-FOA. The second row of cells demonstrates the effect of transformation of an empty pRS315-LEU plasmid. The cells are viable on rich medium at 22ºC due to the presence of the RNA14-pRS316 vector prior to selection but incubation with 5-FOA results in cell death due to the lack of wild type RNA14 gene. Transformation of the pRS315 yeast shuttle vector containing wild type RNA14 results in cell viability on both rich and minimal medium containing 5-FOA at 22ºC. This demonstrates transformation and subsequent selection of cells containing only the RNA14-pRS315 construct is successful. The final row present in the top two panels represents transformation of the HA-tagged version of the RNA14-pRS315 construct. Growth of cells is evident on both the rich and minimal media at 22ºC. Growth of the HA-tagged and untagged versions of RNA14 pRS315 are comparable demonstrating that the HA-tag does not interfere with RNA14 function within the cell. Transformation of the RNA15-pRS315-LEU construct was carried out in the same way as the RNA14-pRS315-LEU transformation. The appropriate controls were included and neither the Rna15-pRS316-URA strain nor cells containing the empty pRS315-LEU plasmid were able to grow on 5-FOA. However, this phenotype was rescued by transformation of both the RNA15 HA-pRS315-LEU and RNA15-pRS315-LEU constructs (Figure 4.2).

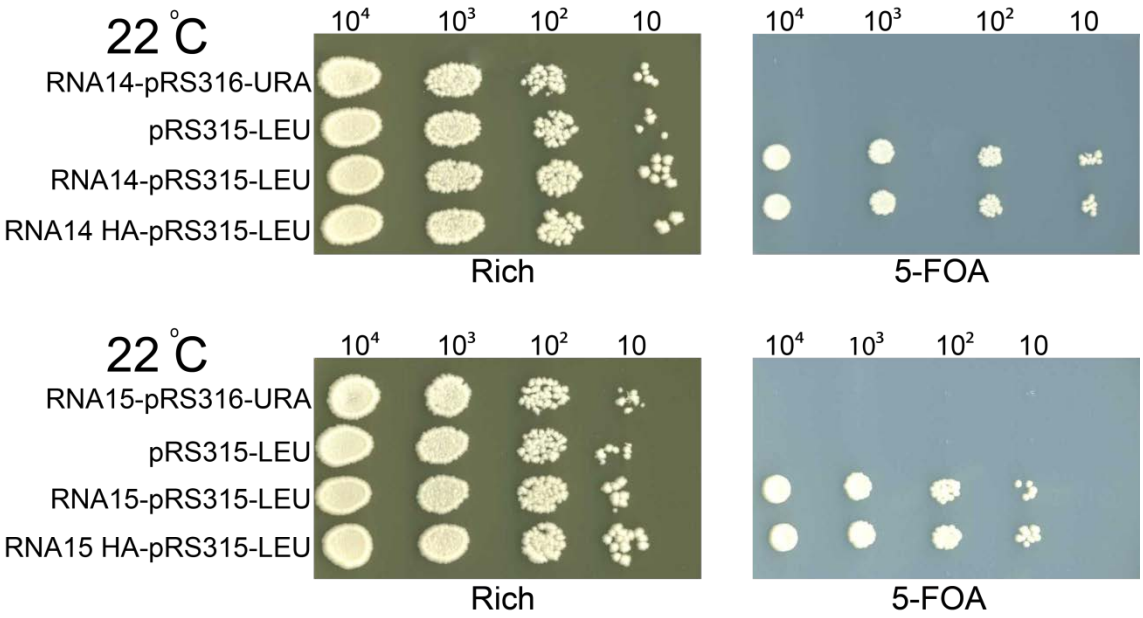


Figure 4.2 Testing the Rna14 and Rna15 *S. cerevisiae* strains. Transformation of Rna14-pRS315-LEU and Rna15-pRS315-LEU constructs into the Δ*rna14* and Δ*rna15 S. cervisiae* deletion strain respectively. Cells are plated as "spots" in a dilution series. Plating on 5-FOA media selects against the RNA14/RNA15-pRS316-URA plasmid making the cells reliant on the RNA14/RNA15-pRS315-LEU for viability.

4.3 Analysis of Rna15 RRM mutations on growth phenotype of *S. cerevisiae*

The Rna15 RRM point mutations (Table 4.1) were introduced into the Rna15-HA-pRS315-LEU construct in order to determine the effects of these mutations on the growth phenotype of *S. cerevisiae*. Again mutant constructs were transformed into the *S. cerevisiae* strain supported by presence of the RNA15-pRS316-URA plasmid and selected for by the 5-FOA assay. Western analysis of a number of transformed Rna15 RRM mutations were performed to demonstrate that all mutations were stably expressed and mutagenesis did not result in degradation. Detection of each Rna15 construct was achieved by employing an antibody able to recognise and bind the HA-tag present on all Rna15 constructs. A subset of Rna15 mutants are shown to express similarly to wild type (Figure 4.3). The band corresponding to the Rna15 Δ16-94 resolves at a smaller molecular weight upon comparison with wild type Rna15 due to deletion of the RRM.

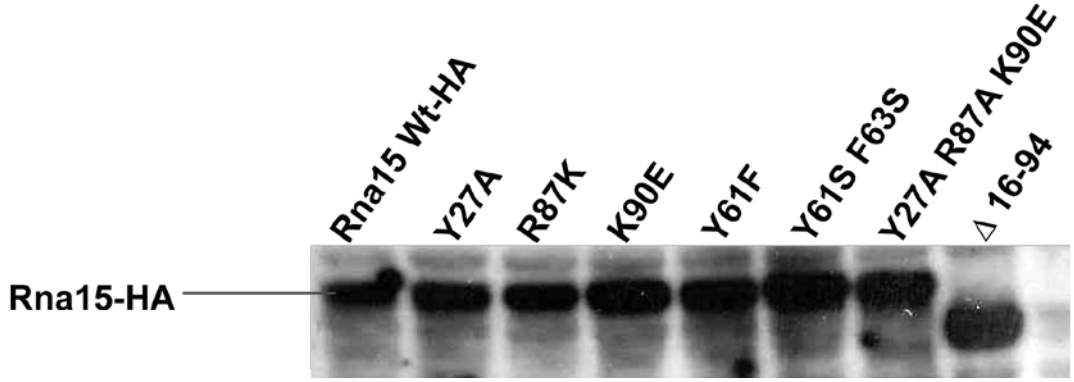


Figure 4.3 Western analyses of Rna15 wild type and mutant constructs. The highlighted bands correspond to Rna15-HA wild type and mutant constructs. All proteins are similarly expressed demonstrating that mutation does not affect stability of the protein. Due to the slightly smaller molecular weight of Rna15 Δ16-94, this band resolves slightly further down the gel.

RNA15 RRM point mutation	Position
Y27F	Site I
Y27A	Site I
R87K	Site I
R87A	Site I
S24T	Site I
S24A	Site I
Y21F	Site II
Y21A	Site II
Y61F	Site III
Y61A	Site III
Y61S	Site III
F63Y	Site III
F63A	Site III
F63S	Site III
Y61F, F63Y	Site III
Y61A, F63A	Site III
Y61S, F63S	Site III
K90R	Surface lysine near site I and II
K90E	Surface lysine near site I and II
K90A	Surface lysine near site I and II
K48R	Surface lysine near site III
K48E	Surface lysine near site III
K48A	Surface lysine near site III
K59R	Surface lysine near site III
K59E	Surface lysine near site III

Table 4.1 List of RNA15 RRM point mutations.

All Rna15-RRM point mutations tested gave identical results therefore only a subset of results are shown (Figure 4.4). The RNA15-pRS316-URA strain was able to grow on rich media at 22ºC and as expected demonstrated no temperature sensitivity at 37ºC. However, selection on 5-FOA rendered the RNA15-pRS316-URA strain non-viable at both 22ºC and 37ºC. Transformation of the empty pRS315 plasmid followed the same pattern. Transformation of the wild type RNA15-pRS315-LEU construct supported growth on both rich and 5-FOA at 22ºC and displayed no temperature sensitvity at 37ºC. Introduction of point mutations into the RRM of Rna15 and the subsequent transformation and selection resulted in growth comparable to the wild type RNA15-pRS315-LEU (Figure 4.4). The reduction in RNA binding affinites described in chapter 3 do not result in any detrimental effect on growth of *S. cerevisiae.* Fluorescence spectroscopy experiments demonstrated that mutation K90E completely abolished Rna15 mediated RNA binding. However, introduction of this mutation into *S. cerevisiae* had no effect on cell growth (Figure 4.5).

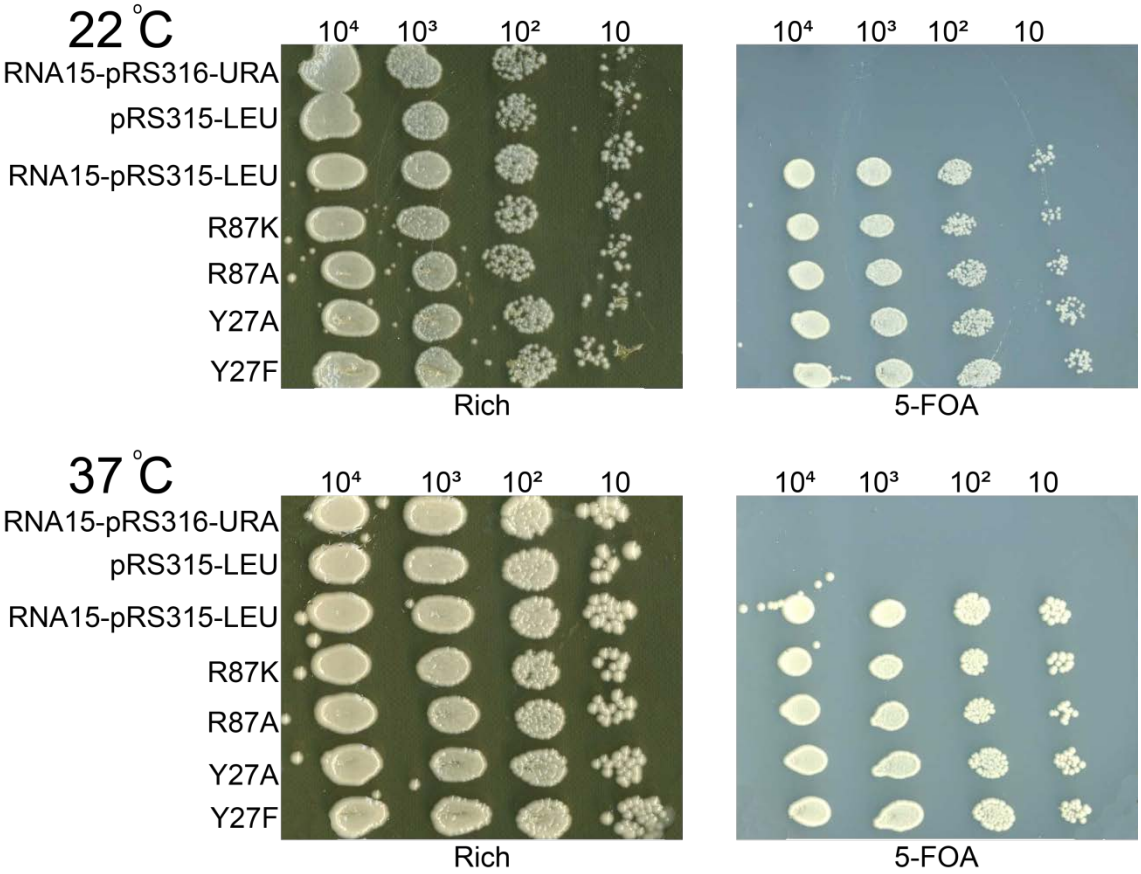


Figure 4.4 5-FOA selection assay of point mutations within the Rna15 RRM. 5-FOA plasmid shuffle assay to determine the effect of mutations within the RRM of Rna15 on phenotype of *S. cerevisiae.* Cells have been plated as "spots" on both rich and minimal media supplemented with 5-FOA. Each spots represent a serial dilution from 10^4-10^1. The top panels show yeast plated and held at a permissive temperature of 22°C whilst the bottom panels demonstrate yeast incubated at the higher temperature of 37°C in order to look for temperature sensitivity in the mutants. As observed in previous 5-FOA assays cells supported by either the RNA15-pRS316-URA construct or the empty pRS315 plasmid are not viable on 5-FOA. However, this phenotype is rescued upon transformation of the wildtype RNA15-pRS315-LEU construct at both 22ºC and 37ºC. Transformation of Rna15 RRM site I mutants results in no growth defect at both 22ºC and 37ºC upon comparison with wildtype.

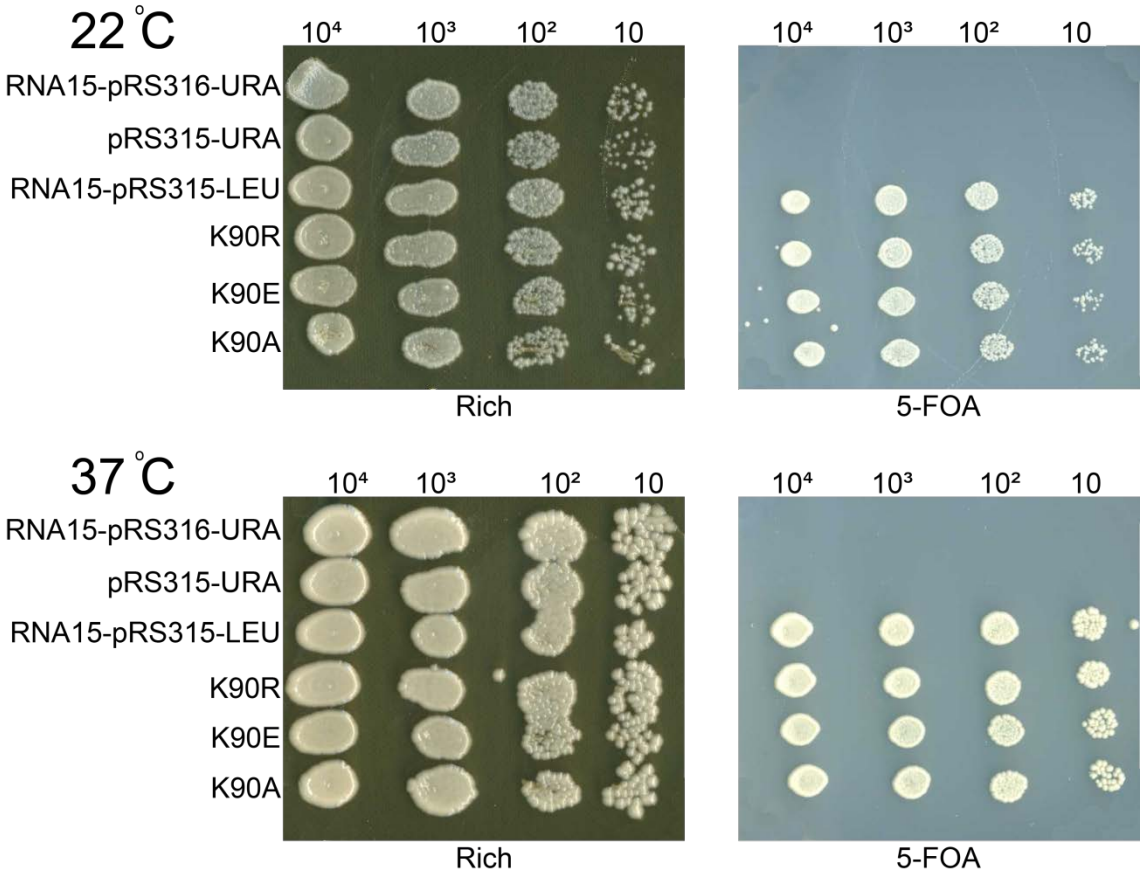


Figure 4.5 5-FOA selection assay demonstrating the effect of point mutations in the RRM of Rna15. Cells were plated as spots on both rich and minimal media containing 5-FOA. The top two panels represent incubation at 22ºC whilst the bottom two panels demonstrate the results of incubation at 37ºC. Spots represent a dilution series ranging from 10^4-10.

Previously published data demonstrated a defect in growth of a R87D *S. cerevisiae* mutant when grown on medium containing 1.5% formamide (140). Therefore, following 5-FOA selection, all mutants were grown in the presence of 1.5% formamide (v/v) in order to determine whether this effect could be replicated. As expected, the wild type Rna15-pRS315-LEU construct was able to grow on both rich and minimal media containing 1.5% formamide at 30ºC (Figure 4.6). However, wild type colonies grown on rich and minimal media containing 1.5% formamide are slightly smaller than those grown in the absence of 1.5% formamide. This is likely due to the increase in stress in growth conditions due to the presence of the formamide. Mutations R87K, R87A, Y27F, Y27A and S24T display growth comparable to wild type *S. cerevisiae* on both rich and minimal media containing 1.5% formamide at 30ºC. All other point mutations grew similarly to wild type. The temperature of incubation was raised to 40ºC as an additional stress condition in order to determine whether point mutations within the RRM displayed temperature sensitivity when grown on formamide (Figure 4.7). Growth of all constructs, wild type and mutant, were observed on rich medium and on minimal medium at 40ºC but colonies were smaller than observed in the previous growth assay (Figure 4.7) demonstrating a slight sensitivity of *S. cerevisiae* at 40ºC. Growth is severely affected on minimal media at 40ºC upon comparison with previous assays. A slight effect in growth is observed in the R87A mutant where no growth is observed in the last spot of the dilution series. This may represent a slight defect in growth in the R87A mutant at 40ºC on rich media. In the presence of 1.5% formamide, both wild type and mutant *S. cerevisiae* were not able to grow at 40ºC. Again, the same result was demonstrated for all Rna15 RRM point mutations.

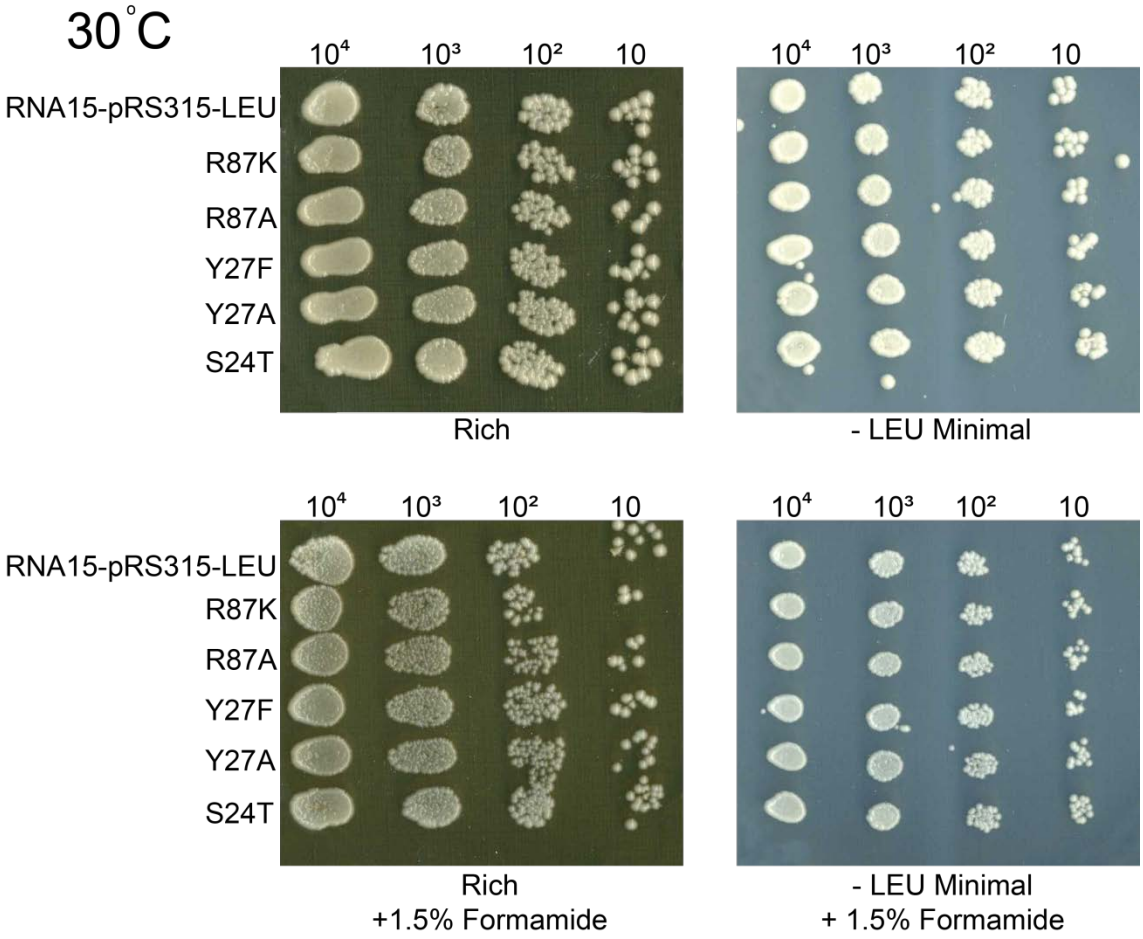


Figure 4.6 Growth of *S. cerevisiae* Rna15 RRM point mutations on minimal media contaning 1.5% formamide. Cells were plated as spots in a dilution series ranging from 10^4-10. The top two panels demonstrate the wild type Rna15-pRS315-LEU construct and Rna15 RRM site I point mutations incubated on rich and minimal media at 30ºC. The bottom two panels demonstrate wild type and mutant yeast grown on rich and minimal media both containing 1.5% formamide. Growth of both wild type and mutant *S. cerevisiae* is observed in all panels.

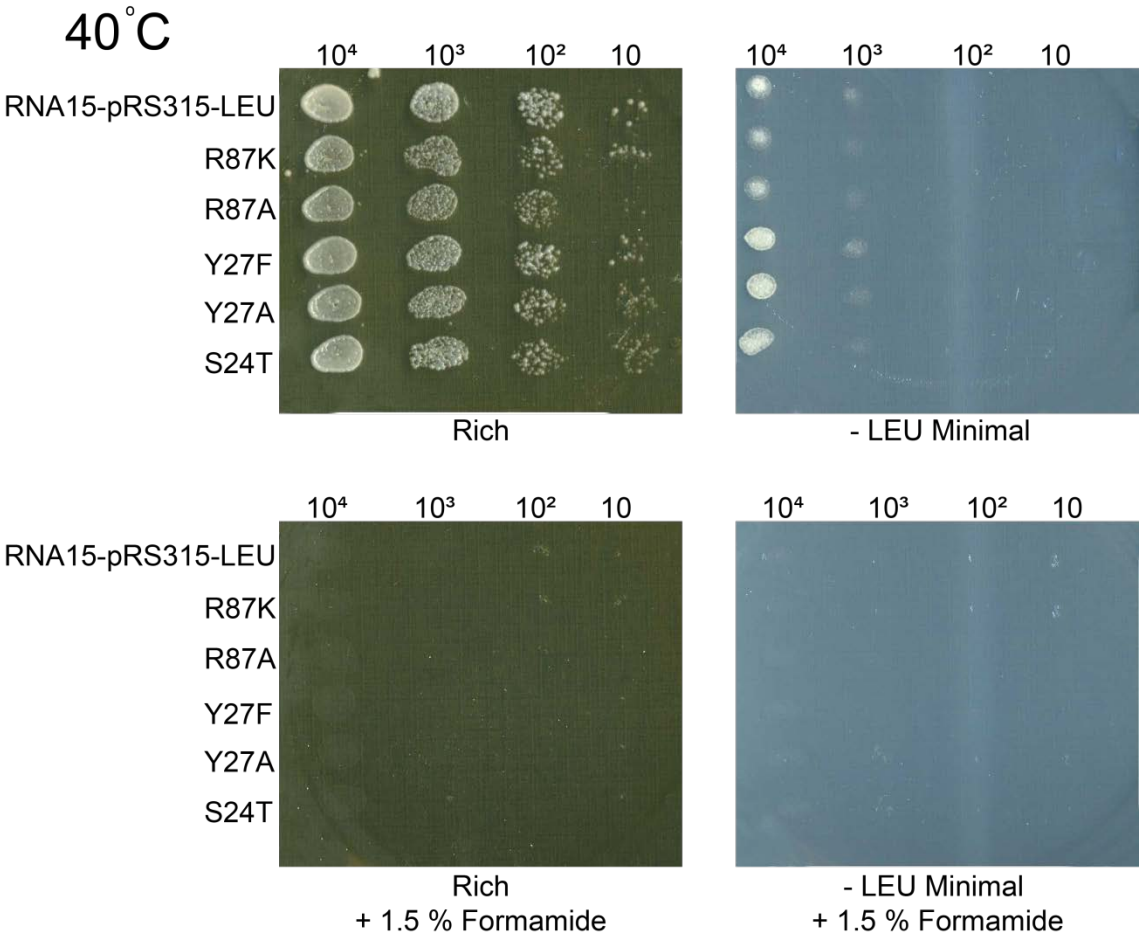


Figure 4.7 Growth assay to determine the effect of temperature and 1.5% formamide on growth of Rna15 RRM site I mutants. Cells are plated as spots in a dilution series ranging from 10^4-10. The top two panels demonstrate incubation of wild type Rna15 and site I mutants on rich and minimal media at 40ºC. The bottom two panels demonstrate growth of wild type and site I mutant *S. cerevisiae* on rich and minimal media containing 1.5% formamide at 40ºC. Both wild type and mutant *S. cerevisiae* are viable at 40ºC on rich and minimal media, however, both wild type and mutant *S. cerevisiae* display temperature sensitivity at 40ºC on 1.5% formamide.

The assay was repeated again at 37ºC with 1.5% formamide. Yeast containing the wild type Rna15-pRS315-LEU construct were able to grow on rich and minimal media containing 1.5% formamide at 37ºC (Figure 4.8). At 37ºC all site I point mutations grew on rich and minimal media as demonstrated in previous experiments. Growth was also comparable to wild type *S. cerevisiae* showing that the point mutations do not have any effect on growth phenotype. A slight decrease in growth is observed in the final spot of the dilution series corresponding to wild type *S. cerevisiae* on minimal media (top panels, Figure 4.8). Growth of both wild type and mutant *S. cerevisiae* is clearly affected by incubation at 37ºC on media containing 1.5% formamide (bottom panels, Figure 4.8). A slower growth is demonstrated by the smaller colonies shown in the bottom panels of Figure 4.8. This phenotype is demonstrated in both wild type and Rna15 RRM site I mutants demonstrating that under temperature stress, mutations in site I of the RRM that lead to an decrease in RNA binding affinity in *in vitro* experiments, do not have any effect on the growth phenotype of *S. cervisiae*.

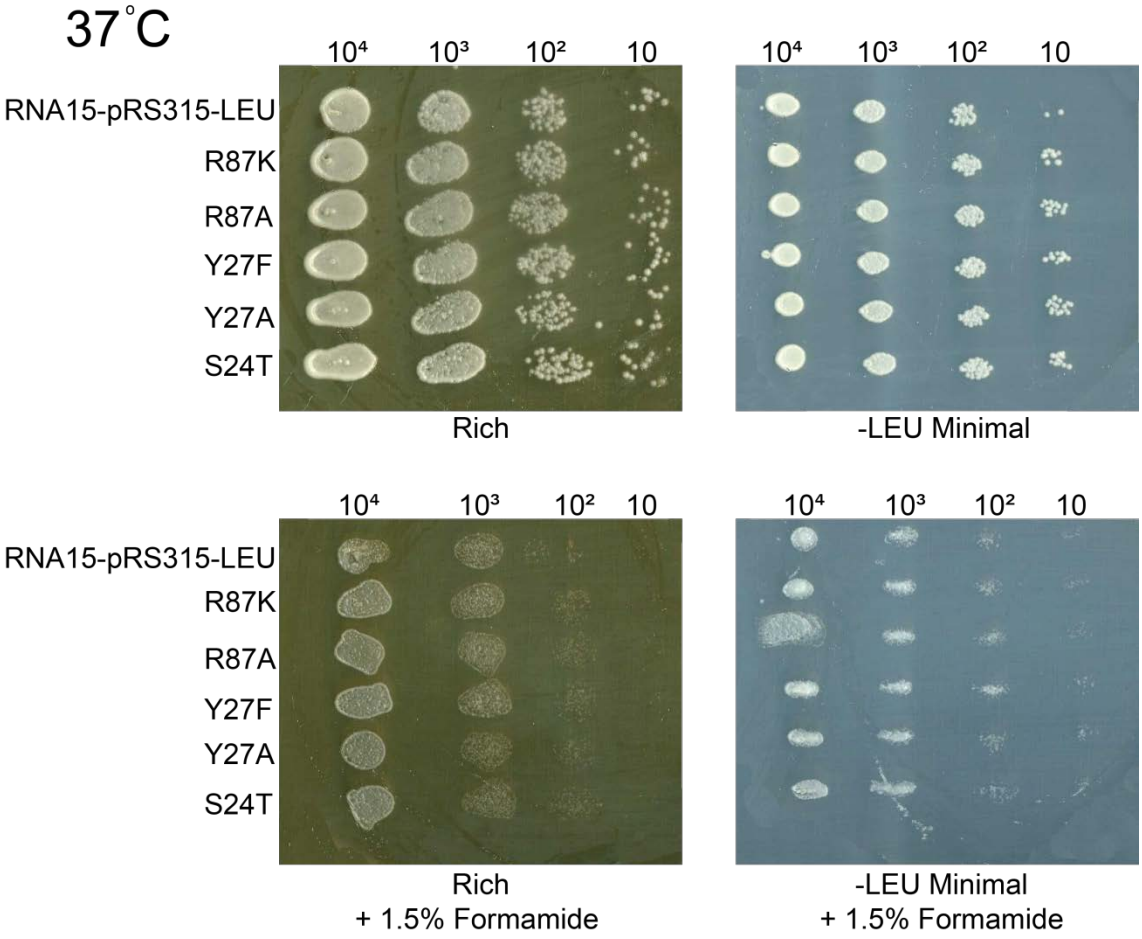


Figure 4.8 Growth of site I mutants on 1.5% formamide at 37ºC. Cells were plated as spots in a dilution series. The top two panels demonstrate wild type and Rna15 site I mutants grown on rich and minimal media at 37ºC. Growth of both wild type and mutant is observed. The bottom two panels demonstrate incubation on rich and minimal media both containing 1.5% formamide at 37ºC. Wild type and mutant *S. cerevisiae* show affected growth on 1.5% formamide upon comparison to the top two panels.

4.4 Deletion of the RRM is lethal on 5-FOA at 37°C

Point mutations in the RRM of Rna15 failed to produce an effect on growth phenotype of *S. cerevisiae* despite differences in measured RNA binding. In order to maximise the chances of seeing an effect further mutagenesis was undertaken combining the mutations Y27A, R87A and K90E. These were selected as they resulted in the largest decreases in RNA binding affinity. Two double mutations, Y27A/K90E and R87A/K90E alongside a triple mutation, Y27A/R87A/K90E were prepared. In addition to the double and triple mutations the RRM of Rna15 was also deleted from the RNA15-pRS315-LEU construct.

The double, triple and deletion mutants were transformed into the appropriate haploid *S. cerevisiae* strain as previously discussed. Selection for the mutant constructs was achieved by the 5-FOA assay. Following selection, the mutants were plated again to establish the growth phenotype in comparison with wild type. The double, Y27A/K90E, R87A/K90E and triple Y27A/R87A/K90E mutants were viable on both rich and minimal media supplemented with 5-FOA with growth comparable to wild type at 22°C (Figure 4.9). The Rna15 RRM deletion displayed a very slight decrease in growth upon comparison with wild type on 5-FOA at 22°C (Figure 4.9).

At 37ºC, the Y27A/K90E and R87A/K90E double mutants and the triple mutant, Y27A/R87A/K90E do not demonstrate temperature sensitivity on rich media. However, on minimal media supplemented with 5-FOA a decrease in growth at 37ºC is observed in the 10^4 and 10^3 serial dilutions upon comparison with the wild type (Figure 4.9). Deletion of the RRM in Rna15 results in a decrease in growth in *S. cerevisiae* on rich media at 37ºC. On minimal media supplemented with 5-FOA at 37ºC deletion of the RRM is lethal (Figure 4.9).

In order to determine whether the addition of 5-FOA to the minimal medium was responsible for the non-viable phenotype displayed by the Rna15 RRM deletion, the Rna15Δ16-94-pRS315-LEU strain was incubated on a variety of different media at 37ºC (Figure 4.10). The deletion of the RRM does not result in cell death on other media as demonstrated on 5-FOA. However, growth is inhibited on both rich and minimal media upon comparison with wild type at 37ºC. Incubation of the deletion mutant on rich and minimal media supplemented with 1.5% formamide demonstrates a severe decrease in growth of *S. cerevisiae* upon comparison with wild type. Therefore although the deletion of the RRM does not appear to be lethal when 5-FOA is absent it severely inhibits growth.

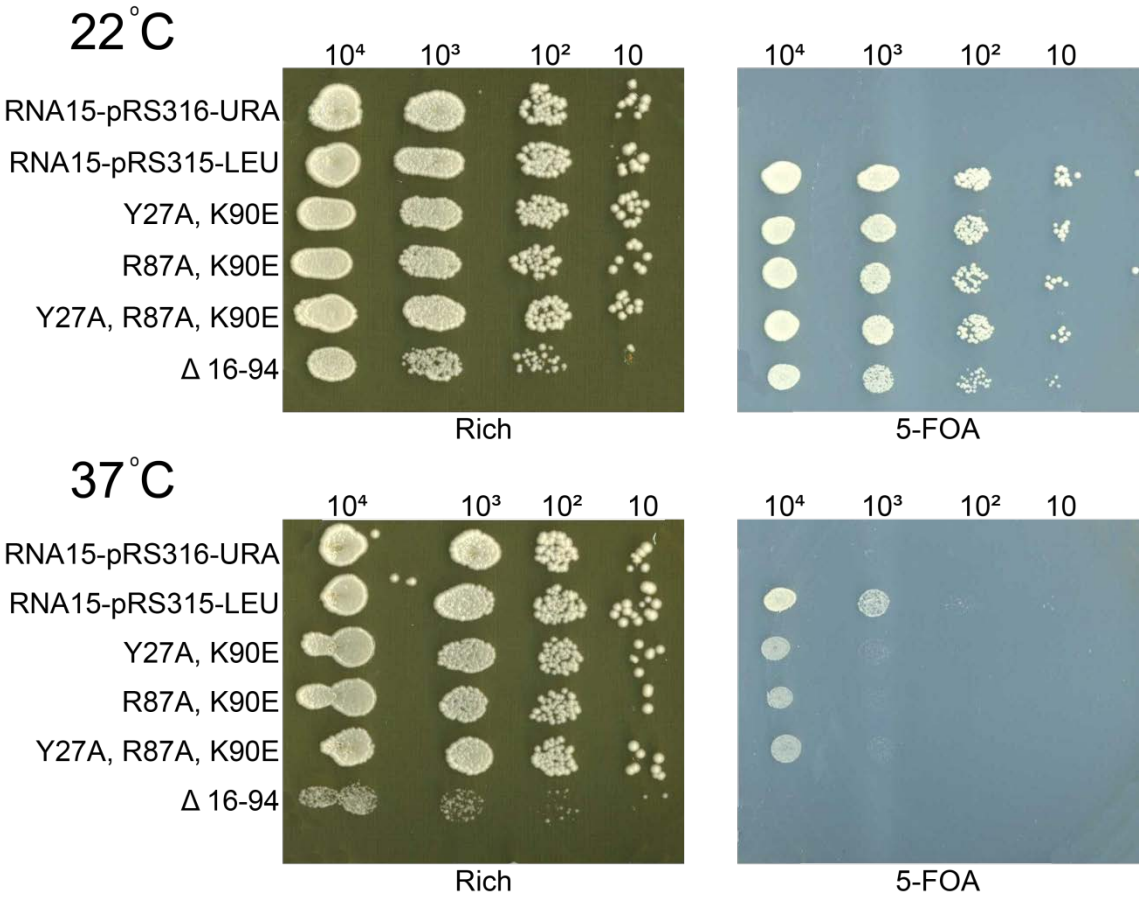


Figure 4.9 5-FOA assay of Rna15 double, triple and deletion mutants. Cells were plated as spots in a dilution series ranging from 10^4-10. The top two panels demonstrate growth of wild type, double, triple and deletion mutants on rich and minimal media containing 5-FOA at 22ºC. The bottom two panels demonstrate incubation at 37ºC.

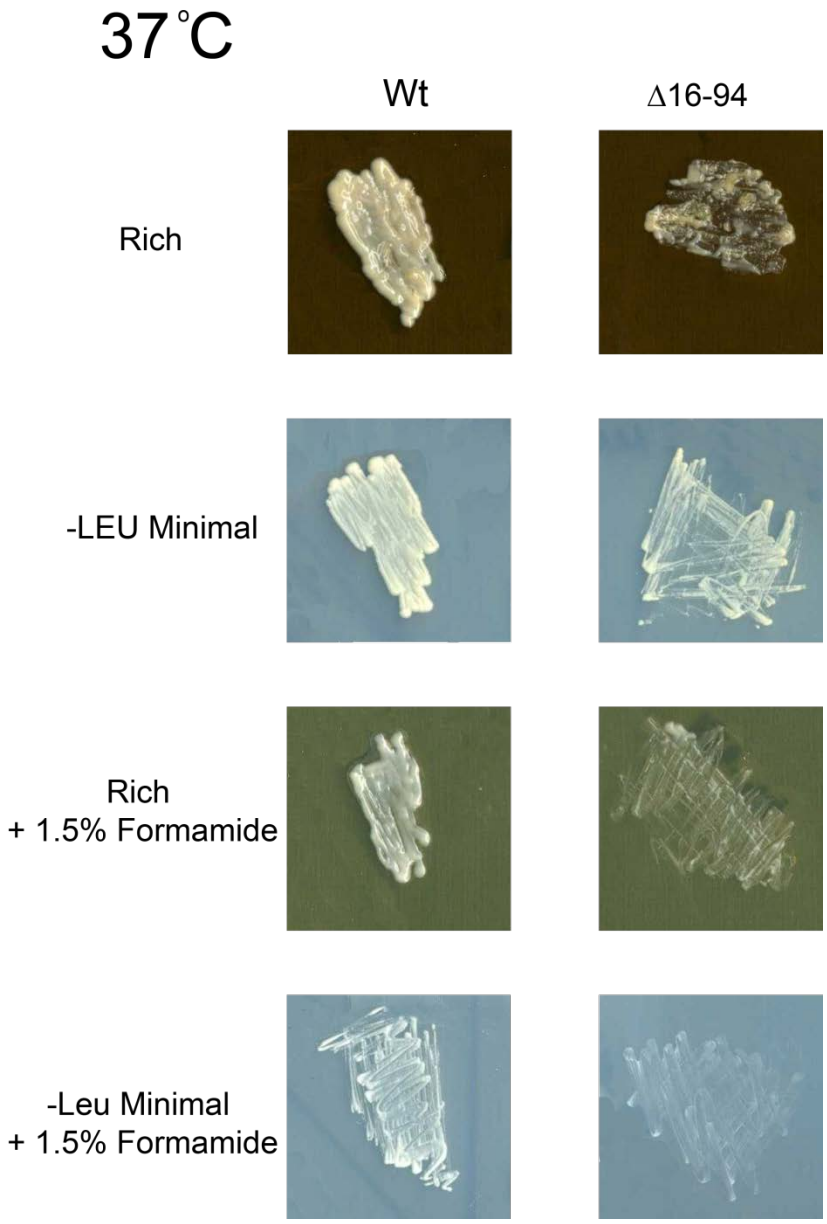


Figure 4.10 Analysis of deletion of the RRM on growth phenotype of *S. cerevisiae.* Deletion of the RRM resulted in inhibited growth of *S. cerevisiae* in all conditions upon comparison with wild type but did not result in lethality as observed on minimal media supplemented with 5-FOA.

In order to determine whether extensive mutation of the RRM had any effect on the growth rate of *S. cerevisiae*, a growth rate experiment was performed. The growth of the Y27A/K90E and R87A/K90E double mutants alongside the triple Y27A/R87A/K90E and the RRM (Δ16-94) deletion mutant were monitored over a 11 hour period in rich media at 30ºC. Triplicate cultures inoculated with the wildtype and four mutants were incubated overnight at 30ºC, shaking 200rpm. The following morning, cells were counted using a haemocytometer and each culture was diluted to $1x10^6$ cells/ml. Cultures were incubated at 30ºC with shaking at 200rpm and the growth monitored over an 11 hour period. Growth of the Y27A/K90E and Y27A/R87A/K90E mutants (Figure 4.11, A) demonstrates a slight decrease in growth rate upon comparison with wild type. The growth rate of the R87A/K90E mutant, however, is comparable to wild type (Figure 4.11, B). These results demonstrate that double and triple mutation of the RRM fails to severely affect growth rate. Deletion of the RRM (residues 16-94) results in a severe decrease of growth rate (Figure 4.11, B). As cells were counted on a haemocytometer, there is a possibility that not all the cells counted are viable. Therefore, a viable cell count was also performed. Cells of known concentration were plated onto minimal media and incubated at 30ºC for 2-3 days. Following incubation, colonies of wild type, the Y27A, R87A, K90E and Δ16-94 mutants were counted and recorded. The Y27A, R87A, K90E triple mutant grew similarly to wild type (Figure 4.11, C). However, growth of the deletion mutant was severely impaired (Figure 4.11, C). Around a 7-fold decrease in cell number is observed in the deletion mutant demonstrating that loss of the RRM greatly impairs growth of *S. cerevisiae*.

Despite the large decrease in RNA binding affinites exhibited by both the Y27A and R87A mutations and the lack of binding by K90E demonstrated during *in vitro* experiments, they demonstrate very little effect on cell growth (Table 4.2). Deletion of the RRM had a more severe effect with reduced growth in rich media at 30ºC and

inviability at 37ºC on 5-FOA (Table 4.2). It is likely that this mutant has a defect in polyadenylation. Also, although no effect on phenotype is observed with the triple, double and point Rna15 RRM mutations it is still possible that polyadenylation is altered with respect to wild type. In addition due to the tightly coupled nature of transcription and polyadenylation it is likely that these mutations will also affect gene expression levels. The next sections aim to characterise the effect of these mutations on polyadenylation and transcription.

Rna15 Construct	RNA binding affinity (μM)	Viability of *S. cerevisiae*	Growth rate assays
Wild type	6.4 ± 0.5	Viable	-
R87A	43.6 ± 10.8	Viable	-
Y27A	76.7 ± 16	Viable	-
K90E	X	Viable	-
Y27A, K90E	-	Viable	Slight decrease in growth rate
R87A, K90E	-	Viable	Growth comparable to wild type
Y27A, R87A, K90E	-	Viable	Slight decrease in growth rate
Δ 16-94 (RRM)	-	Lethal at 37ºC on 5-FOA	Growth severely inhibited

Table 4.2 Summary of Rna15 wild type and mutant RNA binding affinities, viability and growth rate assays in *S. cerevisiae*. The - symbol denotes cases where the experiment was not carried out. The X symbol represents immeasurable RNA binding by fluorescence spectroscopy in the Rna15 K90E mutant.

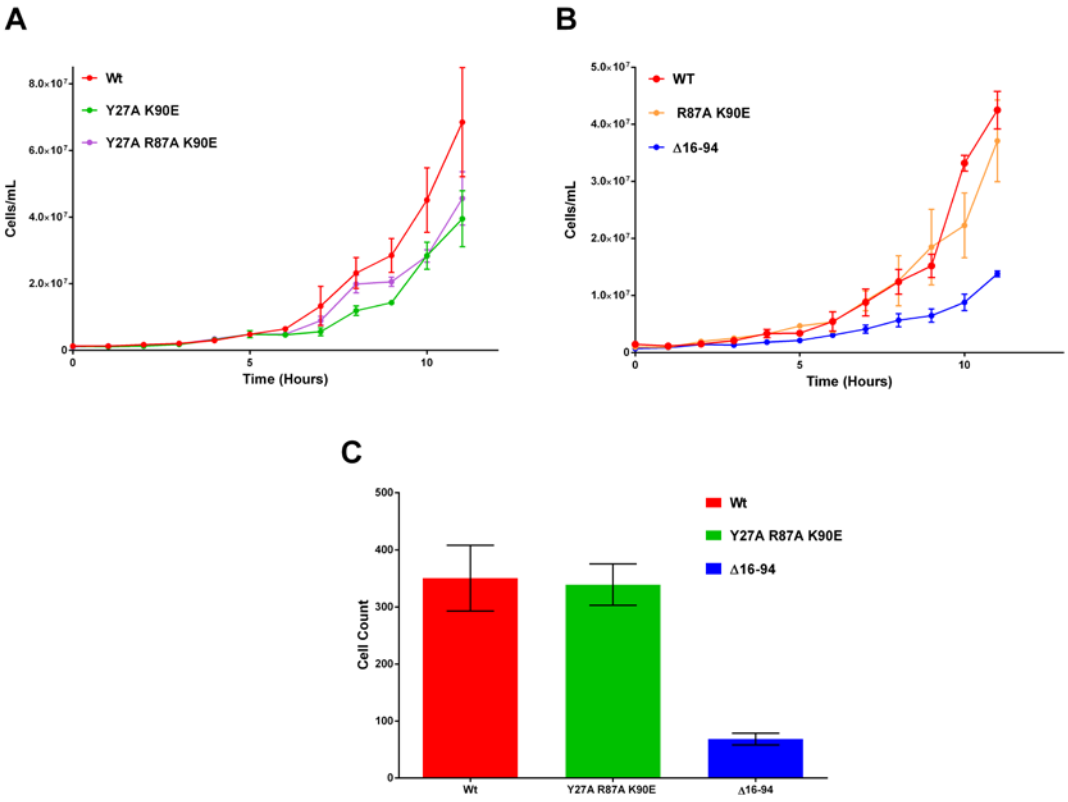


Figure 4.11 Growth of the *S. cerevisiae* mutants. A. Growth curves of wild type, Y27A, K90E and Y27A, R87A, K90E mutants. Growth was monitored over an 11 hour period and cells were counted every hour using a haemocytometer. **B.** Growth curves demonstrating growth of wild type *S. cerevisiae*, the R87A, K90E and RRM deletion mutants. **C.** Viable cell count of the triple mutant and the deletion mutant.

4.5 Analysis of transcription in Rna15 mutants

The effect of deletion of the RRM on expression of a subset of genes was studied using qRT-PCR. Total mRNA was purified from both wild type *S. cerevisiae* and the rna15 Δ16-94 deletion mutant. The aim of this experiment was to quantify the amount of ACT1, TDH2, YPT1, CYC1 and ADH1 mRNA relative to wild type. Transcription of all genes examined was affected in the *rna15 Δ16-94* mutant (Figure 4.12). Since transcription of genes by RNA polymerase II is tightly linked to polyadenylation, raw data need to be normalised to data unbiased by the presence of mutations in RNA15 to provide a more accurate experimental value. Therefore, measured values were normalised to the amount of mRNA produced by the SNR52 gene. This encodes for a small nucleolar RNA and is transcribed by RNA polymerase III and therefore is not polyadenylated. Where 100% of mRNA represents wild type levels, all genes either increased or decreased as a result of deletion of the RRM. Expression of ACT1 decreased by almost 20% relative to wild type whilst YPT1 decreased 11%. A more significant decrease in expression levels is observed in genes, TDH2 and ADH1. Amount of TDH2 mRNA decreased by 40% relative to wild type and expression of ADH1 decreased by 32%. Levels of mRNA corresponding to CYC1 seem to increase by 17% signalling an apparent increase in expression of this gene. As previously stated the RRM of Rna15 is required for selection of G/U rich sequences. Therefore, it is possible that these changes in amount of mRNA are due to usage of the distal polyadenylation sites rather than the proximal polyadenylation sites. In order to determine whether this was the case, RNA sequencing was used to generate a transcription library to assess globally the effect deletion of the RRM has on transcription and poly(A) site selection.

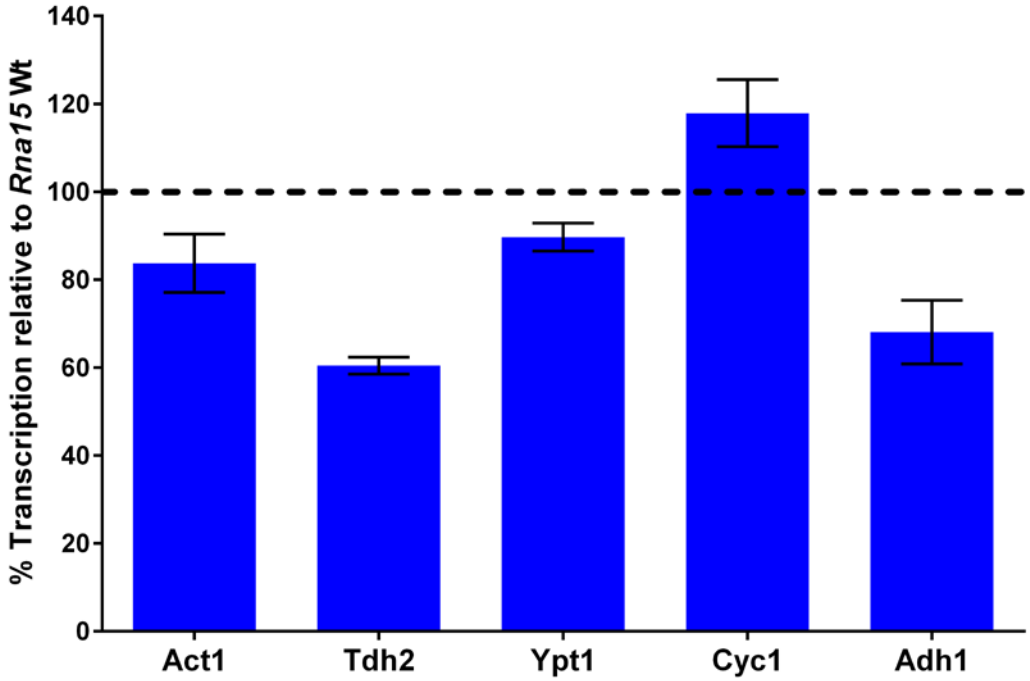


Figure 4.12 Transcription of a subset of genes in the *rna15 Δ16-94* mutant. Transcription of each gene is relative to wild type where % of transcription of wild type is equal to 100%. RNA was quantitated three times in each qRT-PCR plate and the entire experiment was also repeated in triplicate. Raw data was normalised to amount of RNA polymerase III transcribed SNR52 snoRNA.

4.6 Preparation of libraries for RNA sequencing

Following 5-FOA selection, wild type, Δ16-94, Y27A and Y27A/R87A/K90E mutant strains of *S. cerevisiae* were grown at 30ºC in rich media until reaching a cell density of 1 X 10^7 cells/ml. Cells were pelleted and the total mRNA was purified. Ribosomal RNA (rRNA) was removed using the RiboZero™ Magnetic Gold (Yeast) kit following manufacturer's instructions (epicentre, illumina). Following removal of rRNA, an RNA sequencing library was generated using the ScriptSeq™ v2 RNA-Seq Library Preparation Kit following the maufacturer's instructions (epicentre, illumine). The kit works by fragmenting purified RNA samples and creating cDNA using a randomised hexamer with a tagging sequence. The short RNA sequences produced by fragmenting inhibits contamination of neighbouring colonies during PCR amplification on the flow cell surface. This tagging sequence acts to tag the 5' end of the resulting cDNA. Tagging of the cDNA at the 3' end requires the tag sequence fused to a randomised sequence with a dideoxy nucleotide present at the 3' end. PCR results in generation of the tag by amplifying in the antisense direction. Generation of the second strand is achieved by use of primers that pair to the tags present at the 5' and 3' ends. A barcode sequence can be included that is unique to mRNA purified from different samples (i.e. wild-type and mutant) (Figure 4.13).

Following library preparation, sequencing was carried out using the Illumina platform. The method works by attaching the cDNA fragments to a chip surface. A bridging PCR reaction is carried out generating "colonies" of identical cDNA sequence (Figure 4.14). The sequencing reaction then proceeds whereby all four dye labelled reversible terminator nucleotides are added along with primers and DNA polymerase enzyme. The sequencing reaction reads a single nucleotide at a time by laser excitation which captures the emitted fluorescence from each colony. The identity of the first base

is recorded and following removal of the dye, the second cycle can begin on the next nucleotide. Again, all four labelled reversible terminators, along with DNA polymerase, are added to the flow cell. Laser excitation results in the fluorescence emission of the next nucleotide along. The signal emitted from every colony is recorded and the nucleotide identified. This is repeated until ~80 bases have been sequenced in a given fragment (Figure 4.15). The resulting data can then be collated for analysis.

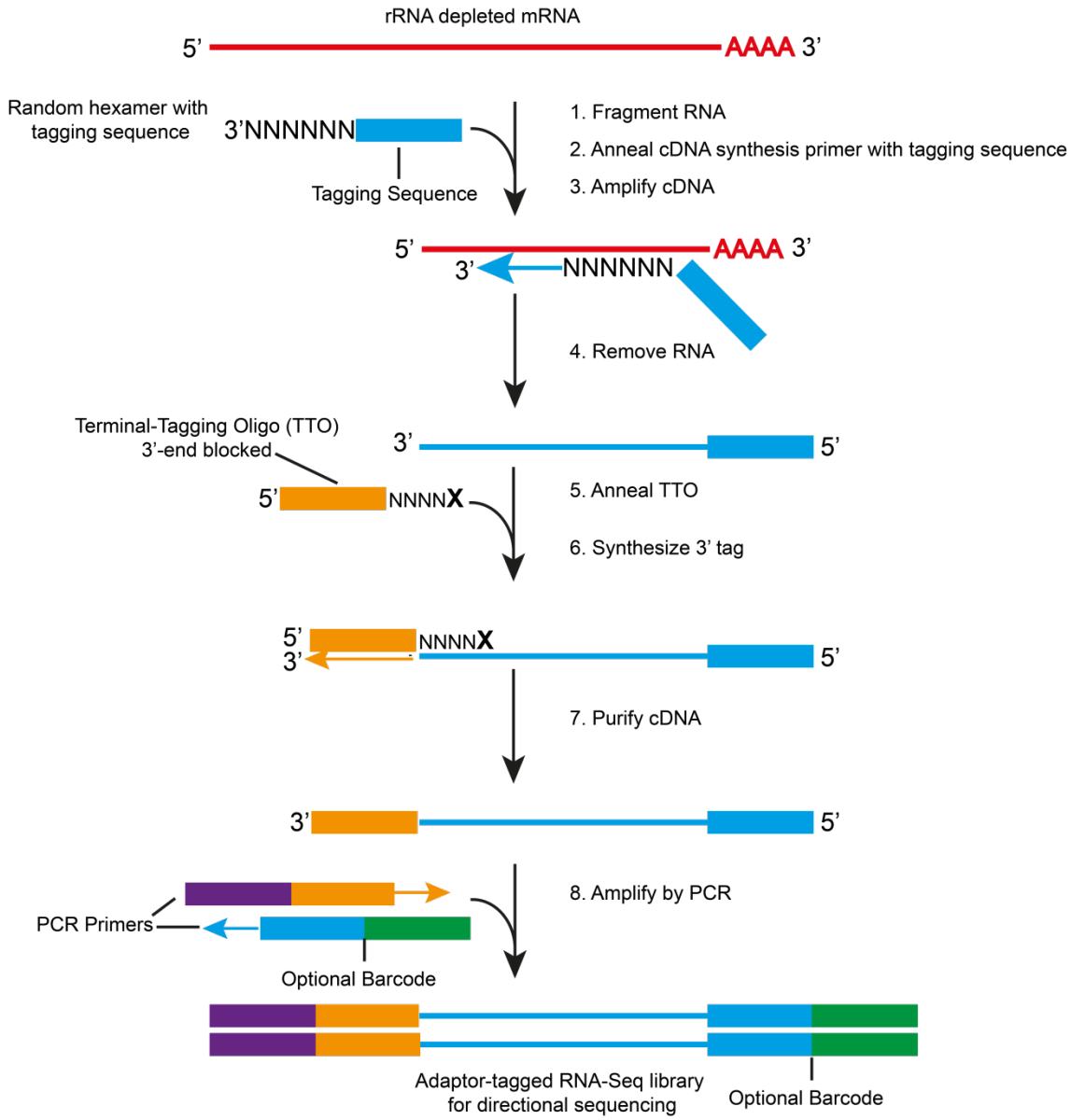


Figure 4.13 ScriptSeq™ v2 RNA-Seq Library Preparation Kit Protocol.

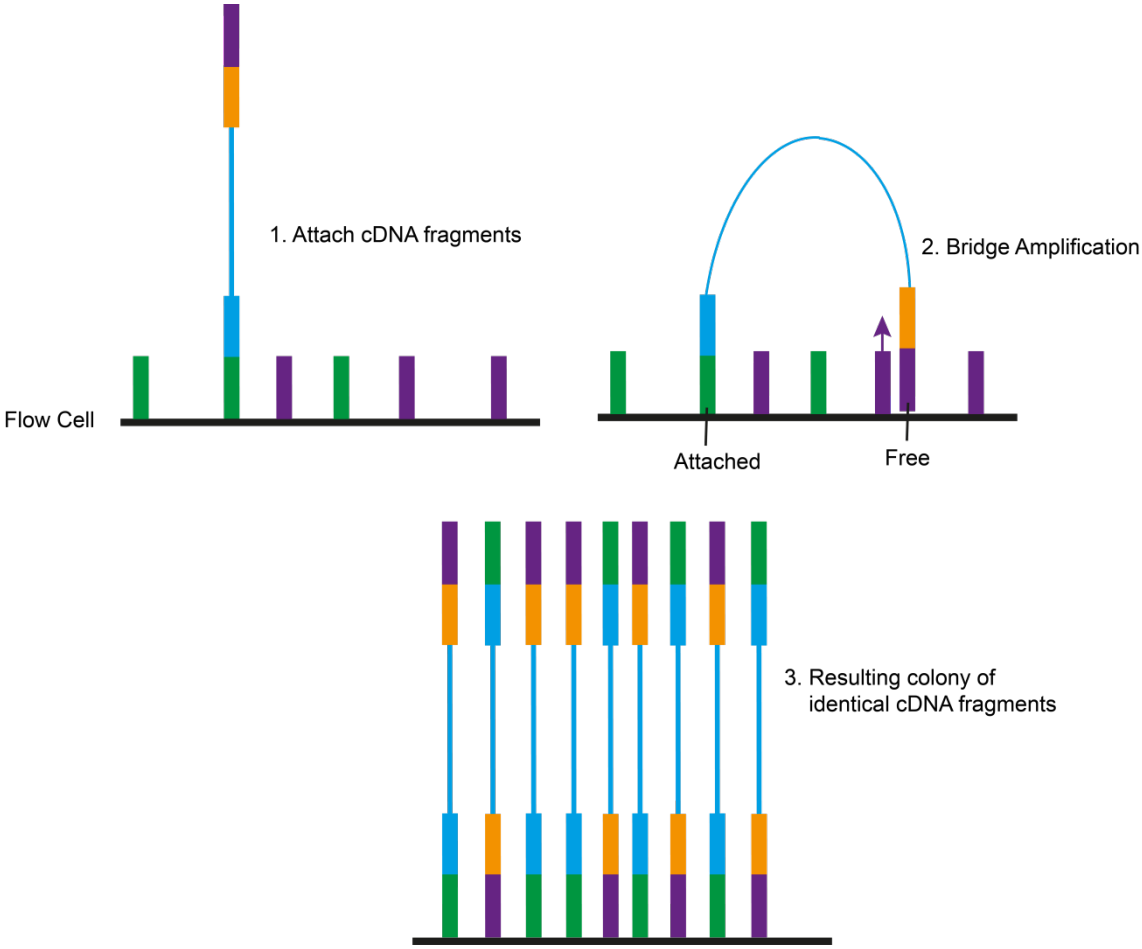


Figure 4.14 Formation of cDNA colonies on a flow cell. The resulting cDNA fragments generated using the ScriptSeq™ v2 RNA-Seq Library Preparation Kit are attached to the flow cell. Bridging amplification results in formation of a "colony" where all cDNA fragments present are of identical sequence. The above schematic demonstrates only one colony. Typically, a flow cell will contain many colonies, each representative of different cDNA fragments.

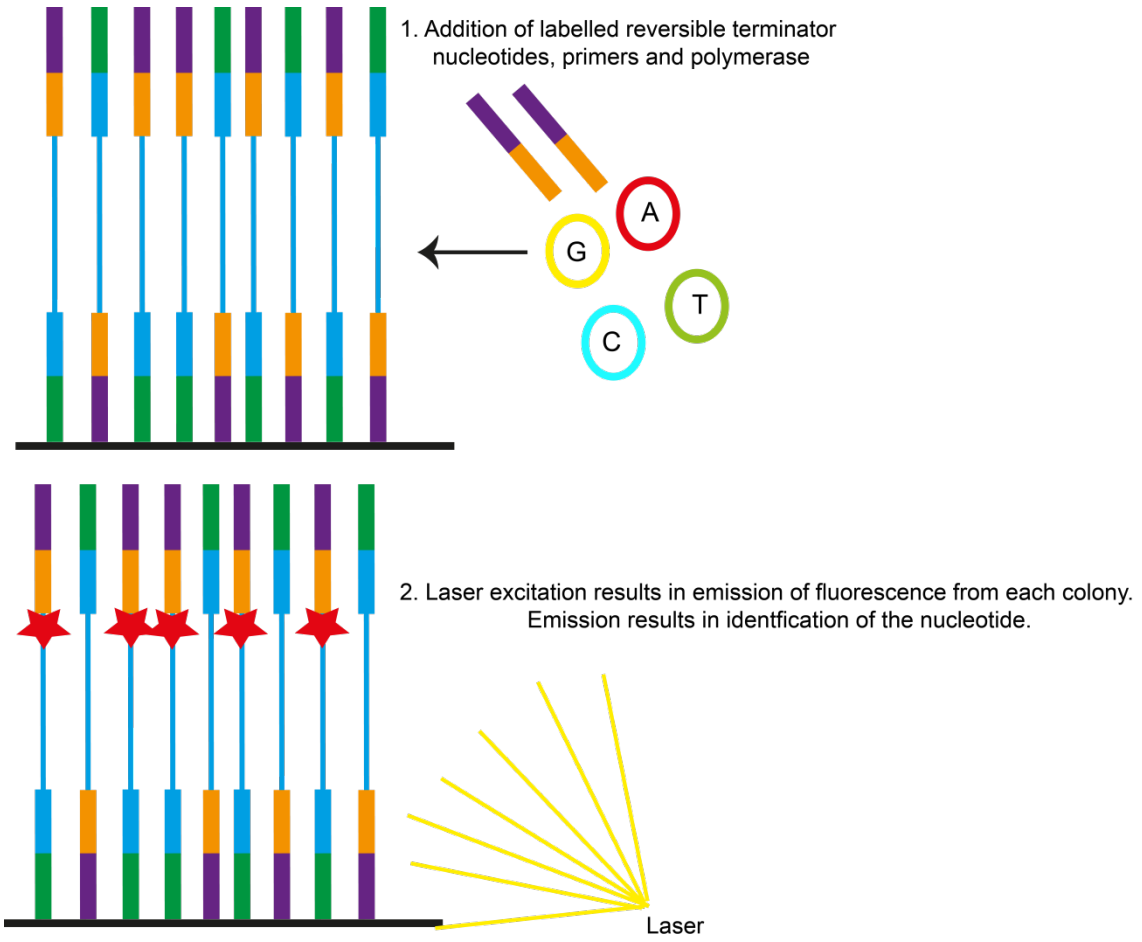


Figure 4.15 Illumina sequening of cDNA colonies. The schematic presented is representative of one colony present on a flow cell. Addition of labelled reversible terminator nucleotides, primers and DNA polymerase is required for the sequencing reaction. Each nucleotide is labelled differently resulting in different emission colour following excitation. Excitation results in fluorescence emission of labelled nucleotides present in each colony. The emission is recorded and corresponds to a labelled nucleotide. This process is repeated until every nucleotide is recorded within a fragment generating the cDNA sequence.

4.7 Analysing RNA sequencing data

Data analysis was split into two main areas. A global analysis of expression was undertaken by Nickolay Nickolov (N.I.M.R). Separately, processing activity within the 3' UTR of 40 genes was investigated. These genes were separated into classes based on their cellular function, i.e metabolism, cell cycle, cell wall maintenance and translation. In addition, poly(A) site readthrough was analysed in genes ACT1, TDH2, ADH1 and YPT1 where the position of poly(A) sites within the 3'UTR are known.

4.7.1 Analysis of expression of genes using RNA sequencing

A global analysis revealed that expression of 131 genes was significantly affected by deletion of the RRM of Rna15. Under the same parameters of analysis 47 genes are affected in the Y27A/R87A/K90E mutant while no genes where shown to be affected in the Y27A mutant. However, it is important to note that stringent statistical analysis was performed on each mutant (FDR value <0.05) and it is possible that a larger number of genes in each mutant are affected by deletion of the RRM. In the Δ16-94 mutant the 131 genes were grouped according to their role in various cellular processes in order to determine whether a trend in expression emerged specific to gene role. 39 genes involved in various metabolic processes within the cell were affected by deletion of the RRM (Figure 4.16). This represents 4.3% of the total number of genes within the *S. cerevisiae* genome that have specific metabolic roles. Deletion of the RRM appears to affect a larger proportion of metabolic genes than is observed in other groups. The effect on expression displays no obvious trend where an increase in expression is observed in 20 and decrease in expression is observed in 19. A high proportion of genes involved in maintaining the integrity of the cell wall are also affected by deletion of the RRM (Figure 4.17, A). Expression of 4% of the total amount of genes present in *S. cerevisiae* that encode for products involved in maintaining cell

wall integrity are affected in the mutant. The majority of genes in this group display an increase in expression in the mutant while only 2 of the 7 demonstrate a decrease in expression. Only 2.8% of total genes involved in the transport of molecules in the cell are altered by deletion of the RRM (Figure 4.17, B). Deletion of the RRM appears to lead to decrease in expression in the majority of genes involved in molecular transport and an increase is expression is observed in only 2 of the 12 genes affected. Genes that encode for products required for RNA and protein modifications display differences in expression in the mutant (Figure 4.18). Around 1.3% of the total number of genes required for RNA modifications are affected by deletion of the RRM (Figure 4.18, A). In general, a decrease in expression is observed in the majority of genes in this group. Expression of 1.8% of total genes involved in mediating protein modifications are influenced significantly by deletion of the RRM of Rna15 (Figure 4.18, B). A less obvious trend is observed in this group where expression of 4 genes increases while 6 decrease. Expression of 2.5% of genes induced in response to stress factors is also affected by deletion of the RRM (Figure 4.19, A). As cells were grown in rich media at 30°C it is not surprising that these genes are expressed at a relatively low level upon comparison with other genes as demonstrated by the normalised read count value for each gene. However, expression of 3 of these genes is shown to increase while a decrease in expression is demonstrated in 4. In general, differences in expression are demonstrated in genes that cannot be assigned to previous groupings but are involved in various other cellular processes (Figure 4.19, B). In addition, expression of genes that encode proteins with no known function are also shown to be affected by deletion of the RRM (Figure 4.20).

In conclusion, this analysis reveals that deletion of the RRM of Rna15 has significant effects on the transcription of 131 genes. A larger proportion of genes are

affected in processes where expression rate is generally higher upon comparison with other groupings (i.e. cell wall and metabolism). However, no trend is observed in expression in each gene in each group when wild type and mutant are compared. Fluctuations in expression appear random from gene to gene however a significant difference in expression of each is observed in the mutant. However, it is important to note that observed decreases in expression in the rna15 Δ16-94 mutant could be due to decreased efficiency in polyadenylation leading to unstable transcripts that are readily degraded. In this instance, observed decreases in mRNA abundance are an indirect product of decreased stability due to inefficient 3' end processing.

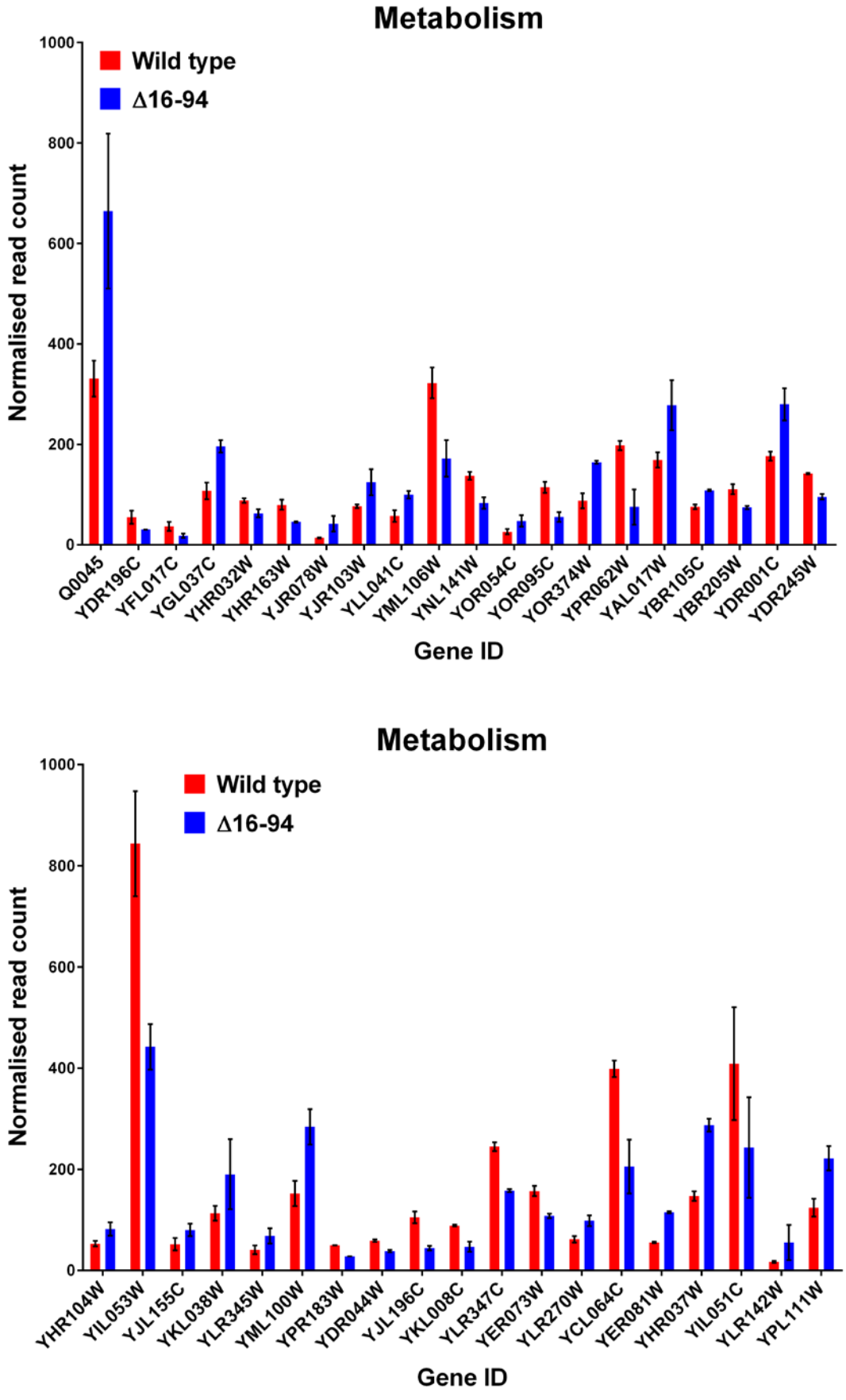


Figure 4.16 Expression of genes involved in metabolism.

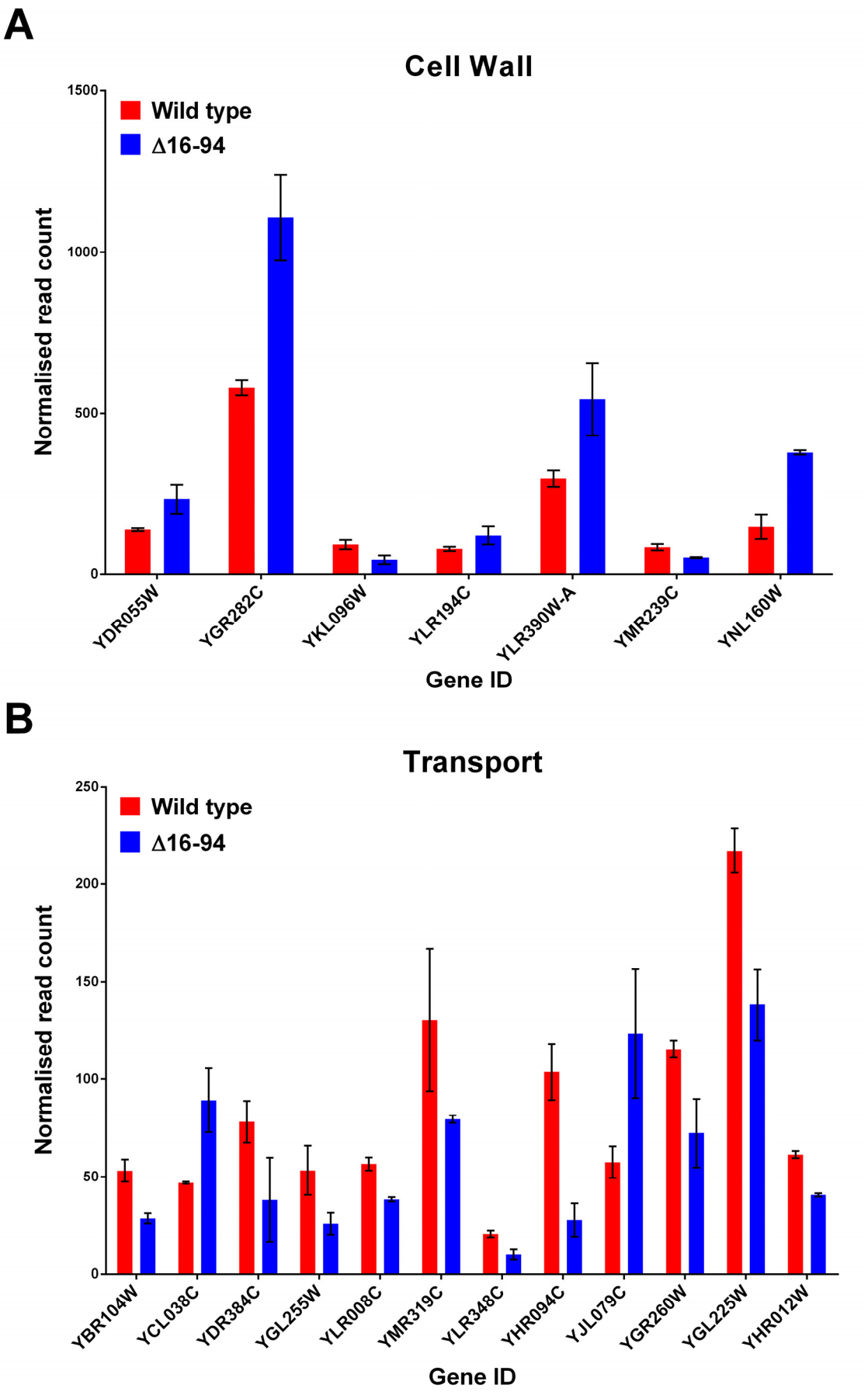


Figure 4.17 Expression of genes involved in maintaining the cell wall (A) and in transport of molecules (B).

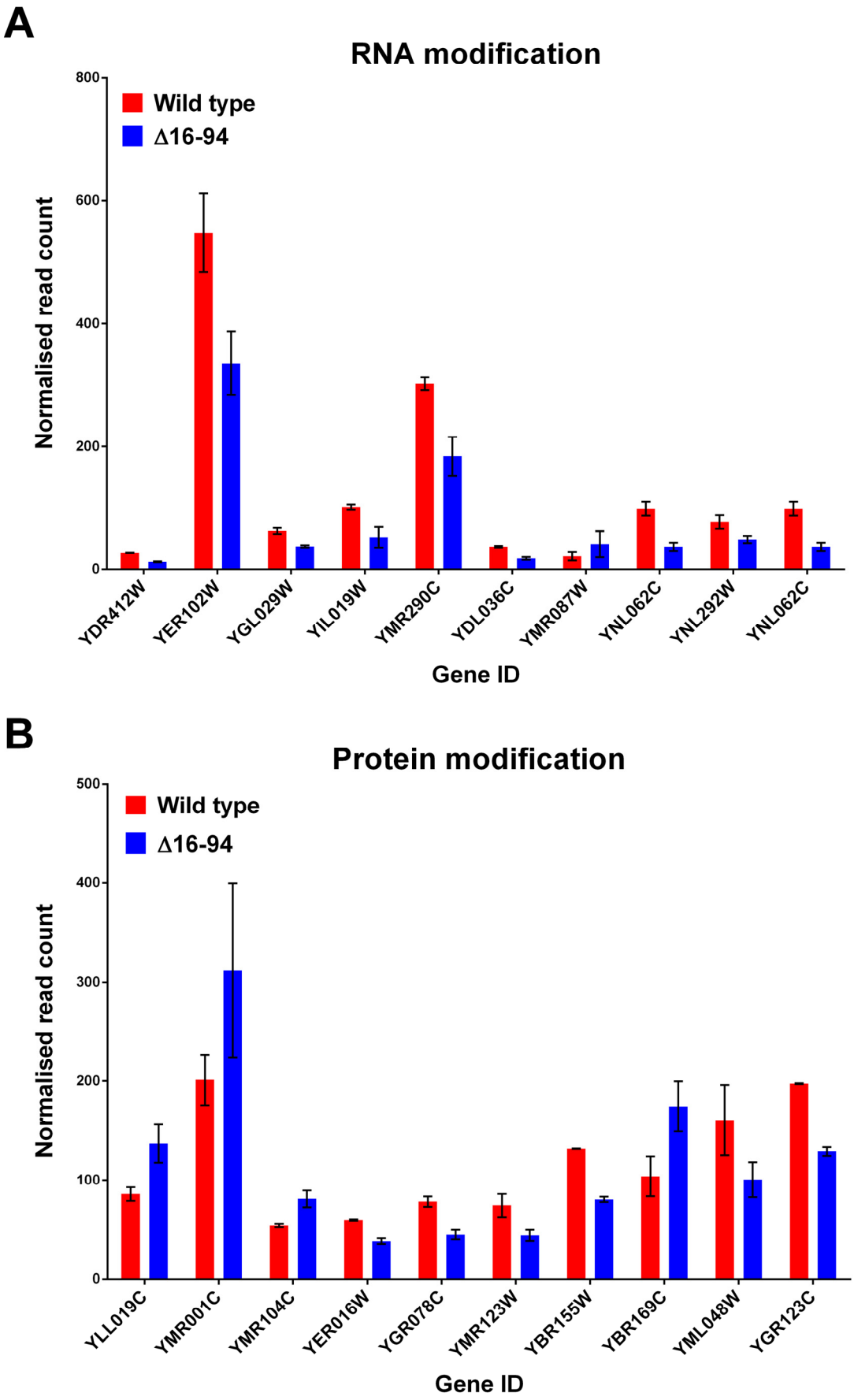


Figure 4.18 Expression of genes involved in modifying RNA molecules (A) and in mediating protein modification (B).

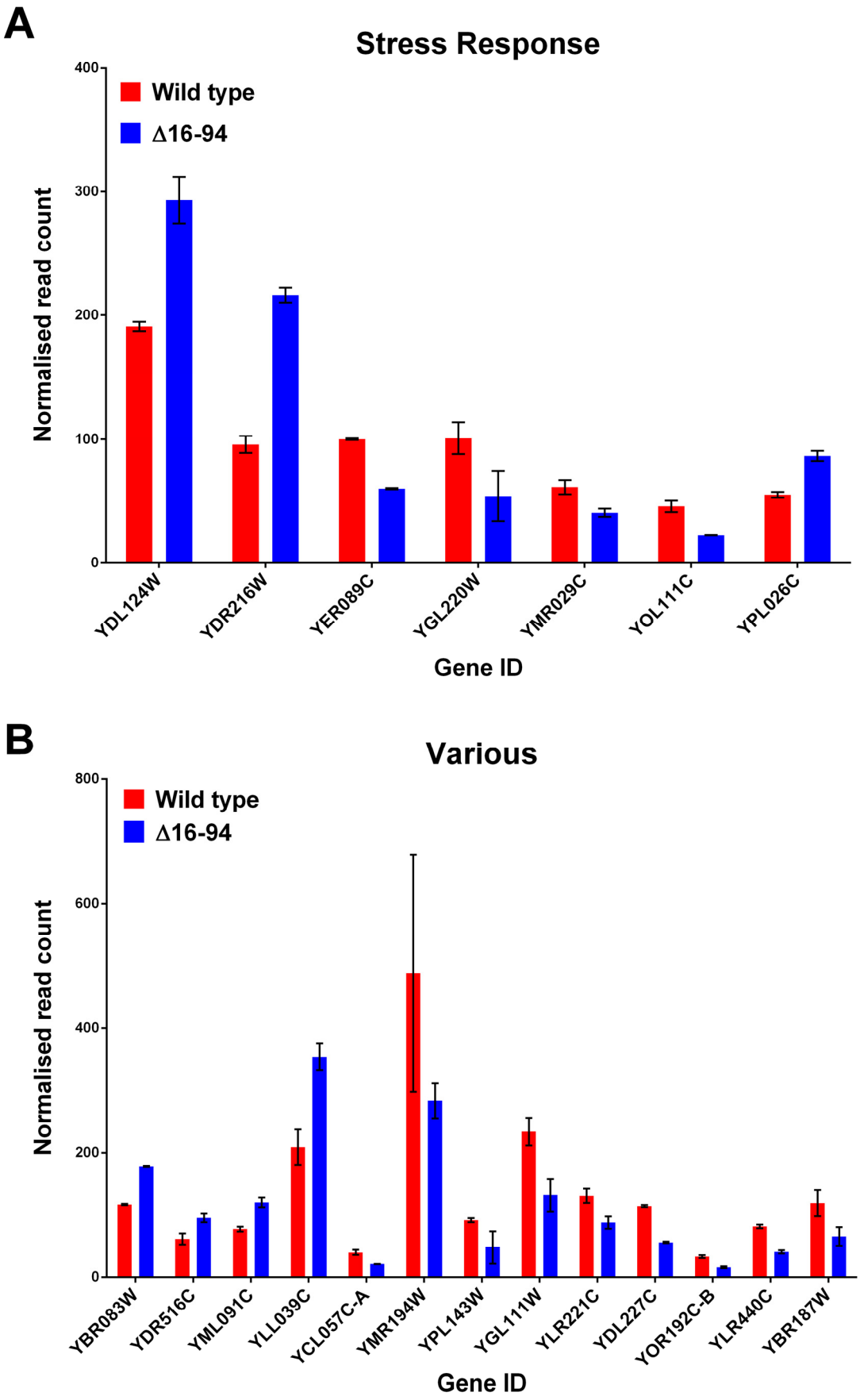


Figure 4.19 Expression of genes induced by stress factors **(A)** and required for numerous cellular processes **(B)**.

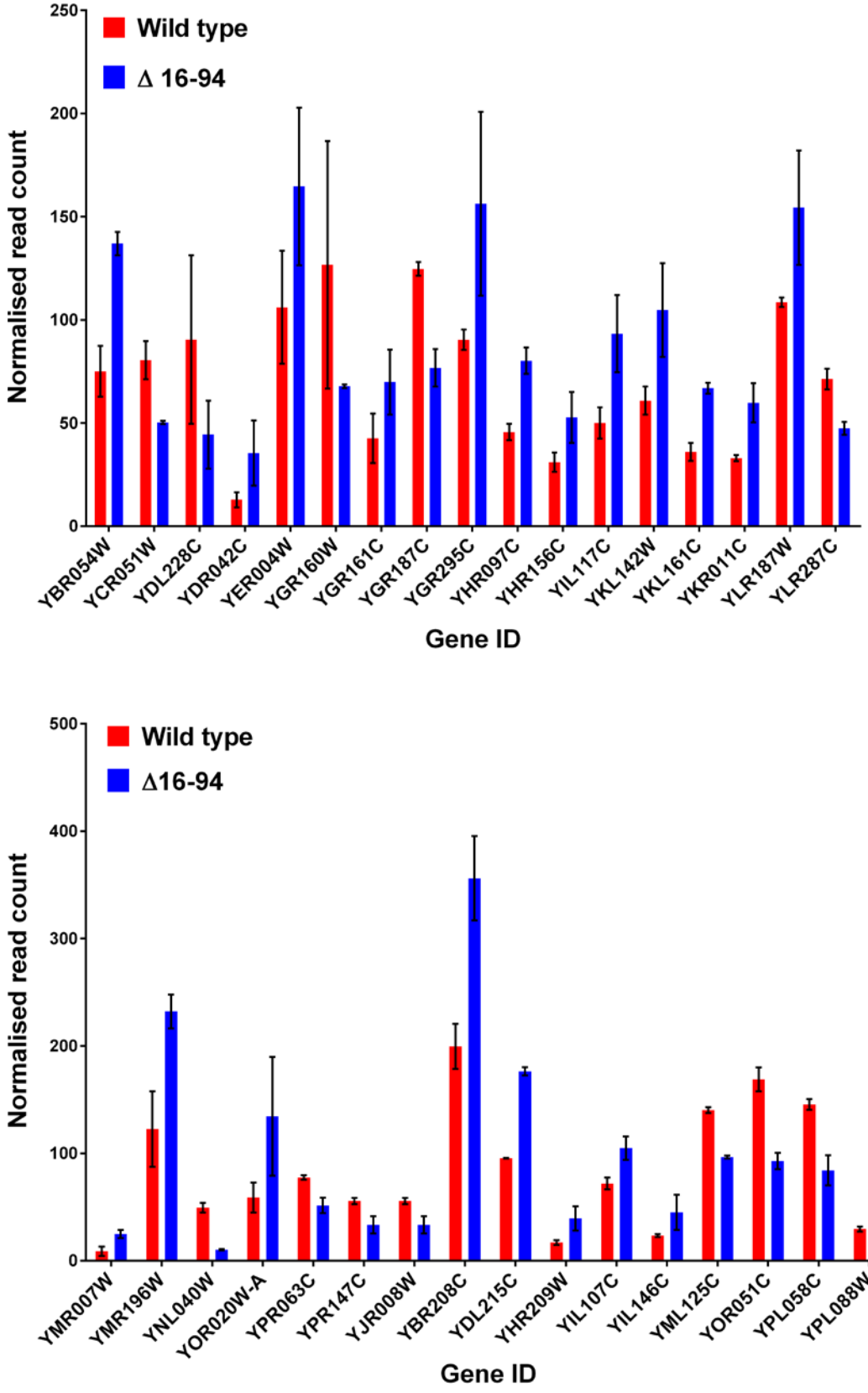


Figure 4.20 Expression of genes of unknown function.

4.7.2 Analysis of the 3’ UTR of genes using RNA sequencing

Prior to analysis, sequencing data from each mutant was collated and aligned to reference genome using the software programs, Tophat and Bowtie (295). This generated an aligned data file for each of the triplicate samples for each mutant. Both wildtype and mutant alignments were visualised using Integrative genomics viewer (IGV) (figure 4.21) which shows the annotated *S. cerevisiae* genome and the aligned sequencing reads. An accompanying plot provided an overview of the number of sequencing reads at any given point allowing analysis of processing activity within the 3' UTRs of genes.

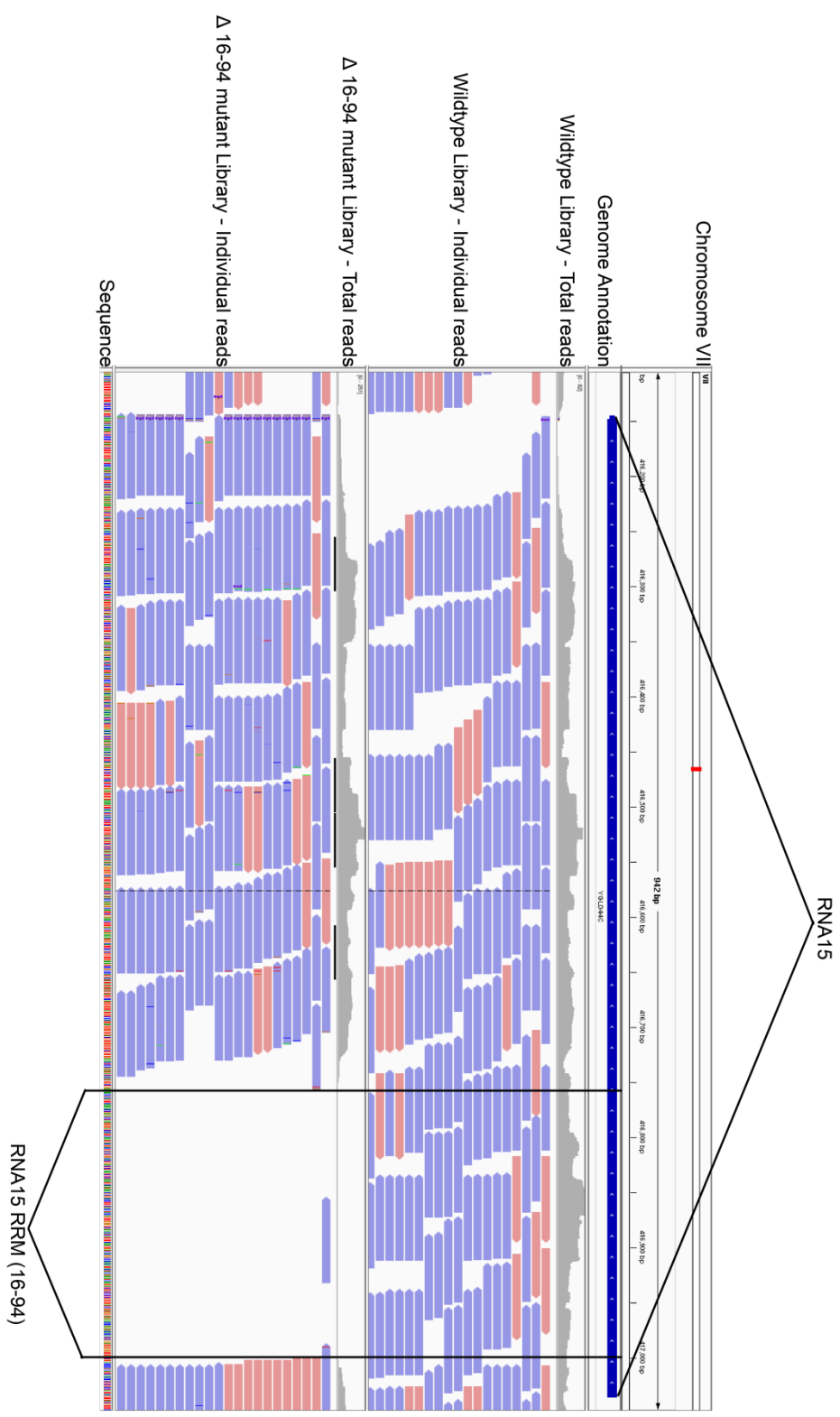


Figure 4.21 Representative aligned RNA sequencing data. The deletion of the RRM is visible in the Δ16-94 mutant where no reads are mapped to the RRM region. Individual reads are shown in blue and pink where pink represents sense reads and blue represents antisense reads.

By measuring the amount of reads in three arbitrary sections of the 3' UTR for each class of gene in the Δ16-94 mutant and wildtype, potential readthrough is inferred by differences in read count. Where read count is high, a processing event is unlikely to have occurred. Polyadenylation within the 3' UTR coincides with a sharp reduction in mRNA amount. The number of reads was taken from triplicate samples of wildtype and duplicate samples of the Rna15 Δ16-94 mutant.

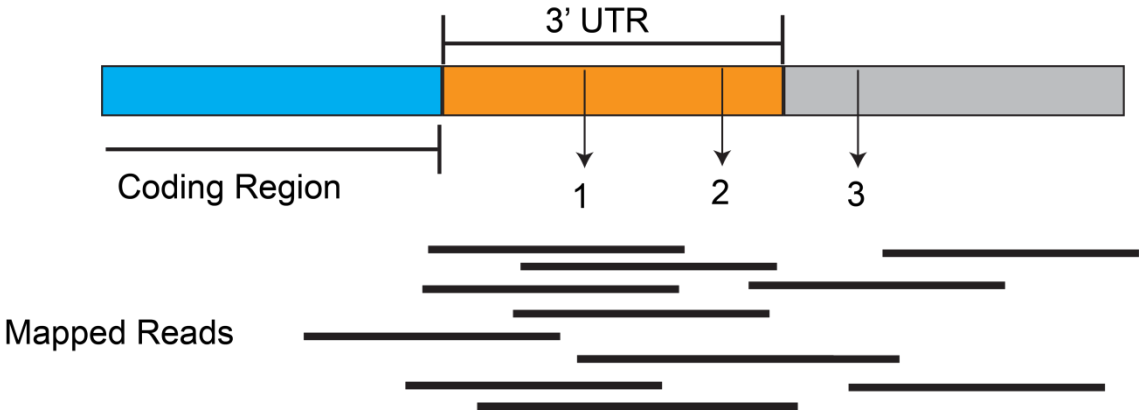


Figure 4.22 Analysis of the 3’ UTR. Three sites are chosen outside the coding region of the target gene. The read count for each site is calculated for both mutant and wild type.

4.8 Analysis of processing events using RNA sequencing

4.8.1. Analysis of processing in genes involved in metabolism and translation

In general, processing in the 3’ UTR of genes involved in cell metabolism remain largely unaffected by deletion of the RRM of Rna15 (Figure 4.23, A). Details of genes analysed and their role in metabolism are detailed in Table 4.3. The largest difference in mRNA amount is in the 3’ UTR of TDH1. This gene encodes a protein involved in glycolysis and gluconeogenesis within the cell. Although the difference in amount of mRNA in the 3'UTR is clear, the error is quite large suggesting this result is ambiguous. Amounts of mRNA observed in genes including RAT1, YCK2 PSK1, remain similar to that of wildtype. Therefore, it is clear that deletion of the RRM does affect 3' end processing in these genes, however, these results suggest that differences are relatively small upon comparison with wild type.

The amounts of mRNA in the 3’ UTR of genes involved in translation follows much the same pattern (Figure 4.23, B). A list of genes involved in translation and their role is detailed in Table 4.4. Large differences in the 3’ UTR is observed in RPG1 which encodes a subunit of the translation initiation factor, eIF3. However, relatively small changes are observed in other genes. As observed in genes involved in metabolism, deletion of the RRM does not affect the 3’ UTRs of these genes in a distinct pattern based on their role in translation.

Gene Name	Systematic Name	Encodes	Role
PSK1	YAL017W	PAS domain	Carbohydrate metabolism
PDX3	YBR085C	Phosphate oxidase	Nitrogen metabolism
ARA1	YBR149W	$NADP^+$ dehydrogenase	Carbohydrate metabolism
PAP2	YOL115W	Poly(A) polymerase 2	RNA metabolism
CYR1	YSL005W	Adenylate cyclase	cAMP pathway (required for metabolism)
PLB2	YMR006C	Phospholipase B	Lipid metabolism
RAT1	YOR048C	Exonuclease	RNA metabolism
YCK2	YNL154C	Casein kinase 1	Glucose metabolism
PHO91	YNR013C	Phosphate transporter	Phosphate metabolism
TDH1	YJL052W	Dehydrogenase	Glycolysis and gluconeogenesis

Table 4.3 Genes involved in metabolism.

Gene Name	Systematic Name	Encodes	Role
EFB1	YAL003W	Translation elongation factor 1 beta	Translation elongation
FUN12	YAL035W	Translation initiation factor	Translation initiation
NCL1	YBL024W	tRNA methyltransferase	Translation elongation
HEK2	YBL032W	RNA binding protein	Translation regulation
OLA1	YBR025C	ATPase	Translation regulation
RPG1	YBR079C	eIF3a subunit of initiation factor eIF3	Translation initiation
PAB1	YER165W	Poly(A) binding protein	Promotes translation initiation
TRM1	YDR120C	tRNA methyltransferase	Translation elongation
NOB1	YOR056C	Endoribonuclease	Required for ribosome biogenesis
TRM112	YNR046W	tRNA methyltransferase	Translation regulation

Table 4.4 Genes involved in translation.

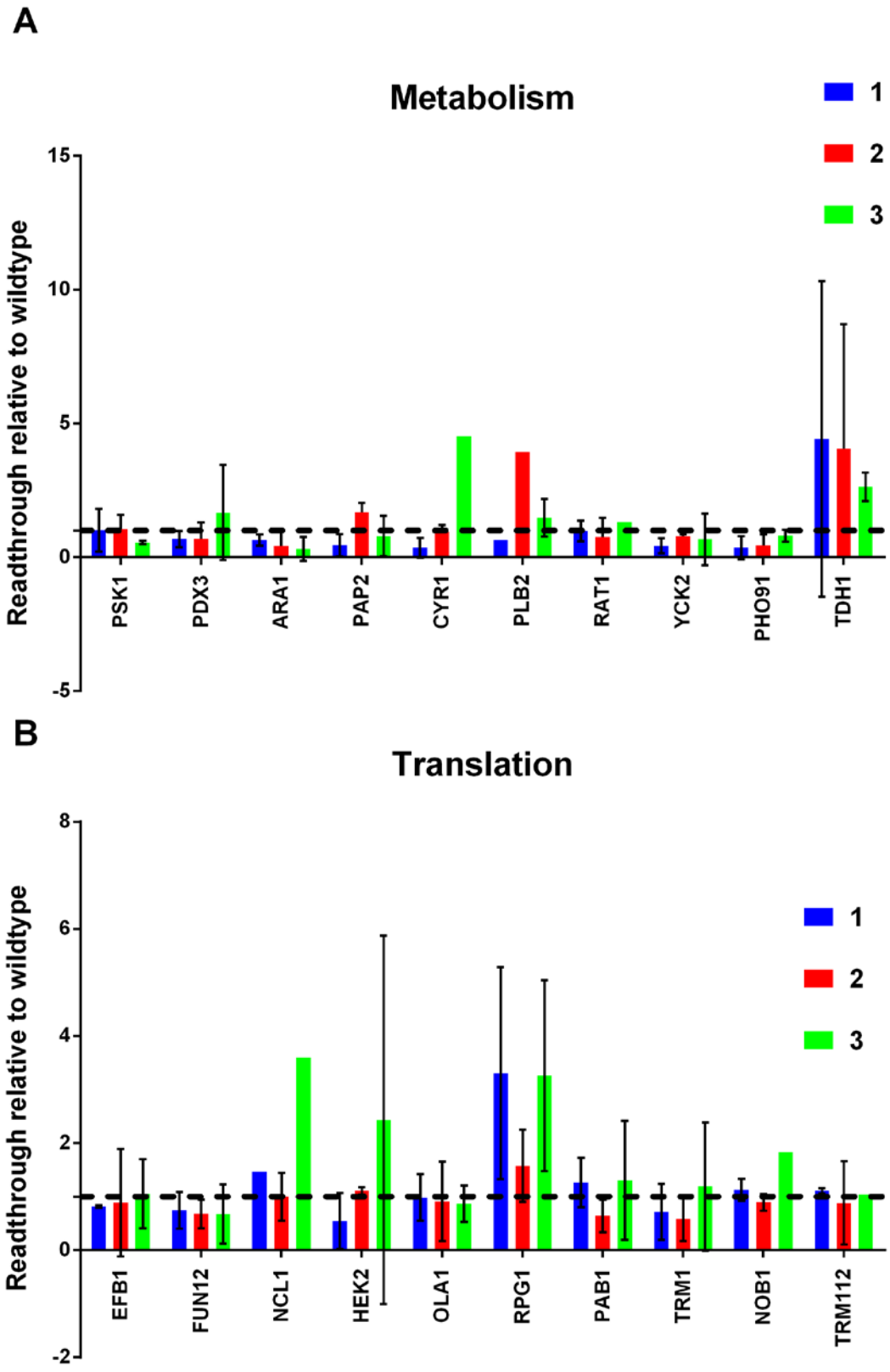


Figure 4.23 mRNA amounts relative to wild type at three distinct sites within the 3'UTR of genes.

The dotted black line denotes wild type mRNA levels.

4.8.2. Analysis of processing in genes involved in cell cycle and cell wall maintenance

The 3' UTR of genes involved in maintaining the integrity of the cell wall are affected by the deletion of the RRM in Rna15 (Figure 4.24, A). Genes analysed based on their involvement in maintaining cell wall integrity are detailed in Table 4.5. Increases in the distal mRNA within the 3' UTR is observed in genes NBP2, RHO1 and ACK1. In other genes such as SKG1 and DSE1 there are decreases in the distal mRNA in the 3' UTR. Again these differences do not follow any distinct pattern and therefore demonstrate that deletion of the RRM does not affect processing in these genes based on their role in cell.

In the case of genes that encode for proteins that regulate and maintain the cell cycle, the amount of mRNA in the 3' UTR is affected by deletion of the RRM (Figure 4.24, B). Genes included in analysis based on their role in cell cycle regulation are detailed in Table 4.6. In the case of CDC28 a large increase is observed in the third site within the 3' UTR. Increases in the amount of mRNA in the 3' UTR is observed in genes, CDC34 and UNG1. However, a general decrease in mRNA amount is observed in genes, ATG1 and MRC1 while MIH1 and CDH1 remain largely unaffected by deletion of the RRM. In general, as observed in other gene groupings, deletion of the RRM does not produce an effect that is distinctive with genes involved in cell cycle regulation. This is demonstrated by the random differences in mRNA amount between genes.

Gene Name	Systematic Name	Encodes	Role
PKC1	YBL150C	Serine/threonine kinase	Cell wall remodelling
FLC2	YAL053W	FAD transporter	Cell wall maintenance
TOS1	YBR162C	Target of Sbf	Associated with cell wall
DSE1	YER124C	Daughter cell specific protein	Regulates cell wall integrity after separation
SKG1	YKR100C	Transmembrane protein	Maintain cell wall polymer composition
ACK1	YDL203C	Functions in cell wall integrity pathway	Maintain cell wall composition
NBP2	YDR162C	Regulates cell wall integrity pathway	Maintain cell wall composition
RHO1	YPR165W	GTP binding protein	Regulates cell wall synthesising enzyme
YGP1	YNL160W	Glycoprotein	Maintain cell wall composition
PRS1	YKL181W	Pyrophosphate synthase	Required for cell wall integrity pathway

Table 4.5 Genes involved in maintaining the composition of the cell wall.

Gene Name	Systematic Name	Encodes	Role
CDC28	YBR160W	Cyclin dependent kinase	Regulates cell cycle progression
MRC1	YCL061C	S phase checkpoint protein	Regulates cell cycle progression
PPS1	YBR276C	Phosphatase	Required during cell cycle DNA replication step
ROK1	YGL171W	RNA dependant ATPase	Regulates cell cycle progression
ATG1	YGL180W	Serine/threonine kinase	Regulates cell cycle progression
MIH1	YMR036C	Tyrosine phosphatase	Regulates cell cycle progression
CDH1	YGL003C	Directs ubiquitination of cyclins	Regulates cell cycle progression
UNG1	YML021C	Uracil DNA glycosylase	Required for DNA repair during checkpoint
CDC34	YDR054C	Ubiquitin conjugating enzyme	Regulates cell cycle progression
PSY4	YBL046W	Subunit of phosphatase PP4	Required for DNA repair during checkpoint

Table 4.6 Genes involved in regulation and progression of the cell cycle.

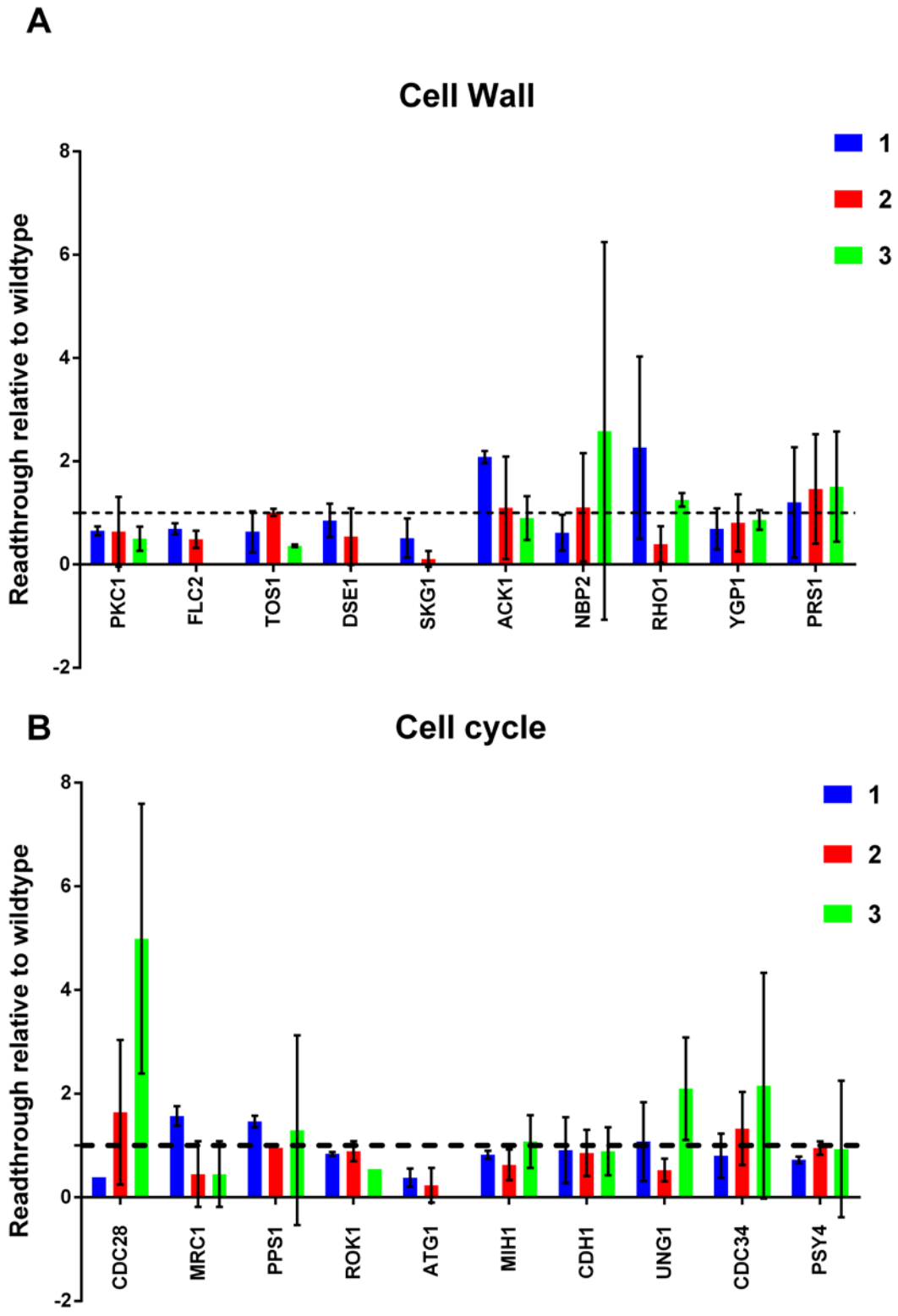


Figure 4.24 Relative amounts of mRNA for genes involved in maintenance of the cell wall and cell cycle. The black dotted line denotes wild type amounts of mRNA.

4.8.3 Analysis of poly(A) site selection in ACT1, TDH2, ADH1 and YPT1

Using previously annotated poly(A) sites provided on the *Sacchromyces cerevisiae* genome database (SGD), poly(A) site selection in genes ACT1, TDH2, ADH1 and YPT1 was investigated. In the same way as before, read count number was taken at the annotated poly(A) site for each triplicate of wild type and duplicate of the Δ16-94 mutant. The amount was normalised by an average expression of each gene (average expression was defined by averaging three read counts within the open reading frame of each gene) and the total amount of processing was calculated by subtracting the read count from poly(A) site 1 from the amount of mRNA present within the 3' UTR prior to processing. This was repeated for subsequent poly(A) sites within each gene. The resulting values were expressed as a percentage of the total amount of mRNA processed.

In the ACT1 gene, slight differences in the amount of processing at all sites are observed (Figure 4.25, A). An increase in the amount of processing activity is observed at the proximal site (site 1) upon comparison with wild type. This is observed again at distal site 3, where a slight increase in processing activity occurs. Conversely, a decrease in processing activity is observed at sites 2 and 4 within the ACT1 gene, however the error demonstrated in the in the mutant at site 4 suggests this result is highly unreliable. In the case of TDH2 (Figure 4.25, B), a similar pattern occurs where processing activity decreases at the proximal site (site 1) but increases at the second distal processing site. A final decrease in processing activity is observed at the third site, again highlighting slight differences in processing activity within the Δ16-94 mutant. In ADH1, slight differences in processing are observed at all three sites within the 3' UTR (Figure 4.26, A). Again this is characterised by a slight decrease at site 1 followed by an

increase in processing at sites 2 and 3. In YPT1, it is important to note that results are ambiguous due to the large error produced at site 2. Nevertheless, wild type processing occurs at the first site implying that deletion of the RRM does not affect processing in YPT1 (Figure 4.26, B). However, a reduction in processing activity is observed at site 2 followed by increases in processing activity at sites 3 and 4.

Analysis of the sequences that direct polyadenylation in genes ACT1, ADH1, TDH2 and YPT1 reveal that cleavage occurs at the expected site in all genes, $Py[A]_3$. The composition of the polyadenylation signal and efficiency element present upstream from the poly(A) site are in agreement with previously defined sequences. In all genes the positioning element is present as either AAUAAA or AAAAA, however, the exact position of this sequence differs from gene to gene. Also, there appears no relationship between occurrence of the different positioning elements and changes in polyadenylation observed in the Rna15 mutant. The efficiency element consists of the expected UAUAUA sequence at sites 1, 2 and 2 of genes ACT1, TDH2 and ADH1 respectively. At other sites this element is present with at least one mutation. Interestingly, there appears to be a correlation of increased processing at sites where UAUAUA is present in the Rna15 Δ16-94 mutant. It is possible that the loss of the RRM in Rna15 enhances Hrp1 binding to this efficiency element sequence. Indeed, SELEX experiments have shown that Hrp1 preferentially binds this sequence during polyadenylation (150). However, this is only observed in these genes presented and a global analysis of polyadenylation in the Δ16-94 mutant is needed to make a confident assessment of the relationship between sequences that direct processing and loss of the RRM of Rna15.

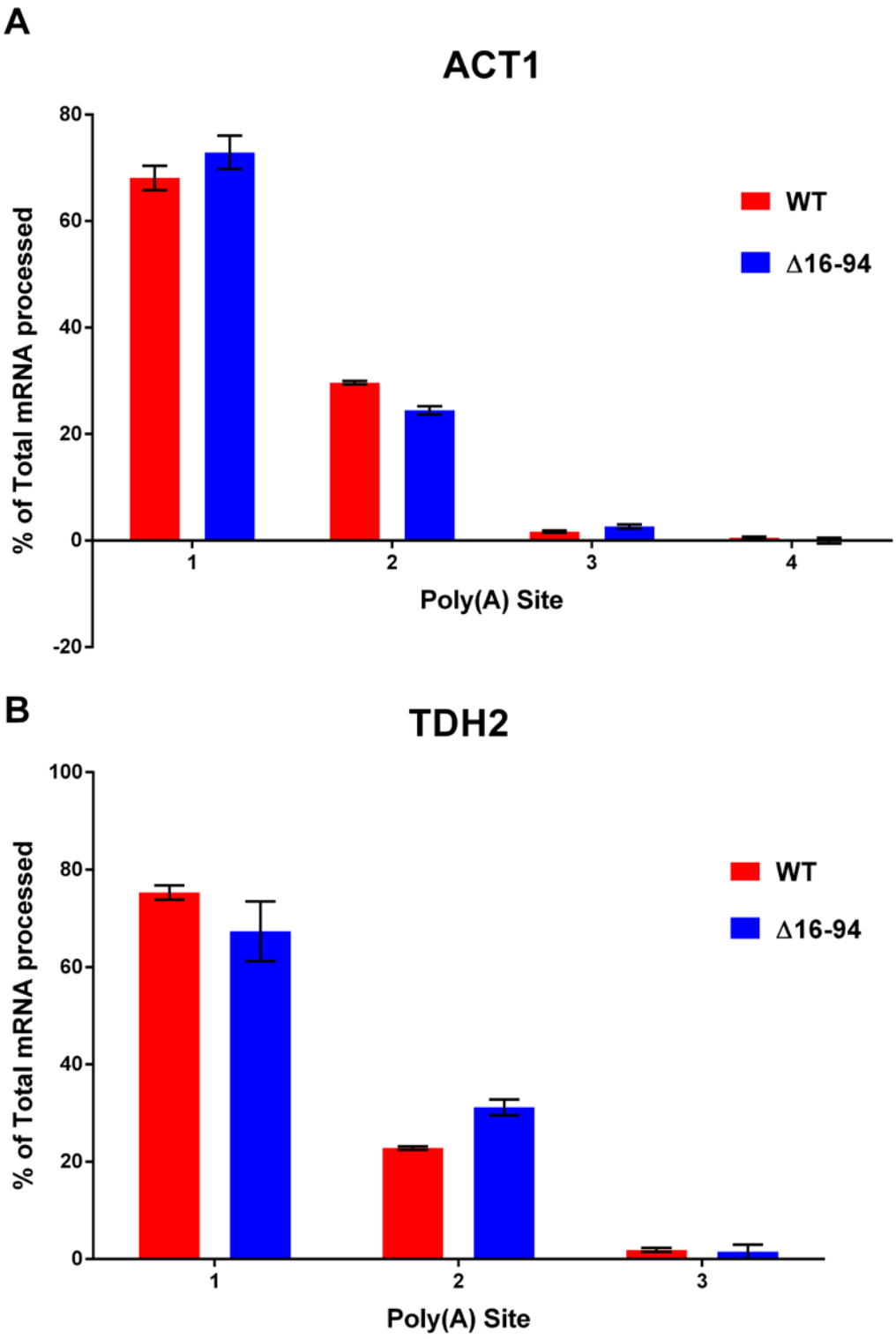


Figure 4.25 Poly(A) site selection in the ACT1 gene (A) and the TDH2 gene (B).

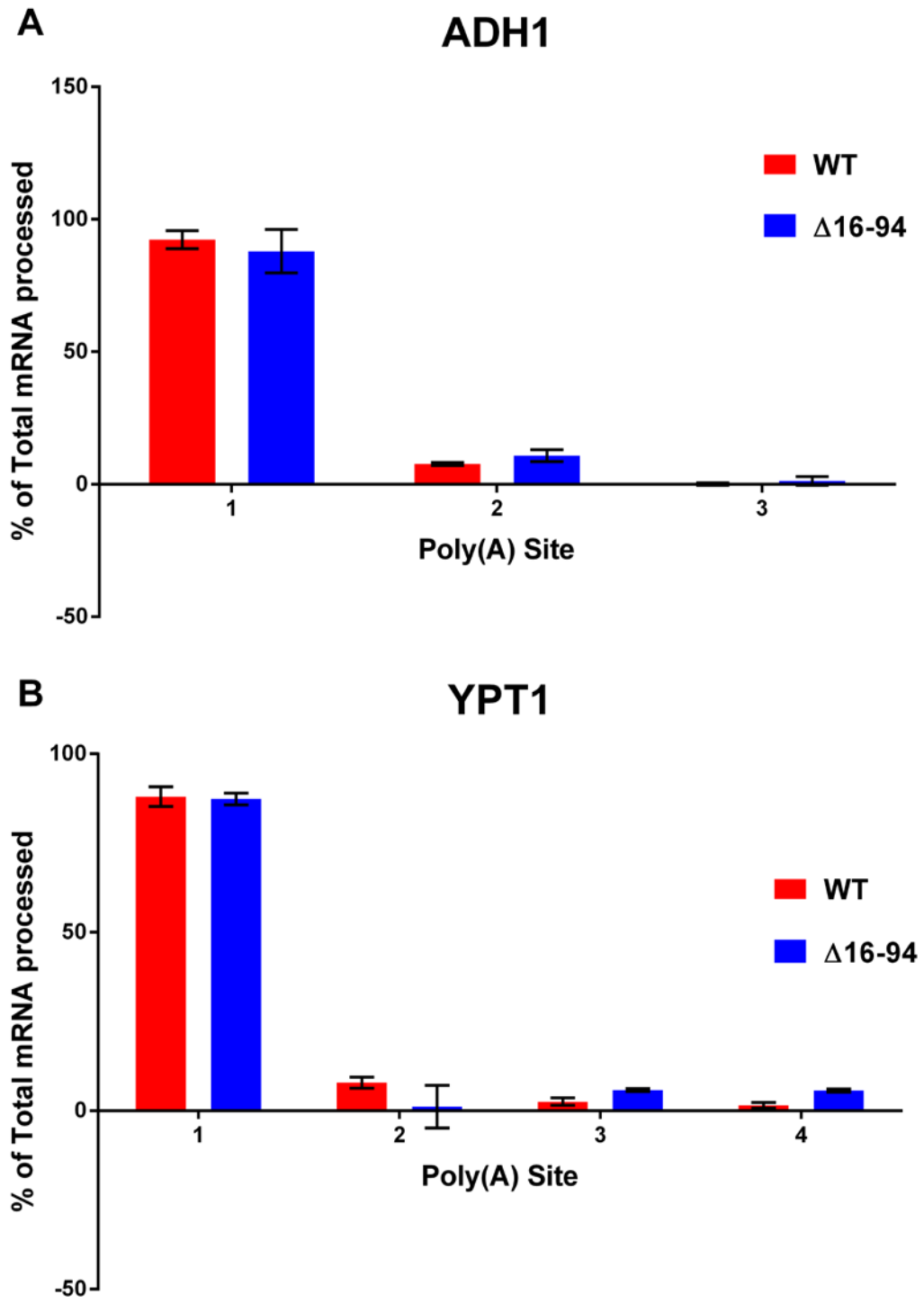


Figure 4.26 Poly(A) site selection in the ADH1 gene (A) and the YPT1 gene (B).

4.8.4 RNA sequencing conclusions

Global analysis of expression in the *S. cerevisiae* Δ16-94 mutant reveals alterations in expression of genes required for a wide variety of cellular processes. It appears a higher proportion of genes that are highly expressed are affected by deletion of the RRM. However, aside from this observation there are no distinct similarities between differences in expression between groupings. Overall expression both increases and decreases in each class of gene. This effect demonstrates that the overall process of transcription is affected by deletion of the RRM. However, it is important to note that the 131 genes that are affected by deletion of the RRM of Rna15 represent 2% of the total *S. cerevisiae* genome. These 131 genes are representative of those chosen based on very stringent statistical analysis (FDR value <0.05). This means that these genes are likely to be the greatest affected by deletion of the RRM, however, it is possible that a larger group of unambiguous genes are affected that do not fall into the FDR <0.05 category.

In general, it appears that deletion of the RRM in Rna15 does result in widespread differences in processing activity within the 3' UTRs of genes. This is shown by the differences in read number from wild type to mutant from genes within each class. Another common feature is the random nature of mRNA amount in the 3' UTRs of genes from gene to gene within each class and from class to class also. This shows that deletion of the RRM affects each gene differently and the mutation does not result in biased changes within genes based on their specific role within the cell. Without annotated poly(A) sites for all genes analysed it is difficult to say for certain how processing at poly(A) sites differ between wild type and mutant. However, it is clear that there are differences in distal mRNA within the 3' UTR and therefore, processing in these genes has clearly been affected. It is possible that these differences

are due to differences in expression level of genes in the mutant. However, amounts differ from site to site in all genes within the 3' UTR possibly due to differences in processing.

Using previously annotated poly(A) sites for ACT1, YPT1, TDH2 and ADH1, it was possible to calculate the processing activity in both wild type and the Δ16-94 mutant. Although a difference in processing activity was observed in all genes, the amount of processing activity was not significantly different from wild type. These results suggest that deletion of the RRM in Rna15 does not result in major differences in poly(A) site selection within the 3' UTRs of these genes.

4.9 Summary

This chapter has aimed to characterise the effect of mutation of the RRM of Rna15 on both viability and mRNA 3' end processing activity in *S. cerevisiae.* In general, point mutations that affect RNA binding *in vitro* do not affect cell viability *in vivo* and combination of the most severe mutations that result in loss of RNA binding *in vitro* do not have any effect on cell viability. However, deletion of the RRM of Rna15 does result in lethality at 37ºC on minimal media containing 5-FOA. Furthermore, growth curve experiments demonstrated that deletion of RRM resulted in retarded growth upon comparison with wild type in rich media at 30ºC. Further investigation into cellular processes within *S. cerevisiae* demonstrated that transcription of a small subset of genes is strongly affected by deletion of the RRM. RNA sequencing results demonstrate that expression of 131 genes are significantly affected by deletion of the RRM of Rna15 when assessed using strict statistical analysis. In addition, RNA sequencing results demonstrate that processing at previously annotated poly(A) sites is

affected by the deletion. Defects in polyadenylation are further evidenced by RNA sequencing analysis of 50 genes that display differences in mRNA amount within the 3' UTR of genes. However, differences in transcription and polyadenylation in ACT1, TDH2, YPT1 and ADH1 appear small upon comparison with wild type. A larger investigation into polyadenylation is required in order to ascertain whether these small differences are representative of processing within the entire transcriptome. However, if these small affects were representative of global polyadenylation in the Δ16-94 mutant then it is possible to understand why a more severe effect on cell viability is not observed. It is possible that other RNA binding proteins part of the polyadenylation machinery, such as, Hrp1, are able to complement mutations in rna15 and direct the 3' end processing machinery to the 3' UTR so that processing can be successfully carried out. However, it is important to note that although the process of polyadenylation may still occur in these strains, the positioning of the cleavage reaction may be compromised where there is loss of Rna15-RNA binding.

5.1 Introduction and Overview

This chapter aims to characterise how additional factors affect Rna15-RNA binding. These include the interaction with Rna14 and the type of RNA sequence favoured by Rna15. Increasing data regarding the mechanism by which Rna15 binds RNA has led to debate over sequence specificity. Previous studies have identified that Rna15 displays sequence specificity for A rich sequences when in complex with Rna14 and Hrp1 (139). However, when alone, Rna15 displays extremely weak RNA binding for A-rich sequences resulting in a K_d value in the 10^{-5} M range (141). By contrast, the crystal structure demonstrates specificity in the site I binding pocket for G and U nucleotides. In addition, fluorescence spectroscopy experiments demonstrate that Rna15 (16-111) binds a G/U rich RNA oligonucleotides in the 10^{-6} M range (141). Experiments presented determine whether Rna15 is able to select preferential RNA binding sites without the assistance of other protein factors resulting in tighter RNA association. Finally, the orientation of the bound RNA was explored using NMR spectroscopy. In addition, Rna14 binds to Rna15 tightly and forms a self-associative tetramer via Rna14-Rna14 interactions (107,108,137). This tetramer makes contacts with Hrp1 and Rna14 acts as a scaffold between Hrp1 and Rna15 when bound to RNA during polyadenylation (151).

Therefore, the aim of this chapter is, initially, to distinguish whether Rna15 displays nucleotide preference using a novel NMR based method (SIA). Binding analyses are presented using Rna14-Rna15 protein to determine whether presence of Rna14 protein and formation of the Rna14-Rna15 tetramer affects Rna15-RNA binding affinity. Finally the importance of the Rna14-Rna15 interaction is demonstrated by analysing what effect deletion of the interaction domain of Rna14 has on the growth phenotype of *S. cerevisiae*.

5.2 Generating a Rna15 consensus sequence

Scaffold Independent Analysis (SIA) was used to generate a potential consensus sequence for Rna15. This technique was developed by Dr A. Ramos (N.I.M.R) and allows unbiased analysis of the target protein binding preference to a pool of short randomised RNA sequences .

The RNA sequences tested comprise 16 pools of RNA oligonucleotides. In each pool there are 64 pentameric RNAs where four out of the five nucleotides are random whilst one nucleotide is constant in position and type of nucleotide base. In position 2, each pool has either an adenine, uracil, guannine or cytosine present in the second position while position 1, 3, 4 and 5 are random. This pattern continues in positions 3, 4 and 5 to generate the 16 pools required for SIA (Table 5.1).

	Position 2	Position 3	Position 4	Position 5
A pools	NANNN	NNANN	NNNAN	NNNNA
C pools	NCNNN	NNCNN	NNNCN	NNNNC
G pools	NGNNN	NNGNN	NNNGN	NNNNG
U pools	NUNNN	NNUNN	NNNUN	NNNNU

Table 5.1 RNA pools employed in SIA experiments. The above pools were titrated with Rna15 (2-103) protein in order to determine a nucleotide preference and generate a consensus sequence.

In SIA, to assess the protein-RNA binding, the target protein is titrated against each of the 16 pools of RNA oligonucleotides and analysed by correlation ^{15}N-^{1}H heteronuclear NMR spectroscopy. Binding affinity for the 16 pools is evaluated by comparison of the ^{15}N-^{1}H HSQC spectra generated from each titration. In a SIA experiment, the target protein is titrated with each randomised 64 sequence RNA pool employing increasing concentrations of RNA at the ratios, 1:0, 1:1 and 1:3 (protein:RNA). The chemical shift position is reflective of the average of the 64 RNAs in the pool bound by the target protein. Analysis of RNA binding is dependent on

movement of the same protein resonances in each of the ^{15}N-^{1}H HSQC spectra taken from each RNA pool and that each move in an concerted direction. This is to ensure a fair comparison as it is possible that movement of protein resonances may become altered when titrated with each RNA pool as chemical shift movements are extremely sensitive to change in the surrounding environment. However, small differences can be averaged out following simultaneous evaluation of a set of peaks. The shifts observed in peaks that are in fast exchange can then be related to the molar fraction of bound protein. This relationship is exploited to rank the preference of the protein for a ribonucleotide in a given position. Whilst this comparative analysis is successfully exploited to generate an insight into binding specificity, it is important to consider the limitations of analysing chemical shift perturbations in protein-ligand interactions. Indeed, chemical shift movements are indicative of protein-ligand interaction, however, introduction of a ligand can result in chemical shift perturbations that are not due to direct binding of the ligand. This is due to amide chemical shifts that are sensitive to structure, hydrogen bonding and solvent accessibility. Therefore, chemical shift perturbations induced upon ligand binding may be remote from the actual binding site and represent a conformational change in the protein itself rather than direct ligand binding.

In this SIA experiment, Rna15 (2-103) was titrated with each of the 16 RNA pools and the spectra for each was recorded (Figure 5.1). Comparative analysis of each spectrum was achieved by measuring the change in the chemical shift ($\Delta\delta$) of 16 individual protein resonances upon titration of the RNA pools. Under the conditions of the experiment saturated 1:3 spectra were excluded from analysis as Rna15 (2-103) was saturated at these concentrations so a preference in binding could not be determined. The $\Delta\delta$ of each individual peak was normalised against the largest $\Delta\delta$ observed between the corresponding four RNA pools with each nucleotide present at a single position (ie

nAnnn, nCnnn, nGnnn and nUnnn). The normalised peak positions were then averaged to generate an SIA score relating to preference of that particular nucleotide.

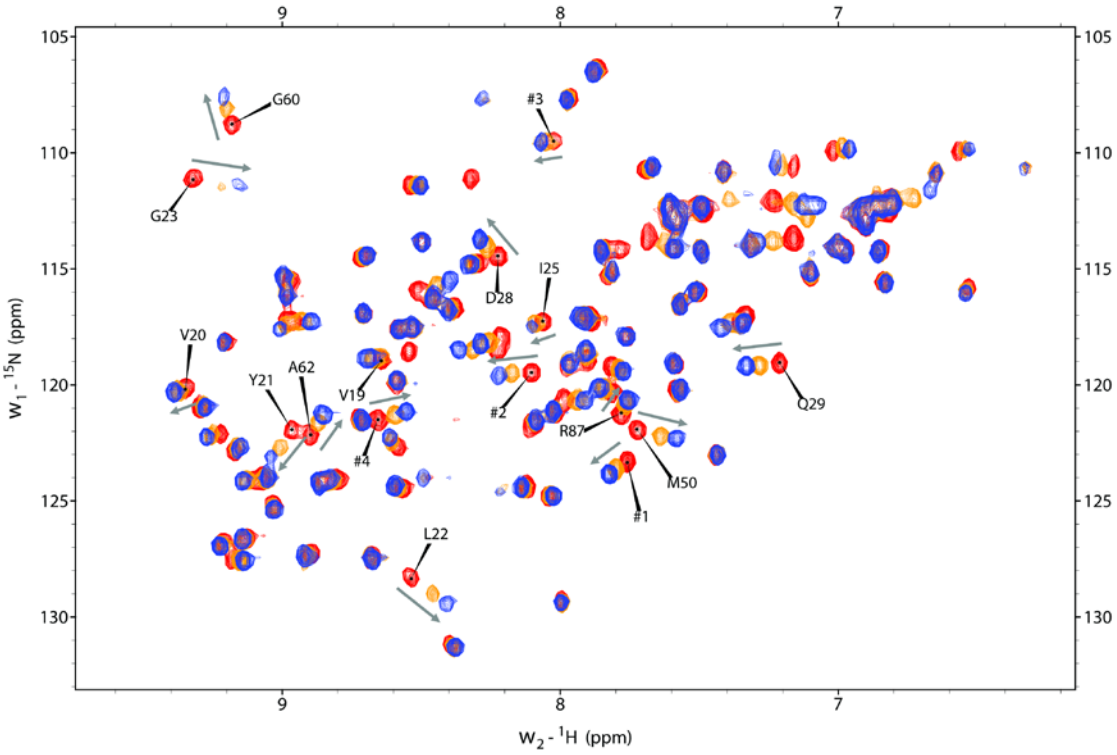


Figure 5.1 Scaffold independent analysis. Example of NMR spectra during an SIA experiment. Rna15 (2-103) protein was added to each pool of RNA at a stoichiometry of 1:0, 1:1 and 1:3 (protein:RNA). Red peaks correspond to free protein, yellow peaks represent 1:1 ratio of protein:RNA and the blue peaks correspond to the 1:3 ratio of protein:RNA. Grey arrows demonstrate the direction of movement of the labelled peaks that were analysed.

The resulting SIA values demonstrate a clear preference for G/U nucleotides (Table 5.2). However, the results indicate that Rna15 (2-103) is still able to bind the RNA pools with A and C present in positions one, two, three and four but the affinity for these RNAs are lesser than that observed for the G and U RNA pools. At position 1 the SIA values for A and C are 0.49 and 0.45 respectively. For both G and U nucleotides the values are 0.91 and 0.95 respectively indicating a clear preference. At position 2, U is favoured over all other nucleotides with an SIA value of 0.98 with G next (0.78) and finally C (0.62) and A (0.54). A G nucleotide is preferred at position 3 demonstrated by the SIA value of 0.97 and again both A and C RNA pools have low SIA values, 0.44, whilst U displays an intermediate value, 0.73. At position 4, the SIA value for the A RNA pool is somewhat higher than at position 1-3. However, the U pool is favoured with an SIA value of 0.94. At position 4, the G pool demonstrates a high affinity for Rna15 2-103 indicated by the SIA value of 0.85. Again the C pool shows weak preference with a SIA value of 0.54.

	1	2	3	4
A	0.49	0.54	0.44	0.73
C	0.45	0.62	0.44	0.54
G	0.91	0.78	0.97	0.85
U	0.95	0.98	0.73	0.94

Table 5.2 SIA results showing a clear consensus for G/U rich nucleotides. The chemical shift variation ($\Delta\delta_{av}$) between 16 corresponding peaks for each ratio (1:0, 1:1 and 1:3) was measured. The resulting values were then normalised and the results of which are shown above. The above values are repsentative of apparent movement of peaks induced upon introduction of RNA at the ratio of 1:1.

The results clearly demonstrate a strong preference of G and U nucleotides over A and C nucleotides in every position bar position 4, allowing the generation of a consensus sequence. Two consensus sequences are demonstrated from the SIA data, 5' UUGU 3' and 5' GUGU 3' (Figure 5.2). The presence of G/U rich sequences at the

poly(A) site has been demonstrated in the 3' UTRs of *S. cerevisiae* (145,146). Therefore, it is possible that the G/U rich sequences are the preferred binding site for Rna15 to anchor CF1A to the 3' UTR.

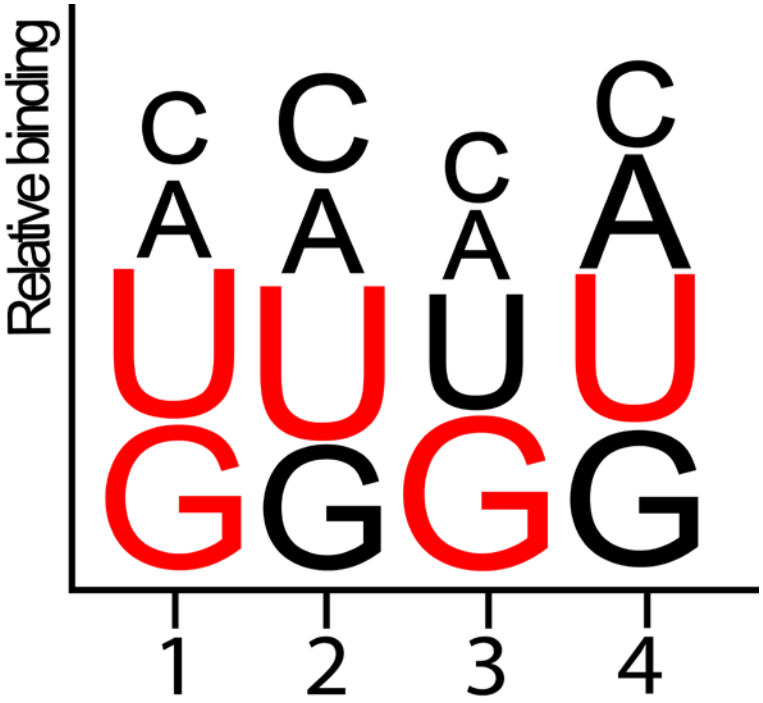


Figure 5.2 SIA generated consensus sequence. The figure highlights the sequence generated by SIA that is the preferential binding of Rna15. The bases that were favoured over others are highlighted in red. The size of the letters also relate to their score in binding after normalisation.

To verify the SIA derived consensus, sequences GUGUGU and GUUUGU were employed in fluorescence spectroscopy experiments. Both were labelled at the 5' terminus with TET. Previous fluorescence experiments demonstrated that a six nucleotide sequence was required for optimal signal detection. Therefore, a dinucleotide of GU was added to the 5' end of the RNA oligonucleotides to maintain optimum signal. The dissociation constant for the 5'TET-GUGUGU oligonucleotide was 1.4 μM ± 0.4 μM whilst the dissociation constant for the 5' TET-GUUUGU 3' was 2.2 μM ± 0.9 μM (Table 5.3). The results demonstrate that Rna15 is able to bind both SIA generated consensus sequences with a higher binding affinity than observed with the orginal 5' TET-UGUUGU 3' RNA oligonucleotide. The combined SIA and fluorescence spectroscopy results demonstrate that Rna15 does display binding specificity with an overall G/U preference that results in an increase in Rna15 RNA binding affinity as demonstrated by tighter association of Rna15 (2-103) with the SIA derived sequences.

RNA oligonucleotide	K_d ($x10^{-6}$ M)
5' TET-UGUUGU 3'	6.4 ± 0.5
5' TET-GUGUGU 3'	1.4 ± 0.4
5' TET-GUUUGU 3'	2.2 ± 0.9

Table 5.3 Binding constants of Rna15 (2-103) with SIA generated RNA oligonucleotides. Fluorescence titrations were carried out in triplicate and the standard deviation of the mean calculated. Rna15 (2-103) wild type was used in these titration experiments.

5.3 Identifying the orientation of the RNA oligonucleotide

Previous mutational and NMR analysis experiments demonstrated the likelihood that RNA binds across the surface of the RRM through interactions with the side chains of residues present on the surface (141). In order to provide further detail into the mechanism of Rna15 mediated RNA binding NMR spectroscopy experiments were employed to attempt to distinguish the orientation of the RNA oligonucleotide. Two RNA oligonucleotides were employed, one labelled with TET at the 5' termini and the other unlabelled. The sequence of the oligonucleotides was that generated by SIA, 5' GUGUGU 3'. The assigned HSQC spectrum was obtained from previous NMR experiments using Rna15 (16-111) (141) which had identified the third binding site (Figure 5.3). Then the presence of the 5' label was exploited in order to identify the orientation of the RNA oligonucleotide. Rna15 2-103 was titrated with labelled and unlabelled RNA oligonucleotide, 5' GUGUGU 3' in separate titrations in order to identify the chemical shift peaks that behave differently upon presence of the 5' TET label.

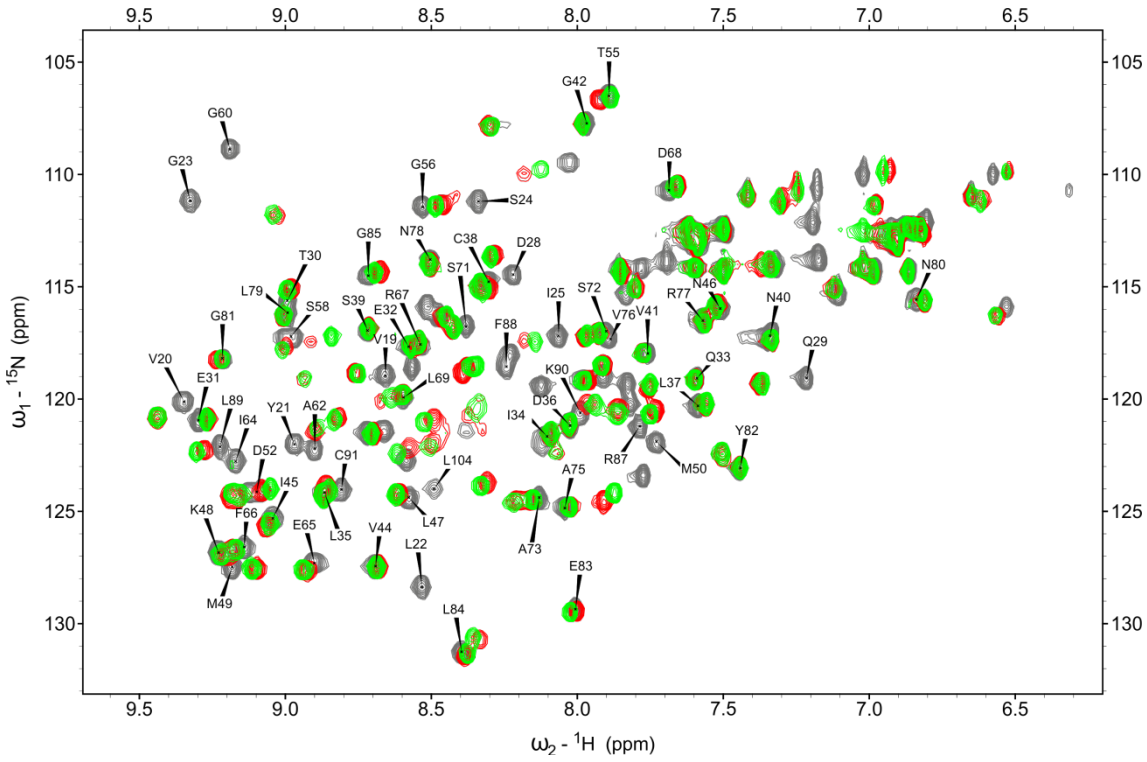


Figure 5.3 Assigned 2D-NMR spectra of Rna15 (2-103). 2D spectra showing the movement of chemical shift peaks in relation to presence or absence of 4-tetrachlorofluoroscein label on the RNA sequence GUGUGU. Assigned peaks are labelled. The grey peaks represent the free protein, the green peaks represent Rna15 (2-103) bound to unlabelled RNA and the red peaks correspond to Rna15 (2-103) bound to RNA where the dye is present on the 5' end.

In the experiment, residues Q83, G81, L89, K90, S58 and T55 that are present on one face of the RRM display the largest change in presence of the label. Movements of residue Q83 was only observed when unlabelled RNA was titrated with Rna15 (2-103) and the chemical shifts of G81 and T55 are perturbed only upon introduction of the labelled RNA. Chemical shift peaks previously assigned to L89, K90 and S58 move further with the unlabelled RNA than was observed with the labelled RNA. When mapped onto the free Rna15 structure it is apparent that all residues affected by the 5' TET label are positioned on the top face of the structure thus highlighting the orientation of the ribo-oligonucleotide and the label. Moreover, the label itself is connected to the 5' nucleotide by a highly flexible C_6 linker. This degree of flexibility allows residues present on the top of the RRM on both sides of site I to be perturbed (Figure 5.4). The lack of affected residues in the remainder of the protein (Figure 5.4, D) could potentially mean that the dye sits slightly forward and rotates through its flexible carbon chain to affect the residues highlighted in blue.

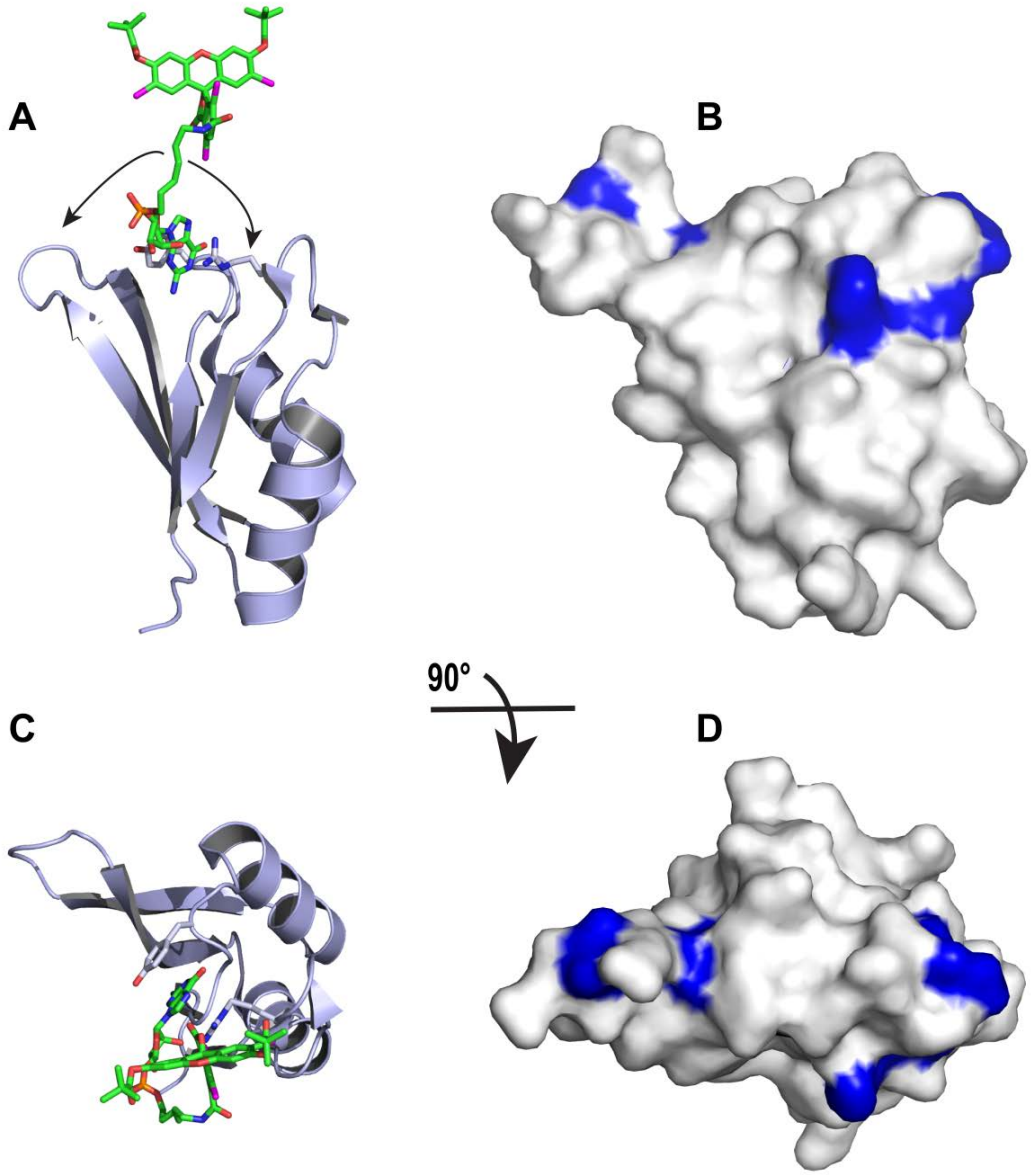


Figure 5.4 Orientation of the RNA oligonucleotide bound to the RRM of Rna15. A. A model of how the 5' TET label may be positioned when the 6 nucleotide long RNA is bound to the Rna15 RRM. The dye is attached to the 5' nucleotide through a highly flexible 6 carbon chain allowing free movement of the dye around the top part of the RRM. **B.** The residues that contact the TET label are highlighted in blue on this structure. **C.** The structure shown in (A) has been rotated 90° to highlight further the part of Rna15 affected by the dye. **D.** The surface model of Rna15 has been rotated to clearly indicate the affected residues. The lack of affected residues at the back of the protein

In order to determine how the protein moves in solution when bound to the labelled RNA oligonucleotide, anisotropy experiments were performed. Polarised fluorescence was detected upon excitation of the RNA in the presence of an excess amount of Rna15 2-103 protein. Polarisation of the fluorescent label was derived using the physical theory of Brownian motion by Perrin:

$$\left(1+\frac{\tau}{\phi}\right)=\frac{\left(\frac{1}{P}-\frac{1}{3}\right)}{\left(\frac{1}{Po}-\frac{1}{3}\right)}$$

where Po represents the fundamental polarisation of the fluorophore (0.45 for TET), τ is the excited state lifetime of the fluorophore (4.1 ns for TET) and ϕ is the rotational correlation time. The measured polarisation (0.24) corresponds to a rotational correlation time of 4ns. Previous NMR experiments demonstrated that free Rna15 (2-94) had a correlation time of 7ns (139). The increase in correlation time presented in this study is likely due to the 5'-TET label that is able to move freely due to the flexible carbon linker. In addition, the similar value generated for both the free and bound Rna15 constructs demonstrate that the RNA is tightly associated with the protein.

As TET labelled RNA oligonucleotides have been used in all binding experiments, the relationship between the TET label and the protein was investigated by a series of competition assays. The aim of these experiments were to exclude the possibility that the TET label may interfere with Rna15 (2-103)-RNA binding. Since K90 was one of the residues shown to be affected by the dye during NMR experiments K90R was also included in the competition assays. Initially competition assays were completed using wild type Rna15 (2-103) protein with unlabelled and labelled RNA oligonucleotide, 5' GUGUGU 3'. Inclusion of the unlabelled RNA oligonucleotide

demonstrates whether the presence of the TET label affects the dissociation constant measured during fluorescence titrations employing the labelled RNA oligonucleotide. The dissociation constant for the 5' TET labelled oligonucleotide was previously demonstrated to be 1.4 µM ± 0.4µM (Figure 5.5, Table 5.3). The dissociation assay employing unlabelled RNA 5' GUGUGU 3' revealed a very similar dissociation constant of 2µM ± 0.4µM (Figure 5.5). This result demonstrates that addition of the 5' TET label does not significantly influence the apparent binding affinity demonstrated by the protein for the RNA oligonucleotide. In addition, competition assays were performed using the Rna15 2-103 K90R construct. This was included as mutation at this residue (K90E) resulted in abolished RNA binding demonstrating direct contact of this residue with the RNA. Subsequent NMR analysis demonstrated that the chemical shift peak corresponding to K90 was affected upon addition of the labelled RNA (5'-TET-GUGUGU-3') compared to the unlabelled RNA (5' GUGUGU 3'). Therefore, in order to establish whether the dye was affecting the abrogation of binding observed upon mutation of this residue, competition assays were performed. Fluorescence titration of the K90R mutant with the 5'-TET-GUGUGU-3' RNA oligonucleotide demonstrated a dissociation constant of 1.5µM (Figure 5.5). The dissociation assay employing the unlabelled 5'-GUGUGU-3' oligonucleotide results in a dissociation constant of 2.2µM. This result demonstrates the TET label does not affect binding affinity and the measured dissociation constant is reflective of Rna15 2-103 K90R-RNA binding.

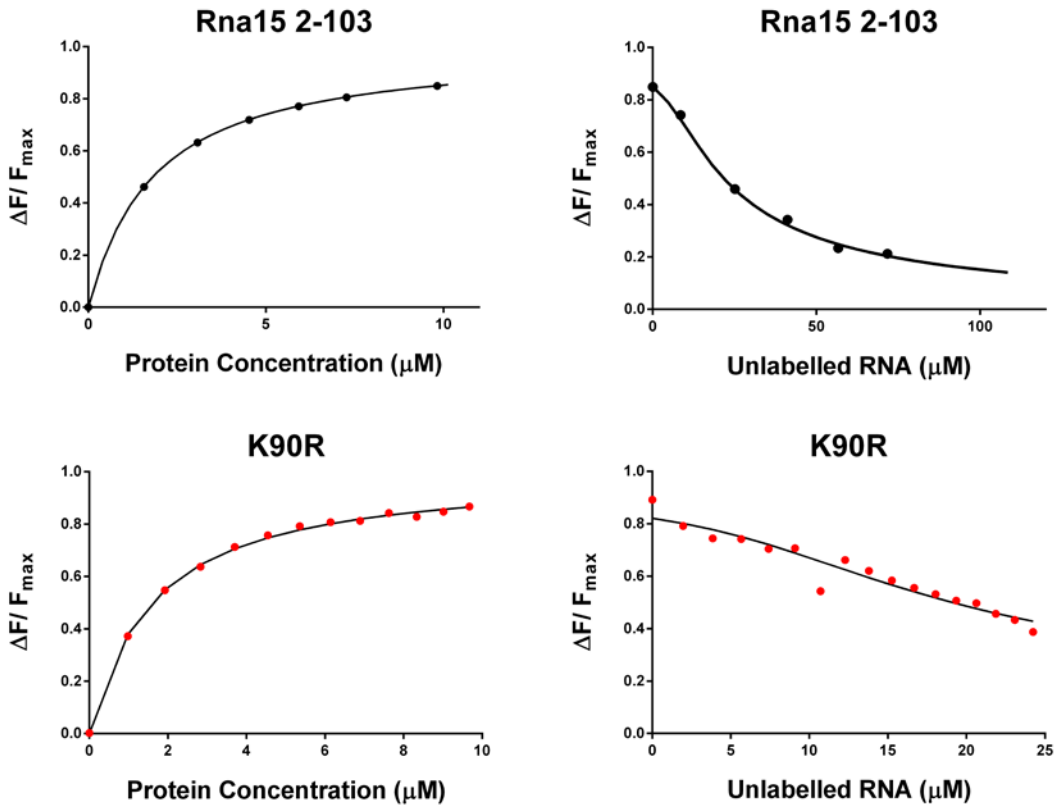


Figure 5.5 Analysis of the effect of the TET dye on RNA binding. Association experiments for both wild type and K90R mutants are displayed in the left top and bottom panels. Dissociation curves for both wild type and K90R mutants are demonstrated in the right top and bottom panels.

5.4 The effect of the Rna14-Rna15 tetramer on RNA binding

5.4.1 Expression of Rna14 and Rna15

RNA14 and RNA15 present in a pETduet1 vector (supplied by Dr. L. Arnold) allow co-expression of both Rna14 and Rna15. Purification of both was achieved via a 6X histidine tag present on the N-terminal of Rna14. Subsequent purification by size exclusion chromatography revealed the presence of two bands within the peak fractions (Figure 5.6). Rna14 is represented in the peak fractions by the clearly defined band present around the 62kDa marker. Rna15 is represented by the clear protein band present around the 38kDa marker band.

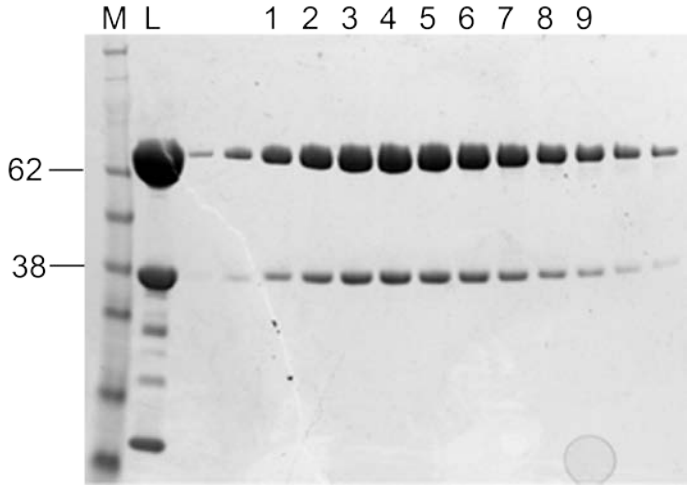


Figure 5.6 Size exclusion purification of Rna14 and Rna15 protein after overexpression in *E. coli*. SDS analysis of purification: M, Marker (SeeBlue® Plus 2, Invitrogen); L sample load. Peak fractions pooled following size exclusion chromatography are shown in lanes 1-9.

5.4.2 Fluorescence spectroscopy of the Rna14-Rna15 tetramer with RNA

Electron microscopy experiments involving Rna14 demonstrated that the Rna14-Rna14 homodimer formed a rod like structure with a kink of ~110º (Figure 5.7, A) (107). Indeed, the structure of the CstF-77 homodimer (murine homologue of Rna14) adopts the same structure with a small bend present at the interface (Figure 5.7, B) (135). The proposed architecture of the Rna14-Rna15 tetramer is shown in the schematic of the Rna14-Rna15 tetramer (Figure 5.7, C). This structure would orientate the RRMs of Rna15 to allow them to freely access the RNA transcript allowing a very tight binding affinity. A tighter association with RNA has been demonstrated when both Rna14 and Rna15 are present. It was thought that this was due to the formation of the Rna14-Rna15 tetramer resulting in binding of one tetramer in two places on the RNA which results in an increase in RNA binding affinity. However, although an increase in Rna15 RNA binding affinity has been observed when Rna14 is present, the mode of binding has not yet been studied. This next section aims to determine how the Rna14-Rna15 tetramer interacts with the RNA to bind it.

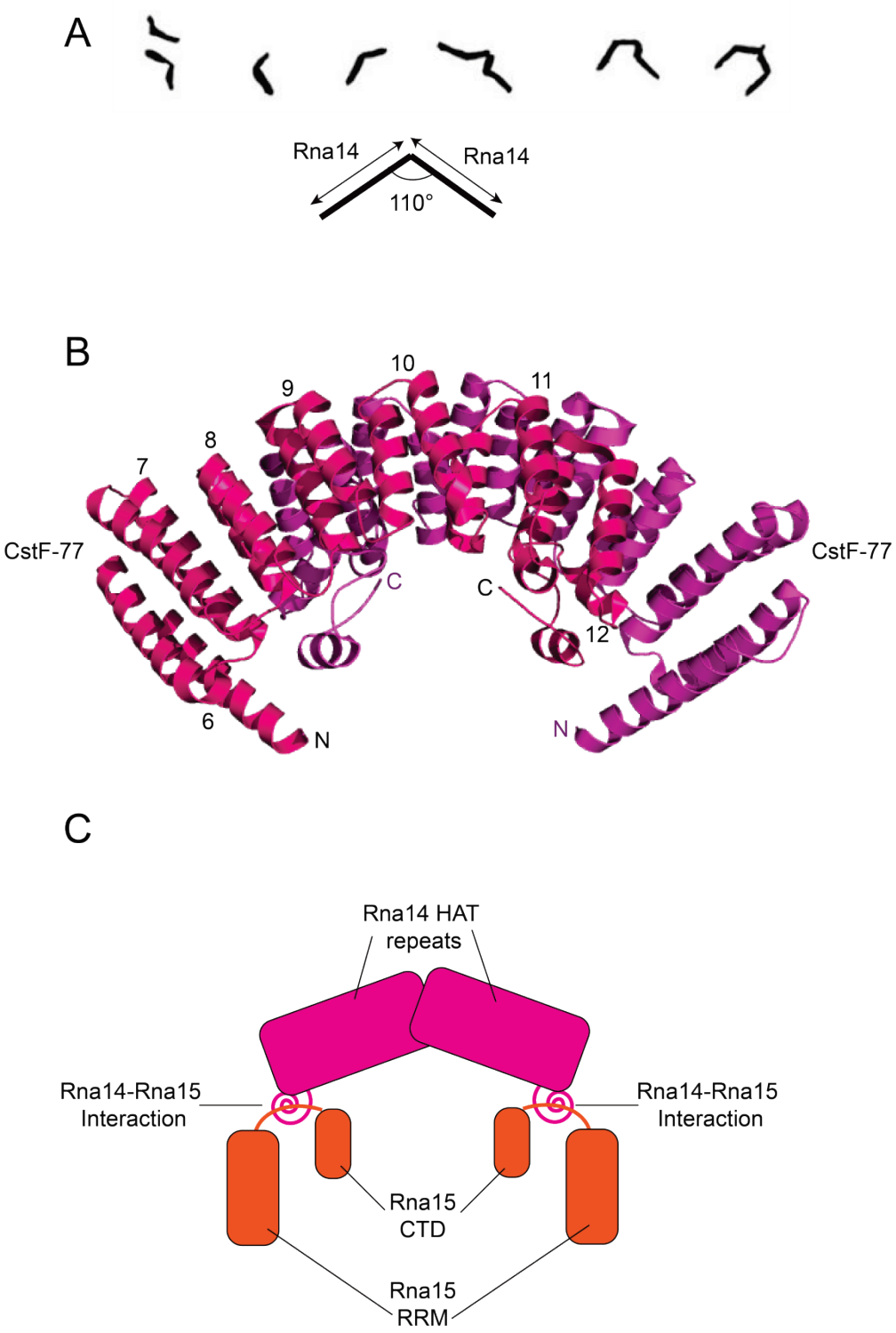


Figure 5.7 The Rna14-Rna15 tetramer. A. Electron microscopy of the Rna14-Rna14 homodimer reveals the presence of rod like particles present with a 110º kink. **B.** Crystal structure of the CstF-77 homodimer in mouse. **C.** The hypothetical structure of the Rna14-Rna15 tetramer is demonstrated as a schematic. The two Rna14 monomers interact via HAT repeats. Each Rna14 monomer is then free to form the monkeytail-hinge like structure with an Rna15 monomer resulting in formation of the tetramer.

Fluorescence spectroscopy was employed to study the mechanism of tetramer-RNA binding. It was hypothesised that formation of the tetramer would result in tighter binding than the monomeric Rna15-RNA dissociation constant of 1.4μM. The presence of two Rna15 RRM domains in the tetramer would allow binding to an RNA oligonucleotide at two distinct sites (Figure 5.8). In order to test this hypothesis and elucidate the tetramer mode of RNA binding, three RNA oligonucleotides were designed that differed in terms of nucleotide length (Table 5.4). All oligonucleotides contained two separate Rna15 binding sites made up of the consensus sequence determined by SIA, GUGUGU, both separated by 4, 12 or 20 nucleotides of A/C rich sequence.

Rna14/Rna15 protein was kept at a constant concentration of 30μM during the titration experiments while an excess (355μM) of the unlabelled RNA oligonucleotide spiked with labelled RNA was titrated into the protein and the change in fluorescence signal detected. The unlabelled RNA oligonucleotide was spiked with a small volume of the same labelled version of the oligonucleotide ~5μM and titrated. A control experiment was run alongside the binding experiments using the same amount of RNA oligonucleotide but did not include any protein. This control titration was then subtracted from the binding titration during analysis.

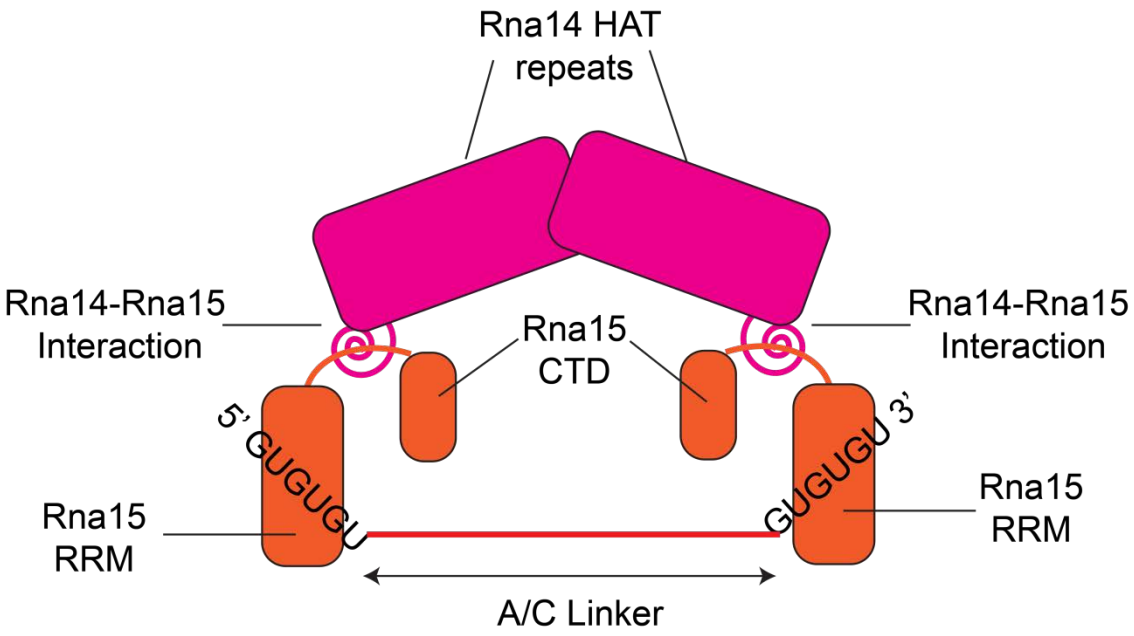


Figure 5.8 Hypothetical tetramer-RNA binding. The diagram shows a schematic representation of the hypothetical mechanism of Rna14-Rna15 tetramer-RNA binding. The two RRMs of Rna15 are free to bind two distinct binding sites present in the oligonucleotide. The red line represents the A/C rich linker of unknown length to facilitate tetramer RNA binding. Binding in this way would result in a large increase in binding affinity from the dissociation constant of 1.4μM observed in the Rna15 2-103 monomer.

Alongside the 16, 24 and 32 RNA oligonucleotides, the original 5'-GUGUGU-3' was also employed in titrations. Titration of all RNA oligonucleotides resulted in dissociation constants that were similar despite the differences in number of nucleotides (Table 5.4, Figure 5.9). In addition the measured binding affinities indicated a less tight association with the RNA than was demonstrated by the Rna15 2-103-5'-GUGUGU-3' association.

RNA oligonucleotide	Composition	K_d (x10^{-6} M)
6	5'-GUGUGU-3'	5.1 ± 1.3
16	5'-GUGUGU-CCAC-GUGUGU-3'	5.3 ± 1.6
24	5'-GUGUGU-CCAA-CCAC-CAAC-GUGUGU-3'	7.7 ± 2.2
32	5'-GUGUGU-CCAA-CCAA-CCAC-CAAC-CACC-GUGUGU-3'	9.9 ± 1.8

Table 5.4 Fluorescence spectroscopy of the Rna15-Rna15 tetramer with RNA. The sequence for each oligonucleotide employed in fluorescence titrations with Rna14-Rna15 protein is shown. The measured dissociation constant for each is also demonstrated.

The concentration of the Rna14-Rna15 protein employed in fluorescence titrations should result in 74% of tetramer within the sample given its K_a. The large proportion of Rna14-Rna15 tetramer and lack of increase in RNA binding affinity suggests that the tetramer does not bind any of these oligonucleotides as hypothesised. However, the measured affinity of the Rna14-Rna15 tetramer for the 5'-GUGUGU-3' oligonucleotide demonstrates only a slight decrease in binding affinity than was observed with the Rna15 (2-103) monomer. It is possible that the orientations of the RRMs of Rna15 are such that RNA binding across the two RRMs as hypothesised is obstructed resulting in two RNA oligonucleotides binding to one tetramer. Indeed, electron microscopy of the tetramer resulted in various conformations of the tetramer at the Rna14-Rna14 interface (107). It is possible that where an RNA oligonucleotide is

present at the optimum length for tetramer binding, the RNA itself helps to rearrange the flexible Rna14-Rna14 interface to anchor the tetramer onto the RNA. The slight decrease observed in binding affinity as the length of the RNA oligonucleotide increases (5-9μM) is likely due to the increased entropy of binding to the longer oligonucleotides. The possibility that the RNA oligonucleotide may also form some type of secondary structure has not been explored thus far. Formation of secondary structure within the oligonucleotides may lead to obstruction of at least one of the GU rich binding sites. Indeed, the A/C rich linker may interact with the GU rich binding site so that Rna15 cannot bind resulting in a lack of increase in binding affinity observed with these oligonucleotides.

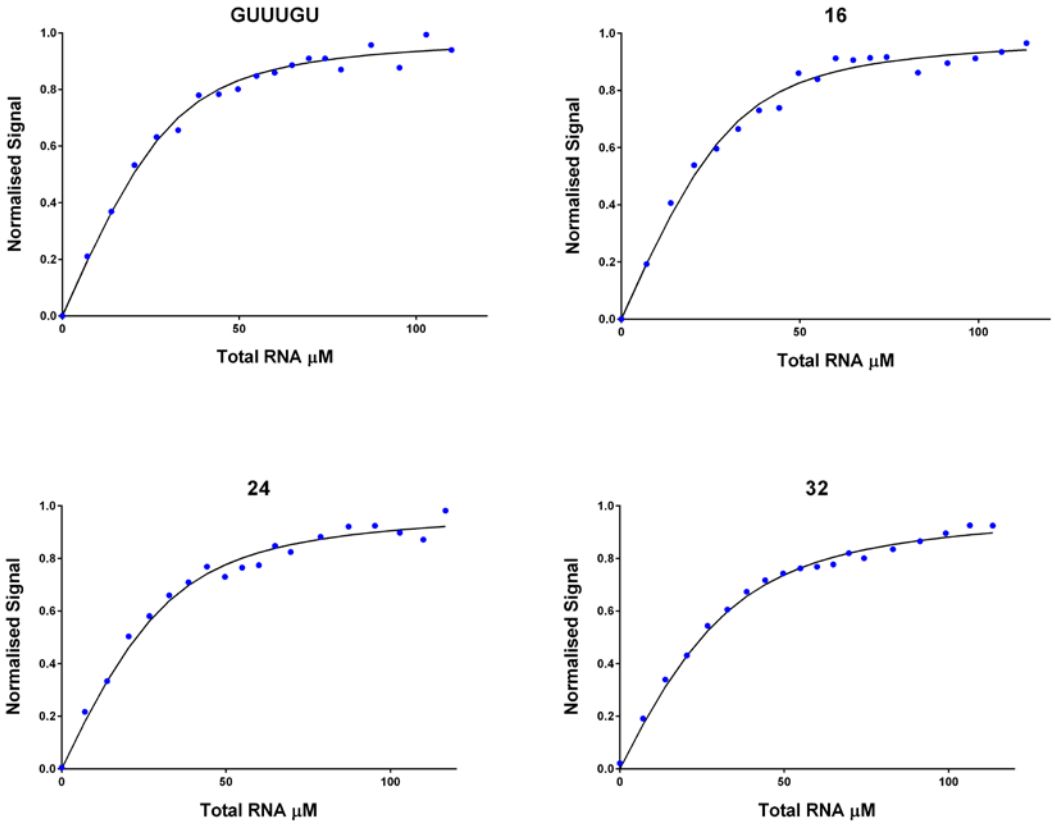


Figure 5.9 Rna14-Rna15-RNA binding curves. The dots represent measured values while the solid line is the fit. Data was fit using a 1:1 binding model (i.e. one Rna15 to one RNA oligonucleotide).

5.5 Deletion of the Rna15 binding interaction domain is lethal in *S. cerevisiae*

The *Δrna14* haploid strain supported by wild type RNA14 present on the pRS316 URA plasmid was employed to study the effects of Rna14 mutation. Previous experiments had demonstrated that the strain was functional and could be used for effective analysis of growth phenotype (chapter 4).

The Rna15 interaction domain of Rna14 in *S. cerevisiae* (residues 626-677) was deleted from the RNA14 HA pRS315-LEU construct. The mutant Rna14 Δ626-677 pRS315-LEU construct was transformed into the Δ*rna14*-RNA14 pRS316-URA and a number of mutant Rna14 Δ626-677 pRS315-LEU transformants were taken for 5-FOA selection (Figure 5.10).

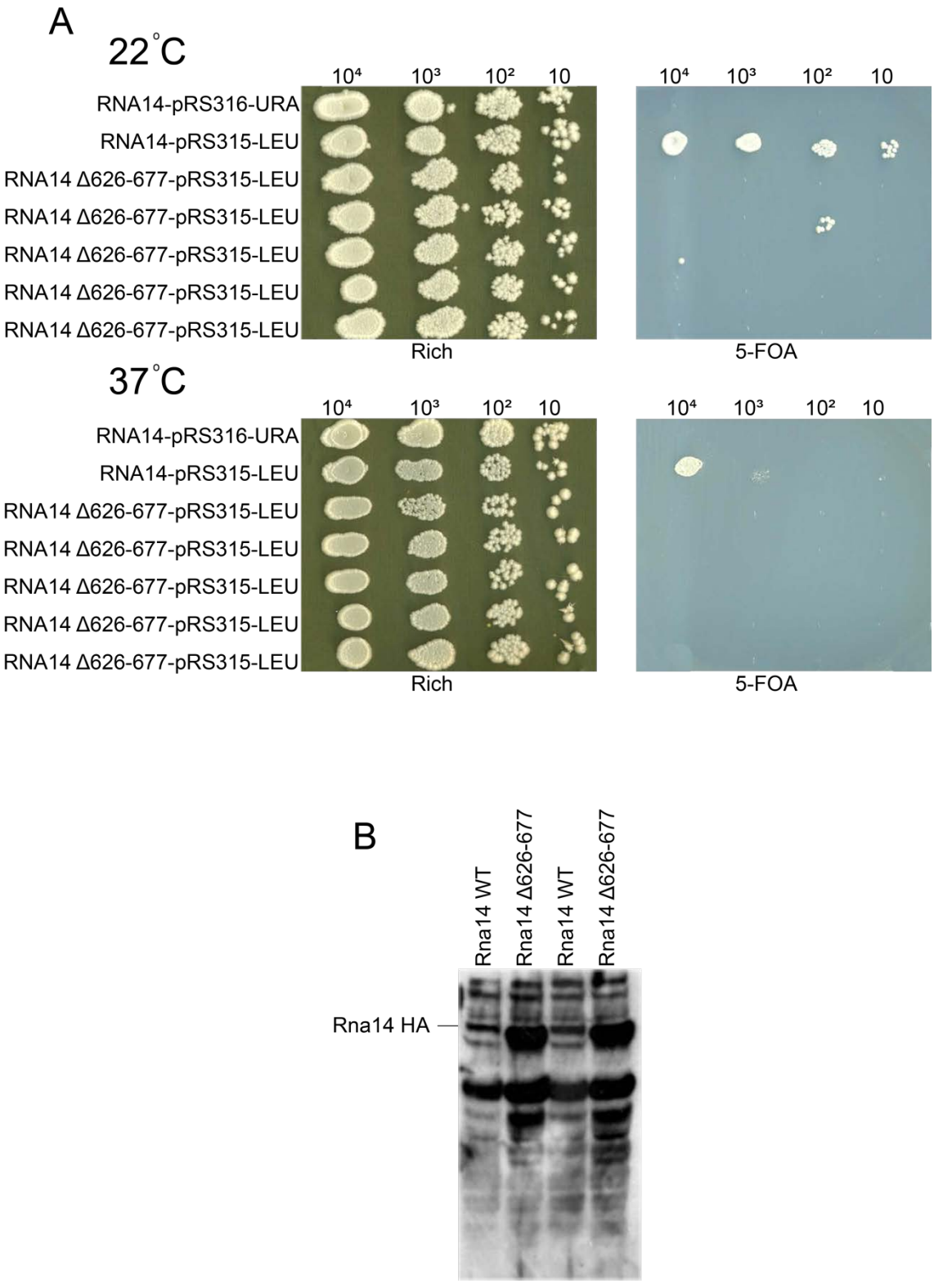


Figure 5.10 Deletion of the Rna15 interaction domain. A. Cells were plated in a dilution series on both rich and minimal media at 22ºC and 37ºC. At 22ºC and 37ºC, this mutant is inviable. **B. Western analyses of Rna14 wild type and Rna14 Δ626-677.** Both Rna14 wild type and mutant proteins are HA tagged allowing detection by western analysis. In this particular experiment there is high background making identification of both wild type and mutant difficult.

The functionality of the 5-FOA selection assay is demonstrated by the growth of the positive Rna14 wildtype LEU^+ and the inviability of the strain containing the RNA14-pRS316-URA construct (Figure 5.10, A). Deletion of the Rna15 interaction domain in Rna14 (Δ626-677) is lethal in *S. cerevisiae*. From the 5 transformants tested none were viable on 5-FOA at either 22ºC or 37ºC. Growth is observed in one mutant at one dilution point on 5-FOA at 22ºC. This is most likely due to cross contamination with the wild type as growth is not observed at the higher concentrations within the dilution series. At 37ºC, growth of wild type *S. cerevisiae* is inhibited on minimal media containing 5-FOA. However growth is observed at both the 10^4 and 10^3 dilution points and is not observed in any of the mutants. This demonstrates that the mutation does result in inviabililty in comparison with wild type as is observed at 22ºC.

In order to determine whether it is the mutation itself that results in inviability and not differences in expression levels in comparison with wild type, western blot analysis was carried out. This experiment exploited the use of the HA tag present on the C-terminal end of the protein. As the Rna14 Δ626-677 mutant is inviable, transformants containing both the HA tagged mutant and untagged Rna14-pRS316-URA constructs were taken for western analysis. However, antibodies employed in the western analysis only recognise the HA tag and therefore only the mutant is detected. The resulting western have a high background but distinct bands are clearly visible in each of the mutant lanes (Figure 5.10, B). Comparison of these bands between wild type and mutant illustrate a difference in expression between the wild type and mutant as other corresponding background bands are of similar intensity. It is possible that this difference in expression is a factor in the deletion mutant lethality. However, the appearance of a second set of bands that resolves further down the gel make it difficult to distinguish which represent Rna14 protein. These bands may represent a degradation product as they have the same molecular weight. Although the western analysis was

repeated a number of times using fresh materials, varying antibody concentration and load concentration, in all cases the background was high.

5.6 Summary

SIA experiments were performed to analyse the nucleotide preference of Rna15. The results provided a clear bias of Rna15 toward G and U nucleotides. Taken together the fluorescence and SIA results suggest that binding to G/U rich sequences is driven by site I residues, Y27 and R87 and by a charge-charge interaction with K90. Further NMR analysis employing labelled and unlabelled RNA oligonucleotides aimed to characterise the orientation of RNA over the face of the RRM. The 5' terminus of the RNA oligonucleotide was revealed to be present at the top of the RRM.

Fluorescence experiments aimed to characterise the mode of tetramer RNA binding resulted in binding affinities similar to those observed with monomer Rna15 (2-103). These results demonstrate that the Rna14-Rna15 tetramer is unable to bind the oligonucleotides in the way hypothesised previously. This result does not exclude the possibility that the Rna14-Rna15 tetramer may bind the oligonucleotide in this way. However, these particular RNA oligonucleotides used are unable to bind in this manner.

The importance of the Rna14-Rna15 interaction both *in vivo* and *in vitro* has also been demonstrated. Deletion of the Rna15 interaction interface in Rna14 is lethal in *S. cerevisiae*. This was demonstrated by the 5-FOA assay where growth of the mutant was not observed on 5-FOA at 37°C. This result suggests that incorporation of Rna15 in the tetramer structure within CF1A is crucial for cell viability.

6. Discussion

6.1 The importance of the C-terminal domain of Rna15

The work presented in chapter 3 characterised the extensive relationship between Rna15 and RNA. Initial experiments aimed to establish whether the region immediately C-terminal to the RRM of Rna15 was capable of regulating Rna15-RNA binding. Residues 95-103 form a short α-helix that has been shown to interact with residues present on the face of the RRM and in human homologue, CstF-64, a C-terminal region has also been shown to pack against the RRM (141,299). It has been hypothesised that this interaction forms the basis of a type of regulatory RNA recognition interaction whereby RNA binding displaces the C-terminal helix allowing access to key RNA binding residues on the RRM (299). Measured binding affinities for both Rna15 2-94 and Rna15 2-103 constructs are essentially the same and therefore it would appear that the C-terminal region of Rna15 does not impede interaction of the 5'-TET-UGUUGU-3' RNA oligonucleotide. However, unlike CstF-64, the C-terminal region of Rna15 is highly unstable and adopts multiple conformations as demonstrated by NMR experiments and the crystal structure (141). In addition NMR experiments demonstrated the transient nature of the interaction between RRM and the C-terminal domain. Therefore, due to the flexibility of the C-terminal region and the weak interaction it makes with the RRM it is highly unlikely that any effect of the C-terminal region would be observed in experiments presented in chapter 3 where a high affinity for RNA is displayed. It is still possible that the C-terminal region does maintain a regulatory role in Rna15-RNA binding but experiments presented here do not provide evidence that such an interaction exists.

The measured binding affinity for the full length Rna15 construct is comparable to the measured binding affinity of both the Rna15 (2-94) and Rna15 (2-103) constructs

demonstrating that residues C-terminal to the RRM and short α-helix do not directly bind RNA. However, residues between 243 and 296 are required for interaction with Pcf11 and this interaction is important for 3’ end processing (140). Deletion of the entire C-terminal domain resulted in viability and subsequent RNA processing experiments employing cell extracts from the viable mutant demonstrated a reduction in polyadenylation at 3' ends (140). Despite the requirement for the full length Rna15 protein for wild type levels of polyadenylation *in vivo*, only the RRM is required for direct RNA binding.

The results provide sufficient evidence that residues 16-94, comprising the RRM domain, are required to mediate Rna15-RNA binding. Previous studies published demonstrate that a Rna15 construct including residues 16-111 binds the 5' TET-UGUUGU-3' RNA oligonucleotide with a dissociation constant of 4.7μM (141). This binding affinity is comparable to that of other constructs presented in this study and further enforces the argument that the RRM alone mediates RNA binding.

6.2 The importance of residues within the RRM of Rna15

In this study mutation of residues across the entire surface of the RRM were shown to have varying effects on RNA binding *in vitro*. These results demonstrate the extent to which residues across the surface of the RRM are involved in direct RNA binding. Previously published results have demonstrated the impact of mutation within the RRM on RNA binding, processing activity and *S. cerevisiae* viability. Mutation of residues F63 and F66 to alanine were reported to have severe detrimental effects on binding to A rich elements in the presence of increasing amounts of salt (KCl) and demonstrated an inability to direct cleavage and polyadenylation on an RNA precursor (139). In addition, transformation of this mutant into *S. cerevisiae* resulted in lethality,

however, western analysis demonstrated that the Rna15 mutant was still able to interact with other components of CF1A (139). Following these results, the crystal structure of the RRM demonstrated that only F63 is present on the surface of the RRM with F66 partially buried. It is possible that this severe decrease in RNA binding and apparent lethality may be due to a decrease in stability of the protein and is not directly due to a decrease in Rna15-RNA binding affinity. In addition, NMR spectroscopy experiments reported that residues between Y60-F63 were perturbed upon introduction of the UGUUGU ribooligonucleotide and not F66 (141). Indeed, double mutation of Y60A/F63A did result in a 6-fold decrease in RNA binding. Mutation of F63 appeared to have more of an effect on RNA binding than point mutation of Y61. Therefore, it does appear that residues involved in the so-called third site are required for Rna15 RNA binding.

Previous studies have suggested that R87 mediates contacts with Hrp1 when both are bound to the 3' UTR of gene GAL7 (142). NMR spectroscopy experiments determined that no unambiguous NOEs were observed between Rna15 and Hrp1 (142). Mutational analysis of R87 failed to generate an effect on the phenotype of *S. cerevisiae*. However, a decrease in growth was observed on 1.5% formamide in previously published results (142). Subsequent *in vitro* cleavage and polyadenylation assays revealed that the cleavage reaction and on occasion the polyadenylation step was severely inhibited (142). Further structural results demonstrated that the R87D mutant was incapable of binding RNA while a R87A mutant was still able to interact with the RNA but with less affinity that observed in wild type. Indeed, Hrp1 was observed to "compete" the anchoring RNA away from the Rna15 R87A mutant. This lead to the hypothesis that contacts between Hrp1 and Rna15 allow simultaneous loading of both proteins onto the RNA (142). However the results presented in this study demonstrate

that a large decrease in binding affinity for the 5'-TET-UGUUGU-3' RNA oligonucleotide is attributed to the R87A mutant. Indeed, the crystal structure showing RNA bound at site I demonstrates the extensive contacts mediated by this residue for the RNA (141). This evidence in addition to the lack of unambiguous NOEs between R87 and D193 of Hrp1 demonstrate that this residue is essential to maintain Rna15-RNA contact.

Alongside R87, Y27 forms extensive contacts with RNA in the site I binding pocket (141). Therefore it is not surprising that the Y27A mutation results in a considerable decrease in RNA binding affinity. This result demonstrates that the site I binding pocket is crucial in maintaining wild type binding affinity for the 5' TET-UGUUGU-3' RNA oligonucleotide.

In addition to site I and site III, the site II binding pocket composed predominantly of Y21 was also investigated using mutational analysis. Mutation of this Y21 to alanine results in a decrease in affinity demonstrating its involvement in RNA binding. The loss of RNA binding is not as severe as observed in the site I binding pocket. This is not unexpected given the crystal structure demonstrates interaction at Y21 is primarily a stacking interaction and lacks the extensive hydrogen bonds formed in site I. The involvement of Y21 in RNA binding is not a novel finding as in addition to the crystal structure, NMR spectroscopy experiments demonstrated that Y21 along with Y61 and F63 were shown to interact with an A/U rich sequence in the 3' UTR of GAL7.

Mutational analysis of lysine residues present on the surface of the RRM reveal the extent to which they directly interact with RNA. Mutation of K48 displays very little

effect on RNA binding affinity. A slight increase in RNA binding affinity is observed in the K48R mutation. NMR spectroscopy experiments demonstrated that in some structures RNA bound in close proximity to the charged side chain of K48 (142). This may explain the slight increase in RNA binding affinity observed in the K48R mutant. Mutation K59E resulted in a slightly larger decrease than observed in K48E. However, mutation K90E abolished RNA binding measured by fluorescence spectroscopy. Indeed, NMR spectroscopy experiments revealed that K90 is able to make contacts with the phosphate group of a bound RNA nucleotide (142).

Data presented in this study is largely in agreement with published data as to the residues that mediate RNA binding. The results presented here demonstrate that RNA binding involves residues over the entire face of the RRM. Additional NMR spectroscopy experiments revealed that TET label present on the 5' end of the RNA oligonucleotide was present near the site I binding pocket demonstrating the orientation of the RNA oligonucleotide. Mutational analysis studies sufficiently demonstrate the extent of involvement of residues present on the surface of the RRM in RNA binding. It is clear that residues Y27, R87 and K90 are essential to maintain wild type Rna15-RNA binding affinity to a U/G rich RNA oligonucleotide while residues Y21, Y61, F63 and potentially K59 form a more modest binding affinity to this RNA oligonucleotide.

6.3 Mutational analysis of the RRM *in vivo*

Despite the severe effect on RNA binding affinity demonstrated by Rna15 RRM point mutations *in vitro*, these mutations fail to exert any detrimental effect on growth phenotype of *S. cerevisiae*. Transformation of all point mutations resulted in growth comparable to wild type. This is possibly not unexpected in those mutations that fail to severely impair RNA binding affinity but it is surprising that mutations such as Y27A,

R87A and K90E do not inhibit growth. It has been demonstrated that transformation of an Rna15 R87D mutation resulted in slow growth on 1.5% formamide (142). However, defects in growth of all Rna15 point mutants were not observed on 1.5% formamide (v/v) in this study. A triple mutant construct, Y27A, R87A and K90E failed to produce any effect on the growth phenotype of *S. cerevisiae*. Only deletion of the RRM (Δ16-94) resulted in lethality on 5-FOA at 37ºC and severely inhibited growth in rich media at 30ºC. These results are surprising given the considerable RNA binding differences observed in the *in vitro* experiments. However, investigation into transcription and site selection of a subset of genes using Q-RT PCR and RNA sequencing provides an explanation as to why a growth defect is not observed. Transcription rate is affected in all genes investigated (ACT1, CYC1, ADH1, YPT1 and TDH2) by Q RT-PCR. Transcription rate of both TDH2 and ADH1 appear to decrease by around 40% of expression observed in wild type. A slightly less obvious effect is observed in ACT1 and YPT1 where only a slight decrease in expression is apparent while a slight increase in expression is apparent in CYC1. Despite these differences from wild type expression, all genes are still expressed and levels are not significantly different to that of wild type. In addition, investigation into poly(A) site readthrough revealed that selection of the proximal site in all housekeeping genes occurs in the Δ16-94 mutant albeit with a slightly higher or lower frequency depending on the gene. This is also true for distal sites where again selection at these sites occurs with a slightly lower or higher frequency depending on each gene. In general, site selection in ACT1, TDH2, ADH1 and YPT1 is relatively unaffected in the deletion mutant.

Although investigation into these genes form a good basis of analysis as they are housekeeping genes and therefore constitutively expressed, a whole global analysis of polyadenylation allows insight into 3' end processing activity of the entire

transcriptome. In this study, expression libraries for wild type and the deletion mutant were generated to allow a global analysis of transcription of all genes. The library sequenced was directional (sequenced in only antisense direction) and so genes transcribed on the antisense or sense strand could be distinguished. This allowed analysis of amount of mRNA in the 3' UTR of genes. Levels of mRNA varied between genes in the same class and also varied between classes demonstrating no pattern in readthrough or poly(A) site selection. Despite previous experiments detailing inhibition of cleavage and polyadenylation in *in vitro* polyadenylation assays (142), the results presented here appear to demonstrate that deletion of the RRM affects each gene differently and that in genes that are consistently expressed poly(A) site selection is not significantly altered. However, these effects in transcription and poly(A) site selection in the deletion mutant may account for the decrease in growth of the deletion mutant.

The discrepancy between the *in vitro* and *in vivo* results raises questions about the extent to which Rna15 directs cleavage and polyadenylation. Indeed, its primary function appears to be in poly(A) site selection, however, deletion of the entire RRM does not result in severe effects in poly(A) site selection when compared with wild type. However, it remains unclear the extent to which other proteins such as Rna14 and Hrp1 influence site selection during polyadenylation *in vivo* and the target binding site for Rna15 is still debated. These are important details that may lead to a greater understanding as to why severe defects in RNA binding *in vitro* do not translate to lethality in *S. cerevisiae.*

6.4 Nucleotide specificity of Rna15

The Rna15 target sequence within the 3' UTR of genes is still debated. A clear consensus sequence for Rna15 is difficult to identify due to the degeneracy of yeast

polyadenylation signals. However, early SELEX (selection amplification) experiments demonstrated a consensus sequence similar to the binding target of Rna15 homologue CstF-64 (143). The sequence was G/U rich and almost identical in sequence to the consensus sequence generated for CstF-64. However, CstF-64 displayed no affinity for the Rna15 consensus sequence demonstrating the similar but distinct nature of sequence recognition from yeast to mammals (143). Indeed, the crystal structure reveals the molecular basis of G/U selectivity whereby only a G or U nucleotide is able to form the Watson-Crick like hydrogen binding to the backbone of Y27 and I25 (141). Furthermore, the site I binding pocket is conserved through evolution and is evident in the CstF-64 homologue of Rna15 thus providing another link between yeast and mammalian 3' end processing systems. Furthermore, the conservation of G/U base selectivity further enforces the specificity of Rna15 for G/U rich sequences within the 3' UTR. Global sequencing analysis have revealed the appearance of U rich sequences within the 3' UTRs of *S. cerevisiae* (145). In addition, insertion of mutation of U rich sequences within the 3' UTRs of CYC1 and ADH1 drastically reduce polyadenylation while introduction of U rich sequences enhance polyadenylation in the GAL7 3' UTR (147,148). SIA results presented in this study further enforce previous finding and demonstrate that Rna15 has an intrinsic preference for G-U nucleotides over A-C nucleotides. In addition, Rna15 association with each pool generates a consensus sequence based entirely on Rna15 nucleotide preference at each position. Accompanying fluorescence spectroscopy experiments demonstrate that Rna15 binds this consensus sequence with a higher affinity.

These results are in disagreement with previously published crosslinking experiments involving the 3' UTR of GAL7 which demonstrate that Rna15 does not show sequence specificity when bound to the 3' UTR. However addition of Rna14 and

Hrp1 was shown to alter Rna15 binding so that it associated with the A-rich positioning element in the 3' UTR of GAL7 (139). Indeed NMR spectroscopy experiments demonstrated the ability of Rna15 to associate with the positioning element of GAL7 when Hrp1 is also associated with the efficiency element (142). Prior to addition of Hrp1, NMR spectra for Rna15 demonstrated weak binding and appeared to associate at different places on the anchor RNA made up of the positioning element and efficiency element found in the 3' UTR of GAL7 (142). However, the NMR structure of the GAL7 anchor RNA on the face of the RRM revealed that chemical shifts corresponding to Y21, Y61 and F63 were perturbed upon addition of the RNA (142). Indeed, Y21 is shown to interact with RNA in the crystal structure and unlike site I does not mediate G/U selectivity as RNA binding at this site predominantly a stacking interaction against the aromatic side chain. Therefore it is possible that if Rna15 associates weakly with A rich positioning elements found in the 3' UTRs of genes then it is possible that deletion of the RRM would not have a significant effect on transcription and 3' end processing as is observed in Q-RT PCR and RNA sequence results.

However, G/U selectivity is an intrinsic feature of Rna15 and conserved through evolution and therefore an important feature of the RRM. The evidence presented provides unambiguous evidence that Rna15 associates with G/U rich sequences within the 3' UTR of genes. Although, the differences in transcription and poly(A) site selection do not appear significantly altered in the deletion mutant when compared with wild type, growth of the mutant is severely impaired at 30ºC in rich media. Furthermore, deletion of the RRM is lethal on 5-FOA at 37ºC. This demonstrates that loss of RNA binding does affect growth phenotype of *S. cerevisiae*. It is possible that discrepancies between *in vitro* and *in vivo* results presented in this work are due to other protein factors involved in poly(A) site selection during 3' end processing events.

6.5 The influence of Rna14 on Rna15-RNA binding

Deletion of the Rna14 "monkey-tail" interaction domain resulted in lethality of yeast at both 22ºC and 37ºC. This result demonstrates that the Rna14-Rna15 interaction is essential for cell viability. Rna14 and Rna15 associate together to form a heterodimer (107,108,137). The NMR structure of the heterodimer demonstrates the formation of a Rna15 helical bundle that an Rna14 monkeytail is able to wrap round (137). Within CF1A, two heterodimers were shown to associate into a tetramer via a Rna14-Rna14 homodimer (107,108). Further contacts have been demonstrated whereby Rna14 contacts Hrp1 and acts as a scaffold between Hrp1 and Rna15 (139). Association of Rna15 with RNA was shown to tether the CF1A complex to the 3' UTR during 3' end processing. However, deletion of the RRM does not demonstrate such a severe phenotype as is observed when the Rna14-Rna15 interaction is disrupted. This result demonstrates that deletion of the RRM alone is not enough to dissociate CF1A from the RNA. However, deletion of the Rna14-Rna15 interface could potentially lead to loss of CF1A interaction at the 3' UTR. Indeed, Rna14 has been shown to associate with Hrp1 in the Rna14-Rna15-Hrp1 scaffold complex, however, it is possible that this interaction is not strong enough to keep CF1A anchored at the 3' end. Also, it is possible that loss of the Rna14-Rna15 interaction destabilises the Rna14-Hrp1 interaction. Rna14 and Rna15 form an extensive interaction with many contacts between the two proteins resulting in a tight interaction, K_d 1.6 μM. The formation of two Rna14-Rna15 interaction domains in the tetramer results in a tight association of CF1A with RNA where two RRMs present in the tetramer contacts the RNA. Deletion of the Rna15 binding interface in Rna14 potentially results in loss of CF1A at the 3' UTR so that other protein factors, holo-CPF, may not be recruited effectively inhibiting both the

cleavage and polyadenylation reactions in all genes. This global affect on the 3' end processing explains the lethality observed in *S. cerevisiae*.

In vitro experiments aimed to characterise the effect of the tetramer on RNA binding did not produce the desired effect. The formation of the Rna14-Rna15 (2:2) tetramer was expected to result in an increase in RNA binding affinity due to the presence of the two RRMs belonging to each of the two Rna15 monomers. However, this was not observed in experiments presented in this study but previously published results demonstrate that disruption of the Rna14-Rna14 interface impairs RNA binding (108). An increase in affinity may not be observed in these studies due to flexibility at the Rna14-Rna14 interface meaning that the RNA oligonucleotide employed in this study were not of optimum length to sufficiently anchor both RRMs of Rna15 in the correct orientation to facilitate binding of one RNA oligonucleotide to one tetramer complex.

6.6 Concluding remarks

In general, the results demonstrated in this thesis show that Rna15 has an intrinsic ability to select G/U nucleotides over A/C nucleotides within the 3' UTR. However, it is possible that in genes such as GAL7, where there is a lack of G/U rich sequences, Hrp1 may influence Rna15 binding through the Rna14 mediated scaffold. Hrp1 has been shown to bind a region in the 3' UTR of genes termed the efficiency element in a number of previously published results demonstrating a concise site for binding (139,142). Hrp1 association with this target site using both RRMs may provide functional redundancy within the Rna15 Δ16-94 mutant where the Rna14-Rna15 interaction is still intact. However, further experiments involving the Rna14-Rna15-Hrp1 in complex with both G/U and A/C rich sequences is required to elucidate whether Hrp1 is able to influence Rna15-RNA binding. A combination of NMR spectroscopy

and fluorescence spectroscopy using long RNA oligonucleotides would aim to elucidate the mode of binding of the Rna14-Rna15-Hrp1 complex. Also, inclusion of A rich oligonucleotides would reveal whether an increase in RNA binding affinity is observed as has been previously stated. In addition a more global analysis of 3' end processing within the Δ16-94 is required in order to determine whether a more severe effect is observed in processing of other genes than those analysed in this study. RNA sequencing of the 3' UTRs of genes only would provide an in depth analysis and precise positioning of the actual polyadenylation event. By sequencing only the 3' UTR of genes in a method described in references 293 and 294 the coverage is such that the polyadenylation event can be precisely mapped. In this way, the effect on processing and in particular poly(A) site selection can be clearly shown.

7.0 Appendix

7.1 Oligonucleotide primers for PCR

7.1.1 LIC primers for cloning *in vitro* constructs

Primer Name	Sequence (5'-3')
Rna15 2 F	**CAGGGACCCGGT**AATAGGCAGAGCGGTGTGAAT
Rna15 94 R	**GGCACCAGAGCGTTA**GGAGTAACCGCATTTCAAAAAC
Rna15 103 R	**GGCACCAGAGCGTTA**TGAAACTCCCGATATATCACT
Rna15 296 R	**GGCACCAGAGCGTTA**AAATGCACCAAATTCTCCCCTTAA

7.1.2 Primers for cloning *in vivo* constructs

Primer Name	Sequence (5'-3')	Restriction site
Rna15 F	GCTGGCG**CCGCGG**ACTTTTGTTCTTTTCTAGATTC	*SacII*
Rna15 R	CGCCAGC**TCTAGA**AAATGCACCAAATTCTCCCCTTAATG	*Xba1*
Rna15 R + HA	CGCCAGC**CTGCAG**AAATGCACCAAATTCTCCCCTTAATG	*Pst1*
Rna14 F	GCTGGCG**CCGCGG**GTACATATATTTTTATGTAAG	*SacII*
Rna14 R	CGCCAGC**GCTAGC**ACCTGACTTGGTGCTCTCAACTG	*Nhe1*
Rna14 R + HA	CGCCAGC**CTGCAG**ACCTGACTTGGTGCTCTCAACTG	*Pst1*

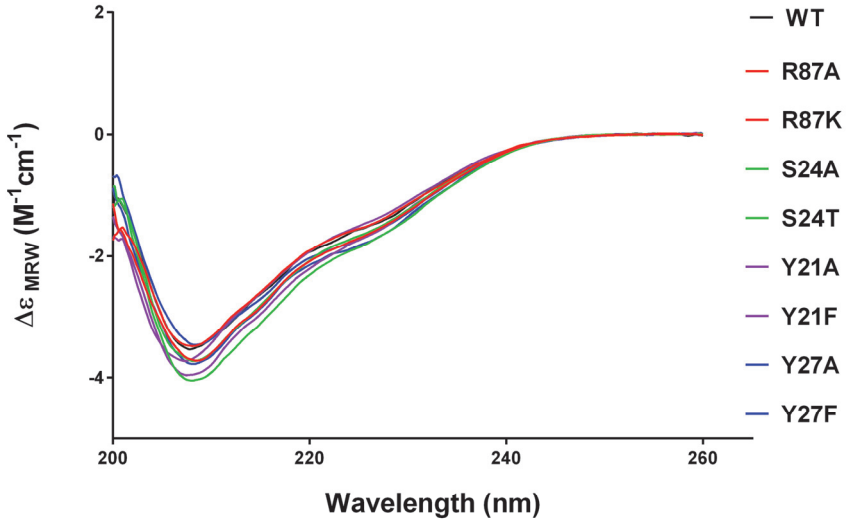
7.1.3 Primers for mutagenesis

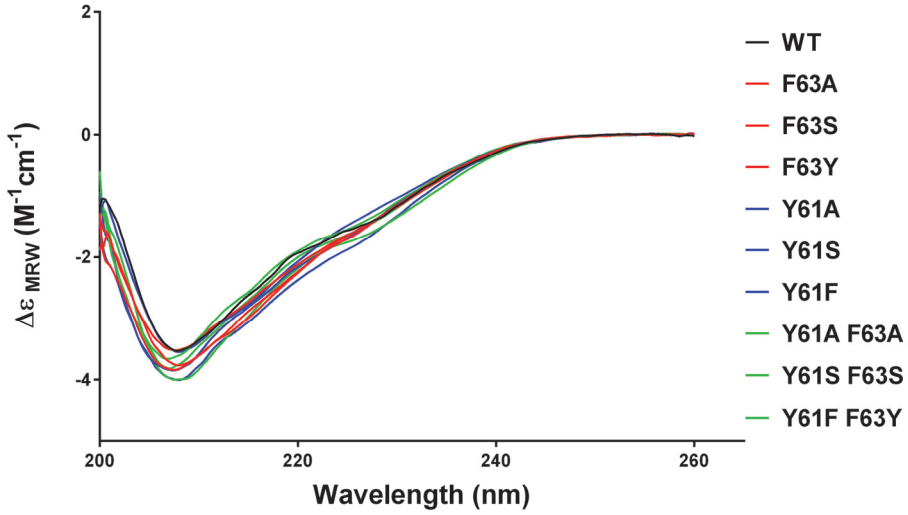
Primer Name	Sequence (5'-3')
Y27F F	CTGGGTTCTATACCA**TTC**GATCAAACAGAGGAG
Y27F R	CTCCTCTGTTTGATC**GAA**TGGTATAGAACCCAG
Y27A F	GTATCTGGGTTCTATACCA**GCG**GATCAAACAGAGGAGCAG
Y27A R	CTGCTCCTCTGTTTGATC**CGC**TGGTATAGAACCCAGATAC
R87A F	GGATACCAATTAGGCTCT**GCG**TTTTTGAAATGCGGTTAC
R87A R	GTAACCGCATTTCAAAAA**CGC**AGAGCCTAATTGGTATCC
R87K F	GGATACCAATTAGGCTCT**AAA**TTTTTGAAATGCGGTTAC
R87K R	GTAACCGCATTTCAAAAA**TTT**AGAGCCTAATTGGTATCC
S24A F	GAGTGGTGTATCTGGGT**GCG**ATACCATACGATCAAAC
S24A R	GTTTGATCGTATGGTAT**CGC**ACCCAGATACACCACTC
S24T F	CGAGTGGTGTATCTGGGT**ACT**ATACCATACGATCAAAC
S24T R	GTTTGATCGTATGGTAT**AGT**ACCCAGATACACCACTCG
K59R F	CAAACTGGTAGGTCG**CGC**GGGTACGCGTTTATTG
K59R R	CAATAAACGCGTACCC**GCG**CGACCTACCAGTTTG
K59E F	CCCAAACTGGTAGGTCG**GAA**GGGTACGCGTTTATTG
K59E R	CAATAAACGCGTACCC**TTC**CGACCTACCAGTTTGGG
K90A F	GGCTCTAGGTTTTTG**GCG**TGCGGTTACTCCAGC
K90A R	GCTGGAGTAACCGCA**CGC**CAAAAACCTAGAGCC
K90R F	GGCTCTAGGTTTTTG**CGC**TGCGGTTACTCCAGC
K90R R	GCTGGAGTAACCGCA**GCG**CAAAAACCTAGAGCC
K90E F	GGCTCTAGGTTTTTG**GAA**TGCGGTTACTCC
K90E R	GGAGTAACCGCA**TTC**CAAAAACCTAGAGCC
Y21A F	CACCATCCCGAGTGGTG**GCG**CTGGGTTCTATACCATAC
Y21A R	GTATGGTATAGAACCCAG**CGC**CACCACTCGGGATGGTG
Y21F F	CATCCCGAGTGGTG**TTT**CTGGGTTCTATAC

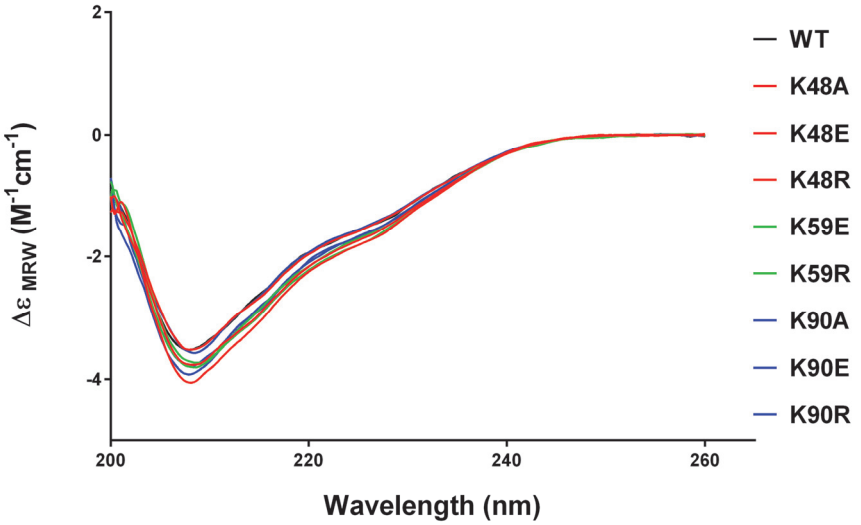
7.1.3 Primers for mutagenesis cont.

Primer Name	Sequence (5'-3')
Y61F F	CTGGTAGGTCGAAAGGG**TTC**GCGTTTATTGAATTTAG
Y61F R	CTAAATTCAATAAACGC**GAA**CCCTTTCGACCTACCAG
Y61A F	CTGGTAGGTCGAAAGGG**GCC**GCGTTTATTGAATTTAG
Y61A R	CTAAATTCAATAAACGC**GGC**CCCTTTCGACCTACCAG
F63A F	GTAGGTCGAAAGGGTACGCG**GCG**ATTGAATTTAGAGATTTAG
F63A R	CTAAATCTCTAAATTCAAT**CGC**CGCGTACCCTTTCGACCTAC
F63Y F	CGAAAGGGTACGCG**TAT**ATTGAATTTAGAG
F63Y R	CTCTAAATTCAAT**ATA**CGCGTACCCTTTCG
Y61F/F63Y F	TGGTAGGTCGAAAGGG**TTC**GCG**TAT**ATTGAATTTAGAGATTTAG
Y61F/F63Y R	CTAAATCTCTAAATTCAAT**ATA**CGC**GAA**CCCTTTCGACCTACCA
Y61A/F63A F	TGGTAGGTCGAAAGGG**GCG**GCG**GCG**ATTGAATTTAGAGATTTAG
Y61A/F63A R	CTAAATCTCTAAATTCAAT**CGC**CGC**CGC**CCCTTTCGACCTACCA
K48R F	CCGTGATCAATTT**GAG**AATGATGTTCGACCC
K48R R	GGGTCGAACATCATT**CTC**AAATTGATCACGG
K48A F	CCCGTGATCAATTT**GGC**GATGATGTTCGACCCCC
K48A R	GGGGGTCGAACATCATC**GCC**AAATTGATCACGGG
K48E F	CCCGTGATCAATTT**GGA**AATGATGTTCGACC
K48E R	GGTCGAACATCATT**TCC**AAATTGATCACGGG
Y61S/F63S F	GGTAGGTCGAAAGGG**TCG**GCG**TCG**ATTGAATTTAGAGATTTAG
Y61S/F63S R	CTAAATCTCTAAATTCAAT**CGA**CGC**CGA**CCCTTTCGACCTACC
Y61S F	GGTAGGTCGAAAGGG**TCG**GCGTTTATTGAATTTAG
Y61S R	CTAAATTCAATAAACGC**CGA**CCCTTTCGACCTACC
F63S F	GTCGAAAGGGTACGCG**TCG**ATTGAATTTAGAGATTTAG
F63S R	CTAAATCTCTAAATTCAAT**CGA**CGCGTACCCTTTCGAC
R87A/K90E F	GGATACCAATTAGGCTCT**GCG**TTTTTG**GAA**TGCGGTTACTCCAGCAATAG
R87A/K90E R	CTATTGCTGGAGTAACCGCA**TTC**CAAAAA**CGC**AGAGCCTAATTGGTATCC
Δ 16-94 F	AGCAATAGTGATATATCGGGAGTTTC
Δ 16-94 R	TGGGTTGTTCTGTACCCCCGCATTC
Rna14 Δ 626-677 F	CTGCAGAGCGGCCGCATCTTTTAC
Rna14 Δ 626-677 R	GAATTTCTTATTGGAGAAGTTTC

7.2 Circular dichroism of Rna15 2-103 mutant constructs

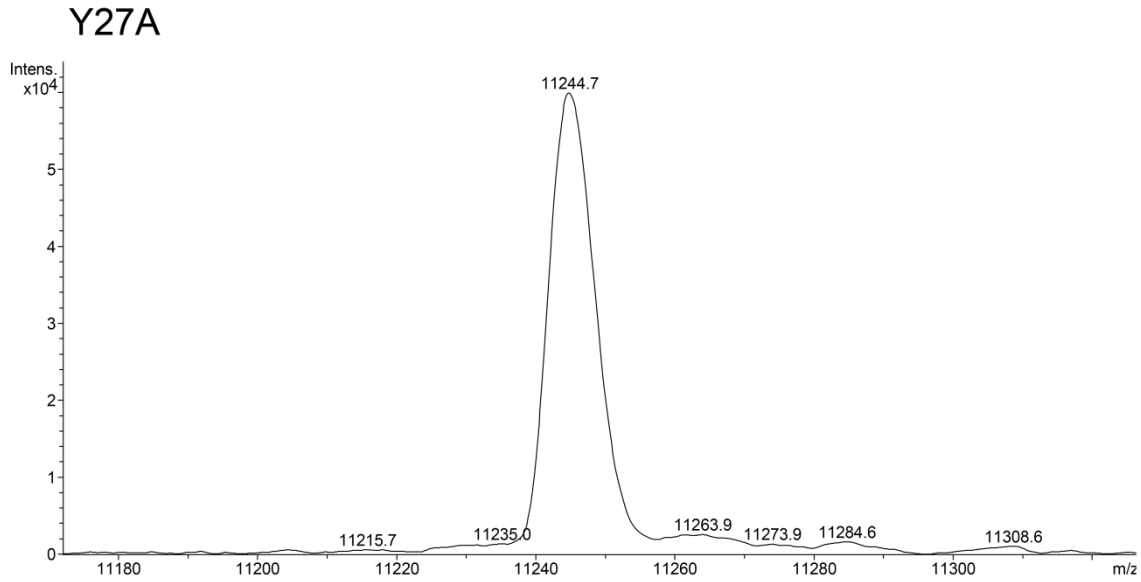


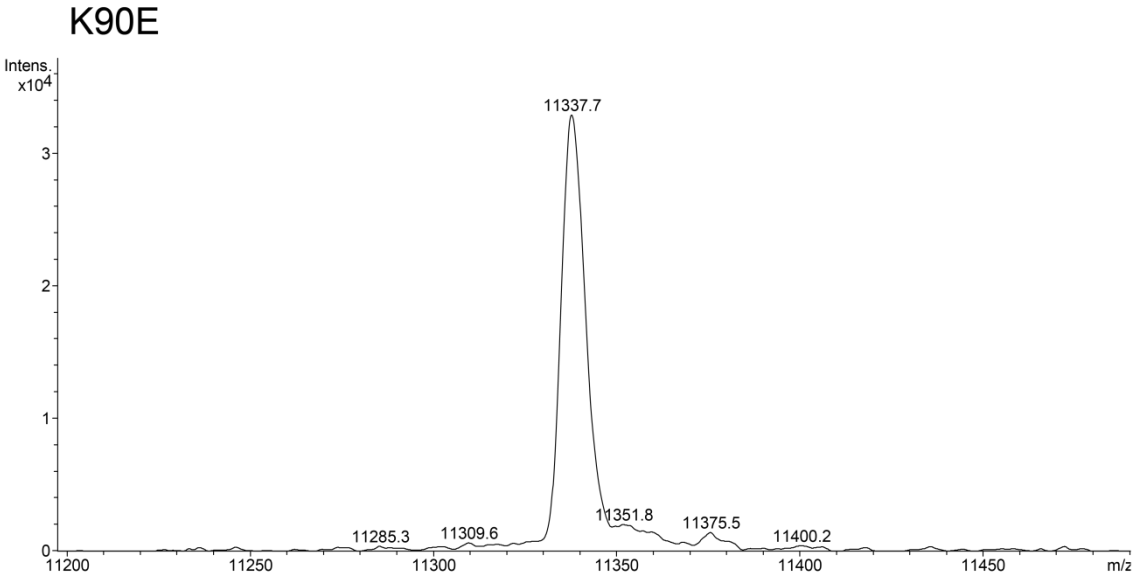




7.3 Mass Spectrometry of Rna15 and mutant constructs

7.3.1 Representative deconvoluted mass spectrum report

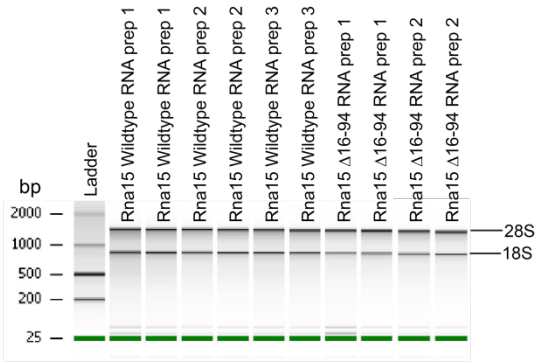


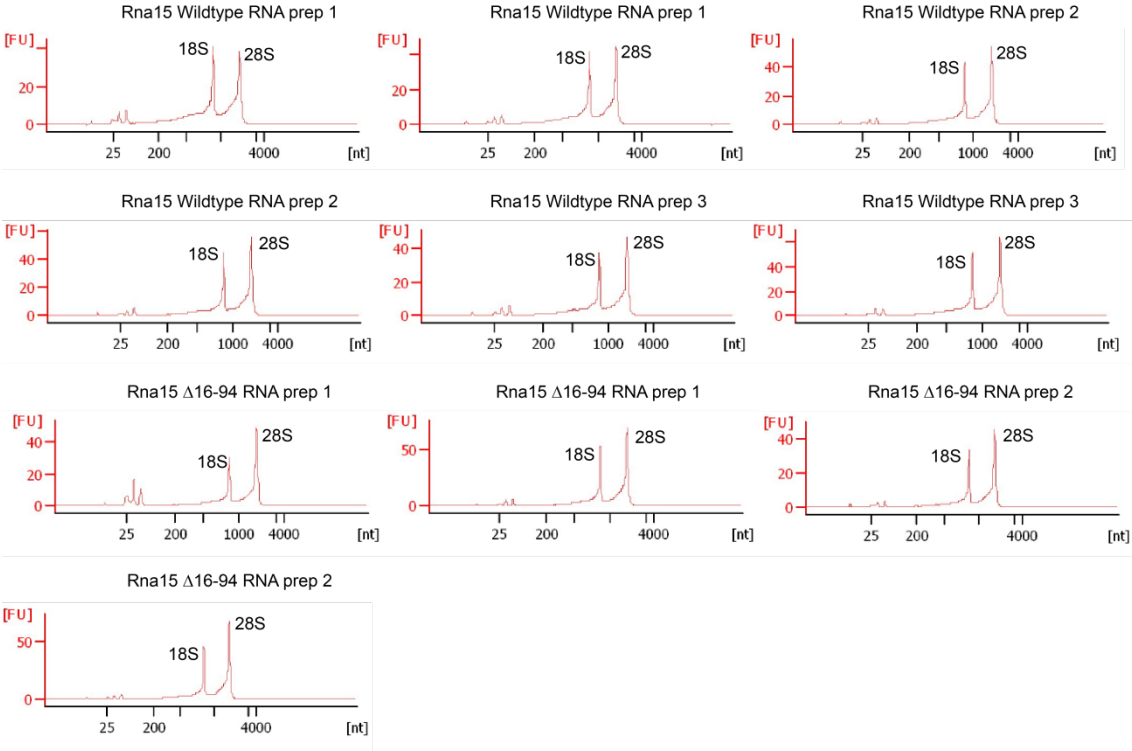


7.3.2 Mass spectrometry results for Rna15 and mutant constructs

Rna15 2-103 construct	Estimated molecular weight (Da)	Mass Spectrometry derived molecular weight (Da)
Wild type	11337.6	11337.1
R87K	11309.6	11309.1
R87A	11252.5	11252.1
Y27F	11321.6	11320.8
Y27A	11245.5	11244.7
S24T	11351.6	11351.0
S24A	11321.6	11321
Y21F	11321.6	11320.8
Y21A	11245.5	11244.9
F63Y	11353.6	11352.6
F63A	11261.5	11251.3
F63S	11277.5	11276.8
Y61F	11321.6	11321.2
Y61A	11245.5	11245.1
Y61S	11261.5	11260.9
Y61F, F63Y	11337.6	11337.0
Y61A, F63A	11169.4	11168.8
Y61S, F63S	11201.4	11200.8
K48E	11338.5	11337.8
K48R	11365.6	11365.0
K48A	11280.5	11280.0
K59E	11338.5	11337.8
K59R	11365.6	11365.0
K90E	11338.5	11337.7
K90R	11365.6	11364.9
K90A	11280.5	11279.6

7.4 Analysis of RNA quality





7.5 Primers and probes employed in qRT-PCR experiments

Primer Name	Sequence (5'-3')	Tm (°C)
ACT1 F	CAACAAATGTGGATCTCAAAACAAG	58.1
ACT1 R	CATAAACATACGCGCACAAAAGCA	59.3
TDH2 F	CATCGTTGATGGTCACAAGATC	58.4
TDH2 R	TGGAGTCAATGGCGATGTCAA	57.9
ADH1 F	CAGAGCTGACACCAGAGAAG	59.3
ADH1 R	GTAAATTTCTGGCAAGGTAGACAAG	59.7
CYC1 F	GTACACAGATGCCAATATCAAGA	57.1
CYC1 R	AACCCACCAAAGGCCATCTT	57.3
YPT1 F	ACACACGCGAGAACATATATACA	57.1
YPT1 R	GCATTAATTGCTGTGGCAGCT	57.9
SNR52 F	TTTCAGAAGGAAGGCAACATAAGT	57.6
SNR52 R	TACTATGATGAATGACATTAGCGTG	58.1

Probe Name	Dye/quencher	Sequence (5'-3'	Tm (°C)
ACT1	6-FAM/MGB	CGAAAGTGGTCCATCTATCGTTCACCACAAGT	68.2
TDH2	6-FAM/MGB	CCAAGAAAGAGACCCAGCTAACTTGCCATG	68
ADH1	6-FAM/MGB	CTTCTTCGCCAGAGGTTTGGTCAAGTCTCCAA	69.5
CYC1	6-FAM/MGB	CATGTCAGAGTACTTGACTAACCCAAAGAAATATATTCC	67.4
YPT1	6-FAM/MGB	CAACCTGAAGGGACAGAGTTTAACCAACACCG	67.4
SNR52	6-FAM/MGB	TTTCTAATCCTAAAATCTTTCGATTTTGTATCAGAGATTG	64.3

7.6 C_T and quantity data from qRT-PCR experiments

7.6.1.Rna15 Wild type

Amplicon	C_T Mean (S.D.) Exp 1	C_T Mean (S.D.) Exp 2	C_T Mean (S.D.) Exp 3
ACT1	20.4 (0.05)	20.1 (0.01)	20.8 (0.04)
TDH2	19.16 (0.11)	16.7 (0.05)	19.6 (0.03)
ADH1	15.8 (0.02)	16.1 (0.1)	14.5 (0.3)
CYC1	21.7 (0.21)	22 (0.3)	23.1 (0.14)
YPT1	23.6 (0.02)	23.8 (0.05)	23.4 (0.15)
SNR52	15.1 (0.26)	17.2 (0.25)	14.8 (0.2)

Amplicon	Quantity Mean (S.D.) Exp 1	Quantity Mean (S.D.) Exp 2	Quantity Mean (S.D.) Exp 3
ACT1	172 (4.7)	229.9 (1.5)	139.1 (2.7)
TDH2	185.2 (1)	214.5 (4.8)	174 (2.2)
ADH1	389.9 (3.7)	458.6 (19.5)	312 (39.4)
CYC1	22.4 (2.2)	37.9 (4.5)	27.2 (1.7)
YPT1	30.2 (0.3)	53.6 (1.2)	34.7 (2.2)
SNR52	911.8 (105.3)	832.1 (82.7)	1029.7 (69.1)

7.6.2.Rna15 Δ16-94

Amplicon	C_T Mean (S.D.) Exp 1	C_T Mean (S.D.) Exp 2	C_T Mean (S.D.) Exp 3
ACT1	20.63 (0.04)	20.3 (0.05)	21 (0.2)
TDH2	20.3 (0.14)	17.9 (0.2)	20.5 (0.02)
ADH1	16.5 (0.06)	16.8 (0.06)	15.2 (0.2)
CYC1	21.2 (0.15)	22 (0.3)	22.5 (0.13)
YPT1	23.7 (0.13)	24 (0.07)	23.5 (0.08)
SNR52	14.9 (0.24)	17.2 (0.22)	14.5 (0.17)

Amplicon	Quantity Mean (S.D.) Exp 1	Quantity Mean (S.D.) Exp 2	Quantity Mean (S.D.) Exp 3
ACT1	149.8 (3.4)	208.3 (4.7)	124.4 (9.6)
TDH2	107.5 (7.5)	132.5 (8.1)	114.8 (0.9)
ADH1	278.5 (8.2)	346.1 (9)	219 (18.4)
CYC1	27.2 (1.9)	47.4 (5.6)	35.5 (2)
YPT1	29 (1.8)	49.5 (1.4)	33.9 (1.1)
SNR52	983.9 (107.9)	825.1 (68.6)	1161.4 (76.7)

8.0 Bibliography

1. Hsu CL, Stevens A. Yeast cells lacking 5'-->3' exoribonuclease 1 contain mRNA species that are poly(A) deficient and partially lack the 5' cap structure. Mol Cell Biol. 1993 Aug;13(8):4826–35.

2. Walther TN, Koning THW, Schümperli D, Müller B. A 5'-3' exonuclease activity involved in forming the 3' products of histone pre-mRNA processing in vitro. RNA. 1998 Sep 1;4(9):1034–46.

3. Tarun SZ, Sachs AB. Association of the yeast poly(A) tail binding protein with translation initiation factor eIF-4G. EMBO J. 1996 Dec 16;15(24):7168–77.

4. Wakiyama M, Imataka H, Sonenberg N. Interaction of eIF4G with poly(A)-binding protein stimulates translation and is critical for Xenopus oocyte maturation. Curr Biol. 2000 Sep 14;10(18):1147–50.

5. Hocine S, Singer RH, Grunwald D. RNA Processing and Export. Cold Spring Harb Perspect Biol [Internet]. 2010 Dec [cited 2014 Feb 5];2(12). Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2982171/

6. Shuman S, Hurwitz J. Mechanism of mRNA capping by vaccinia virus guanylyltransferase: characterization of an enzyme--guanylate intermediate. Proc Natl Acad Sci U S A. 1981 Jan;78(1):187–91.

7. Shuman S. Capping enzyme in eukaryotic mRNA synthesis. Prog Nucleic Acid Res Mol Biol. 1995;50:101–29.

8. Shatkin AJ, Manley JL. The ends of the affair: Capping and polyadenylation. Nat Struct Mol Biol. 2000 Oct;7(10):838–42.

9. Komarnitsky P, Cho E-J, Buratowski S. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes Dev. 2000 Oct 1;14(19):2452–60.

10. Hsin J-P, Manley JL. The RNA polymerase II CTD coordinates transcription and RNA processing. Genes Dev. 2012 Oct 1;26(19):2119–37.

11. Cho E-J, Takagi T, Moore CR, Buratowski S. mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. Genes Dev. 1997 Dec 15;11(24):3319–26.

12. McCracken S, Fong N, Rosonina E, Yankulov K, Brothers G, Siderovski D, et al. 5′-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. Genes Dev. 1997 Dec 15;11(24):3306–18.

13. Yue Z, Maldonado E, Pillutla R, Cho H, Reinberg D, Shatkin AJ. Mammalian capping enzyme complements mutant Saccharomyces cerevisiae lacking mRNA guanylyltransferase and selectively binds the elongating form of RNA polymerase II. Proc Natl Acad Sci U S A. 1997 Nov 25;94(24):12898–903.

14. Cho E-J, Rodriguez CR, Takagi T, Buratowski S. Allosteric interactions between capping enzyme subunits and the RNA polymerase II carboxy-terminal domain. Genes Dev. 1998 Nov 15;12(22):3482–7.

15. Ho CK, Schwer B, Shuman S. Genetic, Physical, and Functional Interactions between the Triphosphatase and Guanylyltransferase Components of the Yeast mRNA Capping Apparatus. Mol Cell Biol. 1998 Sep;18(9):5189–98.

16. Fresco LD, Buratowski S. Active site of the mRNA-capping enzyme guanylyltransferase from Saccharomyces cerevisiae: similarity to the nucleotidyl attachment motif of DNA and RNA ligases. Proc Natl Acad Sci U S A. 1994 Jul 5;91(14):6624–8.

17. Mao X, Schwer B, Shuman S. Mutational analysis of the Saccharomyces cerevisiae ABD1 gene: cap methyltransferase activity is essential for cell growth. Mol Cell Biol. 1996 Feb;16(2):475–80.

18. Schwer B, Shuman S. Mutational analysis of yeast mRNA capping enzyme. Proc Natl Acad Sci U S A. 1994 May 10;91(10):4328–32.

19. Wang SP, Shuman S. Structure-Function Analysis of the mRNA Cap Methyltransferase of Saccharomyces cerevisiae. J Biol Chem. 1997 Jun 6;272(23):14683–9.

20. Wang SP, Deng L, Ho CK, Shuman S. Phylogeny of mRNA capping enzymes. Proc Natl Acad Sci U S A. 1997 Sep 2;94(18):9573–8.

21. Shuman S, Schwer B. RNA capping enzyme and DNA ligase: a superfamily of covalent nucleotidyl transferases. Mol Microbiol. 1995 Aug;17(3):405–10.

22. Håkansson K, Doherty AJ, Shuman S, Wigley DB. X-Ray Crystallography Reveals a Large Conformational Change during Guanyl Transfer by mRNA Capping Enzymes. Cell. 1997 May 16;89(4):545–53.

23. Furuichi Y, Shatkin AJ. Viral and cellular mRNA capping: past and prospects. Adv Virus Res. 2000;55:135–84.

24. Shuman S. Structure, mechanism, and evolution of the mRNA capping apparatus. Prog Nucleic Acid Res Mol Biol. 2001;66:1–40.

25. Cong P, Shuman S. Covalent catalysis in nucleotidyl transfer. A KTDG motif essential for enzyme-GMP complex formation by mRNA capping enzyme is conserved at the active sites of RNA and DNA ligases. J Biol Chem. 1993 Apr 5;268(10):7256–60.

26. Niles EG, Christen L. Identification of the vaccinia virus mRNA guanyltransferase active site lysine. J Biol Chem. 1993 Nov 25;268(33):24986–9.

27. Shibagaki Y, Gotoh H, Kato M, Mizumoto K. Localization and In Vitro Mutagenesis of the Active Site in the Saccharomyces cerevisiae mRNA Capping Enzyme. J Biochem (Tokyo). 1995 Dec 1;118(6):1303–9.

28. Shuman S, Liu Y, Schwer B. Covalent catalysis in nucleotidyl transfer reactions: essential motifs in Saccharomyces cerevisiae RNA capping enzyme are conserved in Schizosaccharomyces pombe and viral capping enzymes and among polynucleotide ligases. Proc Natl Acad Sci U S A. 1994 Dec 6;91(25):12046–50.

29. Higman MA, Christen LA, Niles EG. The mRNA (guanine-7-)methyltransferase domain of the vaccinia virus mRNA capping enzyme. Expression in Escherichia coli and structural and kinetic comparison to the intact capping enzyme. J Biol Chem. 1994 May 27;269(21):14974–81.

30. Mao X, Shuman S. Intrinsic RNA (guanine-7) methyltransferase activity of the vaccinia virus capping enzyme D1 subunit is stimulated by the D12 subunit. Identification of amino acid residues in the D1 protein required for subunit association and methyl group transfer. J Biol Chem. 1994 Sep 30;269(39):24472–9.

31. Yamada-Okabe T, Doi R, Shimmi O, Arisawa M, Yamada-Okabe H. Isolation and characterization of a human cDNA for mRNA 5'-capping enzyme. Nucleic Acids Res. 1998 Apr 1;26(7):1700–6.

32. Lima CD, Wang LK, Shuman S. Structure and mechanism of yeast RNA triphosphatase: an essential component of the mRNA capping apparatus. Cell. 1999 Nov 24;99(5):533–43.

33. Kobor MS, Greenblatt J. Regulation of transcription elongation by phosphorylation. Biochim Biophys Acta. 2002 Sep 13;1577(2):261–75.

34. Schroeder SC, Zorio DAR, Schwer B, Shuman S, Bentley D. A Function of Yeast mRNA Cap Methyltransferase, Abd1, in Transcription by RNA Polymerase II. Mol Cell. 2004 Feb 13;13(3):377–87.

35. Staley JP, Guthrie C. Mechanical Devices of the Spliceosome: Motors, Clocks, Springs, and Things. Cell. 1998 Feb 6;92(3):315–26.

36. Hoskins AA, Moore MJ. The spliceosome: a flexible, reversible macromolecular machine. Trends Biochem Sci. 2012 May;37(5):179–88.

37. Jurica MS, Moore MJ. Pre-mRNA Splicing: Awash in a Sea of Proteins. Mol Cell. 2003 Jul;12(1):5–14.

38. Wahl MC, Will CL, Lührmann R. The Spliceosome: Design Principles of a Dynamic RNP Machine. Cell. 2009 Feb 20;136(4):701–18.

39. Nilsen TW, Graveley BR. Expansion of the eukaryotic proteome by alternative splicing. Nature. 2010 Jan 28;463(7280):457–63.

40. Fabrizio P, Dannenberg J, Dube P, Kastner B, Stark H, Urlaub H, et al. The Evolutionarily Conserved Core Design of the Catalytic Activation Step of the Yeast Spliceosome. Mol Cell. 2009 Nov 25;36(4):593–608.

41. Steitz JA, Dreyfuss G, Krainer AR, Lamond AI, Matera AG, Padgett RA. Where in the cell is the minor spliceosome? Proc Natl Acad Sci U S A. 2008 Jun 24;105(25):8485–6.

42. Lardelli RM, Thompson JX, Yates JR, Stevens SW. Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. RNA. 2010 Mar 1;16(3):516–28.

43. Brow DA. Allosteric Cascade of Spliceosome Activation. Annu Rev Genet. 2002;36(1):333–60.

44. Anderson K, Moore MJ. Bimolecular Exon Ligation by the Human Spliceosome. Science. 1997 Jun 13;276(5319):1712–6.

45. Beyer AL, Osheim YN. Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. Genes Dev. 1988 Jun 1;2(6):754–65.

46. Zhang G, Taneja KL, Singer RH, Green MR. Localization of pre-mRNA splicing in mammalian nuclei. Nature. 1994 Dec 22;372(6508):809–12.

47. David CJ, Boyne AR, Millhouse SR, Manley JL. The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65-Prp19 complex. Genes Dev. 2011 May 1;25(9):972–83.

48. Gu B, Eick D, Bensaude O. CTD serine-2 plays a critical role in splicing and termination factor recruitment to RNA polymerase II in vivo. Nucleic Acids Res. 2013 Feb;41(3):1591–603.

49. McCracken S, Fong N, Yankulov K, Ballantyne S, Pan G, Greenblatt J, et al. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature. 1997 Jan 23;385(6614):357–61.

50. Misteli T, Spector DL. RNA Polymerase II Targets Pre-mRNA Splicing Factors to Transcription Sites In Vivo. Mol Cell. 1999 Jun;3(6):697–705.

51. Hirose Y, Tacke R, Manley JL. Phosphorylated RNA polymerase II stimulates pre-mRNA splicing. Genes Dev. 1999 May 15;13(10):1234–9.

52. Iglesias N, Stutz F. Regulation of mRNP dynamics along the export pathway. FEBS Lett. 2008 Jun 18;582(14):1987–96.

53. Rondón AG, Jimeno S, Aguilera A. The interface between transcription and mRNP export: From THO to THSC/TREX-2. Biochim Biophys Acta BBA - Gene Regul Mech. 2010 Aug;1799(8):533–8.

54. Huertas P, Aguilera A. Cotranscriptionally Formed DNA:RNA Hybrids Mediate Transcription Elongation Impairment and Transcription-Associated Recombination. Mol Cell. 2003 Sep;12(3):711–21.

55. Aguilera A. Cotranscriptional mRNP assembly: from the DNA to the nuclear pore. Curr Opin Cell Biol. 2005 Jun;17(3):242–50.

56. Chavez S, Beilharz T, Rondon AG, Erdjument-Bromage H, Tempst P, Svejstrup JQ, et al. A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in Saccharomyces cerevisiae. EMBO J. 2000 Nov 1;19(21):5824–34.

57. Jimeno S, Rondon AG, Luna R, Aguilera A. The yeast THO complex and mRNA export factors link RNA metabolism with transcription and genome instability. EMBO J. 2002 Jul 1;21(13):3526–35.

58. Huertas P, Garcia-Rubio ML, Wellinger RE, Luna R, Aguilera A. An hpr1 Point Mutation That Impairs Transcription and mRNP Biogenesis without Increasing Recombination. Mol Cell Biol. 2006 Oct;26(20):7451–65.

59. Zenklusen D, Vinciguerra P, Wyss J-C, Stutz F. Stable mRNP Formation and Export Require Cotranscriptional Recruitment of the mRNA Export Factors Yra1p and Sub2p by Hpr1p. Mol Cell Biol. 2002 Dec;22(23):8241–53.

60. Abruzzi KC, Lacadie S, Rosbash M. Biochemical analysis of TREX complex recruitment to intronless and intron-containing yeast genes. EMBO J. 2004 Jul 7;23(13):2620–31.

61. Masuda S, Das R, Cheng H, Hurt E, Dorman N, Reed R. Recruitment of the human TREX complex to mRNA during splicing. Genes Dev. 2005 Jul 1;19(13):1512–7.

62. Zenklusen D, Vinciguerra P, Strahm Y, Stutz F. The Yeast hnRNP-Like Proteins Yra1p and Yra2p Participate in mRNA Export through Interaction with Mex67p. Mol Cell Biol. 2001 Jul;21(13):4219–32.

63. Strasser K, Hurt E. Yra1p, a conserved nuclear RNA-binding protein, interacts directly with Mex67p and is required for mRNA export. EMBO J. 2000 Feb 1;19(3):410–20.

64. Sträßer K, Hurt E. Splicing factor Sub2p is required for nuclear mRNA export through its interaction with Yra1p. Nature. 2001 Oct 11;413(6856):648–52.

65. Sträßer K, Masuda S, Mason P, Pfannstiel J, Oppizzi M, Rodriguez-Navarro S, et al. TREX is a conserved complex coupling transcription with messenger RNA export. Nature. 2002 May 16;417(6886):304–8.

66. Henry M, Borland CZ, Bossie M, Silver PA. Potential RNA binding proteins in Saccharomyces cerevisiae identified as suppressors of temperature-sensitive mutations in NPL3. Genetics. 1996 Jan;142(1):103–15.

67. Duncan K, Umen JG, Guthrie C. A putative ubiquitin ligase required for efficient mRNA export differentially affects hnRNP transport. Curr Biol CB. 2000 Jun 15;10(12):687–96.

68. Hieronymus H, Silver PA. Genome-wide analysis of RNA-protein interactions illustrates specificity of the mRNA export machinery. Nat Genet. 2003 Feb;33(2):155–61.

69. Gwizdek C, Iglesias N, Rodriguez MS, Ossareh-Nazari B, Hobeika M, Divita G, et al. Ubiquitin-associated domain of Mex67 synchronizes recruitment of the mRNA export machinery with transcription. Proc Natl Acad Sci U S A. 2006 Oct 31;103(44):16376–81.

70. Gwizdek C, Hobeika M, Kus B, Ossareh-Nazari B, Dargemont C, Rodriguez MS. The mRNA Nuclear Export Factor Hpr1 Is Regulated by Rsp5-mediated Ubiquitylation. J Biol Chem. 2005 Apr 8;280(14):13401–5.

71. Hobeika M, Brockmann C, Gruessing F, Neuhaus D, Divita G, Stewart M, et al. Structural Requirements for the Ubiquitin-associated Domain of the mRNA Export Factor Mex67 to Bind Its Specific Targets, the Transcription Elongation THO Complex Component Hpr1 and Nucleoporin FXFG Repeats. J Biol Chem. 2009 Jun 26;284(26):17575–83.

72. Qu X, Lykke-Andersen S, Nasser T, Saguez C, Bertrand E, Jensen TH, et al. Assembly of an Export-Competent mRNP Is Needed for Efficient Release of the 3?-End Processing Complex after Polyadenylation. Mol Cell Biol. 2009 Oct;29(19):5327–38.

73. Lei EP, Stern CA, Fahrenkrog B, Krebber H, Moy TI, Aebi U, et al. Sac3 is an mRNA export factor that localizes to cytoplasmic fibrils of nuclear pore complex. Mol Biol Cell. 2003 Mar;14(3):836–47.

74. Gallardo M, Luna R, Erdjument-Bromage H, Tempst P, Aguilera A. Nab2p and the Thp1p-Sac3p complex functionally interact at the interface between transcription and mRNA metabolism. J Biol Chem. 2003 Jun 27;278(26):24225–32.

75. Fischer T, Strasser K, Racz A, Rodriguez-Navarro S, Oppizzi M, Ihrig P, et al. The mRNA export machinery requires the novel Sac3p-Thp1p complex to dock at the nucleoplasmic entrance of the nuclear pores. EMBO J. 2002 Nov 1;21(21):5843–52.

76. Rodríguez-Navarro S, Fischer T, Luo M-J, Antúnez O, Brettschneider S, Lechner J, et al. Sus1, a Functional Component of the SAGA Histone Acetylase Complex and the Nuclear Pore-Associated mRNA Export Machinery. Cell. 2004 Jan 9;116(1):75–86.

77. Zhao J, Hyman L, Moore C. Formation of mRNA 3? Ends in Eukaryotes: Mechanism, Regulation, and Interrelationships with Other Steps in mRNA Synthesis. Microbiol Mol Biol Rev. 1999 Jun;63(2):405–45.

78. Bentley DL. Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. Curr Opin Cell Biol. 2005 Jun;17(3):251–6.

79. Brown CE, Sachs AB. Poly(A) Tail Length Control in Saccharomyces cerevisiae Occurs by Message-Specific Deadenylation. Mol Cell Biol. 1998 Nov 1;18(11):6548–59.

80. Mayer A, Heidemann M, Lidschreiber M, Schreieck A, Sun M, Hintermair C, et al. CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. Science. 2012 Jun 29;336(6089):1723–5.

81. Vinciguerra P, Stutz F. mRNA export: an assembly line from genes to nuclear pores. Curr Opin Cell Biol. 2004 Jun;16(3):285–92.

82. Huang Y, Carmichael GG. Role of polyadenylation in nucleocytoplasmic transport of mRNA. Mol Cell Biol. 1996 Apr;16(4):1534–42.

83. Wickens M, Anderson P, Jackson RJ. Life and death in the cytoplasm: messages from the 3′ end. Curr Opin Genet Dev. 1997 Apr;7(2):220–32.

84. Ford LP, Bagga PS, Wilusz J. The poly(A) tail inhibits the assembly of a 3'-to-5' exonuclease in an in vitro RNA stability system. Mol Cell Biol. 1997 Jan;17(1):398–406.

85. Wells SE, Hillner PE, Vale RD, Sachs AB. Circularization of mRNA by Eukaryotic Translation Initiation Factors. Mol Cell. 1998 Jul;2(1):135–40.

86. Maniatis T, Reed R. An extensive network of coupling among gene expression machines. Nature. 2002 Apr 4;416(6880):499–506.

87. Mandel CR, Bai Y, Tong L. Protein factors in pre-mRNA 3?-end processing. Cell Mol Life Sci CMLS. 2008 Apr;65(7-8):1099–122.

88. Liu D, Brockman JM, Dass B, Hutchins LN, Singh P, McCarrey JR, et al. Systematic variation in mRNA 3?-processing signals during mouse spermatogenesis. Nucleic Acids Res. 2007 Jan;35(1):234–46.

89. Higgs DR, Goodbourn SE, Lamb J, Clegg JB, Weatherall DJ, Proudfoot NJ. Alpha-thalassaemia caused by a polyadenylation signal mutation. Nature. 1983 Nov 24;306(5941):398–400.

90. Orkin SH, Cheng TC, Antonarakis SE, Kazazian HH Jr. Thalassemia due to a mutation in the cleavage-polyadenylation signal of the human beta-globin gene. EMBO J. 1985 Feb;4(2):453–6.

91. Gieselmann V, Polten A, Kreysing J, von Figura K. Arylsulfatase A pseudodeficiency: loss of a polyadenylylation signal and N-glycosylation site. Proc Natl Acad Sci U S A. 1989 Dec;86(23):9436–40.

92. Mayr C, Bartel DP. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. Cell. 2009 Aug 21;138(4):673–84.

93. Calzado MA, Sancho R, Muñoz E. Human immunodeficiency virus type 1 Tat increases the expression of cleavage and polyadenylation specificity factor 73-kilodalton subunit modulating cellular and viral expression. J Virol. 2004 Jul;78(13):6846–54.

94. Kumar GR, Glaunsinger BA. Nuclear import of cytoplasmic poly(A) binding protein restricts gene expression via hyperadenylation and nuclear retention of mRNA. Mol Cell Biol. 2010 Nov;30(21):4996–5008.

95. Nemeroff ME, Barabino SM, Li Y, Keller W, Krug RM. Influenza virus NS1 protein interacts with the cellular 30 kDa subunit of CPSF and inhibits 3'end formation of cellular pre-mRNAs. Mol Cell. 1998 Jun;1(7):991–1000.

96. Weng K-F, Li M-L, Hung C-T, Shih S-R. Enterovirus 71 3C protease cleaves a novel target CstF-64 and inhibits cellular polyadenylation. PLoS Pathog. 2009 Sep;5(9):e1000593.

97. Chen J, Moore C. Separation of factors required for cleavage and polyadenylation of yeast pre-mRNA. Mol Cell Biol. 1992 Aug 1;12(8):3470–81.

98. Preker PJ, Ohnacker M, Minvielle-Sebastia L, Keller W. A multisubunit 3′ end processing factor from yeast containing poly(A) polymerase and homologues of the subunits of mammalian cleavage and polyadenylation specificity factor. EMBO J. 1997 Aug 1;16(15):4727–37.

99. Ohnacker M, Barabino SML, Preker PJ, Keller W. The WD-repeat protein Pfs2p bridges two essential factors within the yeast pre-mRNA 3′-end-processing complex. EMBO J. 2000 Jan 4;19(1):37–47.

100. Brown CE, Tarun SZ Jr, Boeck R, Sachs AB. PAN3 encodes a subunit of the Pab1p-dependent poly(A) nuclease in Saccharomyces cerevisiae. Mol Cell Biol. 1996 Oct;16(10):5744–53.

101. Boeck R, Tarun S Jr, Rieger M, Deardorff JA, Müller-Auer S, Sachs AB. The yeast Pan2 protein is required for poly(A)-binding protein-stimulated poly(A)-nuclease activity. J Biol Chem. 1996 Jan 5;271(1):432–8.

102. Kessler MM, Zhao J, Moore CL. Purification of the Saccharomyces cerevisiae Cleavage/Polyadenylation Factor I SEPARATION INTO TWO COMPONENTS THAT ARE REQUIRED FOR BOTH CLEAVAGE AND POLYADENYLATION OF mRNA 3′ ENDS. J Biol Chem. 1996 Oct 25;271(43):27167–75.

103. Minvielle-Sebastia L, Preker PJ, Keller W. RNA14 and RNA15 proteins as components of a yeast pre-mRNA 3'-end processing factor. Science. 1994 Dec 9;266(5191):1702–5.

104. Amrani N, Minet M, Wyers F, Dufour ME, Aggerbeck LP, Lacroute F. PCF11 encodes a third protein component of yeast cleavage and polyadenylation factor I. Mol Cell Biol. 1997 Mar;17(3):1102–9.

105. Minvielle-Sebastia L, Preker PJ, Wiederkehr T, Strahm Y, Keller W. The major yeast poly(A)-binding protein is associated with cleavage factor IA and functions in premessenger RNA 3?-end formation. Proc Natl Acad Sci U S A. 1997 Jul 22;94(15):7897–902.

106. Kessler MM, Henry MF, Shen E, Zhao J, Gross S, Silver PA, et al. Hrp1, a sequence-specific RNA-binding protein that shuttles between the nucleus and the cytoplasm, is required for mRNA 3′-end formation in yeast. Genes Dev. 1997 Oct 1;11(19):2545–56.

107. Noble CG, Walker PA, Calder LJ, Taylor IA. Rna14-Rna15 assembly mediates the RNA-binding capability of Saccharomyces cerevisiae cleavage factor IA. Nucleic Acids Res. 2004;32(11):3364–75.

108. Gordon J, Shikov S, Kuehner JN, Liriano M, Lee E, Stafford W, et al. Reconstitution of CF IA from overexpressed subunits reveals stoichiometry and provides insights into molecular topology. Biochemistry (Mosc). 2011 Nov 29;50(47):10203–14.

109. Minvielle-Sebastia L, Winsor B, Bonneaud N, Lacroute F. Mutations in the yeast RNA14 and RNA15 genes result in an abnormal mRNA decay rate; sequence analysis reveals an RNA-binding domain in the RNA15 protein. Mol Cell Biol. 1991 Jun 1;11(6):3075–87.

110. Kuehner JN, Pearson EL, Moore C. Unravelling the means to an end: RNA polymerase II transcription termination. Nat Rev Mol Cell Biol. 2011 May;12(5):283–94.

111. Kim M, Vasiljeva L, Rando OJ, Zhelkovsky A, Moore C, Buratowski S. Distinct pathways for snoRNA and mRNA termination. Mol Cell. 2006 Dec 8;24(5):723–34.

112. Sadowski M, Dichtl B, Hübner W, Keller W. Independent functions of yeast Pcf11p in pre-mRNA 3' end processing and in transcription termination. EMBO J. 2003 May 1;22(9):2167–77.

113. Richard P, Manley JL. Transcription termination by nuclear RNA polymerases. Genes Dev. 2009 Jun 1;23(11):1247–69.

114. Licatalosi DD, Geiger G, Minet M, Schroeder S, Cilli K, McNeil JB, et al. Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II. Mol Cell. 2002 May;9(5):1101–11.

115. Barillà D, Lee BA, Proudfoot NJ. Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A. 2001 Jan 16;98(2):445–50.

116. Meinhart A, Cramer P. Recognition of RNA polymerase II carboxy-terminal domain by 3′-RNA-processing factors. Nature. 2004 Jul 8;430(6996):223–6.

117. Noble CG, Hollingworth D, Martin SR, Ennis-Adeniran V, Smerdon SJ, Kelly G, et al. Key features of the interaction between Pcf11 CID and RNA polymerase II CTD. Nat Struct Mol Biol. 2005 Feb;12(2):144–51.

118. Noble CG, Beuth B, Taylor IA. Structure of a nucleotide-bound Clp1-Pcf11 polyadenylation factor. Nucleic Acids Res. 2007 Jan 1;35(1):87–99.

119. Ghazy MA, Gordon JMB, Lee SD, Singh BN, Bohm A, Hampsey M, et al. The interaction of Pcf11 and Clp1 is needed for mRNA 3?-end formation and is modulated by amino acids in the ATP-binding site. Nucleic Acids Res. 2012 Feb;40(3):1214–25.

120. Ghazy MA, He X, Singh BN, Hampsey M, Moore C. The Essential N Terminus of the Pta1 Scaffold Protein Is Required for snoRNA Transcription Termination and Ssu72 Function but Is Dispensable for Pre-mRNA 3?-End Processing. Mol Cell Biol. 2009 Apr;29(8):2296–307.

121. Kyburz A, Sadowski M, Dichtl B, Keller W. The role of the yeast cleavage and polyadenylation factor subunit Ydh1p/Cft2p in pre-mRNA 3?-end formation. Nucleic Acids Res. 2003 Jul 15;31(14):3936–45.

122. He X, Khan AU, Cheng H, Pappas DL Jr, Hampsey M, Moore CL. Functional interactions between the transcription and mRNA 3' end processing machineries mediated by Ssu72 and Sub1. Genes Dev. 2003 Apr 15;17(8):1030–42.

123. Gross S, Moore C. Five subunits are required for reconstitution of the cleavage and polyadenylation activities of Saccharomyces cerevisiae cleavage factor I. Proc Natl Acad Sci U S A. 2001 May 22;98(11):6080–5.

124. Dichtl B, Aasland R, Keller W. Functions for S. cerevisiae Swd2p in 3' end formation of specific mRNAs and snoRNAs and global histone 3 lysine 4 methylation. RNA N Y N. 2004 Jun;10(6):965–77.

125. Leipe DD, Wolf YI, Koonin EV, Aravind L. Classification and evolution of P-loop GTPases and related ATPases. J Mol Biol. 2002 Mar 15;317(1):41–72.

126. Ramirez A, Shuman S, Schwer B. Human RNA 5?-kinase (hClp1) can function as a tRNA splicing enzyme in vivo. RNA. 2008 Sep;14(9):1737–45.

127. Weitzer S, Martinez J. The human RNA kinase hClp1 is active on 3′ transfer RNA exons and short interfering RNAs. Nature. 2007 May 10;447(7141):222–6.

128. Haddad R, Maurice F, Viphakone N, Voisinet-Hakil F, Fribourg S, Minvielle-Sebastia L. An essential role for Clp1 in assembly of polyadenylation complex CF IA and Pol II transcription termination. Nucleic Acids Res. 2012 Feb;40(3):1226–39.

129. Dupin AF, Fribourg S. Structural basis for ATP loss by Clp1p in a G135R mutant protein. Biochimie [Internet]. [cited 2014 Feb 12]; Available from: http://www.sciencedirect.com/science/article/pii/S0300908414000352

130. Holbein S, Scola S, Loll B, Dichtl BS, Hubner W, Meinhart A, et al. The P-Loop Domain of Yeast Clp1 Mediates Interactions Between CF IA and CPF Factors in Pre-mRNA 3? End Formation. PLoS ONE [Internet]. 2011 Dec 22 [cited 2014 Feb 12];6(12). Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3245249/

131. Bonneaud N, Minvielle-Sebastia L, Cullin C, Lacroute F. Cellular localization of RNA14p and RNA15p, two yeast proteins involved in mRNA stability. J Cell Sci. 1994 Apr 1;107(4):913–21.

132. Rouillard JM, Brendolise C, Lacroute F. Rna14p, a component of the yeast nuclear cleavage/polyadenylation factor I, is also localised in mitochondria. Mol Gen Genet MGG. 2000 Jan;262(6):1103–12.

133. Preker PJ, Keller W. The HAT helix, a repetitive motif implicated in RNA processing. Trends Biochem Sci. 1998 Jan;23(1):15–6.

134. Takagaki Y, Manley JL. A polyadenylation factor subunit is the human homologue of the Drosophila suppressor of forked protein. Nature. 1994 Dec 1;372(6505):471–4.

135. Bai Y, Auperin TC, Chou C-Y, Chang G-G, Manley JL, Tong L. Crystal Structure of Murine CstF-77: Dimeric Association and Implications for Polyadenylation of mRNA Precursors. Mol Cell. 2007 Mar 23;25(6):863–75.

136. Legrand P, Pinaud N, Minvielle-Sebastia L, Fribourg S. The structure of the CstF-77 homodimer provides insights into CstF assembly. Nucleic Acids Res. 2007 Jul;35(13):4515–22.

137. Moreno-Morcillo M, Minvielle-Sébastia L, Fribourg S, Mackereth CD. Locked Tether Formation by Cooperative Folding of Rna14p Monkeytail and Rna15p Hinge Domains in the Yeast CF IA Complex. Structure. 2011 Apr 13;19(4):534–45.

138. Paulson AR, Tong L. Crystal structure of the Rna14-Rna15 complex. RNA. 2012 Jun;18(6):1154–62.

139. Gross S, Moore CL. Rna15 Interaction with the A-Rich Yeast Polyadenylation Signal Is an Essential Step in mRNA 3?-End Formation. Mol Cell Biol. 2001 Dec;21(23):8045–55.

140. Qu X, Perez-Canadillas J-M, Agrawal S, De Baecke J, Cheng H, Varani G, et al. The C-terminal domains of vertebrate CstF-64 and its yeast orthologue Rna15 form a new structure critical for mRNA 3'-end processing. J Biol Chem. 2007 Jan 19;282(3):2101–15.

141. Pancevac C, Goldstone DC, Ramos A, Taylor IA. Structure of the Rna15 RRM–RNA complex reveals the molecular basis of GU specificity in transcriptional 3′-end processing factors. Nucleic Acids Res. 2010 May 1;38(9):3119–32.

142. Leeper TC, Qu X, Lu C, Moore C, Varani G. Novel protein-protein contacts facilitate mRNA 3'-processing signal recognition by Rna15 and Hrp1. J Mol Biol. 2010 Aug 20;401(3):334–49.

143. Takagaki Y, Manley JL. RNA recognition by the human polyadenylation factor CstF. Mol Cell Biol. 1997 Jul;17(7):3907–14.

144. Beyer K, Dandekar T, Keller W. RNA ligands selected by cleavage stimulation factor contain distinct sequence motifs that function as downstream elements in 3'-end processing of pre-mRNA. J Biol Chem. 1997 Oct 17;272(42):26769–79.

145. Graber JH, Cantor CR, Mohr SC, Smith TF. Genomic detection of new yeast pre-mRNA 3'-end-processing signals. Nucleic Acids Res. 1999 Feb 1;27(3):888–94.

146. Van Helden J, del Olmo M, Pérez-Ortín JE. Statistical analysis of yeast genomic downstream sequences reveals putative polyadenylation signals. Nucleic Acids Res. 2000 Feb 15;28(4):1000–10.

147. Dichtl B, Keller W. Recognition of polyadenylation sites in yeast pre-mRNAs by cleavage and polyadenylation factor. EMBO J. 2001 Jun 15;20(12):3197–209.

148. Minvielle-Sebastia L, Beyer K, Krecic AM, Hector RE, Swanson MS, Keller W. Control of cleavage site selection during mRNA 3' end formation by a yeast hnRNP. EMBO J. 1998 Dec 15;17(24):7454–68.

149. Pérez-Cañadillas JM. Grabbing the message: structural basis of mRNA 3'UTR recognition by Hrp1. EMBO J. 2006 Jul 12;25(13):3167–78.

150. Valentini SR, Weiss VH, Silver PA. Arginine methylation and binding of Hrp1p to the efficiency element for mRNA 3'-end formation. RNA N Y N. 1999 Feb;5(2):272–80.

151. Barnwal RP, Lee SD, Moore C, Varani G. Structural and biochemical analysis of the assembly and function of the yeast pre-mRNA 3' end processing complex CF I. Proc Natl Acad Sci U S A. 2012 Dec 26;109(52):21342–7.

152. Russell I, Tollervey D. Yeast Nop3p has structural and functional similarities to mammalian pre-mRNA binding proteins. Eur J Cell Biol. 1995 Mar;66(3):293–301.

153. Singleton DR, Chen S, Hitomi M, Kumagai C, Tartakoff AM. A yeast protein that bidirectionally affects nucleocytoplasmic transport. J Cell Sci. 1995 Jan;108 (Pt 1):265–72.

154. Liu C, Smith LD. Differential accumulation of mRNA and interspersed RNA during Xenopus oogenesis and embryogenesis. Zygote Camb Engl. 1994 Nov;2(4):307–16.

155. Eckner R, Ellmeier W, Birnstiel ML. Mature mRNA 3' end formation stimulates RNA export from the nucleus. EMBO J. 1991 Nov;10(11):3513–22.

156. Dreyfuss G, Matunis MJ, Piñol-Roma S, Burd CG. hnRNP proteins and the biogenesis of mRNA. Annu Rev Biochem. 1993;62:289–321.

157. Shen EC, Henry MF, Weiss VH, Valentini SR, Silver PA, Lee MS. Arginine methylation facilitates the nuclear export of hnRNP proteins. Genes Dev. 1998 Mar 1;12(5):679–91.

158. Preker PJ, Lingner J, Minvielle-Sebastia L, Keller W. The FIP1 gene encodes a component of a yeast pre-mRNA polyadenylation factor that directly interacts with poly(A) polymerase. Cell. 1995 May 5;81(3):379–89.

159. Russnak R, Nehrke KW, Platt T. REF2 encodes an RNA-binding protein directly involved in yeast mRNA 3'-end formation. Mol Cell Biol. 1995 Mar;15(3):1689–97.

160. Nedea E, He X, Kim M, Pootoolal J, Zhong G, Canadien V, et al. Organization and function of APT, a subcomplex of the yeast cleavage and polyadenylation factor involved in the formation of mRNA and small nucleolar RNA 3'-ends. J Biol Chem. 2003 Aug 29;278(35):33000–10.

161. Sun ZW, Hampsey M. Synthetic enhancement of a TFIIB defect by a mutation in SSU72, an essential yeast gene encoding a novel protein that affects transcription start site selection in vivo. Mol Cell Biol. 1996 Apr;16(4):1557–66.

162. Wu WH, Pinto I, Chen BS, Hampsey M. Mutational analysis of yeast TFIIB. A functional relationship between Ssu72 and Sub1/Tsp1 defined by allele-specific interactions with TFIIB. Genetics. 1999 Oct;153(2):643–52.

163. Schreieck A, Easter AD, Etzold S, Wiederhold K, Lidschreiber M, Cramer P, et al. RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7. Nat Struct Mol Biol. 2014 Feb;21(2):175–9.

164. Jeronimo C, Bataille AR, Robert F. The writers, readers, and functions of the RNA polymerase II C-terminal domain code. Chem Rev. 2013 Nov 13;113(11):8491–522.

165. Schroeder SC, Schwer B, Shuman S, Bentley D. Dynamic association of capping enzymes with transcribing RNA polymerase II. Genes Dev. 2000 Oct 1;14(19):2435–40.

166. Krishnamurthy S, He X, Reyes-Reyes M, Moore C, Hampsey M. Ssu72 Is an RNA polymerase II CTD phosphatase. Mol Cell. 2004 May 7;14(3):387–94.

167. Bataille AR, Jeronimo C, Jacques P-É, Laramée L, Fortin M-È, Forest A, et al. A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes. Mol Cell. 2012 Jan 27;45(2):158–70.

168. Xiang K, Nagaike T, Xiang S, Kilic T, Beh MM, Manley JL, et al. Crystal structure of the human symplekin-Ssu72-CTD phosphopeptide complex. Nature. 2010 Oct 7;467(7316):729–33.

169. Dichtl B, Blank D, Sadowski M, Hübner W, Weiser S, Keller W. Yhh1p/Cft1p directly links poly(A) site recognition and RNA polymerase II transcription termination. EMBO J. 2002 Aug 1;21(15):4125–35.

170. Vo LT, Minet M, Schmitter JM, Lacroute F, Wyers F. Mpe1, a zinc knuckle protein, is an essential component of yeast cleavage and polyadenylation factor required for the cleavage and polyadenylation of mRNA. Mol Cell Biol. 2001 Dec;21(24):8346–56.

171. Barabino SM, Hübner W, Jenny A, Minvielle-Sebastia L, Keller W. The 30-kD subunit of mammalian cleavage and polyadenylation specificity factor and its yeast homolog are RNA-binding zinc finger proteins. Genes Dev. 1997 Jul 1;11(13):1703–16.

172. Barabino SML, Ohnacker M, Keller W. Distinct roles of two Yth1p domains in 3′-end cleavage and polyadenylation of yeast pre-mRNAs. EMBO J. 2000 Jul 17;19(14):3778–87.

173. Bai C, Tolias PP. Cleavage of RNA hairpins mediated by a developmentally regulated CCCH zinc finger protein. Mol Cell Biol. 1996 Dec;16(12):6661–7.

174. Chanfreau G, Noble SM, Guthrie C. Essential yeast protein with unexpected similarity to subunits of mammalian cleavage and polyadenylation specificity factor (CPSF). Science. 1996 Nov 29;274(5292):1511–4.

175. Jenny A, Minvielle-Sebastia L, Preker PJ, Keller W. Sequence similarity between the 73-kilodalton protein of mammalian CPSF and a subunit of yeast polyadenylation factor I. Science. 1996 Nov 29;274(5292):1514–7.

176. Daiyasu H, Osaka K, Ishino Y, Toh H. Expansion of the zinc metallo-hydrolase family of the beta-lactamase fold. FEBS Lett. 2001 Aug 10;503(1):1–6.

177. Callebaut I, Moshous D, Mornon J-P, de Villartay J-P. Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family. Nucleic Acids Res. 2002 Aug 15;30(16):3592–601.

178. Ryan K, Calvo O, Manley JL. Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. RNA N Y N. 2004 Apr;10(4):565–73.

179. Mandel CR, Kaneko S, Zhang H, Gebauer D, Vethantham V, Manley JL, et al. Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. Nature. 2006 Dec 14;444(7121):953–6.

180. Ryan K. Pre-mRNA 3' cleavage is reversibly inhibited in vitro by cleavage factor dephosphorylation. RNA Biol. 2007 Mar;4(1):26–33.

181. Aravind L. An evolutionary classification of the metallo-beta-lactamase fold proteins. In Silico Biol. 1999;1(2):69–91.

182. Lingner J, Radtke I, Wahle E, Keller W. Purification and characterization of poly(A) polymerase from Saccharomyces cerevisiae. J Biol Chem. 1991 May 15;266(14):8741–6.

183. Zhelkovsky A, Helmling S, Moore C. Processivity of the Saccharomyces cerevisiae poly(A) polymerase requires interactions at the carboxyl-terminal RNA binding domain. Mol Cell Biol. 1998 Oct;18(10):5942–51.

184. Ezeokonkwo C, Zhelkovsky A, Lee R, Bohm A, Moore CL. A flexible linker region in Fip1 is needed for efficient mRNA polyadenylation. RNA N Y N. 2011 Apr;17(4):652–64.

185. Kaufmann I, Martin G, Friedlein A, Langen H, Keller W. Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. EMBO J. 2004 Feb 11;23(3):616–26.

186. Lingner J, Kellermann J, Keller W. Cloning and expression of the essential gene for poly(A) polymerase from S. cerevisiae. Nature. 1991 Dec 12;354(6353):496–8.

187. Bard J, Zhelkovsky AM, Helmling S, Earnest TN, Moore CL, Bohm A. Structure of yeast poly(A) polymerase alone and in complex with 3'-dATP. Science. 2000 Aug 25;289(5483):1346–9.

188. Balbo PB, Bohm A. Mechanism of Poly(A) Polymerase: Structure of the Enzyme-MgATP-RNA Ternary Complex and Kinetic Analysis. Structure. 2007 Sep 11;15(9):1117–31.

189. Balbo PB, Meinke G, Bohm A. Kinetic studies of yeast polyA polymerase indicate an induced fit mechanism for nucleotide specificity. Biochemistry (Mosc). 2005 May 31;44(21):7777–86.

190. Amrani N, Minet M, Le Gouar M, Lacroute F, Wyers F. Yeast Pab1 interacts with Rna15 and participates in the control of the poly(A) tail length in vitro. Mol Cell Biol. 1997 Jul;17(7):3694–701.

191. Sachs AB, Davis RW. The poly(A) binding protein is required for poly(A) shortening and 60S ribosomal subunit-dependent translation initiation. Cell. 1989 Sep 8;58(5):857–67.

192. Caponigro G, Parker R. Multiple functions for the poly(A)-binding protein in mRNA decapping and deadenylation in yeast. Genes Dev. 1995 Oct 1;9(19):2421–32.

193. Caponigro G, Parker R. Mechanisms and control of mRNA turnover in Saccharomyces cerevisiae. Microbiol Rev. 1996 Mar;60(1):233–49.

194. Tarun SZ Jr, Sachs AB. A common function for mRNA 5’ and 3’ ends in translation initiation in yeast. Genes Dev. 1995 Dec 1;9(23):2997–3007.

195. Meyer S, Urbanke C, Wahle E. Equilibrium studies on the association of the nuclear poly(A) binding protein with poly(A) of different lengths. Biochemistry (Mosc). 2002 May 14;41(19):6082–9.

196. Kerwitz Y, Kühn U, Lilie H, Knoth A, Scheuermann T, Friedrich H, et al. Stimulation of poly(A) polymerase through a direct interaction with the nuclear poly(A) binding protein allosterically regulated by RNA. EMBO J. 2003 Jul 15;22(14):3705–14.

197. Keller RW, Kühn U, Aragón M, Bornikova L, Wahle E, Bear DG. The nuclear poly(A) binding protein, PABP2, forms an oligomeric particle covering the length of the poly(A) tail. J Mol Biol. 2000 Mar 31;297(3):569–83.

198. Hector RE, Nykamp KR, Dheur S, Anderson JT, Non PJ, Urbinati CR, et al. Dual requirement for yeast hnRNP Nab2p in mRNA poly(A) tail length control and nuclear export. EMBO J. 2002 Apr 2;21(7):1800–10.

199. Kelly SM, Leung SW, Apponi LH, Bramley AM, Tran EJ, Chekanova JA, et al. Recognition of polyadenosine RNA by the zinc finger domain of nuclear poly(A) RNA-binding protein 2 (Nab2) is required for correct mRNA 3’-end formation. J Biol Chem. 2010 Aug 20;285(34):26022–32.

200. Butler JS, Platt T. RNA Processing Generates the Mature 3$^{\prime}$ End of Yeast CYC1 Messenger RNA in vitro. Science. 1988 Dec 2;242(4883):1270–4.

201. Butler JS, Sadhale PP, Platt T. RNA processing in vitro produces mature 3’ ends of a variety of Saccharomyces cerevisiae mRNAs. Mol Cell Biol. 1990 Jun;10(6):2599–605.

202. Zaret KS, Sherman F. DNA sequence required for efficient transcription termination in yeast. Cell. 1982 Mar;28(3):563–73.

203. Russo P, Sherman F. Transcription terminates near the poly(A) site in the CYC1 gene of the yeast Saccharomyces cerevisiae. Proc Natl Acad Sci U S A. 1989 Nov;86(21):8348–52.

204. Russo P, Li WZ, Hampsey DM, Zaret KS, Sherman F. Distinct cis-acting signals enhance 3' endpoint formation of CYC1 mRNA in the yeast Saccharomyces cerevisiae. EMBO J. 1991 Mar;10(3):563–71.

205. Russo P, Li WZ, Guo Z, Sherman F. Signals that produce 3' termini in CYC1 mRNA of the yeast Saccharomyces cerevisiae. Mol Cell Biol. 1993 Dec;13(12):7836–49.

206. Heidmann S, Obermaier B, Vogel K, Domdey H. Identification of pre-mRNA polyadenylation sites in Saccharomyces cerevisiae. Mol Cell Biol. 1992 Sep;12(9):4215–29.

207. Irniger S, Braus GH. Saturation mutagenesis of a polyadenylation signal reveals a hexanucleotide element essential for mRNA 3' end formation in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A. 1994 Jan 4;91(1):257–61.

208. Mayer SA, Dieckmann CL. Yeast CBP1 mRNA 3' end formation is regulated during the induction of mitochondrial function. Mol Cell Biol. 1991 Feb;11(2):813–21.

209. Sadhale PP, Sapolsky R, Davis RW, Butler JS, Platt T. Polymerase chain reaction mapping of yeast GAL7 mRNA polyadenylation sites demonstrates that 3' end processing in vitro faithfully reproduces the 3' ends observed in vivo. Nucleic Acids Res. 1991 Jul 11;19(13):3683–8.

210. Guo Z, Sherman F. 3'-end-forming signals of yeast mRNA. Mol Cell Biol. 1995 Nov;15(11):5983–90.

211. Guo Z, Russo P, Yun DF, Butler JS, Sherman F. Redundant 3' end-forming signals for the yeast CYC1 mRNA. Proc Natl Acad Sci U S A. 1995 May 9;92(10):4211–4.

212. Keller W, Minvielle-Sebastia L. A comparison of mammalian and yeast pre-mRNA 3'-end processing. Curr Opin Cell Biol. 1997 Jun;9(3):329–36.

213. Colgan DF, Manley JL. Mechanism and regulation of mRNA polyadenylation. Genes Dev. 1997 Nov 1;11(21):2755–66.

214. Manley JL, Takagaki Y. The end of the message--another link between yeast and mammals. Science. 1996 Nov 29;274(5292):1481–2.

215. Merrick WC, Hershey JWB. The Pathway and mechanism of eukaryotic protein synthesis in translational control. New York: Cold Spring Harbor Laboratory Press; 31-70 p.

216. Kessler SH, Sachs AB. RNA recognition motif 2 of yeast Pab1p is required for its functional interaction with eukaryotic translation initiation factor 4G. Mol Cell Biol. 1998 Jan;18(1):51–7.

217. Tarun SZ Jr, Sachs AB. Binding of eukaryotic translation initiation factor 4E (eIF4E) to eIF4G represses translation of uncapped mRNA. Mol Cell Biol. 1997 Dec;17(12):6876–86.

218. Tarun SZ Jr, Wells SE, Deardorff JA, Sachs AB. Translation initiation factor eIF4G mediates in vitro poly(A) tail-dependent translation. Proc Natl Acad Sci U S A. 1997 Aug 19;94(17):9046–51.

219. Jackson RJ, Hellen CUT, Pestova TV. The mechanism of eukaryotic translation initiation and principles of its regulation. Nat Rev Mol Cell Biol. 2010 Feb;11(2):113–27.

220. Vagner S, Vagner C, Mattaj IW. The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF 65 to couple 3'-end processing and splicing. Genes Dev. 2000 Feb 15;14(4):403–13.

221. McCracken S, Lambermon M, Blencowe BJ. SRm160 splicing coactivator promotes transcript 3'-end cleavage. Mol Cell Biol. 2002 Jan;22(1):148–60.

222. Berget SM. Exon recognition in vertebrate splicing. J Biol Chem. 1995 Feb 10;270(6):2411–4.

223. Gunderson SI, Polycarpou-Schwarz M, Mattaj IW. U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(A) polymerase. Mol Cell. 1998 Jan;1(2):255–64.

224. Kaida D, Berg MG, Younis I, Kasim M, Singh LN, Wan L, et al. U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. Nature. 2010 Dec 2;468(7324):664–8.

225. Tikhonov M, Georgiev P, Maksimenko O. Competition within Introns: Splicing Wins over Polyadenylation via a General Mechanism. Acta Naturae. 2013 Oct;5(4):52–61.

226. Hammell CM, Gross S, Zenklusen D, Heath CV, Stutz F, Moore C, et al. Coupling of termination, 3' processing, and mRNA export. Mol Cell Biol. 2002 Sep;22(18):6441–57.

227. Grec S, Wang Y, Le Guen L, Negrouk V, Boutry M. Cryptic polyadenylation sites within the coding sequence of three yeast genes expressed in tobacco. Gene. 2000 Jan 25;242(1–2):87–95.

228. Brodsky AS, Silver PA. Pre-mRNA processing factors are required for nuclear export. RNA N Y N. 2000 Dec;6(12):1737–49.

229. Hilleren P, Parker R. Defects in the mRNA export factors Rat7p, Gle1p, Mex67p, and Rat8p cause hyperadenylation during 3'-end formation of nascent transcripts. RNA N Y N. 2001 May;7(5):753–64.

230. Flach J, Bossie M, Vogel J, Corbett A, Jinks T, Willins DA, et al. A yeast RNA-binding protein shuttles between the nucleus and the cytoplasm. Mol Cell Biol. 1994 Dec;14(12):8399–407.

231. Lee MS, Henry M, Silver PA. A protein that shuttles between the nucleus and the cytoplasm is an important mediator of RNA export. Genes Dev. 1996 May 15;10(10):1233–46.

232. Bucheli ME, He X, Kaplan CD, Moore CL, Buratowski S. Polyadenylation site choice in yeast is affected by competition between Npl3 and polyadenylation factor CFI. RNA N Y N. 2007 Oct;13(10):1756–64.

233. Aitchison JD, Blobel G, Rout MP. Kap104p: a karyopherin involved in the nuclear transport of messenger RNA binding proteins. Science. 1996 Oct 25;274(5287):624–7.

234. Bird G, Zorio DAR, Bentley DL. RNA polymerase II carboxy-terminal domain phosphorylation is required for cotranscriptional pre-mRNA splicing and 3'-end formation. Mol Cell Biol. 2004 Oct;24(20):8963–9.

235. Wetterberg I, Zhao J, Masich S, Wieslander L, Skoglund U. In situ transcription and splicing in the Balbiani ring 3 gene. EMBO J. 2001 May 15;20(10):2564–74.

236. Lewis ED, Manley JL. Polyadenylylation of an mRNA precursor occurs independently of transcription by RNA polymerase II in vivo. Proc Natl Acad Sci U S A. 1986 Nov;83(22):8555–9.

237. Sisodia SS, Sollner-Webb B, Cleveland DW. Specificity of RNA maturation pathways: RNAs transcribed by RNA polymerase III are not substrates for splicing or polyadenylation. Mol Cell Biol. 1987 Oct;7(10):3602–12.

238. Smale ST, Tjian R. Transcription of herpes simplex virus tk sequences under the control of wild type and mutant human RNA polymerase I promoters. Mol Cell Biol. 1985 Feb;5(2):352–62.

239. Mapendano CK, Lykke-Andersen S, Kjems J, Bertrand E, Jensen TH. Crosstalk between mRNA 3' end processing and transcription initiation. Mol Cell. 2010 Nov 12;40(3):410–22.

240. Ahn SH, Keogh M-C, Buratowski S. Ctk1 promotes dissociation of basal transcription factors from elongating RNA polymerase II. EMBO J. 2009 Feb 4;28(3):205–12.

241. Ganem C, Devaux F, Torchet C, Jacq C, Quevillon-Cheruel S, Labesse G, et al. Ssu72 is a phosphatase essential for transcription termination of snoRNAs and specific mRNAs in yeast. EMBO J. 2003 Apr 1;22(7):1588–98.

242. Meinhart A, Silberzahn T, Cramer P. The mRNA transcription/processing factor Ssu72 is a potential tyrosine phosphatase. J Biol Chem. 2003 May 2;278(18):15917–21.

243. Knaus R, Pollock R, Guarente L. Yeast SUB1 is a suppressor of TFIIB mutations and has homology to the human co-activator PC4. EMBO J. 1996 Apr 15;15(8):1933–40.

244. Medler S, Al Husini N, Raghunayakula S, Mukundan B, Aldea A, Ansari A. Evidence for a complex of transcription factor IIB with poly(A) polymerase and

cleavage factor 1 subunits required for gene looping. J Biol Chem. 2011 Sep 30;286(39):33709–18.

245. Papantonis A, Cook PR. Genome architecture and the role of transcription. Curr Opin Cell Biol. 2010 Jun;22(3):271–6.

246. Vilar JMG, Saiz L. DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. Curr Opin Genet Dev. 2005 Apr;15(2):136–44.

247. El Kaderi B, Medler S, Raghunayakula S, Ansari A. Gene looping is conferred by activator-dependent interaction of transcription initiation and termination machineries. J Biol Chem. 2009 Sep 11;284(37):25015–25.

248. Singh BN, Hampsey M. A transcription-independent role for TFIIB in gene looping. Mol Cell. 2007 Sep 7;27(5):806–16.

249. Ansari A, Hampsey M. A role for the CPF 3'-end processing machinery in RNAP II-dependent gene looping. Genes Dev. 2005 Dec 15;19(24):2969–78.

250. Al Husini N, Kudla P, Ansari A. A role for CF1A 3' end processing complex in promoter-associated transcription. PLoS Genet. 2013;9(8):e1003722.

251. Birse CE, Minvielle-Sebastia L, Lee BA, Keller W, Proudfoot NJ. Coupling termination of transcription to messenger RNA maturation in yeast. Science. 1998 Apr 10;280(5361):298–301.

252. Russo P. Saccharomyces cerevisiae mRNA 3' end forming signals are also involved in transcription termination. Yeast Chichester Engl. 1995 Apr 30;11(5):447–53.

253. Connelly S, Manley JL. A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. Genes Dev. 1988 Apr;2(4):440–52.

254. Logan J, Falck-Pedersen E, Darnell JE Jr, Shenk T. A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. Proc Natl Acad Sci U S A. 1987 Dec;84(23):8306–10.

255. Greenblatt J, Nodwell JR, Mason SW. Transcriptional antitermination. Nature. 1993 Jul 29;364(6436):401–6.

256. Calvo O, Manley JL. Evolutionarily conserved interaction between CstF-64 and PC4 links transcription, polyadenylation, and termination. Mol Cell. 2001 May;7(5):1013–23.

257. Rosonina E, Kaneko S, Manley JL. Terminating the transcript: breaking up is hard to do. Genes Dev. 2006 May 1;20(9):1050–6.

258. Osheim YN, Proudfoot NJ, Beyer AL. EM visualization of transcription by RNA polymerase II: downstream termination requires a poly(A) signal but not transcript cleavage. Mol Cell. 1999 Mar;3(3):379–87.

259. Osheim YN, Sikes ML, Beyer AL. EM visualization of Pol II genes in Drosophila: most genes terminate without prior 3’ end cleavage of nascent transcripts. Chromosoma. 2002 Mar;111(1):1–12.

260. Aranda A, Pérez-Ortín JE, Moore C, del Olmo ML. Transcription termination downstream of the Saccharomyces cerevisiae FBP1 [changed from FPB1] poly(A) site does not depend on efficient 3’end processing. RNA N Y N. 1998 Mar;4(3):303–18.

261. Kim M, Krogan NJ, Vasiljeva L, Rando OJ, Nedea E, Greenblatt JF, et al. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. Nature. 2004 Nov 25;432(7016):517–22.

262. Luo W, Johnson AW, Bentley DL. The role of Rat1 in coupling mRNA 3’-end processing to transcription termination: implications for a unified allosteric-torpedo model. Genes Dev. 2006 Apr 15;20(8):954–65.

263. Zhang Z, Fu J, Gilmour DS. CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3’-end processing factor, Pcf11. Genes Dev. 2005 Jul 1;19(13):1572–80.

264. Gilmartin GM, Nevins JR. Molecular analyses of two poly(A) site-processing factors that determine the recognition and efficiency of cleavage of the pre-mRNA. Mol Cell Biol. 1991 May;11(5):2432–8.

265. Murthy KG, Manley JL. The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3’-end formation. Genes Dev. 1995 Nov 1;9(21):2672–83.

266. Rüegsegger U, Beyer K, Keller W. Purification and characterization of human cleavage factor Im involved in the 3’ end processing of messenger RNA precursors. J Biol Chem. 1996 Mar 15;271(11):6107–13.

267. Tian B, Hu J, Zhang H, Lutz CS. A large-scale analysis of mRNA polyadenylation of human and mouse genes. Nucleic Acids Res. 2005;33(1):201–12.

268. Wickens M, Stephenson P. Role of the conserved AAUAAA sequence: four AAUAAA point mutants prevent messenger RNA 3’ end formation. Science. 1984 Nov 30;226(4678):1045–51.

269. Simonsen CC, Levinson AD. Analysis of processing and polyadenylation signals of the hepatitis B virus surface antigen gene by using simian virus 40-hepatitis B virus chimeric plasmids. Mol Cell Biol. 1983 Dec;3(12):2250–8.

270. Gil A, Proudfoot NJ. Position-dependent sequence elements downstream of AAUAAA are required for efficient rabbit beta-globin mRNA 3’ end formation. Cell. 1987 May 8;49(3):399–406.

271. McLauchlan J, Gaffney D, Whitton JL, Clements JB. The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3’ termini. Nucleic Acids Res. 1985 Feb 25;13(4):1347–68.

272. Chou ZF, Chen F, Wilusz J. Sequence and position requirements for uridylate-rich downstream elements of polyadenylation signals. Nucleic Acids Res. 1994 Jul 11;22(13):2525–31.

273. McDevitt MA, Hart RP, Wong WW, Nevins JR. Sequences capable of restoring poly(A) site function define two distinct downstream elements. EMBO J. 1986 Nov;5(11):2907–13.

274. Zarkower D, Wickens M. A functionally redundant downstream sequence in SV40 late pre-mRNA is required for mRNA 3'-end formation and for assembly of a precleavage complex in vitro. J Biol Chem. 1988 Apr 25;263(12):5780–8.

275. Chen F, MacDonald CC, Wilusz J. Cleavage site determinants in the mammalian polyadenylation signal. Nucleic Acids Res. 1995 Jul 25;23(14):2614–20.

276. Gilmartin GM, Fleming ES, Oetjen J, Graveley BR. CPSF recognition of an HIV-1 mRNA 3'-processing enhancer: multiple sequence contacts involved in poly(A) site definition. Genes Dev. 1995 Jan 1;9(1):72–83.

277. MacDonald CC, Wilusz J, Shenk T. The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location. Mol Cell Biol. 1994 Oct;14(10):6647–54.

278. Sheets MD, Ogg SC, Wickens MP. Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. Nucleic Acids Res. 1990 Oct 11;18(19):5799–805.

279. Hu J, Lutz CS, Wilusz J, Tian B. Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation. RNA N Y N. 2005 Oct;11(10):1485–93.

280. Arhin GK, Boots M, Bagga PS, Milcarek C, Wilusz J. Downstream sequence elements with different affinities for the hnRNP H/H' protein influence the processing efficiency of mammalian polyadenylation signals. Nucleic Acids Res. 2002 Apr 15;30(8):1842–50.

281. Bagga PS, Ford LP, Chen F, Wilusz J. The G-rich auxiliary downstream element has distinct sequence and position requirements and mediates efficient 3' end pre-mRNA processing through a trans-acting factor. Nucleic Acids Res. 1995 May 11;23(9):1625–31.

282. Oberg D, Fay J, Lambkin H, Schwartz S. A downstream polyadenylation element in human papillomavirus type 16 L2 encodes multiple GGG motifs and interacts with hnRNP H. J Virol. 2005 Jul;79(14):9254–69.

283. Dalziel M, Nunes NM, Furger A. Two G-rich regulatory elements located adjacent to and 440 nucleotides downstream of the core poly(A) site of the intronless melanocortin receptor 1 gene are critical for efficient 3' end processing. Mol Cell Biol. 2007 Mar;27(5):1568–80.

284. Gromak N, West S, Proudfoot NJ. Pause sites promote transcriptional termination of mammalian RNA polymerase II. Mol Cell Biol. 2006 May;26(10):3986–96.

285. Cocquet J, Chong A, Zhang G, Veitia RA. Reverse transcriptase template switching and false alternative transcripts. Genomics. 2006 Jul;88(1):127–31.

286. Zhang H, Lee JY, Tian B. Biased alternative polyadenylation in human tissues. Genome Biol. 2005;6(12):R100.

287. Zhang H, Hu J, Recce M, Tian B. PolyA_DB: a database for mammalian mRNA polyadenylation. Nucleic Acids Res. 2005 Jan 1;33(Database issue):D116–120.

288. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, et al. The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing. Science. 2008 Jun 6;320(5881):1344–9.

289. Ozsolak F, Kapranov P, Foissac S, Kim SW, Fishilevich E, Monaghan AP, et al. Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. Cell. 2010 Dec 10;143(6):1018–29.

290. Schlackow M, Marguerat S, Proudfoot NJ, Bähler J, Erban R, Gullerova M. Genome-wide analysis of poly(A) site selection in Schizosaccharomyces pombe. RNA N Y N. 2013 Dec;19(12):1617–31.

291. Venkataraman K, Brown KM, Gilmartin GM. Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. Genes Dev. 2005 Jun 1;19(11):1315–27.

292. Graber JH, McAllister GD, Smith TF. Probabilistic prediction of Saccharomyces cerevisiae mRNA 3'-processing sites. Nucleic Acids Res. 2002 Apr 15;30(8):1851–8.

293. Mangone M, Manoharan AP, Thierry-Mieg D, Thierry-Mieg J, Han T, Mackowiak SD, et al. The landscape of C. elegans 3'UTRs. Science. 2010 Jul 23;329(5990):432–5.

294. Johnson SA, Kim H, Erickson B, Bentley DL. The export factor Yra1 modulates mRNA 3' end processing. Nat Struct Mol Biol. 2011 Oct;18(10):1164–71.

295. Robinson MD, Smyth GK. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. Biostat Oxf Engl. 2008 Apr;9(2):321–32.

296. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012 Mar;7(3):562–78.

297. Sreerama N, Woody RW. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. Anal Biochem. 2000 Dec 15;287(2):252–60.

298. Applied Biosystems User Bulletin #2- ABI Prism 7700 Sequence Detection System. [Internet]. Life Technologies, U.S.A.; 1997. Available from: http://www.appliedbiosystems.com/

299. Pérez Cañadillas JM, Varani G. Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein. EMBO J. 2003 Jun 2;22(11):2821–30.