National Institute for Medical Research
Division of Molecular Structure

The Role of Clp1 and Pcf11 in Transcription and pre-mRNA 3'-end Processing

A thesis submitted by

Joseph James Hedden

In partial fulfilment of the requirements of
University College London
for the degree of Doctor of Philosophy in Structural and Molecular Biology

August 2012

Declaration

I, Joseph James Hedden, declare that the work presented in this thesis is my own. This work was carried out in the laboratory of Dr. Ian 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.

Acknowledgments

I would like to begin by thanking Dr. Ian Taylor for giving me the opportunity to work for my PhD in his lab. He found the right balance of showing me how to do things properly whilst letting me take the project in my own direction, which has been most beneficial to me and for which I am truly grateful. He was always on hand to offer help and advice, despite the many other scientists queued up outside his office door looking for answers. His colourful language and 'subtle' humour also brought homely qualities that made me feel I had never left my local pub, and indeed, it will be difficult to leave…

I owe so much to many people along the way. I would like to thank Dr. Dave Goldstone for his tuition in all things protein, the many snooker matches and for being a good friend during my time at the NIMR. I would like to thank Valerie Ennis-Adeniran for also being a great teacher in the lab and for bringing laughter, pilfered lab equipment and hot chilli sauce, but not for supporting a certain second-rate North London football club. I am indebted to Dr. Marco Geymonat, who took me under his wing and taught me enough about yeast genetics to allow me to build the second half of my project. Without Marco the project would not have been possible. I am also very thankful for the help and guidance provided by Dr. Simon Pennel in radioactive matters, and for the critical reading of this thesis.

To all of the members of the Taylor lab (past and present), you are wonderful people. Your support, encouragement and thirst for banter know no limits. To Laura, thank you for at times being a metaphorical punching bag and I am sorry for the constant stream of Scots-related abuse. However, I think we can all agree you give as good as you get. To Laurence, thank you for always bringing a sociable atmosphere, for getting me hooked on tasty ale, but not so much for accelerating the decline in the health of my liver.

Finally I would like to thank my friends and family, who have been an essential and unwavering source of love and support throughout.

Abstract

Eukaryotic transcripts require a number of complex cotranscriptional modifications and processing events before translation to protein. Clp1 and Pcf11 are subunits of cleavage factor IA (CFIA), an essential component of the *Saccharomyces cerevisiae* pre-mRNA 3'-end processing machinery. The crystal structure of a Clp1-Pcf11 complex was determined previously and revealed the binding of ATP to a highly-conserved P-loop motif and a tight Pcf11-Clp1 interaction facilitated by a number of highly-conserved Pcf11 residues. Nonetheless, the biological function of both Clp1-ATP binding and the Pcf11-Clp1 interaction was not well understood. The work in this thesis combines an *in vitro* and *in vivo* investigation of the Clp-ATP and Clp-Pcf11 interactions in an effort to understand the function of these factors in transcription and pre-mRNA 3'-end processing. It is demonstrated that the interaction of ATP and Pcf11 with Clp1 are linked events: Loss of Clp1-ATP binding results in the abrogation of the Pcf11-Clp1 interaction and leads to Clp1 instability *in vitro*, and similarly, mutations that directly uncouple the Pcf11-Clp1 interaction also disrupt Clp1-ATP binding and cause Clp1 instability *in vitro*. An *in vivo* mutational analysis in *S. cerevisiae* revealed that both Clp1-ATP binding and the Pcf11-Clp1 interaction are essential for yeast survival. Further cell and immunoprecipitation studies demonstrated that one essential function of Clp1 is as a chaperone of Pcf11, and RT-qPCR analysis of mRNA from a sample set of yeast genes points to a role for these proteins in transcription and transcription termination rather than in poly(A) site selection.

Table of contents

List of figures

List of tables

List of abbreviations

Abbr.	Definition	Abbr.	Definition
3'-UTR	3' untranslated region	kDa	Kilodalton
5-FOA	5-fluoroorotic acid	LS	Light scattering
ADP	Adenosine diphosphate	MALLS	Multi-angle laser light scattering
AMP	Adenosine monophosphate	mAU	Milli-absorbance units
ATP	Adenosine triphosphate	M_r	Relative molecular mass
AU	Absorbance units	mRNA	Messenger ribonucleic acid
AUC	Analytical ultracentrifugation	mRNP	Messenger ribonucleoprotein
bp	Base pair	NIMR	National Institute for Medical Research
CD	Central domain/circular dichroism	NMR	Nuclear magnetic resonance
cDNA	Complementary DNA	NPC	Nuclear pore complex
CFIA/B	Cleavage factor IA/IB	nt	Nucleotide
CFIIm	Mammalian cleavage factor II	NTD	N (amino)-terminal domain
CFIm	Mammalian cleavage factor I	OD	Optical density
ChIP	Chromatin immunoprecipitation	ORF	Open reading frame
ChIP-chip	ChIP with microarray analysis	PAGE	Polyacrylamide gel electrophoresis
ChIP-seq	ChIP with DNA sequencing	PCR	Polymerase chain reaction
CID	CTD-interacting domain	PDB	Protein Data Bank
coIP	Coimmunoprecipitation	PE	Positioning element
CPF	Cleavage and polyadenylation factor	pI	Isoelectric point
CPM	Counts per million	PNK	Polynucleotide kinase
CPSF	Cleavage and polyadenylation specificity factor	PolII	RNA polymerase II
CstF	Cleavage stimulatory factor	qPCR	Quantitative real-time PCR
C_T	Threshold cycle	RNA	Ribonucleic acid
CTD	C (carboxy)-terminal domain	RNA-seq	RNA sequencing
Da	Dalton	RP-HPLC	Reverse-phase high-performance liquid chromatography
DMSO	Dimethyl sulfoxide	RRM	RNA recognition motif
DNA	Deoxyribonucleic acid	RT-qPCR	Reverse-transcription quantitative real-time PCR
DNA-seq	DNA sequencing	SDM	Site-directed mutagenesis
dRI	Differential refractive index	SDS	Sodium dodecyl sulfate
DTT	Dithiothreitol	SEC	Size-exclusion chromatography
EDTA	Ethylenediaminetetraacetic acid	snRNA	Small nuclear ribonucleic acid
EE	Efficiency element	snoRNA	Small nucleolar ribonucleic acid
GMP	Guanine monophosphate	siRNA	Small interfering ribonucleic acid
GST	Glutathione-S-transferase	TAP	Tandem affinity purification
HA	Human influenza haemaglutinin	TCEP	*Tris*(2-carboxyethyl)phosphine
ID	Interaction domain	TREX	Transcription and export factor complex
IPTG	Isopropyl β-D-1-thiogalactopyranoside	UV	Ultra-violet
ITC	Isothermal titration calorimetry	YPD	Yeast peptone-dextrose medium

Saccharomyces cerevisiae nomenclature

All chapters in this thesis discuss both the results of *in vitro* experiments with recombinantly-purified *S. cerevisiae* proteins, and the findings of *in vivo* yeast genetics experiments. Thus, it is important that the text clearly distinguishes between references to genes and proteins. Wildtype genes are denoted by using upper-case italics (e.g. *CLP1*) and mutant alleles are designated in lower-case italics followed by identification of the mutation (e.g. *clp1* T137A or *pcf11-9*). Wildtype proteins are denoted by lower-case standard text with a first-letter capital (e.g. Clp1 or Pcf11), mutants are designated in the same way but with an identification of the mutation (e.g. Clp1 T137A or Pcf11 RW-A). All references to proteins concern *S. cerevisiae* proteins unless another organism is specified.

1. Introduction

Eukaryotic gene expression is a complex process that commences with deciphering the information encoded within genes and results in the production of mature gene products. Protein coding genes are transcribed into messenger ribonucleic acid molecules (mRNAs) by RNA polymerase II (PolII) in the nucleus. mRNAs are subsequently translated into proteins in the cytoplasm by the ribosome. There are many factors that influence protein expression, beginning with the regulation of chromatin conformation and transcription initiation, and ending with post-translational modification and protein degradation. At the level of transcription, it is not only the rate of transcription that determines how well expressed a gene is. Cotranscriptional modification of mRNA also plays an important part in the control of gene expression by dictating the post-transcriptional fate of mRNA.

1.1. Eukaryotic pre-mRNA processing

Nascent transcripts, or "pre-mRNAs", must undergo a maturation process through a series of complex cotranscriptional modifications before relocation to the cytoplasm for translation. These maturation steps introduce many additional opportunities for the regulation of gene expression. The next four sections introduce the four modification steps of pre-mRNA maturation: 5'-end capping, splicing, 3'-end processing and export. The processes of capping, splicing and export are illustrated in Figure 1.1.

1.1.1. Pre-mRNA 5'-end Capping

A methylguanosine cap is added to the 5'-end of pre-mRNA after PolII has transcribed the first 25-30 nucleotides (nt) of a gene. Capping plays a major part in ensuring mRNA stability, protecting against nuclear and cytoplasmic degradation of transcripts by 5'->3' exonucleases [1, 2]. The binding of the translation factor complex eIF-4F to the 5' cap permits circularisation of pre-mRNAs through the interaction of eIF-4F with poly(A) binding protein (Pab1) bound to the 3' poly(A) tails of pre-mRNAs. Circularisation of transcripts promotes translation and translation reinitiation [3, 4]. In the budding yeast *Saccharomyces* cerevisiae, the capping process involves three enzymatic activities carried out by three distinct proteins (see Figure 1.1A) [reviewed in 5]: Firstly, the RNA triphosphatase (RT) Cet1 removes the γ-phosphate from the 5'-terminal nucleotide of the transcript [6]. Guanine monophosphate (GMP) is then transferred to the free 5'-end

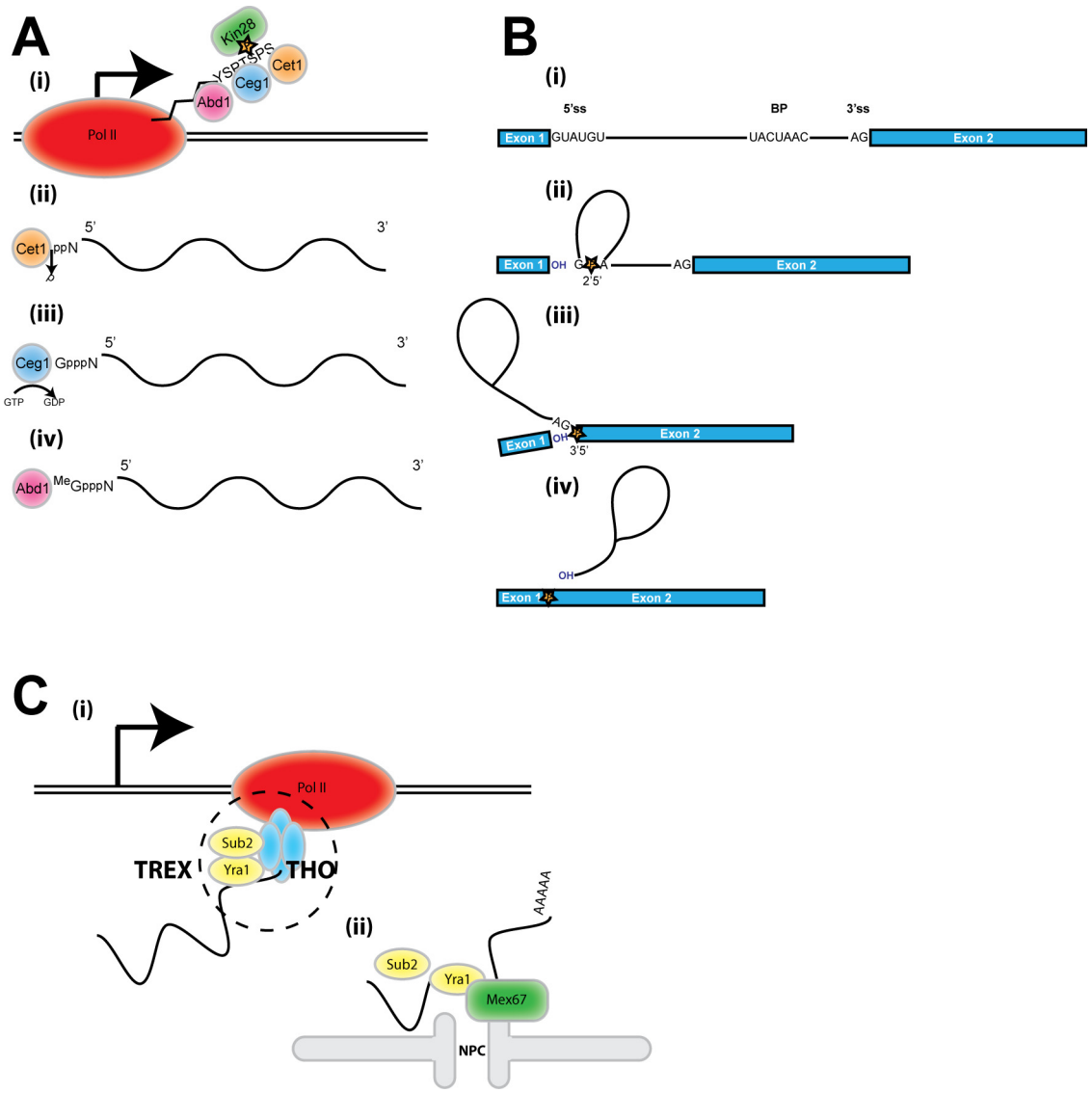


Figure 1.1. A summary of the 3'-end processing steps. (**A**) 5'-end capping involves phosphorylation of PolII CTD at Ser5 (i). The γ-phosphate of the terminal nucleotide is removed (ii) followed by transfer of a GMP molecule to the 5'-end (iii), which is subsequently methylated (iv). (**B**) Introns are removed by splicing. Specific sequences in pre-mRNA direct the splicing machinery (i). The branch point (BP) performs a nucleolytic attack on the 5' splice site (5'ss) (ii), creating a lariat intermediate (iii). The free 3' hydroxyl of the 5' exon attacks the 3' splice site (3'ss) (iii) leading to exon ligation and intron release (iv). (**C**) Nuclear export of mRNAs. The THO complex is recruited to PolII during transcription and loaded onto the nascent pre-mRNA. Subsequent Yra1 and Sub2 binding forms TREX (circled black) (i). Mex67 displaces Sub2 from Yra1 and escorts the mRNP to the nuclear pore complex (NPC) (ii).

β-phosphate by the guanylyltransferase (GT) Ceg1 [7]. Finally, N7 of the guanine base is methylated by the RNA 7-methyltransferase (MT) ABD1 [8]. All three proteins are recruited to the pre-mRNA by PolII C-terminal domain (PolII CTD), only after phosphorylation at Ser5 by the TFIIH subunit Kin28 at the promoter [9, 10]. The interaction of capping proteins with PolII stabilises the transcription complex, and Ceg1 also promotes early transcription elongation [11, 12].

1.1.2. Pre-mRNA splicing

Mammalian genes typically contain multiple, long non-coding intronic sequences that fragment the protein coding sequences into short "exons." The precise splicing of pre-mRNAs is fundamental in ensuring that exons are brought together accurately to create correctly encoded gene transcripts. Alternative splicing of pre-mRNAs allows the estimated 20-25,000 genes of the human genome to code for around 100,000 different proteins [13, 14]. In contrast, there are only ~6000 annotated protein-coding genes in *S. cerevisiae* and only 5% of these contain introns, with genes that contain multiple introns being rarer still [15]. Despite this, around 50% of the steady-state RNA population in yeast is derived from intron-containing genes [16, 17]. Splicing is achieved by the spliceosome, a large macromolecular machine composed of a core of five small uridine-rich nuclear ribonucleoproteins (UsnRNPs) that are highly conserved throughout eukaryotes, with many other associated proteins also involved (reviewed in [18-21]). Spliceosome assembly is directed by the association of UsnRNPs and other protein factors with three *cis* sequence elements: the 5' splice site (5'ss, consensus GUAUGU), branch point (BP, consensus UACUAAC), and 3' splice site (3'ss, consensus AG) [22-26]. In yeast, the mature spliceosome is formed in a stepwise fashion, firstly with the recruitment of U1 (UsnRNP) to the 5'ss and BP of the nascent pre-mRNA. A conformational change allows U2 recruitment followed by recruitment of the tri-UsnRNP U4/U5/U6. Subsequent structural rearrangements generate a catalytically competent spliceosome [27]. The splicing reaction consists of two transesterification reactions (summarised in Figure 1.1B): Firstly, the 2' hydroxyl of the branch point adenosine performs a nucleolytic attack on the phosphate of the 5' terminal nucleotide at the 5'ss, forming a 2'5' phosphodiester bond which pushes the intron into a lariat intermediate. The free 3' hydroxyl of the 5' exon then attacks the 3'ss forming a second phosphodiester bond between the 5' and 3' exons, releasing the intron (reviewed in [18-20, 28]). Aside from the obvious function of removing non-coding RNA from

transcripts, splicing is linked to transcription and other pre-mRNA processing steps, providing another level for regulation of gene expression. The presence of introns in both mammalian and yeast genes increases polymerase density and the removal of splicing signals and introns results in a reduction of nascent transcription [29]. On intron-containing genes, the export factors Yra1 and Sub2 associate with pre-mRNA only after splicing has occurred, promoting mRNA export [30, 31].

1.1.3. Pre-mRNA 3'-end processing

Export, transcript stability and translation efficiency all depend upon the generation of a polyadenylate tail (polyadenylation) at the 3'-end of transcripts. Two enzymatic processes are required for 3'-end maturation of transcripts: (1) endonucleolytic cleavage in the pre-mRNA 3'-untranslated region (3'-UTR) and (2) polyadenylation. The work presented in this thesis focuses on the protein factors involved in pre-mRNA 3'-end processing and a detailed account of the mechanisms involved is given in section 1.2.

1.1.4. mRNA export

Export of the message from the nucleus is often seen as the final step of mRNA maturation, although in reality export is dependent upon the packaging of mRNAs into mRNA ribonucleoprotein particles (mRNPs) whilst the nascent transcript is still undergoing other processing events (reviewed in [32, 33]). Packaging of mRNPs and nuclear export is mediated by a number of different multimeric complexes.

In *S. cerevisiae*, the Hpr1, Mft1, Thp2 and Tho2 proteins associate to form the THO complex, which is recruited to pre-mRNAs at the site of transcription (see Figure 1.1C). The THO complex makes subsequent interactions with Sub2 and Yra1 to form the transcription/export complex (TREX) [34, 35]. The interaction of the export factor Mex67 with the TREX components Sub2 and Yra1 is mutually exclusive, and it is possible Mex67 displaces Sub2 to escort the Yra1-associated mRNP to the nuclear pore complex (NPC) for export [36-39]. THO is essential for transcription elongation. In the absence of THO, the nascent RNA is free to hybridise with the non-template DNA strand, forming R-loops. These hinder the next round of transcription elongation by stalling the next PolII molecule, and can cause recombination leading to genetic instability [40, 41].

The TREX-2 complex is another important protein complex that serves as a connection between transcription and mRNA export. The SAGA component of TREX-2 is a histone acetylase required for transcription activation, and its association with the other TREX-2 components has been shown to relocate and anchor genes being actively transcribed to the NPC. It has also been shown that Mex67 and the NPC can interact to facilitate "gene-anchoring" [42-44]. It is currently not known whether the TREX and TREX-2 complexes act synergistically or gene-specifically when modulating transcription, mRNP formation and nuclear export via gene anchoring at the NPC.

Two other proteins, Npl3 and Nab2 have also been identified as adaptors for Mex67. Nab2 has been shown to interact directly with the NPC component Mlp1 [45-47]. Npl3 and Nab2 are believed to represent two non-redundant, major mRNA export pathways in *S. cerevisiae* (reviewed in [32]). Intimate links exist between mRNP formation and pre-mRNA 3'-end processing, and some key examples of this are presented in section 1.6.2.

1.1.5. Timing of processing events

It is convenient to think of capping, splicing, cleavage/polyadenylation and export as temporally distinct processes that are all necessary for mRNA maturation. In the past two decades much evidence has been provided to demonstrate that pre-mRNA processing events are temporally coupled with transcription and are therefore "cotranscriptional". Transcriptional coupling of pre-mRNA processing is often facilitated by the "loading" of protein factors responsible for processing onto the C-terminal domain of PolII prior to and during transcription. Consequentially, PolII CTD is often referred to as a "landing pad" or "assembly platform" for pre-mRNA processing machineries (reviewed in [48-53]).

1.2. Pre-mRNA 3'-end processing

The body of work undertaken for this PhD thesis concentrates on specific proteins belonging to the *S. cerevisiae* Cleavage Factor IA complex (CFIA), and their contributions to pre-mRNA 3'-end processing (hereafter referred to as 3'-end processing) and transcription. As stated previously, generation of mature 3'-ends requires two enzymatic activities: (1) endonucleolytic cleavage of pre-mRNA in the 3'-UTR, and (2) polyadenylation.

For gene transcripts, possession of a polyadenylate (poly(A)) tail is imperative for a number of reasons. The poly(A) tail serves to recruit protein complexes that assist in export of mRNA to the cytoplasm. The link between 3'-end processing and export of mRNA is described further in section 1.6. Once in the cytoplasm, there is an implicit requirement for mRNA to be protected from degradation. In *S. cerevisiae,* the major mRNA degradation pathways are first dependent on the deadenylation of the poly(A) tail (reviewed in [54, 55]). With this in mind, it serves well to use both a 5' cap and a long poly(A) tail as a buffer to protect the genetic code within the transcript. The poly(A) tail and factors that associate with it have also been shown to be required for efficient translation [3, 4, 56]. Finally, the cleavage and polyadenylation machinery is intimately linked with transcription, the implications of which are considered in section 1.7.

Cleavage and polyadenylation is facilitated by the highly-coordinated interaction of two multi-subunit protein complexes with an array of loosely conserved *cis* sequence elements in the 3'-UTRs of pre-mRNAs (reviewed in [57-61]). Sections 1.2.1. to 1.4.2. comprise a comprehensive account of the composition and function of the sequence elements and protein complexes necessary for 3'-end formation *in S. cerevisiae*. A model for *S. cerevisiae* pre-mRNA 3'-end processing is outlined in section 1.5 and Figure 1.8B.

1.2.1. Sequence elements for mRNA 3'-end formation

Several sequence elements required for cleavage/polyadenylation were initially identified in mammalian mRNAs. These sequences were found to show a degree of conservation throughout mammalian gene transcripts, exemplified by the highly-conserved hexanucleotide AAUAAA, termed the polyadenylation signal [62]. It took

significantly longer to find equivalent sequences in yeast mRNAs, due to a high amount of degeneracy and redundancy in yeast 3'-end forming sequence elements [58, 60].
The mechanism of 3'-end formation in *S. cerevisiae* was elucidated by the discovery that whole-cell yeast extract was able to cleave iso-1-cytochrome c (*CYC1*) pre-mRNA at a specific site, and add a poly(A) tail of 60-80 adenylate residues [63, 64]. Previously, it had been found that a 38 bp deletion in the 3'-UTR of the *CYC1* gene (*cyc1-512*) had a negative effect on *CYC1* mRNA stability and translational efficiency [65]. The subsequent analysis of eight revertant mutants of *cyc1-512* led to the identification of three separate sequence elements required for efficient 3'-end processing: (1) A positioning element (PE); (2) an efficiency element (EE) and (3) the cleavage/poly(A) site [66, 67].

The PE is located 10-30 nt upstream of the cleavage/poly(A) site and functions to determine the physical location of cleavage and polyadenylation rather than influence how efficiently the mRNA is processed [67, 68]. Initially the sequences AAGAA and UAAGAAC were identified as PEs in *CYC1* mRNA, although saturation mutagenesis of these elements revealed that AAUAAA and AAAAAA are optimal sequences [68]. Indeed, there is a high amount of degeneracy in sequences that can form effective PEs and the range of effects observed when the PE is deleted is wide. For instance, in *CYC1* deletion of the PE shifts the major cleavage/poly(A) site to a weaker downstream site [66, 67]. On the other hand, deletion of the PE in the *GAL7* gene abolishes cleavage and polyadenylation. This is likely to be because *GAL7* contains no redundant PEs [69].

The EE is found at variable distances upstream of the PE and functions to enhance the use of the preferred sites located downstream, thereby modulating the efficiency of 3'-end processing. The EE sequences UAUAUA, UACAUA and the bipartite UAG…UAUGUA were initially identified in *cyc1-512* revertant mutants [66, 67]. The bipartite UAG…UAUGUA has also been found in *TRP1, ARO4* and *TRP4* genes and the UAUAUA in both *GAL7* and *MRP2* genes [69-71]. Another sequence element, UUUUUAUA, has been identified as an EE in *CYC1*, *GCN4*, *PHO5* and *ADH1* genes [70, 72, 73]. These sequences point to an AU-rich preference for EEs in yeast genes, allowing a certain amount of degeneracy in the exact sequence. A consensus sequence for the EE was determined from studies of the plant poly(A) sequence UAGUAUGUA found embedded in a cauliflower mosaic virus (CaMV) sequence. This sequence was

also able to direct 3'-end processing in yeast. Mutation of this sequence reduced the efficiency of 3'-end formation to 15% of wildtype levels [74] and saturation mutagenesis of this sequence revealed UAUAUA produced the most effective EE [75]. A subsequent computer analysis of 1017 nuclear genes with known 3'-UTRs found that the TATATA hexanucleotide was present in the 3'-UTRs of 52% of these genes [72].

The cleavage/poly(A) site marks the location at which the pre-mRNA is physically cleaved, generating a free 3' hydroxyl group to which poly(A) polymerase 1 (Pap1) adds a 60-80 nt poly(A) tail. The cleavage/poly(A) site is a short sequence element consisting of a pyrimidine nucleotide followed by multiple adenosines ($Py(A)_N$). It is cleaved 3' of an adenosine and the PE tightly controls its location [76, 77].

More recently, a fourth uridine-rich yeast poly(A) signal has been uncovered. The existence of uridine-spanning sequences immediately flanking the cleavage/poly(A) site was revealed by comparison of 1352 expressed sequence tags (ESTs) to genomic sequences [78-80]. Remarkably, an *in vitro* assay showed that these U-rich sequences were capable of directing cleavage of short *CYC1* and *ADH1* substrates in the absence of both the EE and PE. Furthermore, mutation of the U-rich elements reduced cleavage activity [81].

The history of poly(A) signal definition illustrates the redundancy and degeneracy in many of the sequence elements involved. It is consequently difficult to hypothesise one model encompassing the function of each type of sequence element. Instead, it is highly likely that sequence elements are utilised in a gene-specific manner: All sequence elements on a transcript used in concert would ensure efficient 3'-end processing of mRNA, or selective use of sequence elements could be a method by which to finely-tune gene expression.

1.2.2. Protein factors for pre-mRNA 3'-end processing

Definition of the factors required for 3'-end processing began with the demonstration that *S. cerevisiae* cell extract could generate mature 3'-ends on *CYC1* pre-mRNA in an *in vitro* assay [63, 64]. Individual protein complexes responsible for the two enzymatic steps of 3'-end formation were later characterised by separating them from cell extract using ammonium sulphate fractionation and ion exchange chromatography: four separate fractions contained cleavage factor I (CFI), cleavage factor II (CFII),

polyadenylation factor I (PFI) and poly(A) polymerase (Pap1) respectively [82-84]. Reconstitution of the cleavage reaction *in vitro* was shown to require CFI and CFII, whereas polyadenylation was facilitated by the combination of CFI, PFI and Pap1. Interestingly, cleavage of *GAL7* and *CYC1* pre-mRNA was dependent on the presence of their EE sequences, uncovering for the first time sequence specific protein-RNA interactions in the 3'-end processing machinery [82]. It has since been recognised that CFII, PFI and Pap1 together with an array of other proteins, form a larger complex termed the cleavage and polyadenylation factor (CPF) [85, 86]. The subunit compositions and important protein-protein and protein-RNA interactions made by CFI and CPF will be discussed in sections 1.3. and 1.4. respectively.

1.3. Cleavage factor I (CFI)

The first separation of 3'-end processing complexes identified CFI as essential for reconstitution of both cleavage and polyadenylation. Further purification of CFI from cell extract led to the discrimination of two different components, CFIA and CFIB, both of which were essential for cleavage and polyadenylation of *GAL7* pre-mRNA [87]. A variety of genetic and biochemical approaches were used to illuminate the identities of the CFIA and CFIB subunits. CFIA was shown to contain four subunits, Rna14, Rna15, Pcf11 and Clp1, whereas CFIB is a single protein, Hrp1 [88-91]. The stoichiometry of CFIA subunits has since been determined as 2:2:1:1 (Rna14:Rna15:Pcf11:Clp1) [92]. several studies have looked at the importance of CFIA and CFIB subunits to pre-mRNA 3'-end processing. The key findings for each subunit are detailed in the following sections (1.3.1-1.3.5).

1.3.1. Rna14 (CFIA)

Rna14 was first identified as a 3'-end processing factor from a genetic screen in which the temperature sensitive *rna14-1* allele was combined with a temperature sensitive allele of Pap1. The combination of these two temperature sensitive alleles caused a synergistic, lethal phenotype indicative of a physical or functional interaction between these two proteins [93, 94]. Furthermore, whole-cell extract from yeast bearing the *rna14-1* allele was unable to both cleave and polyadenylate pre-mRNA *in vitro*. Biochemical complementation of *rna14-1* with partially purified CFI restored the ability of the extracts to cleave and polyadenylate the *GAL7* precursor, thus confirming the role of Rna14 in pre-mRNA 3'-end formation [88]. Later attempts to unmask the CFIA subunits resulted in the isolation of Rna14 from yeast whole-cell extract, showing it to

be a 76 kDa protein [87]. Secondary structure analysis of the Rna14 protein coding sequence demonstrated the bulk of the protein is made up of half-TPR (HAT) repeats [95]. HAT repeats are formed by three hydrophobic residues with conserved spacing and may form similar super-helical structures to those observed in HEAT repeats [96]. It has been postulated these structures could provide interaction surfaces for other proteins. Furthermore, other HAT-containing proteins have been identified as mRNA processing factors [95].

There is strong evidence to show that Rna14 interacts with other 3'-end processing factors. Firstly, Rna14 was shown to coimmunoprecipitate (coIP) with Rna15 [87]. This interaction was later characterised by electron microscopy and analytical ultra-centrifugation (AUC), revealing a stable heterotetramer of two Rna14 and two Rna15 molecules, dependent upon the Rna14-mediated association of two Rna14-Rna15 heterodimers [97]. Recently, the solution structure of the Rna14-Rna15 interface was solved by nuclear magnetic resonance (NMR). The structure showed Rna14 bound Rna15 using a "monkeytail" domain located between C-terminal residues 626-677. The Rna14 monkeytail domain binds to Rna15 by wrapping around a core of four alpha helices, with N- and C-terminal helices from Rna15 stabilising the interaction [98]. A direct Rna14-Pcf11 interaction was also confirmed by glutathione-S-transferase (GST) pulldown experiments, even in the absence of other CFIA subunits [99]. Rna14 has also been identified as the bridge between CFIA and CFIB (Hrp1) based on the observation of synthetic lethality *in vivo* when double mutants of *hrp1/rna14* or *hrp1/rna15* were made [91]. Following on from this finding, an *in vitro* pulldown experiment using recombinant CFIA components and Hrp1 showed that only Rna14 bound to Hrp1 [99].

Large-scale analyses of protein-protein interactions has been made possible in yeast by tandem affinity purification (TAP), where one protein is "TAP-tagged" and purified in combination with its binding partners and other proteins in an associated complex. This approach is often used to isolate whole macro-molecular protein complexes, whose subunits can be identified by protein sequencing or mass spectrometry [100, 101]. When CPF subunits were TAP-tagged, a weak interaction between CPF and Rna14 was identified, demonstrating communication between 3'-end processing complexes [102]. This was consistent with earlier findings from a variety of affinity capture experiments which revealed that Cft1, Fip1, Pti1 and Pfs2 subunits of CPF all directly interact with

Rna14 [86, 103-105]. Yeast two-hybrid analysis and an *in vitro* pull down assay showed Rna14 also binds to the C-terminal domain of PolII (CTD) when PolII CTD is phosphorylated [106]. In contrast to the Pcf11-CTD interaction (discussed in section 1.7.1.), the significance of the Rna14-PolII interaction is still unclear.

Aside from its function as an essential pre-mRNA processing factor in the nucleus, Rna14 has also been discovered localised in the mitochondria. Further to this, the *rna-14-1* mutant (which does not cleave or polyadenylate pre-mRNA) displays no defects in mitochondrial function, suggesting Rna14 might have a distinct, but redundant function in mitochondrial metabolism [107, 108].

Despite having an array of interaction partners, no specific activity (enzymatic, RNA-binding etc.) has been suggested for Rna14. Instead, evidence points towards Rna14 as an architectural component of CFIA that bridges Rna15 and Hrp1, both of which bind to pre-mRNA (see sections 1.3.2 and 1.3.5.). Indeed, it has been shown *in vitro* that Rna14 increases the affinity of both Rna15 and Hrp1 for pre-mRNA [109].

1.3.2. Rna15 (CFIA)

The importance of Rna15 in pre-mRNA 3'-end processing was unveiled in parallel with Rna14 by exploiting the *rna15-1* temperature sensitive allele using the same genetic and biochemical approaches that were used with Rna14 [88, 93, 94]. Rna15 was first isolated from yeast extract, revealing it to be a 38 kDa protein that was essential for reconstitution of both cleavage and polyadenylation reactions *in vitro* [88]. Sequence homology and secondary structure analysis show Rna15 contains an N-terminal RNA-recognition motif (RRM) together with an essential CTD.

Rna15 is essential for pre-mRNA 3'-end maturation because it is the RNA-binding component of CFIA, and the position to which it binds to pre-mRNA indirectly defines where the transcript will be cleaved and polyadenylated. The RNA-binding capability of Rna15 was first confirmed in an *in vitro* assay, when irradiation with UV light cross-linked Rna15, but not other CF1 subunits, to a radiolabelled *GAL7* RNA substrate [87]. Many experiments following on from this finding endeavoured to determine the specificity of Rna15 for pre-mRNA sequence elements, and whether other CFI components are necessary for recognition of pre-mRNA by Rna15. Initially, a preference for uridine-rich sequences was suggested, when *in vitro* translated Rna15

was pulled-down by poly(U)-agarose, but not poly(A/C/G)-agaroses. This did not require the presence of any other CFI components [110]. In contrast, subsequent UV cross-linking experiments demonstrated that recombinant Rna15 bound to a *GAL7* precursor non-specifically. However, when combined with Rna14 and Hrp1, Rna15 was UV cross-linked specifically to the A-rich *GAL7* PE, inferring that only in a stable Rna15-Rna14-Hrp1 complex could the specificity of Rna15 for certain RNA sequences be conferred [111]. One study that expanded on this discovery used surface plasmon resonance and gel retardation experiments to characterise the Rna14-Rna15-RNA interaction. These fresh approaches enabled lower protein:RNA concentrations to be used than in UV cross-linking or affinity-capture experiments, helping to lower the possibility of non-specific protein-RNA binding. This showed that an Rna14-Rna15 complex interacted strongly with a *GAL7* 3'-UTR-derived oligonucleotide, but Rna15 alone did not interact. This highlighted the necessity of at least Rna14 as a binding partner to promote the Rna15-RNA interaction [97].

Perhaps the most robust evidence to date for Rna15-RNA sequence specificity was provided from three crystal structures of the Rna15 RRM [112]. The structure of the Rna15 RRM alone, together with two structures of Rna15-RNA complexes revealed a conserved mechanism for recognition of GU-rich RNAs. The structures demonstrate that the Rna15 RRM comprises a canonical RRM fold (β1-α1-β2-β3-α2-β4-β5) that binds to a GUUGU penta-ribonucleotide through two nucleotide-binding sites termed Site-I and Site-II (see Figure 1.2A). Specifically, residues located in Site-I of the RRM make important base-stacking interactions with guanine or uracil bases, with additional hydrogen bonding stabilising the interaction (see Figure 1.2B and legend). Crucially, functional groups on the edges of the U and G bases form Watson-Crick-like hydrogen bonds with the backbone of Site-I residues. These interactions would not be possible with A or C bases in the orientation observed in the crystal structure, signifying Site-I discriminates against A and C bases in favour of interactions with U or G bases. The base-discrimination mechanism was supported by measurement of the binding affinities of the Rna15 RRM for ribo-oligonucleotides of different base compositions. In these experiments, the RRM showed a clear preference for GU-rich or U-rich RNA-oligos over RNA-oligos containing C or A bases. The crystal structures also revealed that a second site (Site-II) is important for RNA binding, although this site is probably not for base-discrimination. Instead, the aromatic side chain of Tyr21 in the Site-II pocket makes stacking interactions with the bases of a GU di-ribonucleotide. Furthermore,

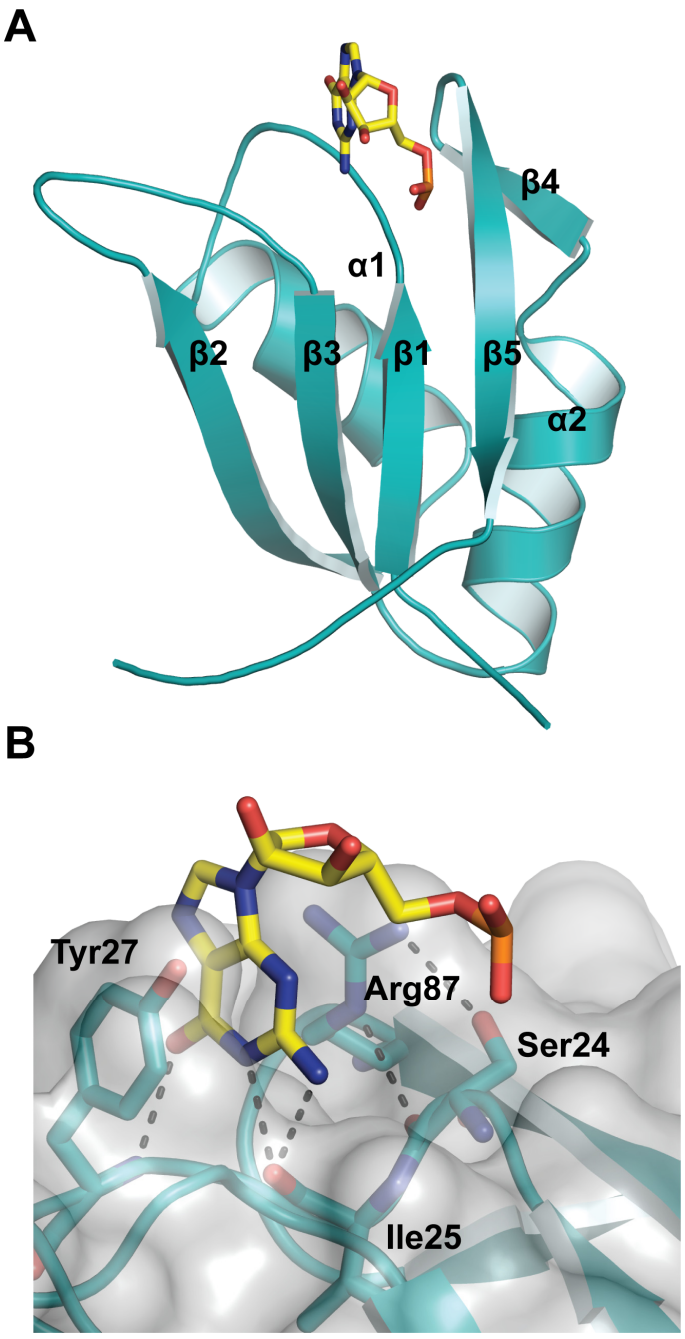


Figure 1.2. Structure of the Rna15 RRM and Site-I base-binding pocket. (**A**) The Rna15 RRM (cartoon representation) adopts a β1-α1-β2-β3-α2-β4-β5 topology (secondary structure elements are labelled). The guanine base (yellow sticks) is bound within a base-binding site located on the loop connecting β1 and α1. (**B**) Site-I interactions with the guanine base. The guanine base (yellow sticks) is stacked between the side chains of Tyr27 and Arg87. The conformation of the base-binding pocket is maintained by hydrogen bonding (dashed black lines) between the side chain of Arg87 and both the main carbonyl and side chain hydroxyl of Ser24. Binding is further stabilised by hydrogen bonds between functional groups on the edges of the guanine base and the main chains of Tyr27 and Ile25. The Rna15 RRM is depicted as a cartoon with semi-transparent surface representation. Important residues are illustrated (shown in stick representation and labelled). *PDB codes: 2X1A, 2X1B, 2X1F.*

site-directed mutagenesis of key residues in both Site-I and Site-II reduced RNA binding drastically, confirming the importance of these nucleotide-binding sites in RNA recognition [112]. The authors also draw attention to the abundance of U-rich sequences flanking the cleavage/poly(A) site throughout the yeast genome [78-80, 112]. They liken this to previous findings that the mammalian homologue of Rna15, CstF64, also displays a preference for GU-rich sequences that surround the poly(A) signals of mammalian genes [113]. Alignment of the CstF64 RRM and Rna15 RRM structures demonstrates strong sequence and structure conservation of Sites-I and II, confirming that the mechanism of nucleobase-discrimination is also conserved. The conservation of a mechanism conferring GU specificity taken with the prevalence of U-rich sequences flanking yeast cleavage/poly(A) signals makes it highly likely that it is the U-rich elements that are the preferred Rna15 target sequences [112].

There is a wealth of evidence to suggest that the specificity of the Rna15-RNA interaction is dependent on Rna15 incorporation into CFIA [97, 111]. Rna15 residues 127-232 contain a core of four α-helices that comprise the Rna14 interaction domain (described in section 1.3.1) [98]. It was previously shown that Rna14-Rna15 complexes formed A_2B_2 heterotetramers, in which the association of individual Rna14-Rna15 heterodimers was mediated by the homodimerisation of two Rna14 molecules [97, 98]. It recently came to light that the stoichiometry of Rna14-Rna15 complexes is important for correct pre-mRNA 3'-end formation. Mutation of the Rna14 dimerisation domain prevented the formation of Rna14-Rna15 heterotetramers *in vitro*. Yeast extract from a strain bearing the same mutation (*rna14ES*), showed reduced cleavage and polyadenylation of a pre-mRNA substrate, signifying that for full 3'-end processing function two copies of Rna15 are required [92]. Rna15 also binds to Pcf11 within CFIA [87]. The Pcf11 binding site has been mapped to the conserved Rna15 CTD by *in vitro* pulldown assays. Truncation of Rna15, or mutation of Rna15 CTD residues abolished the Rna15-Pcf11 interaction and deletion of the Rna15 CTD caused a restrictive-growth phenotype in yeast (*rna15-t8* temperature sensitive mutant). The *rna15-t8* mutant displayed a marked reduction in cleavage and polyadenylation, but chromatin-immunoprecipitation analysis (ChIP) revealed there was no defect in transcription termination [114]. This points to a requirement for the Rna15-Pcf11 interaction in efficient pre-mRNA 3'-end processing but not for transcription termination.

There have also been reports of other important binding partners for Rna15 outside the CFIA complex. Rna15 may be important in crosstalk between 3'-processing and transcription by facilitating gene looping. These findings will be discussed in greater detail in section 1.7.2.

1.3.3. Pcf11 (CFIA)

Initial separation of CFI components revealed five proteins, two of which had been previously determined to be Rna14 and Rna15 [87, 88]. Yeast two-hybrid screens in which either Rna14 or Rna15 were used as bait, identified genetic interactions with a 70 kDa CFI subunit, subsequently termed Pcf11 [89]. Analysis of the Pcf11 protein sequence reveals a central stretch of 20 glutamines, and two zinc finger motifs located in the CTD. The N-terminal 140 amino acids of Pcf11 comprise a highly-conserved domain that interacts with PolII CTD, termed the CTD-interacting domain (CID) [115]. The membership of Pcf11 in CFIA was confirmed in the same study by an *in vivo* experiment showing an epitope-tagged version of Pcf11 could copurify with Rna14 and Rna15. The importance of Pcf11 to pre-mRNA 3'-end processing became immediately apparent when yeast extract derived from the *pcf11-1* or *pcf11*-2 temperature sensitive mutants was unable to correctly cleave or polyadenylate mRNA precursors [89].

Purification of recombinant CFI factors enabled affinity-capture experiments designed to elucidate the protein-protein interactions within CFI. These experiments revealed that Pcf11 interacts with all of the other CFIA subunits, but not Hrp1 (CFIB) [99]. A domain map of Pcf11 is shown in Figure 1.3A. The Pcf11-Rna14/ Rna15 interaction domain was mapped by two-hybrid analysis (using various length fragments of Pcf11) to residues 288-400 [89]. More detailed information on the Pcf11-Rna14 and Pcf11-Rna15 interactions can be found in sections 1.3.1. and 1.3.2. Details of the Pcf11-Clp1 interaction have been provided by a crystal structure of a Clp1-Pcf11-ATP complex [116]. The remarkable conservation of key residues in the Pcf11-Clp1 interaction domain (ID) is portrayed along with the structure of the binding interface in Figure 1.4. Residues 475-499 of Pcf11 constitute the Clp1-ID, and form an extended conformation that is entirely bound to Clp1 within a cleft between the Clp1 central domain (CD) and CTD, and a shallow groove in the CD. A surface representation of the crystal structure shows that the Pcf11 fragment is intimately associated with Clp1 (Fig 1.4A). The Pcf11-Clp1 interface is maintained largely by hydrogen bonding of the side and main chains of Pcf11 residues to Clp1 residues located on the β10-β11 loop that lines one

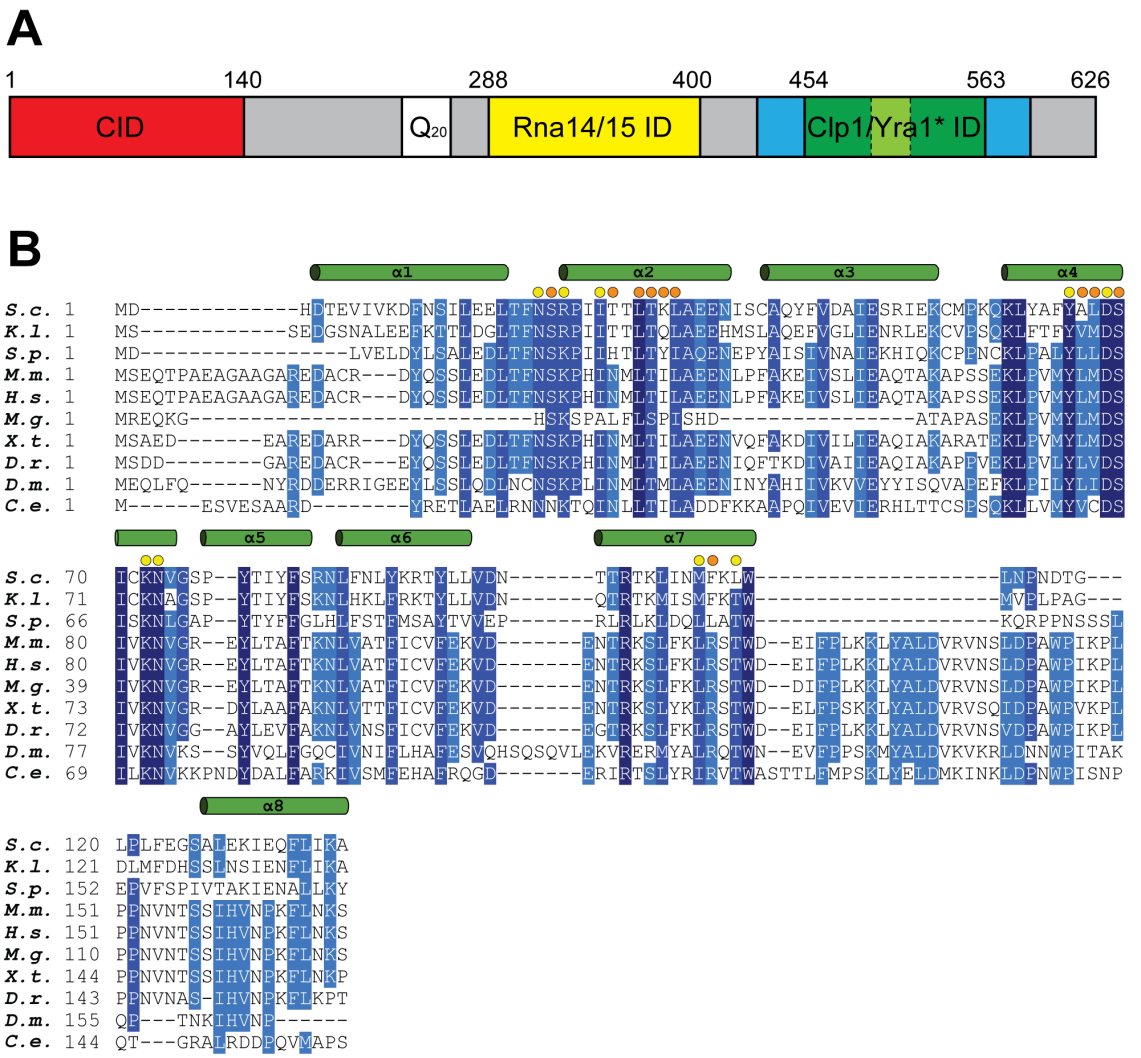


Figure 1.3. Pcf11 contains binding sites for other pre-mRNA 3'-end processing factors. (**A**) Domain map of Pcf11. Interaction domains are displayed in different colours and labelled with residue numbers cited above the map. The two blue areas indicate conserved zinc finger motifs. The light green area within the dashed grey lines contains residues 475-499 which signifies the known Clp1 interaction domain. Q_{20} indicates the polyglutamine tract. Grey areas show where no function has been inferred. (**B**) Sequence alignment of the Pcf11 CID. The strength of residue conservation is highlighted in shades of blue (dark blue most conserved). Secondary structure elements are marked above the alignment in green, yellow dots indicate residues that interact with PolII CTD in the CID-CTD crystal structure, orange dots indicate additional residues that display NMR chemical shift patterns on CTD binding. Species names can be found with their abbreviations in Appendix 8.1.

side of the CD-CTD cleft. Three highly-conserved residues in Pcf11 (Arg480, Trp482 and Trp489) are particularly important for complex formation (see Fig 1.4B). The side chain of Arg480 makes contacts with Ser192 and Gln191 on the β10-β11 loop, and also Tyr332 and Glu331 located on α7. The edge of the Clp1 CD-CTD cleft contains a small hydrophobic cavity in which the aromatic side chain of Trp482 is buried, making solely hydrophobic contacts with the aliphatic side chains of Leu406, Pro426 and Val427. The short α-helix of the Pcf11 fragment contains Trp489, which extends into a pocket on the surface of the Clp1 central domain, making hydrogen bonds with Gln204 and Arg236 and hydrophobic contacts with Met206 and Arg236. Compellingly, a sequence alignment reveals that Arg480, Trp482 and Trp489, are absolutely conserved in other yeast, mammals, birds, amphibians, fish and invertebrates (Figure 1.4C). The function of the Pcf11-Clp1 interaction remains obscure, yet the extensive bonding networks of these three Pcf11 residues combined with their strong conservation suggests that the Pcf11-Clp1 interaction serves an important purpose, and is likely to be essential [116]. Like Rna14, Pcf11 bridges CFIA and CPF complexes through multiple connections with CPF subunits. This was demonstrated in an *in vitro* affinity-capture experiment employing four core subunits of CPF, GST-tagged Cft1, Cft2, Ysh1 and Pta1, that all bound to *in vitro* translated Pcf11 [103].

As previously discussed, Pcf11 is essential for correct 3'-end formation of mRNA transcripts. However, Pcf11 has also been implicated as an essential factor in mRNA export, transcription termination and more recently transcription initiation. The multi-functional nature of Pcf11 has sparked a rich vein of research into how different stages of gene expression are tightly coupled and the key findings portraying the centrality of Pcf11 in gene expression will be elaborated further in sections 1.6 and 1.7.

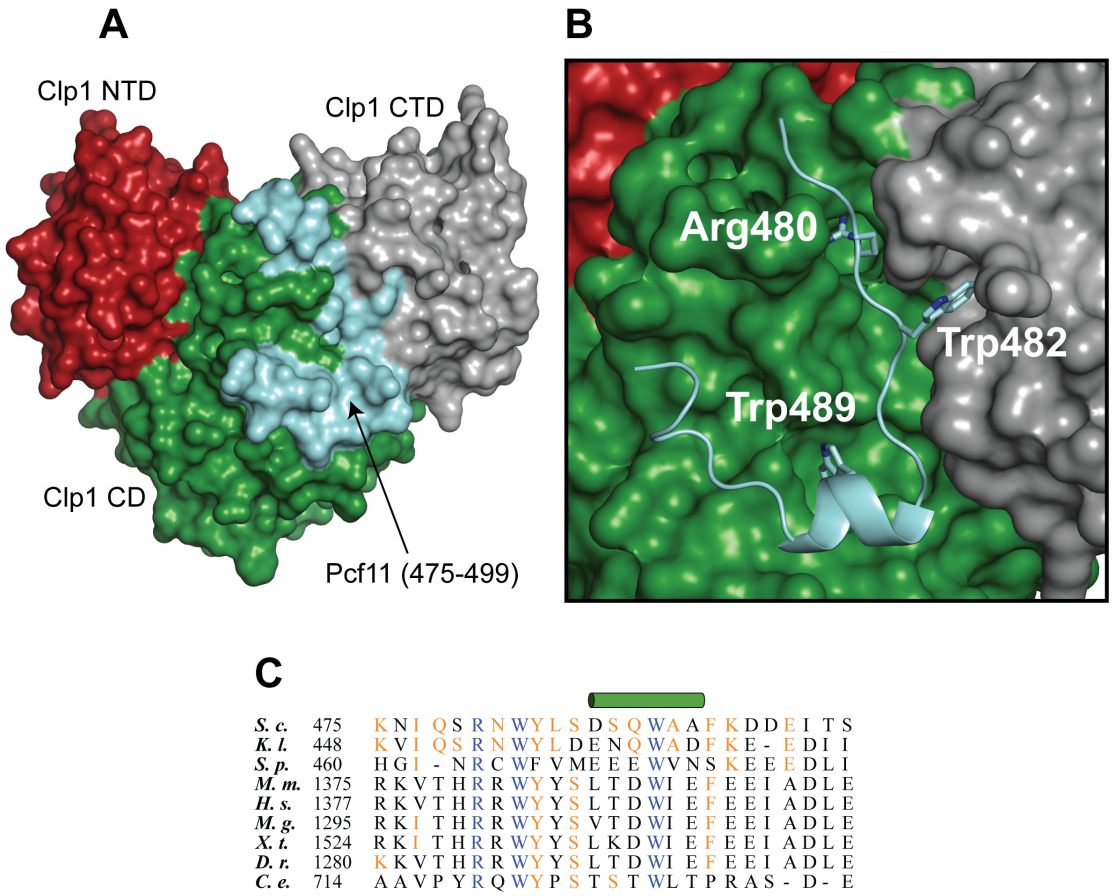


Figure 1.4. The Pcf11-Clp1 interaction. (**A**) Surface representation of the Pcf11-Clp1 heterodimer shows that Pcf11 (cyan) moulds tightly into a cleft between the Clp1 central domain (CD, green) and C-terminal domain (CTD, grey), and extends into a groove in the CD. The N-terminal domain (NTD) is shown in red. (**B**) The Pcf11-Clp1 interface contains three highly-conserved Pcf11 residues: Arg480, Trp482 and Trp489. The side chains of these residues are oriented into pockets within the Clp1 CD-CTD cleft and the groove within the CD. Pcf11 is shown as a cyan cartoon with residues labelled and side chains shown in stick representation. Clp1 is portrayed in surface representation with domain colouring as in (**A**). (**C**) Sequence alignment of the Pcf11-Clp1 ID. Residues totally conserved with *S. cerevisiae* are coloured blue, residues conserved in most species are coloured orange. The green cylinder indicates an α-helix. Species names and their abbreviations can be found in Appendix 8.1. *PDB code: 2NPI.*

1.3.4. Clp1 (CFIA)

With various properties of Rna14, Rna15 and Pcf11 already assigned, the 50 kDa CFIA subunit remained unidentified. One study used a range of genetic and biochemical techniques to address this issue. A vector containing a His-tagged *RNA15* gene was introduced into a yeast strain containing a lethal *rna15* deletion, and CFIA components were then copurified from yeast extract with His-Rna15 using a variety of chromatographic techniques. Peptide microsequencing of the 50 kDa subunit allowed subsequent cloning of the gene open reading frame (ORF), designated *CLP1* for cleavage/polyadenylation factor IA subunit [90]. In an *in vitro* assay using individual recombinant CFIA factors combined with extract-purified CPF, *CYC1* pre-mRNA was only cleaved and polyadenylated when Clp1 was present, indicating it is obligatory for 3'-end processing [99]. Further pull-down experiments using recombinant Clp1 attached to beads with Clp1 antiserum revealed that within CFIA, Clp1 only interacts with Pcf11 and that the Clp1-Pcf11 interaction was obligatory for capture of Rna14 and Rna15 by Pcf11. This suggests that Clp1 is mandatory for CFIA formation [99]. A study investigating the interaction partners of core CPF subunits revealed Clp1 also interacts strongly with Ysh1, and displays a weaker affinity for Cft1 [103]. The Clp1-Ysh1 interaction is extremely interesting, since Ysh1 is suspected to provide the endonuclease activity that cleaves pre-mRNA (see section 1.4.1).

Sequence analysis of the Clp1 human homologue (hClp1) uncovered a Walker A motif, which was shown to be highly-conserved in other eukaryotic Clp1 homologues including *S. cerevisiae* [117]. A multi-species sequence alignment of Clp1 can be found in Appendix 8.1. Walker A motifs are typically found in P-loop NTPase proteins and are required for nucleotide-binding. The motif is composed of a flexible loop between an α-helix and a β-strand, which contains the consensus sequence GxxxxGKT/S and binds the triphosphate moiety of a nucleotide [118].

The crystal structure of a nucleotide-bound Clp1-Pcf11 complex provided a detailed view of the nucleotide-binding site, and raised questions as to the function of the bound nucleotide and Clp1 in general [116]. The structure reveals Clp1 is composed of a large central domain (CD), flanked by two smaller N- and C-terminal domains (Figure 1.5A). The Clp1 CD comprises both α and β secondary structure elements, and is responsible for nucleotide-binding. It is a member of the SIMIBI class of NTPases, containing the characteristic features which are: (1) a P-loop/Walker A motif (herein referred to as P-

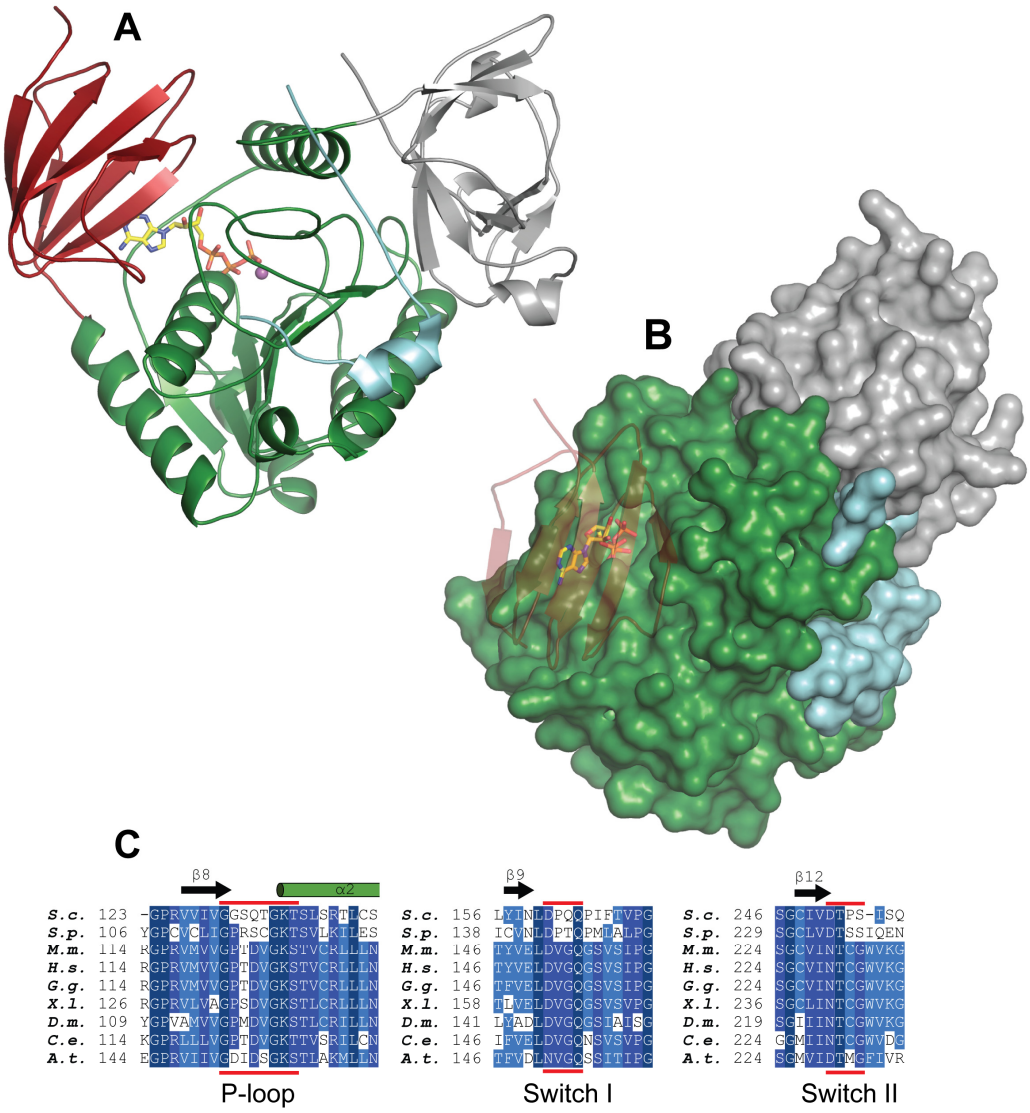


Figure 1.5. Crystal structure of the Clp1-Pcf11-ATP complex. (**A**) Clp1 is composed of a large CD (green cartoon) flanked by a smaller NTD (red cartoon) and CTD (grey cartoon). The CD contains a nucleotide-binding pocket within which an ATP molecule is bound (stick representation, yellow base) and a magnesium ion is present (purple sphere). A small fragment of Pcf11 comprising residues 475-499 (cyan cartoon) is bound to the opposite face of Clp1 from the ATP-binding pocket. (**B**) Clp1 NTD (shown here as a transparent red cartoon), is a seven-stranded β-sandwich that packs tightly against the CD rendering the ATP-binding pocket inaccessible. The colour scheme is as in (**A**). (**C**) Sequence alignments of the Clp1 P-loop, Switch I and Switch II motifs. Residue conservation is illustrated by blue shading (dark blue most conserved). Secondary structure elements are marked above the sequence. Residues that constitute the motifs are underlined in red and labelled. Species names and abbreviations can be found in Appendix 8.1. *PDB code: 2NPI.*

loop) containing a third glycine (consensus GxxGxGKT/S); (2) an aspartate responsible for magnesium ion coordination termed "Switch I"; (3) a second aspartate, "Switch II", which makes interactions required for structural stability of the nucleotide-binding pocket [119]. These three motifs are brought together to form the majority of the nucleotide-binding pocket of Clp1. Close inspection of the P-loop (Figure 1.6A) shows the α-phosphate of ATP is hydrogen bonded only to the main and side-chains of Ser138. The β-phosphate is hydrogen bonded to the main and side-chains of Thr137, and also contacts the main-chains of Thr134, Gly135 and Lys136. The γ-phosphate makes hydrogen bonds with the side-chains of Gln133 and Lys136, as well as a contact outside the P-loop with the ε-amino group of the Lys321 side chain located in the CD. The Clp1 P-loop therefore behaves in a classic manner, positioning the phosphate moiety of ATP.

Outside the P-loop, there is a hydrogen bond between the 2'OH of the ribose and Lys72 in the Clp1 N-terminal domain (NTD). The ε-amino group of Lys72 also contacts the main-chain of Ala315, which helps maintain a close proximity between the adenine base and the base-binding loop (residues 308-312). This allows the main-chains of the base-binding loop residues to make Van der Waals interactions with the adenine base. The base is also bound by Asp33 in the NTD, in a manner that is suggestive of specificity for an adenine base: The exocyclic N6 amino group of the adenine base is contacted by the side-chain of Asp33, whereas the O6 carbonyl oxygen of a guanine base in the same position would not be able to form a hydrogen bond [116]. A summary map of the interactions between Clp1 residues and ATP is presented in Figure 1.6B.

Switch I and Switch II are located adjacent to each other and opposite the P-loop, enclosing the triphosphate moiety of ATP. The P-loop and Switch regions of Clp1 adopt a similar conformation to those observed in the nucleotide-binding site of the NifH iron protein of the *Azobacter vinelandii* nitrogenase. In NifH, the Switch regions contain residues that effect nucleotide hydrolysis. Instead of interacting directly with the bound nucleotide, they coordinate a Mg^{2+} ion via a network of three equatorially positioned water molecules, forming an octahedral geometry of magnesium coordination [120]. The Clp1 structure shows Mg^{2+} ion coordination is similar to that of NifH, although in Clp1 only one water molecule is present, possibly due to lack of

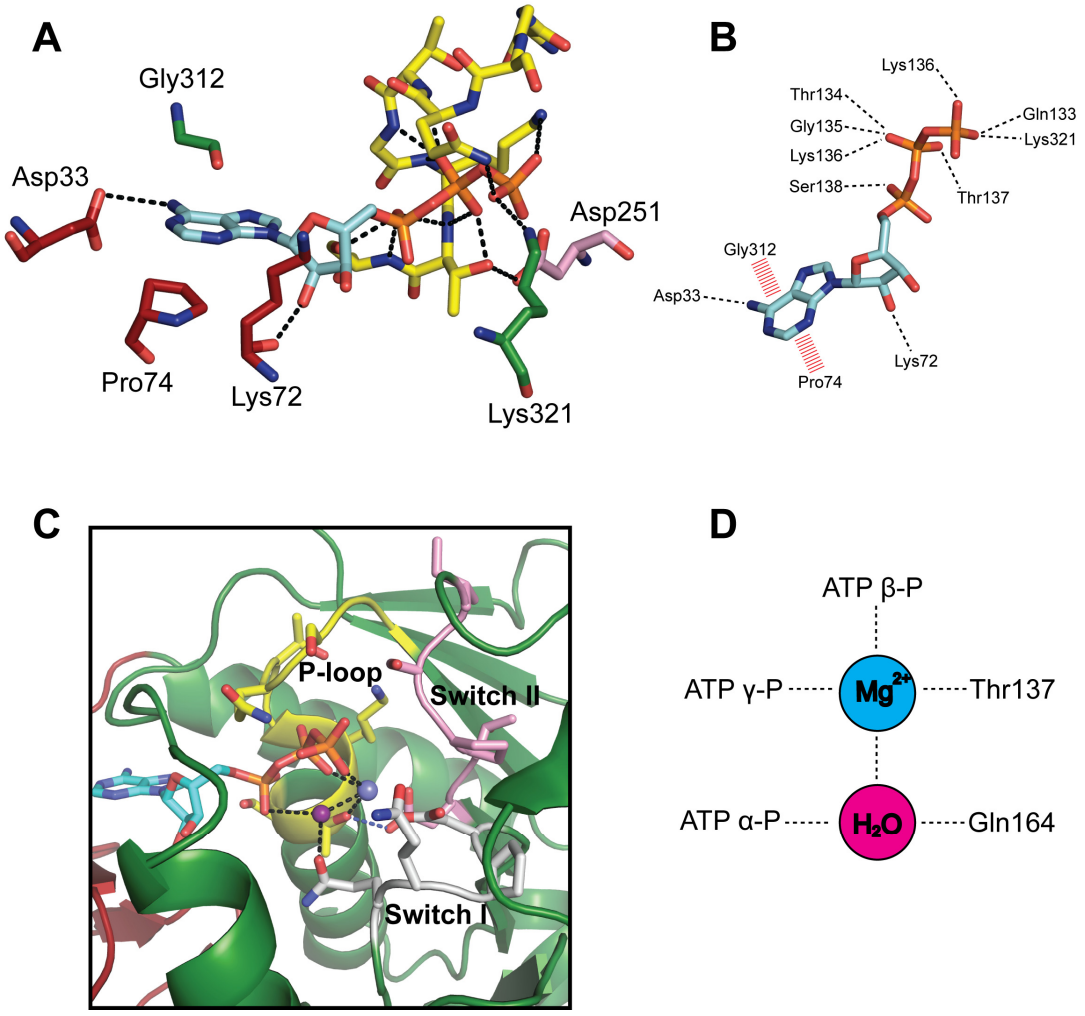


Figure 1.6. Structure of the Clp1 ATP-binding pocket. (**A**) Structural depiction of residues that interact with ATP. ATP is shown as sticks with the base in cyan and the phosphates in orange. The main and side-chains of amino acids involved in ATP-binding are displayed as sticks: P-loop residues are shown in yellow, Switch II residues in pink, CD residues in green and NTD residues in red. Non-P-loop residues are labelled and hydrogen bonds are marked by dashed black lines. (**B**) Summary of Clp1 residues that interact with ATP. Residues are labelled, hydrogen bonds are marked by dashed black lines and Van der Vaals interactions are marked by dashed red lines. (**C**) Structural depiction of magnesium ion coordination. The magnesium ion is shown as a blue sphere and water molecule as a purple sphere. Coordination of the water molecule and magnesium ion is marked by dashed black lines. Clp1 is in cartoon representation with the side-chains of important residues depicted as sticks. The colour scheme is as in (**A**), with Switch I shown in grey. (**D**) A summary of magnesium ion coordination. Coordinating residues/molecules are labelled and coordination is shown by dashed black lines.

resolution in the structure (Figure 1.6C). The Mg^{2+} ion is coordinated by the side-chain of Thr137, both β and γ-phosphates of ATP and a water molecule. The water molecule is positioned by the ATP α-phosphate and the side-chain of Switch I Gln164 (Figure 1.6C, D). Importantly, there is a hydrogen bond between the side-chains of Switch II Asp251 and P-loop Thr137. This interaction is conserved amongst other ATPases and GTPases and acts to bolster the P-loop to Switch II, which is probably crucial for the stability of the nucleotide-binding pocket [116].

The crystal structure of Clp1 also reveals that the NTD consists of a seven-stranded β-sandwich, which packs tightly against the CD through interactions between a conserved loop separating β3-β4 (NTD) and α2 (CD). Figure 1.5B shows this packing completely masks the ATP-binding pocket, meaning a conformational change would be required in order to access it. The CTD contains a mixture of α and β secondary structure, and contains a conserved region in the loop separating β20-β21. This loop forms part of the border with the CD and encompasses the hydrophobic cleft in which the Pcf11 peptide (475-499) is bound (the Pcf11-Clp1 interaction is discussed in section 1.3.3.) [116].

In the same study from which the crystal structure of Clp1 was provided, it was demonstrated that the sole species of nucleotide that could be extracted from Clp1 (expressed and purified from *Escherichia coli*) is ATP. This indicates that ATP bound by Clp1 is stable to hydrolysis. At the time, the authors postulated that the interaction of another protein partner may be required to induce a conformational change in Clp1 to permit ATP hydrolysis [116]. To date, no evidence has been provided to support this theory. Two recent publications have posited that Clp1 does not turn over ATP, but instead ATP can be differentially oriented within the ATP-binding pocket [121, 122]. In this scenario, the orientation of ATP dictates whether Clp1 can interact with Pcf11, which in turn has downstream effects on communication between 3'-end processing machineries and PolII [121]. However, the required structural evidence to validate this mechanism has also not yet been provided.

1.3.5. Hrp1 (CFIB)

Hrp1 was originally identified in a genetic screen that showed a mutant *hrp1-1* allele acted as an extragenic suppressor of the *npl3-1* temperature sensitive allele (*NPL3* encodes Npl3, an mRNA export factor). The *HRP1* gene was subsequently cloned and

sequence analysis revealed a single domain protein containing two RRMs between residues 160-290. These RRMs shared significant sequence similarity with the human hnRNPs A1, A2/B1 and C1 [123]. At around the same time as the *HRP1* gene was identified, the study in which the individual subunits of CFI were purified revealed a 73 kDa protein as the only constituent of CFIB [87]. The 73 kDa CFIB protein was subjected to a tryptic digest, and the protein sequences of two resulting peptides were perfectly matched to the Hrp1 protein sequence [91]. Hrp1 was found to be vital for cleavage and polyadenylation of *GAL7* pre-mRNA *in vitro*, and poly(A) tails isolated from the *hrp1-5* temperature sensitive yeast strain were found to be abnormally short, reinforcing the fact that *hrp1-5* is required for proper polyadenylation [91].

Hrp1 is bridged to the rest of CFI via its interaction with Rna14 (see section 1.3.1). A genetic interaction between Hrp1 and Rna15 was observed by both synergistic lethality between *hrp1-6*/*rna15-1* and *hrp1-8*/*rna15-1* double mutants, and by a yeast two-hybrid approach [91]. However, *in vitro* pull-down experiments using recombinant proteins contradicted the genetic approaches by suggesting these proteins are not binding partners [99].

RNA-binding activity had already been detected within CFI through the Rna15 subunit, and the presence of two RRMs in Hrp1 indicated that it might also interact with pre-mRNAs. Hrp1-RNA binding was confirmed when it was demonstrated recombinant Hrp1 could be UV-crosslinked to a radiolabelled *GAL7* substrate. This interaction was specific for the EE of *GAL7* pre-mRNA (UAUAUA) [91]. Quantitative gel mobility shift assays showed that the Hrp1 binds substrates containing the $(UA)_3$ sequence with nanomolar (nM) affinity [124]. Competition assays reinforced these results, showing *ADH2* pre-mRNA lacking the EE sequence could not compete with a pre-formed Hrp1-*ADH2* RNA (wildtype) complex. However, a sequence containing the *ADH1* gene 3'-UTR (which does not contain the UAUAUA element) could successfully compete for Hrp1 binding with *ADH2* RNA, indicating that although Hrp1-RNA binding is specific, Hrp1 does recognise other sequences in the 3'-UTRs of pre-mRNAs [125]. The structure of the Hrp1-RNA interaction domain in complex with the *GAL7* EE has since been solved by NMR. The structure shows two typical RRM folds (β1-α1-β2-β3-α2-β4-β5) with both involved in binding to the EE. The RRMs interact with each other, creating a cleft that is highly positively charged and within which the majority of the EE sequence binds. The first two bases at the 5'-end of the hexa-ribonucleotide are bound

by RRM2, whereas the next four bases are bound by RRM1. Specificity of the adenine and uracil bases is mediated by both hydrogen bonding and hydrophobic stacking interactions between the most conserved residues in the RRMs and RNA [126, 127]. Taken together, the biochemical and structural findings point to a specific interaction between Hrp1 and the $(UA)_3$ EE in the pre-mRNA 3'-UTRs of *S. cerevisiae* gene transcripts. Although other sequences can also be recognised, it is unclear what their compositions are and how recognition would be conferred at the molecular level.

Although the role of Hrp1 as a pre-mRNA 3'-end processing factor is undisputed, it may not be involved in the processing of transcripts from every yeast gene. An RNA-IP microarray analysis revealed Hrp1 was enriched on transcripts containing the UA-rich EE, and was also enriched on the transcripts of genes involved in metabolism [128]. Furthermore, it has been noticed that Hrp1 is rapidly translocated to the cytoplasm during hypo-osmotic stress. The lack of Hrp1 in the nucleus in this scenario implies it is not required for the processing of all gene transcripts, or is not required for transcriptional processing during the stress response [129]. Taking this into account, it is tempting to speculate that Hrp1 may act within CFIA to indirectly position Rna15 on templates to which it has low affinity (templates with a lack of GU-rich sequence flanking the poly(A) site) using Rna14 as a bridge. This theory is given substance by the fact that Rna15 only recognises the A-rich PE of *GAL7* pre-mRNA specifically as a Hrp1-Rna14-Rna15 complex [111].

1.4. Cleavage and Polyadenylation Factor (CPF)

The pioneer studies of pre-mRNA 3'-end formation in *S. cerevisiae* focused on determining the minimal set of protein factors that are required to reconstitute the cleavage and polyadenylation reactions. These studies showed that in the presence of CFI, CFII provided cleavage activity and PFI directed polyadenylation of pre-mRNA [82, 85]. However, advances in protein purification technology enabled the single-step purification of a stable complex of CFII and PFI, suggesting the individual complexes interact during pre-mRNA processing steps. This large combinatorial protein complex responsible for both cleavage and polyadenylation was termed cleavage and polyadenylation factor (CPF) [86].

The CPF subunits derived from CFII and PFI are considered as the "core" of CPF. After tagging and purifying Ref2 (a non-essential transcription factor) from yeast extract, an even larger "holo" CPF complex was identified. This complex contained all core CPF components, with an additional 6-subunit protein complex, APT (Associated with Pta1). Tagging of different APT subunits indicated holo-CPF was a stable complex, although in some purifications APT subunits were found in substoichiometric amounts, which could signify formation of holo-CPF is only required for transcription and transcript processing of certain genes [102].

All holo-CPF subunits are essential with the exception of Ref2 and Syc1 from APT. There have been many reported cases where the mutation or removal of a single holo-CPF subunit causes abnormal pre-mRNA cleavage and polyadenylation. However, not all holo-CPF subunits have a direct role in pre-mRNA 3'-end processing. Moreover, some are important in assembly of the complex and disrupting them leads to downstream effects on processing.

The subunit composition of holo-CPF is outlined in Table 1.1, with details of sub-complexes and the known functions of each protein. A selection of holo-CPF subunits are described in the context of their direct functions in pre-mRNA cleavage and polyadenylation in the next two sections.

CPF subunit	Mammalian homologue	Size (kDa)	Structural Features	Function	Refs
Cft1/Yhh1 (CFII)	CPSF-160	150	Predicted 3 β-propeller in CD	Binds pre-mRNA cleavage site, binds PolII CTD	[57, 130]
Cft2/Ydh1 (CFII)	CPSF-100	105	β-CASP domain, β-lactamase domain (non metal-binding)	Binds pre-mRNA EE, scaffold for interactions with CPF, PolII CTD and CFIA	[57, 103, 131, 132]
Ysh1 (CFII)	CPSF-73	100	β-CASP domain, β-lactamase domain (metal-binding)	Endonuclease for pre-mRNA 3'-end cleavage	[131, 133]
Yth1 (PFI)	CPSF-30	26	5 zinc fingers	Interacts with U-rich sequences on pre-mRNA, recruits Fip1 and PAP	[134-136]
Fip1 (PFI)	hFip1	55	Central flexible linker	Recruits PAP to CPF	[104, 137, 138]
Pta1 (CFII)	symplekin	88	Unknown	Scaffold protein, binds PolII CTD	[139, 140]
Pfs2 (PFI)	-	53	WD-40 repeats (x7)	Bridges CPF to CFIA	[86]
Mpe1 (PFI)	-	58	Zinc-knuckle motif	Unknown	[102, 141]
Pap1	PAP	63	β-polymerase fold, missing thumb domain	Generates polyadenylate tail	[63, 64, 142-144]
Pti1 (APT)	-	47	N-terminal RRM	CPF processing of pre-mRNA and snoRNAs	[102, 105, 145]
Ref2 (APT)	-	48	Unknown	Binds pre-mRNA with C/U preference. Enhances processing of weak poly(A) sites	[146]
Swd2 (APT)	-	37	WD-40 repeats (x7)	CPF assembly	[147]
Syc1 (APT)	-	21	P-loop and Leucine zipper, both homologous to CtD of Ysh1	Non-essential protein. Negative regulator of cleavage	[148]
Ssu72 (APT)	-	23	CX_5RS motif (tyrosine phosphatases) with aspartate loop	Protein phosphatase, dephosphorylates PolII CTD Ser5	[149-151]
Glc7 (APT)	-	36	Type 1 serine/threonine phosphatase	Protein phosphatase, dephosphorylates Pta1	[152]

Table 1.1. Summary of CPF subunit functions. In the CPF subunit column, details of subcomplex membership are provided in brackets underneath the subunit names.

1.4.1 The role of CPF in cleavage

It is clear that CFIA is essential for cleavage/poly(A) site selection through the Rna15-RNA interaction, and that Hrp1 may be required for site selection on certain gene transcripts. CPF also binds to pre-mRNA after it has been recruited to the correct site by CFIA/B. There are many cross-complex interactions between CFIA and CPF. Those that have been physically characterised are summarised in an interaction map found in Figure 1.7. The association of CPF with pre-mRNA is presumably fundamental in bringing its associated endonuclease activity into close proximity with the cleavage/poly(A) site.

CPF binds pre-mRNA via multiple subunits, including the two largest subunits, Cft1 and Cft2. The 150 kDa Cft1 protein has been shown to bind pre-mRNA in the vicinity of the cleavage/poly(A) site. Analysis of the amino acid sequence of Cft1 revealed the central domain contains predicted β-propeller-forming repeats, which mediate the Cft1-RNA interaction [130]. Experiments in which GST-tagged recombinant Cft1 was incubated with pre-mRNA substrates demonstrated Cft1 bound to sequences containing the cleavage/poly(A) site, although no direct sequence specificity could be inferred. [130]. Yeast extracts from different *yhh1* (Cft1) temperature sensitive mutants were unable to cleave and polyadenylate *GAL7* and *CYC1* pre-mRNAs *in vitro*. Additionally, Northern blotting experiments showed that cleavage/poly(A) site selection on *ACT1* transcripts had shifted to weaker distal sites, that had previously been characterised by examining *ACT1* transcripts from *rna15-2* and *rna14-3* temperature sensitive yeast strains [130, 153]. The next largest subunit of CPF, Cft2, also interacts with pre-mRNA. An *in vitro* experiment showed Cft2 purified from yeast extract could be UV-crosslinked specifically to the $(UA)_3$ EE sequence of *GAL7* pre-mRNA. This interaction was not observed on pre-cleaved RNAs, meaning there is also a requirement for sequences downstream of the cleavage/poly(A) site [132]. Extracts derived from *ydh1* (Cft2) temperature sensitive mutant yeast strains could not cleave or polyadenylate *CYC1* pre-mRNA in an *in vitro* assay. As with the *yhh1* (Cft1) temperature sensitive mutants, *ACT1* mRNA from *ydh1* (Cft2) temperature sensitive mutant yeast strains displayed a shift in poly(A) site usage towards distal sites [103]. In light of this, it is evident that both Cft1 and Cft2 are important in cleavage/poly(A) site selection.

With CPF accurately positioned through interaction with CFIA and the binding of its own subunits to pre-mRNA, the transcript is cleaved at a dinucleotide (usually CA)

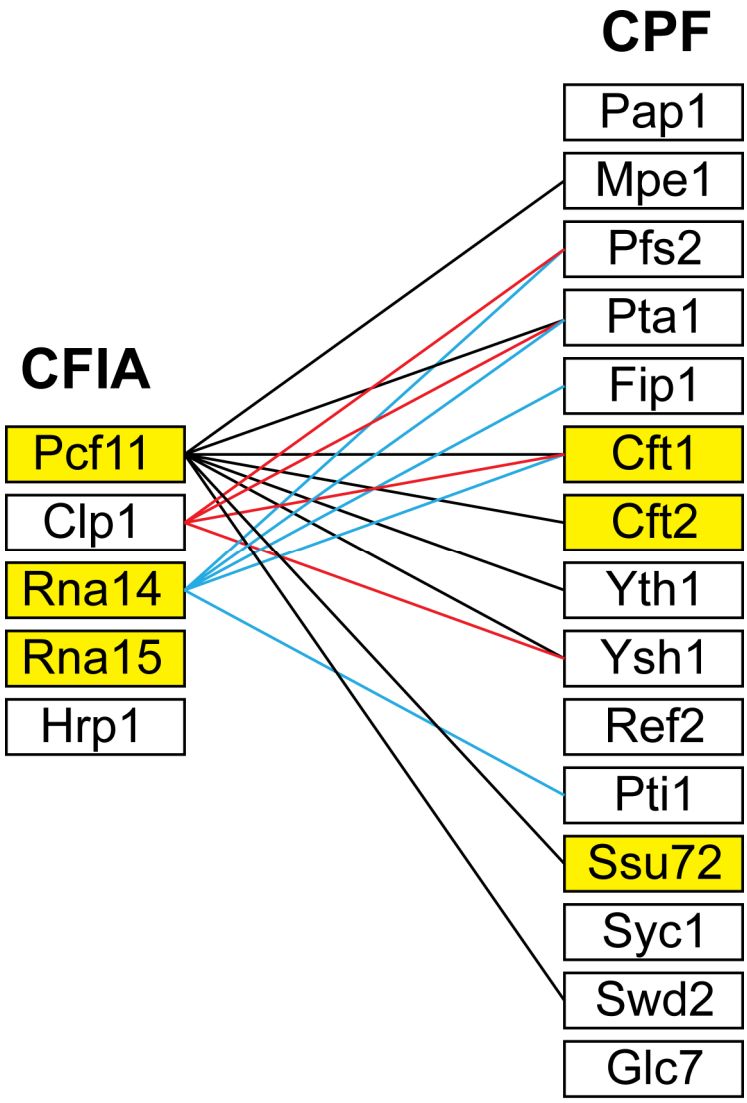


Figure 1.7. CFIA-CPF interaction map. These interactions have been demonstrated by genetic, immunoprecipitation or *in vitro* pulldown experiments. Hits generated by large scale TAP-tagging and mass spectrometry experiments are not shown. To improve the visibility of the map, interactions made by Pcf11 are shown by black lines, Clp1 by red lines and Rna14 by light blue lines. Proteins that bind to PolII CTD are highlighted in yellow.

within the cleavage/poly(A) site. The 100 kDa Ysh1 subunit of CPF is the putative endonuclease that physically cleaves pre-mRNA. Although there is no direct evidence in yeast that Ysh1 effects transcript cleavage, the mammalian homologue of Ysh1, CPSF-73, has been more extensively characterised as the 3'-end endonuclease. After the initial cloning and identification of CPSF-73 as an important pre-mRNA 3'-end processing factor, a sequence analysis revealed bovine CPSF-73 shared 53% sequence identity with the first 500 amino acids of Ysh1 from *S. cerevisiae*. In the same studies it was shown that whole-cell extracts derived from the *brr5-1* yeast strain (a Ysh1 temperature sensitive allele), or extracts in which Ysh1 was immunodepleted, were unable to cleave or polyadenylate *CUP1* and *CYC1* pre-mRNAs respectively [154, 155]. These findings signified the importance of Ysh1 in both steps of 3'-end formation. The crystal structure of CPSF-73 has been solved revealing the protein is a member of the metallo-β-lactamase superfamily of zinc-dependent hydrolases [131]. It contains a metallo-β-lactamase domain with two zinc ions in the active site, and a β-CASP domain. The five signature motifs that metallo-β-lactamase domains usually carry are conserved in CPSF-73, and are ligands to the two active site zinc ions present in the structure. The features and arrangement of the active site insinuated that CPSF-73 was a catalytically competent nuclease. This observation was corroborated in a biochemical experiment where recombinant CPSF-73 alone could cleave a radiolabelled RNA substrate [131]. The fact that Ysh1 also contains predicted metallo-β-lactamase and β-CASP domains combined with the conservation of active site residues with CPSF-73 suggests Ysh1 can also provide nuclease activity to the 3'-end processing machinery. Indeed, mutation of the Ysh1 active site residues was lethal to *S. cerevisiae,* meaning these residues most probably fulfil an active and essential role in 3'-end processing in yeast [133]. Site-specific labelling of a pre-mRNA substrate in conjunction with UV-crosslinking showed CPSF-73 binds directly to the mammalian poly(A) signal [133]. In contrast, no Ysh1-RNA binding activity has yet been detected in yeast, although given that other CPF subunits with which Ysh1 interacts bind pre-mRNA in the vicinity of the cleavage/poly(A) site, it is feasible Ysh1 is also localised in this position.

1.4.2. The role of CPF in polyadenylation

After cleavage, a homo-polymeric tail of 60-80 adenines is added to the newly generated 3'-end of pre-mRNA transcripts [63, 64]. The enzymatic activity required to achieve polyadenylation is confined to a single protein, Pap1. There is some debate as

to whether Pap1 is a member of CPF. Initial fractionation of yeast extracts responsible for cleavage and polyadenylation isolated Pap1 within its own fraction [82]. However, more advanced purification techniques used in later studies found Pap1 to be associated in stoichiometric quantities with the other posited core-CPF factors [102]. For the sake of clarity, Pap1 will be considered as a core-CPF subunit in addition to the CFII and PFI sub-complexes (see Table 1.1) in future discussion.

The enzymatic properties of Pap1 were first characterised after it was purified to homogeneity from yeast extract. It was discovered that *in vitro*, Pap1 could generate large polyadenylate tails on different RNA substrates in the absence of other protein factors. This lead to the presumption that specificity and regulation of tail-length must be conferred through Pap1 interaction with other proteins [84]. Subsequent cloning and sequence analysis of the *PAP1* gene revealed it shares 47% sequence identity with the first 400 amino acids of mammalian Pap1, and biochemical characterisation of the recombinant protein confirmed its enzymatic activity [83]. There are various crystal structures of Pap1 complexed with different ligands now available. They have provided key insights into the mechanism of polyadenylation in *S. cerevisiae.* The first structure revealed a three-domain protein with a large, central cleft [144]. The NTD comprised the classic polymerase "palm", in which the active site is located. Below the palm the CD forms the "fingers", which are shown interacting with the phosphates of a 3'-dATP molecule (a structural analogue of ATP). This nucleotide is equivalent to the predicted position of an incoming nucleotide ready to be incorporated into the pre-mRNA chain. The "thumb" domain present in other polymerases is completely absent whilst the CTD makes up the wall of the cleft opposite the palm in a distal position relative to the active site [144]. Subsequent kinetic analyses revealed Pap1 harbours an induced-fit mechanism that dictates nucleotide specificity [143]. More recently, a ternary structure has been solved of a stalled Pap1 in complex with a five-residue adenylate ribo-oligonucleotide [142]. This structure adopts a "closed" conformation compared to the original structure due to folding of the catalytic NTD permitting contacts with residues on the CTD. The structure reveals that Pap1 makes contact with only the last three residues of the adenylate RNA-oligo within a tunnel created by the closed conformation, with the -1 AMP located close to MgATP in the active site in the correct position for initiating the adenylation reaction. Whilst Pap1 can polyadenylate freshly-cleaved RNA *in vitro* with no apparent specificity for residues at the 3'-end of the transcript, this crystal structure suggests poly(A) RNA and MgATP are both required to

induce a conformational change that shifts Pap1 into the closed complex required for processive polyadenylation [63, 64, 142].

As discussed above, *in vitro* Pap1 does not require protein partners to polyadenylate an RNA substrate. Therefore *in vitro*, polyadenylation is totally unregulated, with large poly(A) tails generated without proper selection of RNA substrate [83, 84]. It has since been shown that the CPF subunit Fip1 displays a degree of control over Pap1 polyadenylation. Interaction of Fip1 and Pap1 was first demonstrated both *in vitro* and *in vivo* using biochemical and genetic approaches. This interaction was found to be crucial for polyadenylation, as yeast extracts from *fip1* mutant strains were able to cleave but not polyadenylate *CYC1* pre-mRNA [104]. Fip1 regulates the processivity of Pap1 *in vitro* by binding Pap1 CTD at a location overlapping a Pap1-RNA binding site. By occluding RNA-binding, Fip1 can drastically limit Pap1 polyadenylation [156]. Biochemical and biophysical analyses of recombinant Fip1 showed that although the protein was soluble, it was nearly completely disordered in the absence of binding partners [157]. In light of these findings, it was not surprising that a flexible linker was found in the centre of Fip1, between the Pap1 and Yth1 interaction domains. Manipulating the length of the linker was found to decrease polyadenylation *in vitro* and *in vivo*, signifying that as well as Fip1 recruiting Pap1 to pre-mRNAs, the linker may also be important in allowing Pap1 to reach the cleaved 3'-end of pre-mRNAs [137].

It has been suggested that another CPF subunit, Yth1, may fulfil an important role in facilitating polyadenylation. Yth1 is a 30kDa protein with a conserved central section comprised of five zinc fingers, and interacts with pre-mRNA around the poly(A) site of *CYC1* pre-mRNA [134, 135]. Functional dissection of Yth1 revealed zinc finger 4 (ZF4) was necessary for the binding of both pre-mRNA and Fip1. A competition assay demonstrated that the efficiency of UV-crosslinking Yth1 to *GAL1* pre-mRNA was decreased dramatically in the presence of increasing concentrations of Fip1, indicating that the binding of pre-mRNAs and Fip1 to Yth1 is mutually exclusive. Considering this, it may be that Yth1 acts as a switch between cleavage and polyadenylation: Yth1 is stably associated with pre-mRNAs prior to cleavage and after cleavage it is released and free to interact with Fip1, which recruits Pap1 to initiate polyadenylation [136].

1.5. A model for pre-mRNA 3'-end processing in *S. cerevisiae*

Following the above detailed introduction to the pre-mRNA sequence elements and protein factors involved in 3'-end processing, it is apt to bring this information together into a simplified, general model, presented in Figure 1.8 and described in the following text. In this model, the CFIA complex first facilitates cleavage/poly(A) site selection through the specific interactions of Rna15 with GU-rich sequences and Hrp1 with the PE. CFIA then recruits CPF via multiple interactions between subunits of these complexes. CPF is stabilised on the transcript by the binding of Cft1 and Yth1 to sequences surrounding the cleavage/poly(A) site, and Cft2 association with the EE. Ysh1 binds directly to the $Py(A)_N$ cleavage/poly(A) site and employs its endonuclease activity to cleave the pre-mRNA at the 3' of an adenosine. A subsequent rearrangement of CPF allows Yth1 to break free from the pre-mRNA and interact with Fip1. The flexible linker of Fip1 positions Pap1 at the newly generated 3'hydroxyl of the cleaved pre-mRNA. Poly(A) polymerase recognises the cleaved 3'-end and begins to polyadenylate. The first few adenylate residues added to the pre-mRNA are bound in the tunnel leading to the Pap1 active site. Together with MgATP in the Pap1 active site, this drives Pap1 into the closed conformation that is required for processive polyadenylation.

This model does not account for how poly(A) tail-length is regulated in *S. cerevisiae.* In the past, it was thought that poly(A) tail-length was controlled by the deposition of the major cytoplasmic poly(A)-binding protein Pab1 on the growing polyadenylate tail, followed by trimming of the tail by a 3'->5' exonuclease, Pab1-dependent poly(A) nuclease (PAN) [90, 158, 159]. It has since emerged that a previously identified mRNA export factor, Nab2, may in fact exert more direct control over poly(A) tail-length. Depletion of Nab2 *in vivo* causes the accumulation of hyper-adenylated mRNAs in the nucleus, and an *in vitro* experiment demonstrated that the addition of recombinant Nab2 to yeast extract stimulates polyadenylation of *CYC1* pre-mRNA, but limits poly(A) tail-length to 60-80 nt [160]. In line with these findings, recombinant Nab2 mutant proteins displayed a decreased affinity for poly(A) RNA-binding *in vitro*, and the corresponding *nab2* mutant yeast strains showed poor poly(A) tail-length control *in vivo* [161]. Crucially, association of Nab2 with the poly(A) tail was also shown to prevent the cleavage and polyadenylation machinery still associated at with the transcript from cleaving the poly(A) tail [162]. From a logical point of view, although Pab1 can shuttle

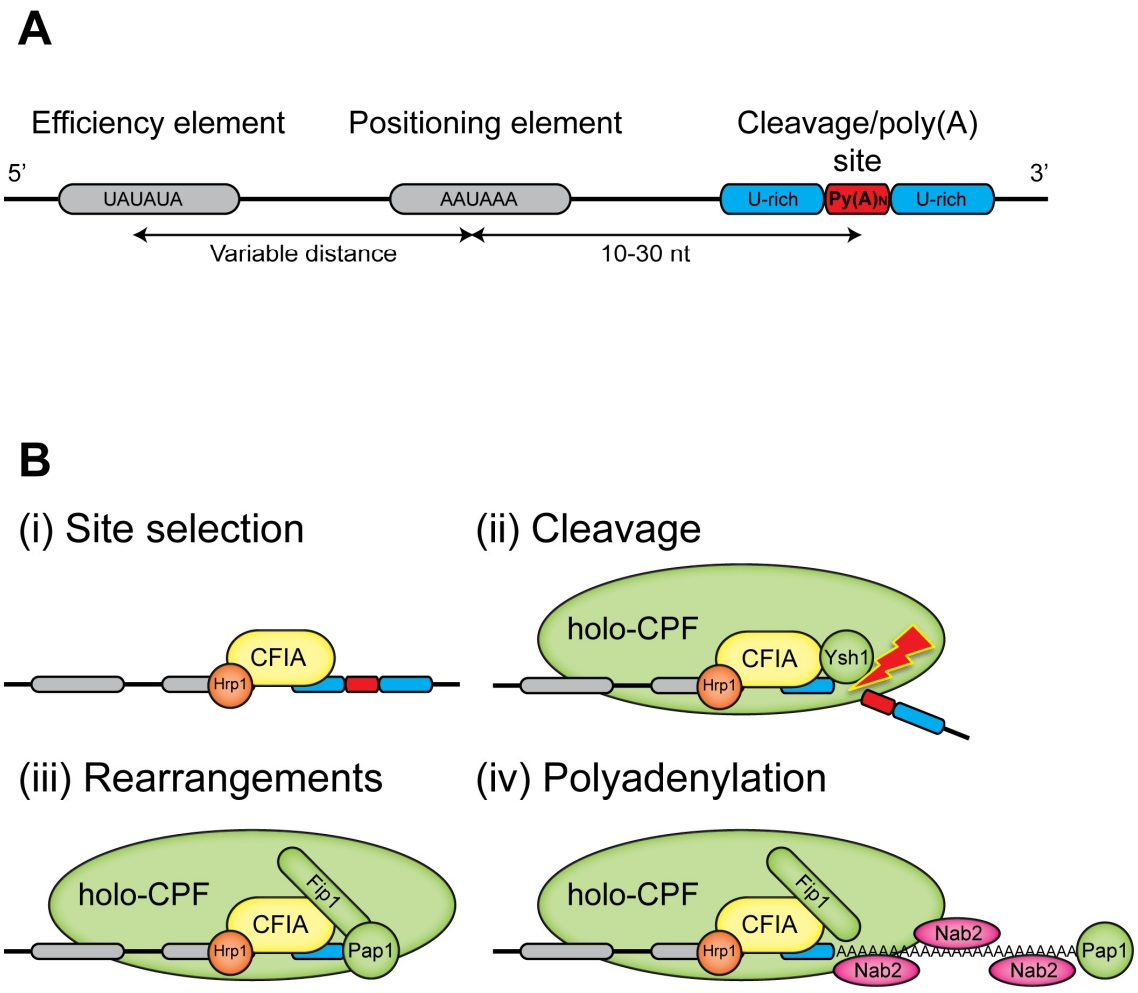


Figure 1.8. Model for pre-mRNA 3'-end processing. (**A**) pre-mRNA sequence elements with details of optimal consensus sequences. (**B**) CFIA and Hrp1 control site selection (i) allowing holo-CPF recruitment and pre-mRNA cleavage by the Ysh1 subunit (ii). Subsequent rearrangement of CPF allows Pap1 to engage the cleaved 3'-end of the pre-mRNA (iii) and initiate polyadenylation, with Nab2 regulating poly(A) tail-length (iv). The red lightning bolt signifies the cleavage reaction.

between the nucleus and cytoplasm [163], the fact that it is predominantly located in the cytoplasm [164] deems it unlikely one of its direct functions is poly(A) control. Research into the role of Nab2 and polyadenylation presents a strong case for Nab2 being the key regulator of poly(A) tail-length, and as a result this has been built into the model portrayed in Figure 1.8 accordingly. However, the nature of the distribution of Nab2 along poly(A) RNA remains enigmatic.

The above model is simplistic in that it does not consider the timing of pre-mRNA 3'-end processing, or the crosstalk between different processing events. The integration of other pre-mRNA processing events and transcription with 3'-end formation will be reviewed in sections 1.6. and 1.7. respectively, with a focus on findings associated with CFIA factors.

1.6. Links between 3'-end processing and other processing events

There are a variety of protein factors that associate with poly(A) tails and the 3'-end processing machinery that play important roles in other processing events. These may serve to couple distinct processing steps temporally, or may be required to determine the fate of mRNAs after nuclear export. Some of the most well documented links between pre-mRNA 3'-end processing and other steps of mature mRNA formation are presented in this section.

1.6.1. 3'-end processing and 5'-end capping

The association of Pab1 with the poly(A) tails of pre-mRNA has been shown to stimulate translation initiation [165]. Following on from this finding, the purification of mRNA from yeast extracts through ^{7m}GDP chromatography (affinity for the 5' cap of pre-mRNA) revealed that Pab1 was found associated with the eIF-4F cap-binding complex [3]. To elucidate the finer molecular details, a mutational analysis showed that the eIF-4G subunit of eIF-4F mediated the Pab1-eIF-4F interaction. Although the interaction was not required for poly(A)-dependent translation in an *in vitro* system, it was shown to synergistically stimulate translation [166]. Combining these facts with other studies in the translation initiation field, a model was proposed in which the poly(A) tail bound to Pab1 stimulates translation by interacting with eIF-4G and enhancing the recruitment of the 40S ribosomal subunit [167]. This model hypothesises

that the Pab1-eIF-4G interaction circularises the mRNA (forming a mRNP). Supporting this hypothesis, atomic force microscopy clearly showed that the combination of eIF-4E, eIF-4G (together eIF-4F complex) and Pab1 with capped and polyadenylated mRNA facilitated circularisation of the transcript into an mRNP [56]. In summary, communication between proteins associated at the 5' cap and 3' poly(A) tail of mRNAs promotes efficient translation.

1.6.2. 3'-end processing and splicing

There is some evidence from experiments with mammalian cells that implies splicing factors can have a stimulatory effect on pre-mRNA 3'-end formation (reviewed in [48]). However, a parallel has not been drawn in *S. cerevisiae*. This could be due to the comparative lack of introns in the *S. cerevisiae* genome.

1.6.3. 3'end processing and export

As discussed in section 1.5, the association of the Nab2 protein with poly(A) RNA controls poly(A) tail-length in *S. cerevisiae.* Recent studies have implied that Nab2 has a second function in mRNA export. Purification of Nab2 by TAP uncovered it is associated with the key export factors Yra1, Mex67 and the THO-TREX complex. Furthermore, Nab2 was shown to be commonly associated with mRNA throughout the *S. cerevisiae* transcriptome [47]. The interaction of Nab2, Yra1 and Mex67 was confirmed *in vitro*, and shown to be necessary to guide mRNA to the NPC, where exchange of mRNA into the cytoplasm is mediated by Nab2 interaction with the NPC subunit Mlp1 [32, 46].

Intriguingly, the Swd2 subunit of CPF has an indirect effect on mRNA export. Immunoprecipitation of HA-tagged Swd2 unearthed the export factors Npl3, Nab2 and Mex67 as binding partners. Binding experiments showed that both formation of the APT sub-complex of CPF, as well as Swd2 association with export factors, was dependent on the ubiquitylation of Swd2 at key residues. Correctly assembled CPF is therefore likely able to recruit components of the mRNA export machinery [168].

Particularly relevant to the work presented in this thesis is the observation that the export factor Yra1 interacts directly with the CFIA complex through the Pcf11 subunit. Yra1 is a member of the RNA and export factor (REF) family of RNA-binding proteins whose role in mRNA export is briefly outlined in section 1.1.4. The Yra1-Pcf11

interaction was discovered in one seminal study using both *in vitro* and *in vivo* approaches [169]. A chromatin immunoprecipitation (ChIP) analysis of Yra1 occupancy on the *PMA1* and *ACT1* genes demonstrated that Yra1 is recruited progressively throughout transcription, and is enriched at the 3'-end of genes. Further ChIP experiments in this study revealed Yra1 occupancy on the *TEF1* gene was almost eliminated in the *pcf11-2* and *pcf11-9* temperature sensitive yeast strains, pointing to a possible Yra1-Pcf11 interaction. The interaction was confirmed when it was shown that HA-tagged Pcf11 could coIP Yra1. Subsequent pull-down assays mapped the Yra1 binding site on Pcf11 to the zinc-finger/Clp1-ID domain. The authors postulated that the early recruitment of export factors to the transcription complex via Pcf11-Yra1 might serve as a quality control mechanism, as only a fully-assembled transcription elongation complex containing CFIA would be competent to recruit Yra1 for its role in subsequent mRNA export [169]. Recently, research into the interplay between Yra1 and Pcf11 has been extended, illuminating a more direct role for Yra1 in pre-mRNA 3'-end processing [170]. Because the Clp1 and Yra1 binding sites on Pcf11 were thought to overlap, an *in vitro* binding experiment was employed to show if Pcf11 bound each protein in a mutually exclusive manner. The experiment clearly showed that recombinant Clp1 and Yra1 compete for the binding of Pcf11. This disrupted CFIA activity, and abolished cleavage and polyadenylation of *GAL7-1* pre-mRNA. The idea that Clp1 may not be incorporated into CFIA at the same time as Pcf11 was novel, and was tested by looking at the ChIP profiles of Pcf11 and Clp1 across a large set of different genes. The occupancy of Clp1 was shown to be distinct from that of Pcf11, with a marked bias towards the poly(A) region of genes compared to the relatively evenly-distributed Pcf11, strengthening the notion that CFIA assembly may not be straightforward. Compellingly, ChIP profiles of Yra1-depleted transcripts revealed early recruitment of Clp1 along genes, indicating the Pcf11-Yra1 interaction may preclude the Pcf11-Clp1 interaction during transcription elongation. Because this could have implications for 3'-end formation, poly(A) site choice across the yeast transcriptome was monitored after Yra1 depletion by RNA-seq (sequencing of total yeast mRNA). Strikingly, poly(A) site choice had been altered in nearly one third of all transcripts compared to the wildtype *S. cerevisiae* transcriptome [170]. It is therefore clear that as well as being an export factor, Yra1 is also a 3'-end processing factor with the ability to assert a certain degree of control over cleavage/poly(A) site selection. However, the mechanism for Yra1-mediated poly(A) site selection remains obscure.

1.7. Links between 3'-end processing and transcription

Because pre-mRNA processing events are targeted to distinct locations on gene transcripts, it is easy to presume they occur successively after transcription is complete. However, in the last two decades it has become apparent that processing events are not only linked with each other, but they are also intimately coupled to transcription. In other words, the pre-mRNA is secured with a 5'cap, and introns spliced from the coding sequence whilst PolII is still transcribing the rest of the gene. This is enabled by the loading of capping and splicing machineries onto PolII at the promoters of genes, and throughout transcription elongation (reviewed in [48, 52, 53, 171]).

It is obvious that maturation of pre-mRNA 3'-ends must happen slightly later than capping and splicing, as the signal sequences for cleavage/poly(A) must first be transcribed. However, it is now known that elements of the 3'-end processing machinery are also recruited at an early stage, and that this is crucial for gene expression. Not only does correct 3'-end processing require PolII, the polymerase also requires interaction with the 3'-end processing machinery for transcription termination. In short, the formation of a mature 3'-end drives the dissociation of PolII from the 3'-UTRs of genes.

PolII CTD is the communication centre of the polymerase, acting as a landing pad for the pre-mRNA processing machineries, including 3'-end processing factors [49]. In *S. cerevisiae*, CTD contains 26 heptad repeats of the sequence YSPTSPS, and is connected to the rest of PolII by a flexible linker [172]. Individual residues in the repeat sequences are phosphorylated in a dynamic manner. For example, the serine at position 5 of the repeat (Ser5) is phosphorylated at promoters but declines 5' to 3' along genes, whereas phosphorylation of the serine at position 2 of the repeat (Ser2) occurs only in coding regions [9]. Dynamic phosphorylation may be gene-specific, and act as a "CTD code" that specifies the recruitment of specific protein factors at precise intervals in the transcription cycle [173].

Components of CFIA and CPF 3'-end processing complexes have been found to interact with CTD (see Figure 1.7). Probing the importance of these interactions by mutagenesis has been key in demonstrating how transcription and 3'-end processing are entwined. Breaking the interactions can cause aberrant 3'-end processing, inefficient

transcription termination, lower levels of transcription and sometimes a combination of these effects. The next two sections will consider specific examples where 3'-end processing factors are linked to transcription termination and initiation respectively, with a focus on CFIA subunits.

1.7.1. 3'-end processing and transcription termination

Attempts to explain the mechanism behind the 3'-end processing and transcription termination connection have culminated in the proposal of two models: the "allosteric" (or "anti-terminator") and "torpedo" models of transcription termination (reviewed in [174]).

The allosteric model (Figure 1.9A) stresses the importance of the cleavage/poly(A) site and PolII CTD. Here, transcription of the cleavage/poly(A) site induces a conformational change, resulting in the depletion of anti-termination factors and recruitment of pro-termination factors to PolII CTD. The end result is the release of the PolII transcription elongation complex (TEC) from the gene [51, 174]. The problem with this model is that cleavage is not a prerequisite for termination, which contradicts the fact that 3'-end processing factors are essential for efficient transcription termination. The torpedo model (Figure 1.9B) is similar to Rho-dependent transcription termination in bacteria, and unlike the allosteric model, the cleavage reaction is fundamental. In the torpedo model, transcription of the cleavage/poly(A) site is followed by the cleavage reaction and release of the transcript. As the TEC continues to transcribe, the free 5'-end of the nascent RNA is critical in allowing entry of Rat1, a 5'->3' exonuclease. Rat1 degrades the nascent RNA, effectively chasing the TEC. When Rat1 catches the TEC, it somehow induces a conformational change that promotes dissociation of the TEC leading to transcription termination [174, 175]. However, a complex RNP-immunoprecipitation experiment showed that loss of transcription termination observed in the *rat1-1* temperature sensitive mutant did not correspond with the accumulation of nascent RNA, suggesting the exonuclease activity of Rat1 alone does not drive transcription termination [176].

In response to the shortcomings of both the allosteric and torpedo models, a third model has now been proposed (Fig 1.9C), in which elements of both previous models are unified [176]. This model was driven by a series of ChIP experiments that

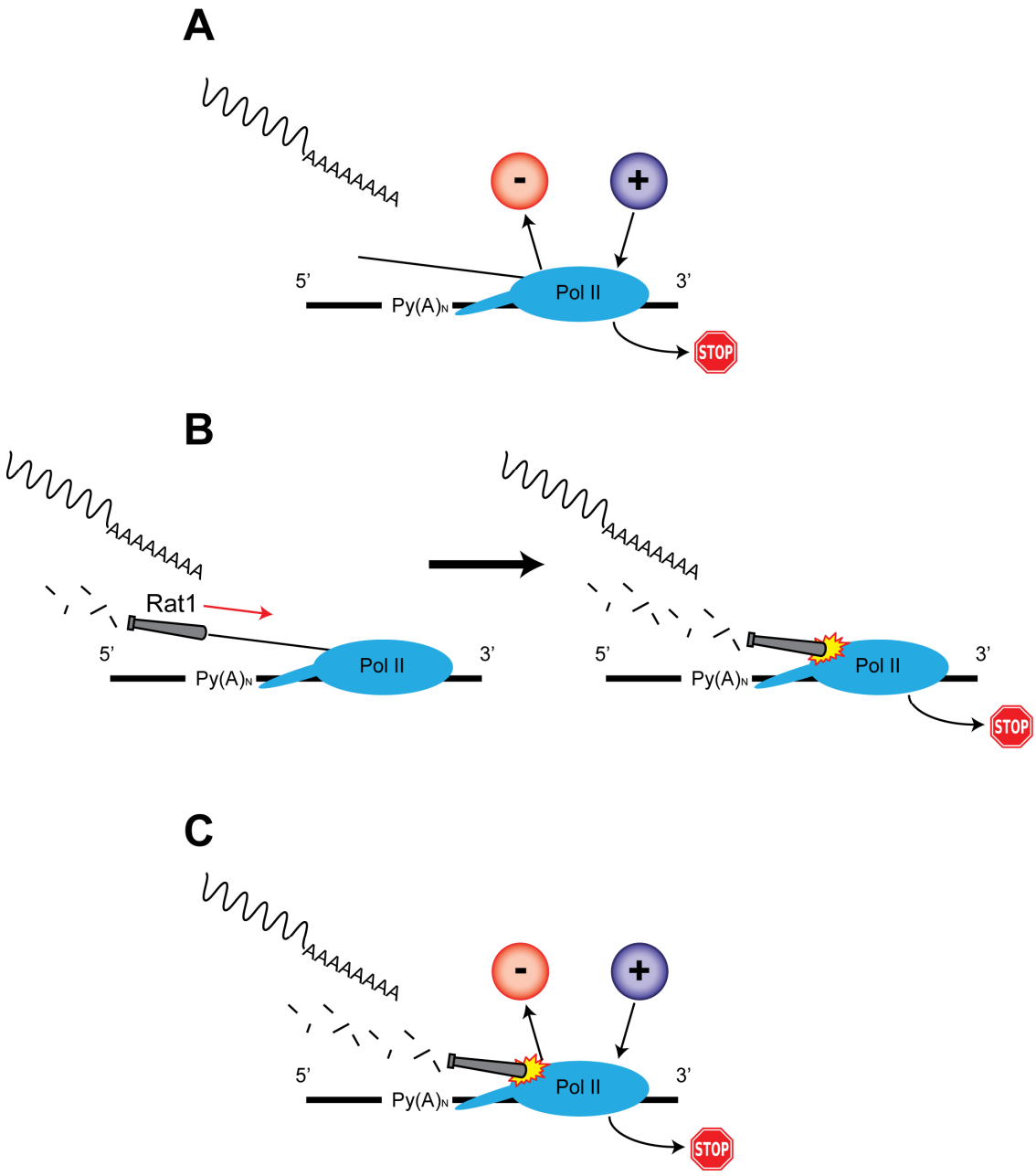


Figure 1.9. Models of transcription termination. (**A**) The "allosteric" model. After transcription of the cleavage/poly(A) site a rearrangement of PolII CTD-associated factors causes TEC release and transcription termination. Red - indicates the departure of anti-termination factors, Blue + indicates the recruitment of pro-termination factors. (**B**) The "torpedo" model. After cleavage, Rat1 degrades nascent RNA 5'->3' and displaces the TEC from the gene causing transcription termination. (**C**) A unified allosteric/torpedo model where both rearrangements of factors on PolII CTD and degradation of nascent RNA leads to TEC release and transcription termination.

demonstrated the colocalisation of Rat1 and the 3'-end processing factors Pcf11 and Rna15 at the 3'-end of the *ADH4* gene. These experiments also showed that *pcf11-2* and *pcf11-9* yeast strains displayed a reduction in Rat1 recruitment, and the *rat1-1* mutation led to a decrease in Pcf11 and Rna15 levels, pointing to a codependence of these factors in their recruitment to the 3'-ends of genes. The unified model proposes that a combination of nascent RNA degradation by Rat1 and conformational changes to the TEC after transcription of the poly(A) site is required for dissociation of the TEC and transcription termination [176].

In *S. cerevisiae*, crosstalk between 3'-end processing and transcription termination was first observed with the CFIA subunits Pcf11, Rna14 and Rna15. Mutated forms of these proteins had all previously been shown to cause aberrant cleavage and polyadenylation *in vitro* and *in vivo* (see section 1.3). Using yeast extracts derived from strains carrying the *rna14-1, rna15-1* and *pcf11-9* alleles, a transcription run-on analysis of the *CYC1* gene revealed transcription failed to terminate at the correct location with all three mutants [110]. Since these observations, a range of other 3'-end processing factors that also affect transcription termination have been identified, but only Cft1 and Ssu72 directly interact with PolII CTD [130, 177]. It is now clear there is a family of 3'-end processing factors that are important for transcription termination. Pcf11 is central to the work presented in this thesis, and as such it is important to consider the attributes that allow it to couple 3'-end processing and transcription.

As mentioned previously, PolII interacts with various RNA-processing factors through its CTD. Pcf11, along with several other factors, contains a conserved CTD-interacting domain (CID). Figure 1.3B illustrates the conservation of the CID amongst Pcf11 homologues. The crystal structure of the Pcf11 CID in complex with a phosphopeptide containing the PolII CTD repeat sequence has been solved [115], and is depicted in Figure 1.10. The CID is composed of eight α-helices in a right-handed superhelical arrangement (Figure 1.10A). The CTD phosphopeptide adopts a β-turn conformation around pSer2-Pro3-Thr4-Ser5 that binds to a conserved groove in CID (Figure 1.10B). Although phosphorylation of Ser2 is required for CID-binding, the phosphor group of the side-chain is actually oriented away from the CID and is not directly involved in the interaction. However, the phosphor group makes hydrogen bonds that stabilise the β-turn conformation and thus is intimately involved in CID recognition. The PolII CTD-

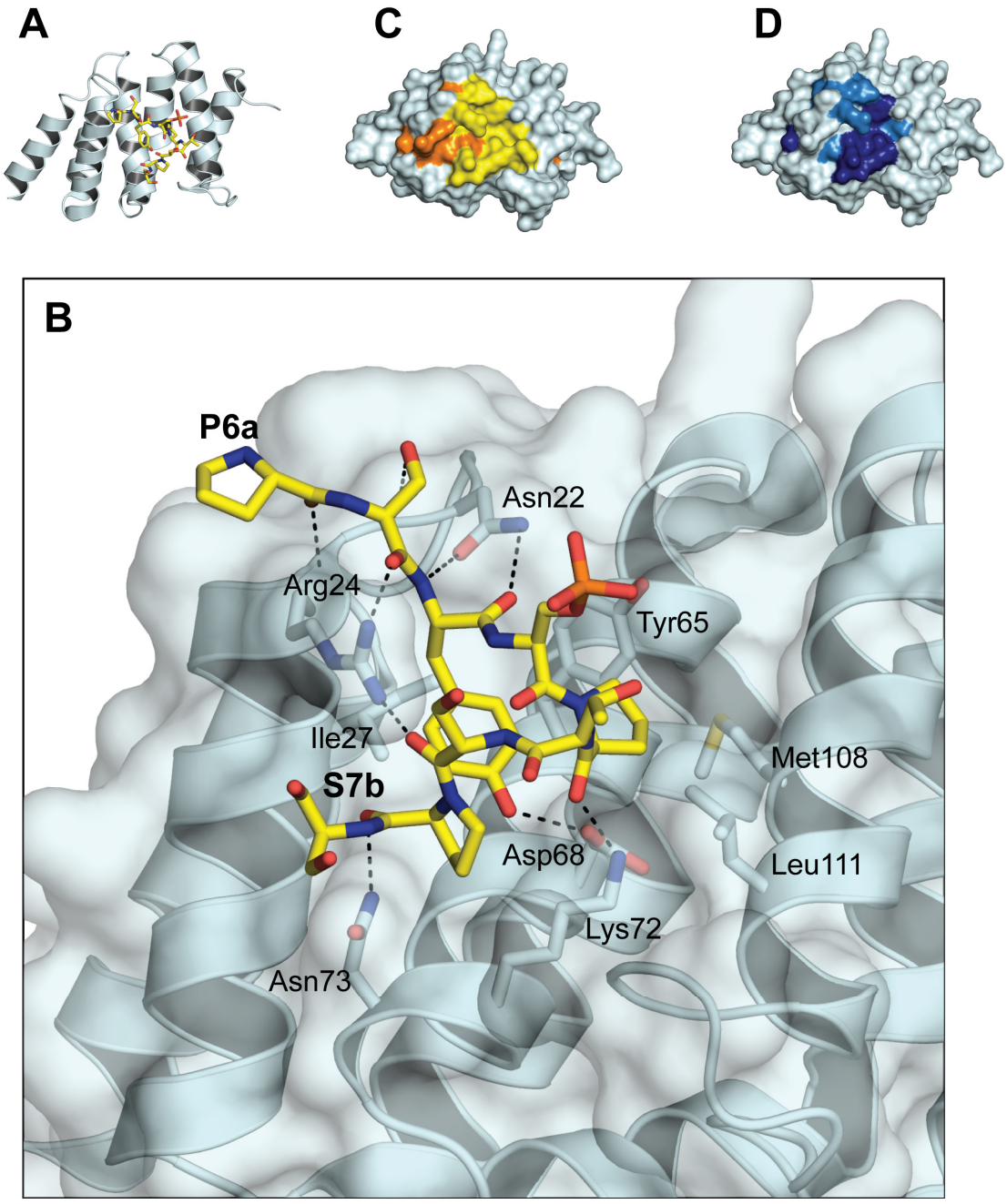


Figure 1.10. The interaction between Pcf11 CID and PolII CTD. (**A**) Crystal structure of the Pcf11 CID-PolII CTD complex. CID is shown as a pale cyan cartoon, CTD as yellow sticks. (**B**) Magnified view of the CID-CTD interface. The CTD phosphopeptide binds Pcf11 CID within a conserved groove. CID is shown as a pale cyan cartoon with a semi-transparent surface and side-chains involved in the interaction are illustrated in stick representation and are labelled. The CTD phosphopeptide is shown as yellow sticks (phosphate in orange). P6a denotes the proline in one copy of the repeat, S7b denotes the terminal serine of the next repeat. (**C**) CTD binding surface of CID. CID is shown in pale cyan surface representation. Residues involved in CTD-binding observed in the crystal are shown in yellow, additional interacting residues identified by NMR chemical shift assays are shown in orange. (**D**) The CTD binding surface of CID is conserved. CID is shown as before, but conserved residues are highlighted in shades of blue (dark blue most conserved), refer to the Pcf11 CID sequence alignment in Figure 1.3. *PDB codes: 1SZA, 1SZ9, 2BF0.*

Pcf11 CID crystal structure is consistent with other *in vitro* studies that have demonstrated that CID prefers Ser2-phosphorylated CTD. It has also been shown by ChIP that Pcf11 recruitment along genes correlates with CTD Ser2 phosphorylation [173, 178]. Comparison of the CID structure with the sequence alignment shows it is mainly completely or near-completely conserved residues that constitute the CID-CTD interaction surface (compare Figure 1.3 with Figure 1.10D). Another seminal study provided a similar structure with an extensive biophysical interrogation of the binding mechanism [179]. Chemical shift mapping by NMR revealed the interaction surface is probably slightly larger than seen in the crystal structure (Figure 1.10C). Although it had been previously suggested that CID-CTD recognition depended on a preformed CTD structure (a heptad repeat β-turn facilitated by Ser2 phosphorylation), a series of biophysical experiments determined that this was not the case. The far-UV circular dichroism (far-UV CD) and ^{31}P NMR spectra of CTD-derived peptides showed that they do not form stable secondary structure elements after phosphorylation of serine residues within the heptad repeat sequences. Furthermore, NMR experiments revealed that CID-CTD binding caused a chemical shift perturbation in the ^{31}P NMR spectrum of a Ser2-phosporylated CTD peptide and that residues within the peptide underwent a conformational transition. These findings pointed to an induced fit mechanism of CID-CTD binding, in which CID forces CTD into the β-turn conformation observed in the crystal structure. The binding affinity measured by isothermal titration calorimetry (ITC) determined that the CTD-CID interaction was weak, although increasing the copy number of the repeat sequence increased the apparent affinity. The authors postulated that the presence of many repeats in the full CTD would elevate the binding affinity by increasing the local concentration of binding sites for CID. Additionally, the same study identified that the CTD prolines exist in a mixed *cis-trans* population with a *trans* dominance, but only the *trans* form of CTD is selected for interaction with CID [179]. Considering these findings, it can be hypothesised that PolII CTD recognition by pre-mRNA processing factors is mediated by two types of site-specific modification (phosphorylation and proline *cis-trans* isomerisation) and by the local concentration of binding sites.

There are conflicting reports on the necessity of the CID-CTD interaction for both 3'-end processing and transcription. Experiments conducted by Sadowski *et al* [106] found mutations in Pcf11 that abrogated CTD binding (*pcf11-9, pcf11-13*) abolished

transcription termination on the *CYC1* gene *in vivo* but not cleavage and polyadenylation of *CYC1* precursor *in vitro*. The reverse was also true, as mutations in Pcf11 that caused defective cleavage and polyadenylation (*pcf11-2*) had no effect on transcription termination. On the other hand, a later study by Luo *et al* [176] showed the *pcf11-2* allele which displays cleavage and polyadenylation defects also does not allow transcription termination. These disparities have not been resolved, and as a result it remains unclear as to whether the 3'-end processing and transcription termination functions of Pcf11 can be separated.

Clp1 has also been recently implicated as a transcription termination factor. A new *pcf11* temperature sensitive allele (*pcf11* RW-A) that abrogates Pcf11-Clp1 binding at the restrictive temperature was identified by work described in this thesis and by others [122]. This work revealed that yeast extracts from *pcf11* RW-A were unable to support mature 3'-end formation in a coupled cleavage and polyadenylation assay *in vitro*. A further *in vitro* experiment demonstrated that this mutation also caused defective transcription termination on *ADH2* and *GAL7* genes [122]. Another study confirmed Clp1 is necessary for transcription termination after *in vivo* depletion of Clp1 reduced termination of transcription on snoRNA genes [121]. Recombinant CFIA containing Clp1 P-loop mutants which remove the Clp1-ATP interaction were found to complement the 3'-end processing and termination defects associated with Clp1 depletion in a coupled *in vitro* assay [180]. This suggests Clp1-ATP binding is not essential for its functions in 3'-end processing and transcription termination. The results section of this thesis presents contrasting data on the topic of Clp1-ATP binding and an analysis of this disparity is provided in the discussion.

1.7.2. 3'-end processing and transcription initiation

The classical view of the PolII transcription cycle is that PolII assembles at the promoter, transcribes a linear piece of DNA encoding a gene, terminates transcription and dissociates from the gene ready to be recruited back to the promoter for another round of transcription. However, in recent years evidence has been provided that suggests transcription may not be as simple as this, and that gene looping may provide a mechanism for transcription re-initiation. Gene looping is defined as the interaction between the promoter and terminator regions of a gene during transcription (reviewed in [181]).

The transcription factor TFIIB is crucial for gene looping, and is able to interact with both the promoter and sequences downstream of the cleavage/poly(A) signal [182]. It is now clear that the interaction of 3'-end processing factors with TFIIB is fundamental to gene looping. The CPF subunit Ssu72 is particularly important. The cross-linking of TFIIB to terminator regions of the *PMA1* and *BLM10* genes requires Ssu72 [182]. Furthermore, Ssu72 is a PolII CTD phosphatase, and its phosphatase activity has been shown to be essential for the looping of *BUD3* and *SEN1* genes during transcription [149, 151, 183]. It is therefore likely that Ssu72 is involved in erasing phosphorylation marks on PolII CTD after one round of transcription to enable the same PolII molecule to cross over to the adjacent promoter region and re-initiate transcription [183]. As well as Ssu72, three subunits of CFIA have been identified as critical for gene looping. A ChIP experiment showed functional Rna15 was obligatory for the recruitment of TFIIB to the terminator region of the *MET16* and *INO1* genes and it was found that mutation of the *rna15* gene corresponded with decreased levels of transcription. This was supported by the observation from chromosome conformation capture analysis [183] that gene looping had been abolished in an *rna15-1* mutant strain [184]. The same analysis has since been applied to Rna14 and Pcf11, as well as Pap1. These factors were also shown to physically interact with TFIIB, and are required for localisation of TFIIB to the terminator and for gene looping [185].

The requirement for CFIA in gene looping suggests it may have an essential function in transcription initiation. The gene looping evidence is perfectly complemented by previous observations that *pcf11* temperature sensitive alleles or depletion of Pcf11 *in vivo* results in a marked decrease in transcription levels [89, 176, 186]. The evidence points to further roles for CFIA and CPF in transcription initiation, in which components of these complexes physically link gene promoters and terminator regions to facilitate PolII re-initiation and subsequent rounds of transcription.

1.8. Mammalian homologues of 3'-end processing factors

Aberrant 3'-end processing in humans has been shown to cause a variety of disease states (reviewed in [187]). The continued use of *S. cerevisiae* as a model organism for 3'-end processing studies is justified by the degree of conservation of 3'-end processing protein factors between yeast and mammals.

Mammalian 3'-end processing is conducted in the same fashion as in *S. cerevisiae*: cleavage and polyadenylation at the 3'-end of gene transcripts. There are mammalian homologues of most yeast 3'-end processing factors, although the composition of the large multi-subunit protein complexes differs (reviewed in [57, 60]). Mammalian homologues of the yeast CPF subunits are listed in Table 1.1, and differences in the composition of 3'-end processing complexes can be viewed in Figure 1.11. Briefly, the Cft1, Cft2, Ysh1 and Yth1 subunits of yeast CPF are found in the cleavage and polyadenylation specificity factor (CPSF) complex in mammals as the 160, 100, 73 and 30 kDa subunits respectively. CPSF also contains hFip1, the mammalian homologue of Fip1. The Pta1 protein of yeast CPF is separated in mammals and is referred to as symplekin. Likewise, the poly(A) polymerase (PAP) is thought of as a separate protein in mammals. The homologues of yeast CFIA factors are split into two different complexes in mammals: The Rna14 and Rna15 homologues are found in the cleavage stimulatory factor complex (CstF) as CstF-77 and CstF-64 subunits respectively. hClp1 and hPcf11 are the Clp1 and Pcf11 homologues, and comprise Cleavage Factor II (CFIIm). The mammalian poly(A) binding protein (PABP) is essential for 3'-end formation in mammalian cells as with yeast, and further additional factors found in yeast holo-CPF are either not present or not yet identified in mammals.

In *S. cerevisiae*, CFIA, CFIB and the CFII activity of CPF are necessary for *in vitro* cleavage, whilst there is an additional requirement for the whole CPF complex and Pab1 for *in vitro* polyadenylation. In contrast, mammalian 3'-end processing requires all factors bar PABP for *in vitro* cleavage, and *in vitro* polyadenylation requires only CPSF, PAP and PABP (see Figure 1.11) [57, 60].

This thesis focuses on the function of Clp1 in 3'-end processing in *S. cerevisiae*. However, important similarities and key differences between Clp1 and mammalian hClp1 have recently come to light, making it necessary to provide a more detailed background for hClp1. hClp1 was first isolated from HeLa cell nuclear extract by a variety of chromatographic techniques, and its identity was assigned after sequence alignments showed it shared 23% sequence identity with yeast Clp1 [117]. A series of *in vitro* experiments demonstrated hClp1 was necessary for cleavage, but not polyadenylation, whereas yeast Clp1 is required for both processes [117, 121]. By identifying cross-factor interactions using immunoprecipitation experiments, it has been suggested the role of hClp1 in mammalian 3'-end processing is to bridge CFIm, CFIIm

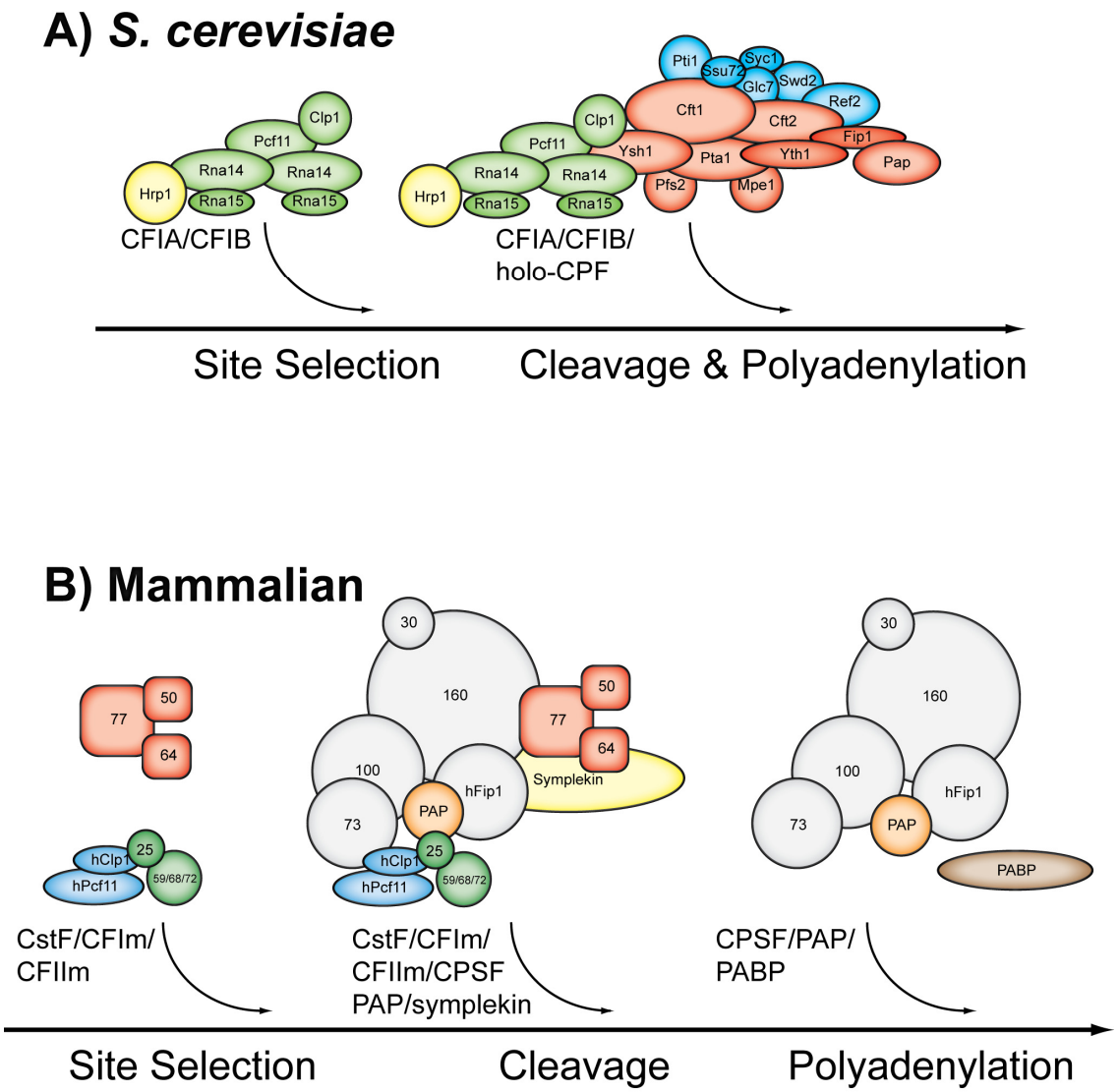


Figure 1.11. Comparison of *S. cerevisiae* and mammalian 3'-end processing protein complexes. (**A**) In *S. cerevisiae*, CFIA (green) and Hrp1 (yellow) are required for site selection. Cleavage and polyadenylation *in vitro* requires the addition of CPF. Core CPF subunits are shown in red, the additional CPF subunits without homologues in mammalian CPSF are shown in blue. (**B**) In mammals, site selection is mediated by CstF (red), CFIm (green) and CFIIm (blue). Cleavage *in vitro* requires the presence of CPSF (grey), PAP (orange) and symplekin (yellow), whereas *in vitro* polyadenylation requires only CPSF, PAP and PABP (brown).

(of which it is a subunit) and CPSF [117]. During purification, hClp1 copurified with components of the splicing and transcription machineries, implying that the role of yeast Clp1 in transcription termination may also be conserved in mammals [117].

A sequence alignment (see Figure 1.5 and Appendix 8.1) reveals the P-loop, Switch domains and base-binding loop are all conserved between yeast Clp1 and hClp1. The key difference is that a catalytic function separate to its role in pre-mRNA processing has been assigned to hClp1, whereas no catalytic activity has been detected in yeast Clp1. hClp1 has been shown to provide the kinase activity required to phosphorylate the 5'-OH groups of synthetic short interfering RNAs (siRNAs) *in vivo*. In the same study, it was shown hClp1 copurified with components of the tRNA ligase machinery responsible for joining tRNA exons. Subsequent *in vitro* experiments revealed hClp1 provided the kinase activity required as a prerequisite for the ligation of tRNA exons [188]. The tRNA kinase activity of hClp1 has since been confirmed *in vivo* in a yeast system. A genetic approach demonstrated that the lethal phenotype caused by elimination of the *trl1* gene (yeast tRNA ligase/kinase) could be reversed by the introduction of hClp1. Furthermore, mutation of the P-loop and Switch I aspartate of hClp1 resulted in a reversion back to the lethal phenotype, indicating the kinase activity of hClp1 was essential in rescue of the yeast *trl1* mutant. In contrast, mutation of the same motifs in yeast *clp1* had no visible effects on yeast cell viability [189]. An *in vitro* kinase assay also showed that whilst hClp1 was able to efficiently phosphorylate the 5'-ends of RNA and DNA substrates, yeast Clp1 displayed no kinase activity [189]. It therefore seems that both yeast Clp1 and hClp1 share functions in 3'-end processing, although hClp1 is also a tRNA kinase which yeast Clp1 is clearly not. Interestingly, hClp1 was unable to complement a *clp1* null mutation in *S. cerevisiae* [189]. Taken together, these findings imply there is an element of functional divergence between yeast Clp1 and hClp1.

1.9. Research objectives

Clp1 and Pcf11 are two essential proteins that are required for efficient 3'-end processing of pre-mRNAs in *S. cerevisiae*. The previously determined crystal structure of a Clp1-Pcf11-ATP complex revealed that Clp1 binds ATP via a set of conserved structural motifs, and that the interaction of Pcf11 with Clp1 is facilitated by highly-

conserved Pcf11 residues. However, although the conserved nature of both of these interactions suggests each may have essential functions, they remain poorly characterised.

The principle objectives of this thesis are to determine the specific functions of Clp1-ATP binding and the Pcf11-Clp1 interaction in transcription, transcription termination and pre-mRNA 3'-end processing. This begins with the *in vitro* characterisation of recombinant Clp1-ATP binding and Pcf11-Clp1 binding mutant proteins using hydrodynamic, biophysical and biochemical techniques. Following this, the effects of mutations are probed using an *in vivo* analysis to determine whether these interactions are essential in *S. cerevisiae*. Finally, mRNA extracted from conditional *clp1* and *pcf11* mutant yeast strains is analysed using experiments designed to determine the precise functions of Clp1-ATP binding and the Pcf11-Clp1 interaction in transcription, transcription termination and pre-mRNA 3'-end processing.

2. Materials and Methods

2.1. Bioinformatics

2.1.1. DNA and protein information

Sequence information for *CLP1* and *PCF11* genes, their flanking regions and the proteins they encode was 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). Sequences were analysed using the SeqMan Pro (v8.0.2) software from the DNASTAR Lasergene 8 sequence analysis suite. Purification of a recombinant protein and subsequent determination of its concentration 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 Protean (v8.0.2) software from DNASTAR Lasergene 8.

2.1.2. Multiple sequence alignments

Protein sequences for multiple sequence alignments of Clp1 and Pcf11 were obtained from the UniProt (http://www.uniprot.org/) and NCBI (http://www.ncbi.nlm.nih.gov/protein/) databases. A full list of organism names for sequences used in alignments is presented in Appendix 8.1. Multiple sequence alignments were performed using MegAlign (v8.0.2) software from DNASTAR Lasergene 8, using the ClustalV method [190-192]. Multiple sequence alignments were annotated using GeneDoc (2.7.0.0) software (http://www.nrbsc.org/gfx/genedoc/).

2.2. Molecular biology techniques

2.2.1. Bacterial strains

E. coli NovaBlue "Singles" competent cells were used for cloning of protein expression constructs. Recombinant proteins were expressed in *E. coli* BL21(DE3) "Singles" competent cells. This strain contains a chromosomal copy of bacteriophage T7 RNA polymerase under control of the *lacUV5* promoter. Induction with isopropyl β-D-1-thiogalactopyranoside (IPTG) results in constitutive production of T7 polymerase. For expression vectors in which DNA encoding the target protein is under control of the T7 promoter, the mRNA is transcribed at high copy-number leading to overexpression of the target protein. The genotypes of strains used in this study can be found in Table 2.1.

E. coli strain name	Genotype	Function
NovaBlue	*endA1 hsdR17* (r_{K12}^{-} m_{K12}^{+}) *supE44 thi-1 recA1 gyrA96 relA1* *lac* F′[*proA*$^{+}$*B*$^{+}$ *lacI*^{q}*ZΔM15*::Tn*10*] (Tet^{R})	Cloning
BL21 (DE3)	*F*$^{-}$*ompT hsdS*$_B$(r_B^{-} m_B^{-}) *gal dcm* (DE3) pLysS (Cam^{R})	Protein expression

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

2.2.2. Yeast strains

The *clp1* and *pcf11* deletion yeast strains were obtained from the European Saccharomyces Cerevisiae Archive for Functional Analysis (EUROSCARF-http://web.uni-frankfurt.de/fb15/mikro/euroscarf/index.html). These are diploid strains in which one chromosomal copy of the *clp1* or *pcf11* gene has been deleted by insertion of a cassette encoding kanamycin resistance. The genotypes of diploid *clp1* and *pcf11* strains are presented in Table 2.2. Some experiments required the creation of haploid strains in which the single chromosomal copy of *clp1* or *pcf11* was deleted and the deleted gene complemented with a copy of *CLP1* or *PCF11* on a plasmid with a selectable marker. The creation of a haploid *pcf11*Δ strain is described in section 2.8.2. The haploid *clp1*Δ strain was supplied by Dr. M. Geymonat (formerly NIMR). The sexes of the haploid strains have not been determined.

S. cerevisiae strain (EUROSCARF accession number)	Genotype	Function
clp1Δ/CLP1 (Y22506)	BY4743; Mat a/α; *his3Δ*1/*his3Δ*1; *leu2Δ*0/*leu2Δ*0; *lys2Δ*0/*LYS2*; *MET15*/*met15Δ*0; *ura3Δ*0/*ura3Δ*0; *clp1*::kanMX4/*CLP1*	Creation of a haploid *clp1*Δ strain
pcf11Δ/PCF11 (Y23587)	BY4743; Mat a/α; *his3Δ*1/*his3Δ*1; *leu2Δ*0/*leu2Δ*0; *lys2Δ*0/*LYS2*; *MET15*/*met15Δ*0; *ura3Δ*0/*ura3Δ*0; *pcf11*::kanMX4/*PCF11*	Creation of a haploid *pcf11*Δ strain

Table 2.2. Genotypes of *S. cerevisiae* strains.

2.2.3. Plasmid construction

The details of wildtype plasmid constructs are presented in Table 2.3, mutant constructs are listed in Appendix 8.2. The Pcf11 456-626 EXP construct was used for expression of Pcf11 456-626 with a C-terminal hexahistidine sequence (His-tag) and was provided by Dr. I. Taylor (NIMR). Other Pcf11 CTD protein fragments were expressed from the pGEX-6P-1 construct (GE Healthcare, see Table 2.3). The Pcf11 pGEX-6P-1

constructs were created by cloning *Bam*HI/*Sal*I fragments containing DNA encoding the relevant residues of Pcf11 into pGEX-6P-1. The Clp1/Pcf11-EXP construct used for expression of the Clp1-Pcf11 (454-563) protein complex was provided by Dr. I. Taylor (NIMR). This construct is a pACYCDuet1 vector (Novagen) containing the full-length *CLP1* gene sequence in multiple cloning site 1 (MCS1), fused to a C-terminal His-tag. A fragment of *PCF11* encoding residues 454-563 is present in MCS2. The *URA3-PCF11* construct was created by cloning a *Bam*HI/*Sal*I fragment containing the *PCF11* gene and promoter (500 bp sequence upstream of *PCF11* gene) into pRS316 (provided by Dr. M. Geymonat, formerly NIMR). This construct was used to create a haploid *pcf11*Δ yeast strain. *LEU2-CLP1* and *LEU2-PCF11* constructs were generated from the pRS315 vector provided by Dr. M. Geymonat. The cloning method used is illustrated in Figure 2.1. These constructs were used to complement *clp1*Δ and *pcf11*Δ yeast strains. *LEU2-CLP1*(HA) and *LEU2-PCF11*(HA) contain in-frame fusions at the 3'-ends of their ORFS, with a 444 nt fragment encoding a 3X repeat sequence of a human influenza haemagglutinin DNA (HA-tag), for use in western blotting applications. *GAL1-CLP1*(HA) and *GAL1-PCF11*(HA) are identical to the aforementioned constructs, but their promoters were replaced by the *S. cerevisiae GAL1* promoter (see Figure 2.1) which is constitutively expressed during yeast growth in the presence of galactose. All constructs were generated using standard cloning techniques: polymerase chain reaction (see section 2.2.5), restriction digestion (see section 2.2.6) and ligation (see section 2.2.7) of DNA sequences. Clp1/Pcf11-EXP as well as *LEU2-CLP1*(HA), *LEU2-PCF11*(HA) (and their *GAL1* promoter derivatives) were all used as templates for site-directed mutagenesis (see section 2.2.9).

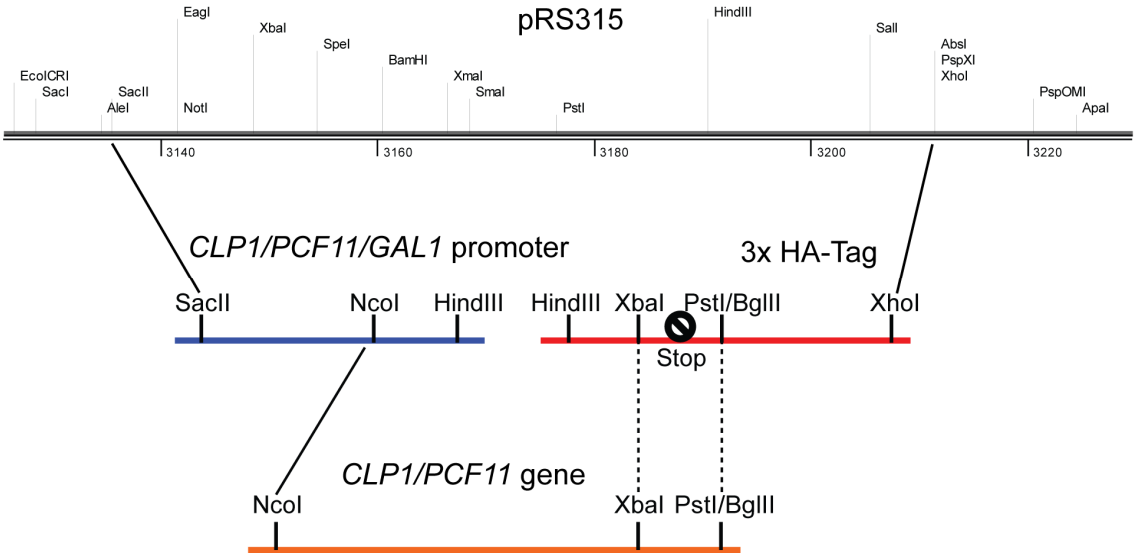


Figure 2.1. Cloning methods for pRS315-based yeast constructs. The *CLP1* and *PCF11* promoters were cloned independently of their gene sequences to enable their replacement with the *GAL1* promoter. Solid black lines between restriction sites indicate where PCR-derived DNA sequences have been inserted by restriction digestion and ligation. The black dashed lines show how different restriction sites were used at the junction with the HA-tag. Cloning of the desired gene sequence fragment using *Nco*I/*Pst*I or *Nco*I/*Bgl*II created an in-frame fusion with the HA-tag allowing the production of a C-terminally HA-tagged protein. Cloning using the *Nco*I/*Xba*I sites introduced a translational stop codon allowing production of an untagged protein.

Construct name	Vector	Organism	Marker	Function
Pcf11 456-626 EXP	pET-22b (Novagen)	*E. coli*	ampR	Expression of Pcf11 456-626
Pcf11 400-626 EXP	pGEX-6P-1 (GE Healthcare)	*E. coli*	ampR	Expression of Pcf11 400-626
Pcf11 416-626 EXP	pGEX-6P-1 (GE Healthcare)	*E. coli*	ampR	Expression of Pcf11 416-626
Pcf11 400-499 EXP	pGEX-6P-1 (GE Healthcare)	*E. coli*	ampR	Expression of Pcf11 400-499
Pcf11 416-499 EXP	pGEX-6P-1 (GE Healthcare)	*E. coli*	ampR	Expression of Pcf11 416-499
Clp1/Pcf11-EXP	pACYCDUET1 (Novagen)	*E. coli*	camR	Expression of Clp1-Pcf11 (454-563) complex
URA3-PCF11	pRS316 (ATCC)	*S. cerevisiae/ E. coli*	*URA3* ampR	Creating haploid *pcf11*Δ yeast strain
LEU2-CLP1	pRS315 (ATCC)	*S. cerevisiae/ E. coli*	*LEU2* ampR	Complementation of *clp1*Δ strain
LEU2-PCF11	pRS315 (ATCC)	*S. cerevisiae/ E. coli*	*LEU2* ampR	Complementation of *pcf11*Δ strain
LEU2-CLP1(HA)	pRS315 (ATCC)	*S. cerevisiae/ E. coli*	*LEU2* ampR	Complementation of *clp1*Δ strain
LEU2-PCF11(HA)	pRS315 (ATCC)	*S. cerevisiae/ E. coli*	*LEU2* ampR	Complementation of *pcf11*Δ strain
GAL1-CLP(HA)	pRS315 (ATCC)	*S. cerevisiae/ E. coli*	*LEU2* ampR	Complementation of *clp1*Δ strain, *CLP1* overexpression
GAL1-PCF11(HA)	pRS315 (ATCC)	*S. cerevisiae/ E. coli*	*LEU2* ampR	Complementation of *pcf11*Δ strain, *PCF11* overexpression

Table 2.3. Wildtype plasmid constructs. Mutant constructs are listed in Appendix 8.2.

2.2.4. DNA manipulation and analysis

Plasmid DNA was purified from overnight cultures of *E. coli* NovaBlue "Singles" Competent Cells using the QIAprep Spin Miniprep and QIAgen Plasmid Midi kits (QIAGEN) following the manufacturer's instructions. Plasmid DNA was eluted in 30-50 µl dH_2O. DNA was visualised by agarose gel electrophoresis as described by Sambrook *et al*. [193]. Final DNA concentrations were determined from UV absorbance at 260 nm using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific).

2.2.5. Polymerase chain reaction.

DNA sequences were amplified by polymerase chain reaction (PCR) using an Eppendorf Mastercycler PCR machine (Eppendorf), and the KOD Hot Start DNA Polymerase kit (Novagen). All reactions were carried out in 0.2 ml PCR tubes (STARLAB) at a final volume of 50 µl. The PCR protocol used was modified from that of the manufacturer (Novagen), an example of which is outlined in Tables 2.4 and 2.5. The addition of dimethyl sulfoxide (DMSO) reduces DNA secondary structure in the template and primers. PCR primers were synthesised by Eurofins MWG and are listed in Appendix 8.3. PCR products were purified using the QIAquick PCR Purification kit (QIAGEN) in accordance with the manufacturer's instructions. DNA was eluted in 30 µl dH_2O.

Reagent	Volume (µl)
KOD Hot Start buffer (10X)	5
$MgSO_4$ (25 mM)	3
dNTPs (2 mM each)	5
Forward oligonucleotide primer (10 µM)	1.5
Reverse oligonucleotide primer (10 µM)	1.5
DMSO (100%)	2.5
Template DNA (plasmid 10 ng/µl) (genomic 100 ng/µl)	1
dH_2O	29.5
KOD Hot Start DNA Polymerase (1 U/µl)	1
Total	50

Table 2.4. Typical PCR setup. DMSO is added to a final concentration of 5% to reduce template and primer secondary structure.

Step	Temperature		Time			
	103 ºC (constant lid temperature)					
1	95 ºC (initial melting)			2 min		
2	95 ºC (melting)			20 s		
3	57 ºC (annealing)			10 s		
4	70 ºC (extension)	Target size:	<0.5 kb	0.5-1 kb	1-3 kb	>3 kb
			10 s/kb	15 s/kb	20s/ kb	25 s/kb
5	Repeat steps 2-4, 25-30 cycles					
6	4 ºC			5 min		

Table 2.5. Typical thermal cycling conditions for PCR. The manufacturer's guidelines were used with slight modifications. Instead of using the lowest primer melting temperature (Tm, ºC) for the annealing step, 57 ºC was found to be effective.

2.2.6. Restriction digests

Double restriction digests were carried out using restriction enzymes from New England Biolabs (NEB). Where possible, a buffer system compatible with both enzymes was chosen in accordance with the NEBuffer Activity Chart for Restriction Endonucleases (see http://www.neb.com/). The reaction conditions for a typical restriction digest are

presented in Table 2.6. Where a common buffer for both enzymes could not be found, a sequential digestion was performed. Restriction digests were carried out in a volume of 30 µl at 37°C for 2 h, sequential digestions were in a final volume of 40 µl at 37°C for a total of 4 h.

Digested DNA products were separated by agarose gel electrophoresis using standard 1% agarose tris-acetate-EDTA (TAE) gels stained with 0.5 µg/ml ethidium bromide (Bio-Rad). Separated digestion products were then visualised on a Benchtop 3UV Transilluminator (UVP).

After restriction digestion, vector DNA was treated with Antarctic phosphatase (NEB) to remove 5'-phosphate groups to prevent vector recircularisation (see Table 2.7). Digested vectors were then separated by electrophoresis and purified directly from agarose gels using the QIAquick Gel Extraction Kit (QIAGEN) following the manufacturer's instructions. Digested PCR products were purified using the QIAquick PCR Purification kit (QIAGEN) in accordance with the manufacturer's instructions. Purified vector DNA and PCR products were eluted in 30 µl dH_2O.

Reagent	Volume (µl)
NEB buffer (10X)	3
BSA (100X)	0.3
DNA (vector 1-3 µg) (total PCR product)	1-3
NEB enzyme A (10 U/µl)	2
NEB enzyme B (10 U/µl)	2
dH_2O to final volume	30

Table 2.6. Typical restriction digestion conditions. For sequential digests DNA was first digested in 30 µl volume with enzyme A for 2 h in compatible buffer, then enzyme B and compatible buffer was added and DNA digested in 40 µl volume for a further 2 h.

Reagent	Volume (µl)
NEB Antarctic phosphatase buffer (10X)	4
Restriction digest	30
dH_2O	4
NEB Antarctic phosphatase (5 U/µl)	2
Total	40

Table 2.7. Typical phosphatase reaction conditions.

2.2.7. Ligation reactions

All ligation reactions were performed for a minimum period of 16 h at 16°C using T4 DNA Ligase (NEB). Reaction conditions can be found in Table 2.8.

Reagents	Volume (μl)
NEB T4 DNA ligase buffer (10X)	2
Digested vector (25 ng/μl)	4
Digested PCR insert (x ng/μl)	x
NEB T4 DNA ligase (400 U/μl)	1
dH_2O to final volume	20

Table 2.8. Ligation reaction conditions. The amount of vector used was always 100 ng. All ligations used a molar ratio of 3:1 (PCR insert:vector), with the exact volume of insert required being dependent on the concentration of the purified PCR product.

2.2.8. Transformations

E. coli NovaBlue "Singles" and *E. coli* BL21(DE3) "Singles" competent cells (Novagen) were used for cloning and protein expression purposes respectively. Transformation of both strains was carried out in the same manner: 1 μl of plasmid DNA (10 ng/μl) or 4 μl of a ligation reaction was added to a 50 μl aliquot of the relevant competent cells and the cells incubated on ice for 5 min. Cells were then heat shocked at 42°C for 30 s, and subsequently placed on ice for a further 2 min. A recovery step was then performed in which 200 μl of Luria-Bertani broth (LB) was added to cells followed by incubation at 37°C for 30 min. The entire transformation volume was then plated on LB agar containing the appropriate antibiotic. All plates were incubated overnight at 37°C. Transformation of yeast strains is described in section 2.8.1.

2.2.9. Site-directed mutagenesis

Site-directed mutagenesis (SDM) was carried out by PCR using an Eppendorf Mastercycler PCR machine (Eppendorf), and the KOD Hot Start DNA Polymerase kit (Novagen). Mutagenic primers (Eurofins MWG) were designed primarily according to the QuikChange protocol (Stratagene), however a non-overlapping primer design was also used in which single reverse primers could be used in conjunction with a choice of forward primers containing different mutations (see Figure 2.2). A list of mutagenic primers is presented in Appendix 8.3. The typical PCR protocol used for mutagenesis was adapted from the KOD Hot Start protocol (Novagen), and is outlined in Tables 2.9 and 2.10.

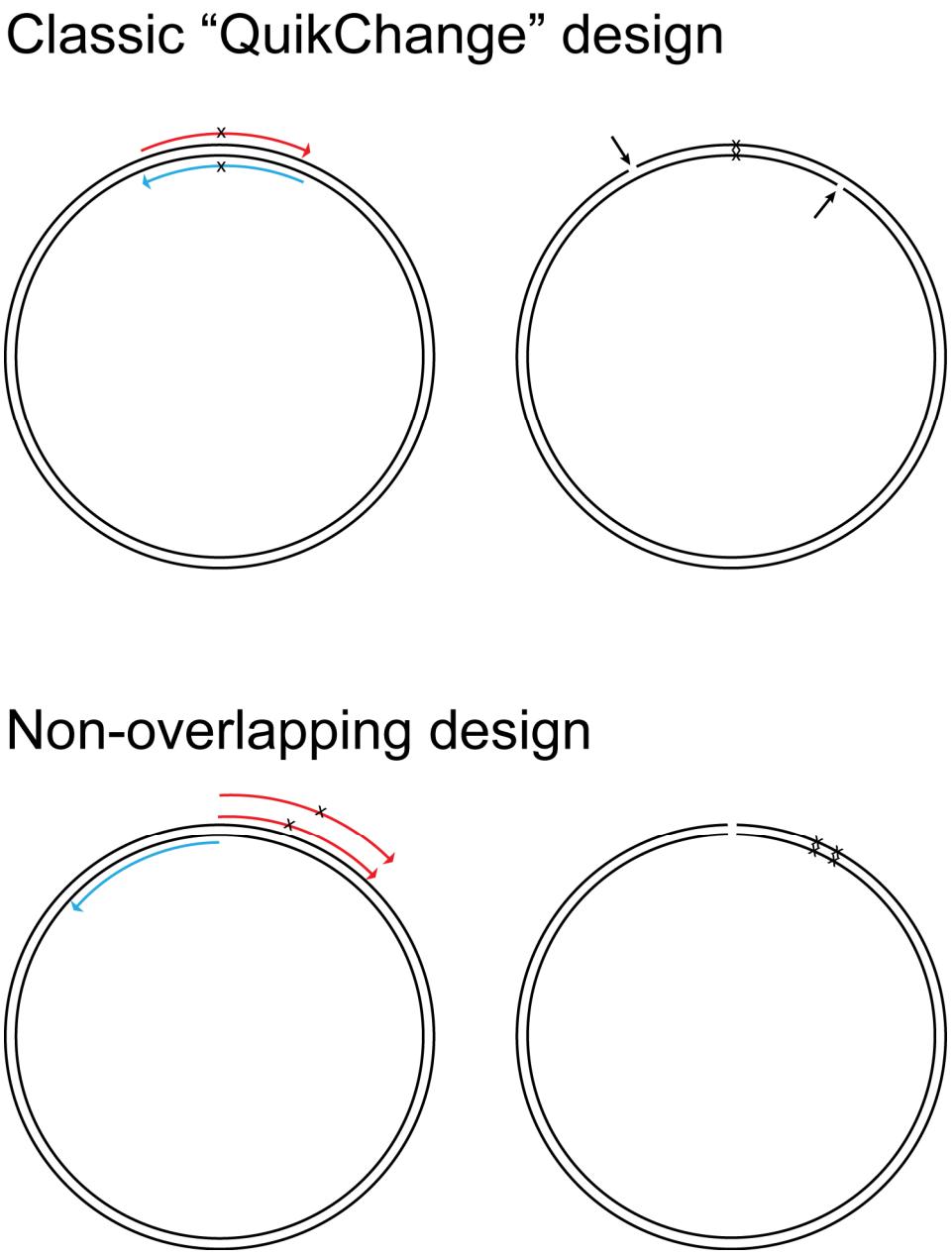


Figure 2.2. Primer design for site-directed mutagenesis (SDM). Most primers for SDM were designed in accordance with the “classic” QuickChange method. The overlapping primers result in amplification of template creating staggered nicks. The complementary sequence between the nicks is sufficient for recircularisation of the vector. With the non-overlapping design, a single reverse primer can be used with a choice of forward primers containing different mutations. The primers are abutting (no sequence overlap) and as a result the nicks on each strand are in the same location. A ligation step is required to facilitate recircularisation.

Reagent	Volume (µl)
KOD Hot Start buffer (10X)	5
$MgSO_4$ (25 mM)	3
dNTPs (2 mM each)	5
Forward oligonucleotide primer (10 µM)	1
Reverse oligonucleotide primer (10 µM)	1
Template DNA (plasmid construct 10 ng/µl)	1
dH_2O	33
KOD Hot Start DNA Polymerase (1 U/µl)	1
Total	50

Table 2.9. Typical SDM PCR setup. This reaction mixture was effective for both mutagenic strategies.

Step	Temperature	Time
1	95 ºC (initial melting)	2 min
2	95 ºC (melting)	20 s
3	57 ºC (annealing)	10 s
4	70 ºC (extension)	25 s/kb
5	(repeat steps 2-4, 20 cycles)	
6	70°C (final extension)	5 min
7	4 ºC	5 min

Table 2.10. Typical SDM PCR thermal cycling conditions.

PCR products were purified using the QIAquick PCR purification kit (QIAGEN), and eluted in 30 µl dH_2O. Methylated parental DNA template was digested with the restriction enzyme *Dpn*I (NEB). The 5'-ends of constructs amplified using non-overlapping primers were phosphorylated with polynucleotide kinase (NEB), and constructs recircularised by ligation using T4 DNA ligase (NEB). The reaction conditions for template digestion, phosphorylation and ligation of constructs are summarised in Tables 2.11, 2.12 and 2.13.

After final purifications, 10 ng of each mutant construct was transformed into *E. coli* Novablue "Singles" competent cells as described in section 2.2.8, and constructs purified as described in section 2.2.4.

QuikChange primer design

Reagent	Volume (µl)
NEB buffer 4 (10X)	4
PCR product	30
dH_2O	4.5
NEB *Dpn*I (20 U/µl))	1.5
Total	40

- Incubation at 37°C 1 h 30 min
- Purification (QIAGEN kit)

Table 2.11. Modification of SDM PCR products after QuikChange mutagenesis. When QuikChange primers have been used *Dpn*I treatment is required before transformation.

Non-overlapping primer design

Reagent	Volume (µl)
NEB T4 DNA ligase buffer (10X)	4
PCR product	30
dH_2O	3.5
NEB *Dpn*I (20 U/µl))	1.5
NEB T4 PNK (10 U/µl)	1
Total	40

- Incubation at 37°C 1 h 30 min
- Purification (QIAGEN kit)

Table 2.12. Modification of SDM PCR products generated using the non-overlapping primer design. PCR products require 5'-end phosphorylation.

Reagent	Volume (µl)
NEB T4 DNA Ligase Buffer (10X)	4
5'-phosphorylated construct PCR product	30
dH_2O	5
NEB T4 DNA ligase (400 U/µl)	1
Total	40

- Incubation at room temperature 30 min
- Purification (QIAGEN kit)

Table 2.13. Ligation reaction conditions to recircularise PCR products after 5'-end phosphorylation.

2.3. Protein expression and purification

2.3.1. Protein expression

Constructs for protein expression were transformed into *E. coli* BL21 (DE3) Singles competent cells (Novagen) as described in section 2.2.8. For protein expression, a single colony from an overnight plate was inoculated into 200 ml LB supplemented

with 34 μg/ml chloramphenicol (Sigma) (pACYCDuet1 constructs used in this study confer chloramphenicol resistance). Cultures were then incubated overnight in a shaking incubator (200 RPM) at 37°C. The following day, 7.5 ml of the starter culture was inoculated into 750 ml Terrific Broth (TB) supplemented with 34 μg/ml chloramphenicol. The culture was incubated with shaking at 37°C and monitored by measurement of OD_{600} using an Ultrospec cell density meter (GE Healthcare). When the culture density reached an OD_{600} of 0.6-0.8 the temperature of the incubator was dropped to 20°C and when the OD_{600} reached 1-1.5, IPTG was added to a final concentration of 1 mM to induce protein expression. The induced culture was incubated at 20°C for a further 16 h and the cells harvested by centrifugation at 4,500*g* for 20 min. Cell pellets were resuspended in 50 ml phosphate-buffered saline solution (PBS), and transferred to 50 ml centrifuge tubes (Corning Life Sciences). Cells were harvested again by centrifugation at 3,000*g* for 20 min. The supernatant was discarded and cells stored at -20°C. Typically, expression of recombinant proteins for *in vitro* studies was carried out at a three litre scale using 4 X 750 ml cultures.

2.3.2. Bacterial cell lysis

All cell lysis procedures were performed on ice. Cell pellets were defrosted and then resuspended at 10 ml/g (buffer/pellet) in the appropriate Buffer A (see Table 2.14 for buffer compositions). In addition, 15 kU/g rLysozyme and 12.5 kU/g Benzonase Nuclease (both Novagen) were added to aid lysis and reduce lysate viscosity. Two Complete Mini EDTA-free Protease Inhibitor Cocktail tablets (Roche) were also added to reduce protein degradation by *E. coli* proteases. The cell suspension was incubated on ice for 30 min. Cells were then lysed by sonication using a Branson Sonifer 450 set at 40% power with six 1 min pulses (1 min rests between each pulse). The insoluble fraction of the lysate was removed by centrifugation at 61,000*g* for 1 h. Samples of the clarified lysate and insoluble pellet (dissolved in 10 M urea) were removed for SDS-PAGE analysis (see section 2.3.6).

2.3.3. Purification of glutathione S-transferase fusion proteins

The Glutathione S-transferase Gene Fusion System (GE Healthcare) enables the expression and purification of fusion proteins produced in *E. coli*. The system is designed to overexpress a protein of interest that has been fused (at its C-terminus) with the 26 kDa Glutathione S-transferase protein (GST) derived from *Schistosoma*

japonicum. The fusion protein can be purified from bacterial lysate by affinity chromatography using immobilised glutathione.

GST purifications were carried out at room temperature. In each protein purification, 2 ml of glutathione sepharose 4B resin (GE Healthcare) was prepared according to the manufacturer's instructions. Clarified bacterial cell lysates (see section 2.3.2) were incubated with prepared glutathione Sepharose in 50 ml Falcon tubes at room temperature for 30 min with gentle agitation. The resin (containing bound proteins) was pelleted by centrifugation at 500*g* for 2 min and the supernatant was discarded. The resin was resuspended with 5 ml GST-A buffer (see Table 2.14) and incubated at room temperature for 10 min with gentle agitation. Resin was then pelleted again by centrifugation at 500*g* for 2 min and the supernatant was discarded. This wash step was repeated twice more and the resin was finally resuspended in 5 ml GST-A. Samples of the insoluble cell pellet, clarified lysate, bound fraction (resin after lysate incubation) and each wash were analysed by SDS-PAGE gel electrophoresis (see section 2.3.6). To remove the GST-tag from target proteins, non-cleavable GST-tagged 3C protease (GE Healthcare) was added to the resin according to the manufacturer's instructions and incubated with gentle agitation overnight at 4°C. The protease and the free GST generated by the cleavage remain bound to the resin. The efficiency of cleavage and cleavage products were analysed by SDS-PAGE (see section 2.3.6).

2.3.4. Nickel affinity purification

His-Tag sequences enable the isolation of recombinant proteins through the affinity of histidine residues to a matrix containing chelated nickel ions. Nickel affinity purification was carried out at room temperature using buffers chilled on ice. A 3 ml Ni-NTA Superflow (QIAGEN) gravity-flow column was equilibrated with 15 ml of Buffer Ni-A (see Table 2.14 for buffer compositions). Clarified lysate (prepared as described in section 2.3.2) was applied to the column and allowed to drain by gravity-flow. The column was washed with 15 ml Ni-A, then 25 ml Ni-Wash Buffer to remove proteins bound non-specifically to the nickel resin. The column was eluted with 25 ml Ni-B into 1 ml fractions. Samples of the lysate flow-through, washes and every other eluted fraction were saved for analysis by SDS-PAGE (see section 2.3.6).

Buffer	Composition	Purpose
GST-A	25 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.5 mM TCEP	Cell lysis and GST purification
Ni-A	50 mM Tris-HCl pH 8.0, 300 mM NaCl, 20 mM Imidazole pH 8.0, 5 mM $MgCl_2$, 0.5 mM TCEP	Cell lysis, binding of target protein to nickel column
Ni-Wash	50 mM Tris-HCl pH 8.0, 300 mM NaCl, 50 mM Imidazole pH 8.0, 5 mM $MgCl_2$, 0.5 mM TCEP	Removal of proteins bound non-specifically to nickel column
Ni-B	50 mM Tris-HCl pH 8.0, 300 mM NaCl, 300 mM Imidazole pH 8.0, 5 mM $MgCl_2$, 0.5 mM TCEP	Elution of 6-His-Tagged proteins
Gel Filtration (GF)	25 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM $MgCl_2$, 0.5 mM TCEP	SEC, protein storage

Table 2.14. Buffer compositions for protein purification.

2.3.5. Size-exclusion chromatography

Size-exclusion chromatography (SEC) allows separation of proteins by size. The matrices of SEC columns contain pores of a specified size, which allow proteins in the mobile phase to differentially penetrate the pores depending on the size of protein molecules. After analysis by SDS-PAGE, protein fractions from nickel affinity purification containing the desired protein product were pooled and concentrated (see section 2.3.7) to a volume of 2-5 ml. The concentrated protein was then loaded onto a HiLoad 16/60 Superdex (75 or 200) Prep grade SEC column linked to an AKTA prime system (all GE Healthcare). The column was eluted at a flow rate of 0.5 ml/min at 4°C. The first 30 ml of eluate corresponding to the column dead volume was allowed to flow to waste. After this, eluate was collected in 1 ml fractions. The protein content of fractions was analysed using the AKTA PrimeView Evaluation software (v5.0 GE Healthcare) and by SDS-PAGE (see section 2.3.6). Fractions containing purified target protein were pooled and concentrated as described in section 2.3.7.

2.3.6. SDS-PAGE

SDS-PAGE was used to analyse target protein purity after GST affinity chromatography, nickel affinity chromatography and SEC. All SDS-PAGE experiments were performed using NuPAGE Bis-Tris precast polyacrylamide gels with 1X NuPAGE MES SDS running buffer (Invitrogen). Gels were run at 200 V for 35 min in an XCell SureLock mini-cell using a PowerEase 500 power supply (both Invitrogen). 10 µl of prepared sample was loaded per lane, composed of an equal mixture of protein sample, 3X SDS loading buffer and 100 mM dithiothreitol (DTT, 30 mM final concentration). Samples were heated to 100°C for 5 min prior to loading. The Novex

Sharp Pre-stained Protein Standard (Invitrogen) was used as a molecular size marker. Gels were developed by staining with a Coomassie solution (50% methanol, 10% acetic acid, 0.05% Coomassie Brilliant Blue R-250) for 5 min, then destained with at least three 5 min washes with a destain solution (10% methanol, 10% acetic acid). Results were recorded with a Canon powershot PC1234 digital camera.

2.3.7. Protein concentration, storage and dialysis

Prior to SEC or storage, proteins were concentrated by centrifugal ultrafiltration at 4,000*g* using 20 ml Vivaspin ultrafiltration concentrators (Vivaproducts) of an appropriate molecular weight cut-off (MWCO). For storage after final concentration, the protein was snap-frozen in liquid nitrogen in aliquots of 30 or 100 µl (depending on yield) and stored at -20°C. Before circular dichroism experiments (see section 2.7), it was necessary to dialyse protein samples into phosphate buffer (10 mM Na_2HPO_4 /NaH_2PO_4 pH 7.4, 150 mM NaCl). A protein sample of typically 100-200 µl volume was defrosted on ice, and diluted with pre-chilled phosphate buffer to a final volume of 500 µl. The protein sample was injected into a 0.1-0.5 ml capacity 10,000 kDa MWCO Slide-A-Lyzer Dialysis Cassette (Thermo Scientific), and the sample dialysed against 2 l of phosphate buffer for 2 h at 4°C. The buffer was then replaced with a fresh 2 l of phosphate buffer and the dialysis continued overnight at 4°C. For CD, dialysed protein was removed from the cassette and concentrated to 0.15 mg/ml by centrifugation at 4,000*g* in a 500 µl capacity 10,000 MWCO Vivaspin 500 ultrafiltration (Vivaproducts).

2.3.8. Determination of protein concentration

Protein concentration was determined spectrophotometrically using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific). For each sample, a UV spectrum (200 nm to 350 nm) was recorded. Concentration was determined using the Beer-Lambert law (see Equation 2.1).

$$C_m = A_{280}/\varepsilon_{280}.l \qquad \text{(Equation 2.1)}$$

Where C_m is the Molar concentration, A_{280} is the absorbance at 280 nm, ε_{280} is the protein molar extinction coefficient at 280 nm (see section 2.1.1 for ε_{280} determination) and l is the optical path length, usually 1 cm.

2.4. Multiangle laser light scattering

Multiangle laser light scattering (MALLS) is a photometric method that uses the measurement of Rayleigh scattering (the scattering of light by molecules far smaller than the wavelength of the incident light beam) of molecules in solution to determine their absolute molar mass. When used in conjunction with SEC (SEC-MALLS), this technique is frequently used to elucidate protein molar mass, oligomerisation states and for the detection of protein aggregates [194]. A common SEC-MALLS laboratory setup uses an analytical SEC column to remove aggregates and separate proteins in a sample by size before the chromatographic peaks are passed through a series of absorbance (UV-Vis), light scattering (LS) and differential refractive index (dRI) detectors. Within the LS detector, a vertically-polarised laser beam is projected into the flow cell through which the eluted chromatographic peaks pass. The light intensity scattered by molecules in the sample (I) is measured by multiple detectors placed at different angular positions around the sample (see Figure 2.3A). The incident light is also measured as a reference (I_0). The detectors provide a response directly proportional to the intensity of the scattered light they receive, which is relayed to a computer for processing. The light scattering data is then used to calculate the Rayleigh ratio ($R\theta$, see Equation 2.2), which is the light scattered by the sample at an angle θ in excess of the light scattered by the pure solvent. This can be expressed as:

$$R\theta = \frac{I_\theta r^2}{I_0} \qquad \text{(Equation 2.2)}$$

where I_0 is the intensity of the incident light beam, I_θ is the total intensity of scattered light (in excess of that scattered by the pure solvent) observed at angle θ at a distance r from the point of scattering. The Rayleigh ratio is an integral part of the equation that is used to determine the absolute molar mass [195]:

$$\frac{KC}{R\theta} = \frac{1}{M} + 2A_2C \qquad \text{(Equation 2.3)}$$

where C is the solute (protein) concentration in g/ml, M is the weight-averaged molar mass, A_2 is the second virial coefficient (a measure of solution non-ideality) and K is an optical parameter equal to Equation 2.4:

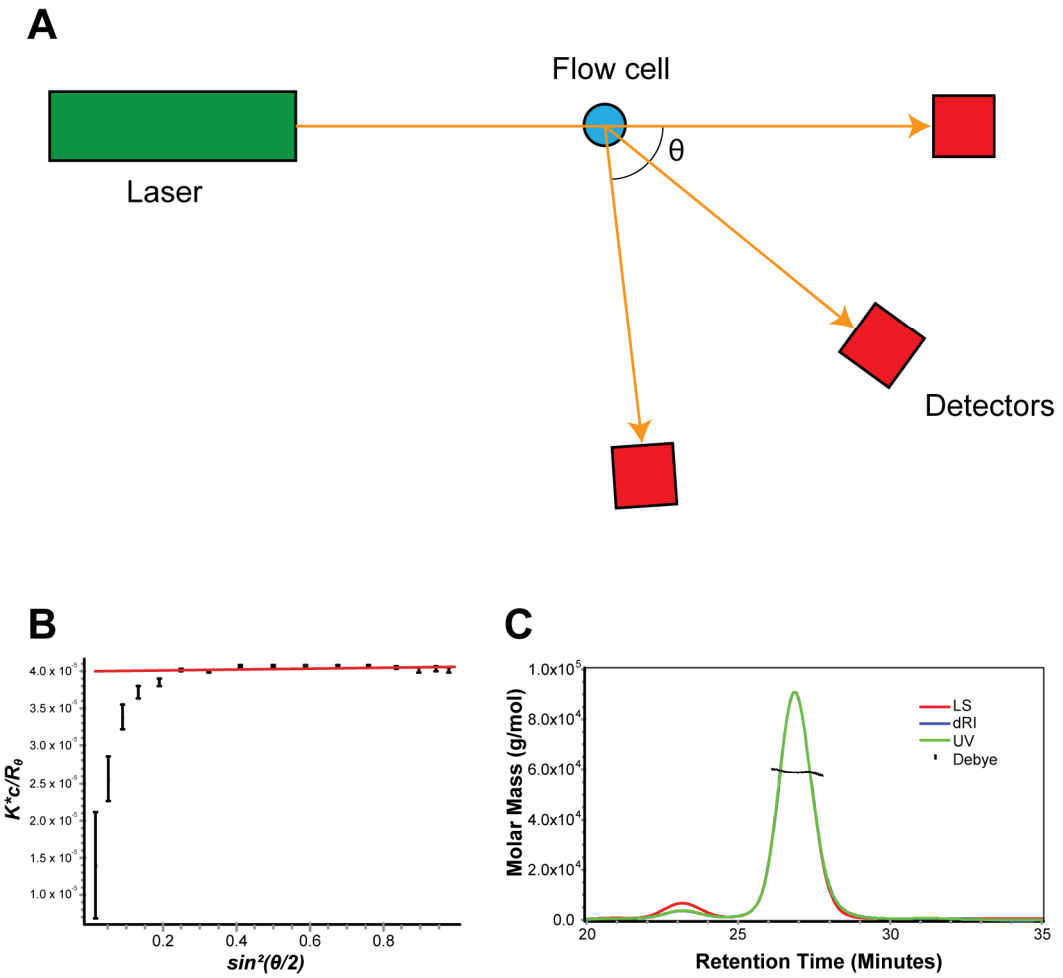


Figure 2.3. Multiangle laser light scattering (MALLS). (**A**) Design of a light scattering instrument. A collimated laser beam is passed into a flow cell (or a static cell) containing the sample. Molecules in the sample scatter the light from the laser in all directions. Detectors placed at different angular positions around the sample provide a response directly proportional to the intensity of scattered light they receive. The analogue signals from detectors are digitised and processed by relevant computer software. (**B**) Example of a Debye plot. Each point of the plot represents the KC/R_θ value for a given detector angle ($sin^2(\theta/2)$). The intercept of the points (red line) is used to calculate the weight averaged molar mass from Equation 2.3. The computer software generates plots or "slices" at short time intervals throughout the elution of a chromatographic peak. Combining and averaging the data of all the plots within a defined peak is used to calculate average molar mass and polydispersity. (**C**) Example of a MALLS chromatogram generated using ASTRA v5.3.4.19 software (Wyatt Technology). Light scattering (LS), differential refractive index (dRI) and UV traces indicate the position of chromatographic peaks (see inset key). The thick black line is composed of individual points that represent the weight-averaged molar mass determined by individual Debye plots as illustrated in (**B**).

$$K = \frac{4\pi^2 n_0^{\,2}(dn/dC)^2}{(\lambda^4 N)} \qquad \text{(Equation 2.4)}$$

where n_0 is the refractive index of the solvent, dn/dC is the refractive index increment of the solute at the same wavelength as the light scattering measurement, N_A is Avogadro's number and λ_0 is the vacuum wavelength of the incident light.

As can be seen in Equation 2.3, estimation of protein molar mass requires knowledge of the protein concentration in a sample. Consequentially, the LS detector is used in conjunction with a differential refractometer, which measures the very small differences between the refractive index of the sample and the solvent (n - n_0), allowing determination of protein concentration from the refractive index increment dn/dC.

In a SEC-MALLS experiment used to determine the absolute molar mass of a protein, the scattered light intensities and corresponding protein concentration are measured at a series of time intervals throughout peak elution and used to construct Rayleigh (KC/R_θ *vs C)* or Debye (KC/R_θ *vs* $sin^2(\theta/2)$) plots. Weight-averaged molar masses can then be determined from the intercept of individual plots, and an overall average molar mass and polydispersity term for the entire protein peak can be calculated by combining and averaging the data from all of the plots. The use of Debye plots to determine protein absolute molar mass is summarised in Figure 2.3B, and an example of a chromatogram demonstrating a series of Debye plots generated through a chromatographic peak is shown in Figure 2.3C.

In a typical SEC-MALLS analysis, samples were applied to a Superdex (75 or 200) 10/300 GL SEC column (GE Healthcare) equilibrated in 25 mM Tris-HCl pH 8.0, 150 mM NaCl, 4.9 mM $MgCl_2$, 0.5 mM TCEP and 3 mM NaN_3 (unless otherwise stated). The column was connected to a Jasco HPLC system equipped with UV detection (Jasco UV-970M) controlled by the Chrompass v1.8.6.1. software package (Jasco). The scattered light intensity of the column eluent was measured at sixteen angles using a DAWN-HELEOS photometer (Wyatt Technology) with a laser wavelength of 658 nm. The protein concentration of the eluate was determined using an Optilab rEX differential refractometer (Wyatt Technology) with its light source also set at 658 nm and a refractive index increment (dn/dC = 0.186 ml/g). The weight-averaged molar

mass of material contained in the chromatographic peaks was determined by combining data from the LS and dRI detectors using the ASTRA v5.3.4.19 software (Wyatt Technology).

2.5. Reversed-phase high-performance liquid chromatography

Reversed-phase high-performance liquid chromatography (RP-HPLC) can be used to separate and purify biomolecules within a sample based upon their hydrophobic properties. The separation mechanism is based upon the hydrophobic interaction of molecules dissolved in an aqueous mobile phase with immobilised hydrophobic ligands (the stationary phase) [196]. Typically, the stationary phase is made up of linear hydrocarbon chains attached to a silica matrix. The addition of an ion-pairing agent such as trifluoroacetic acid can be used to modify the retention time of proteins and peptides on a reversed-phase (RP) column. This is achieved through an ionic interaction between the pairing agent and positively charged basic residues in a protein or peptide, reducing their hydrophilicity. Introduction of an organic modifier lowers the polarity of the aqueous mobile phase and causes elution of biomolecules from the stationary phase. By using an elution gradient in which the polarity of the mobile phase is gradually lowered, it is possible to separate different biomolecules based on their varying degrees of hydrophobicity. By monitoring the eluate from a RP column using the chromophore absorbance, it is possible to visualise "resolved" peaks for further analysis or purification.

2.5.1. Reversed-phase nucleotide assay

Nucleotides were extracted from protein samples by addition of perchloric acid to 0.7%. Samples were immediately neutralised by addition of sodium acetate to 200 mM, and the supernatant containing the extracted nucleotides separated from precipitated protein by centrifugation at (17,000*g*) for 5 min at 4 ºC. The sample (80 µl at a concentration of 25 µM) was applied to a Zorbax SB-C18 column (4.6 x 250 mm, 5 µm pore size, Agilent Technologies) equilibrated in 100 mM K_2HPO_4/KH_2PO_4 (pH 6.5), 8% acetonitrile, 10 mM tetrabutylammonium bromide (TBAB), mounted on a Jasco HPLC system controlled by Chrompass v1.8.6.1. software (Jasco). The column temperature was maintained at 30°C. The retention time of the eluted nucleotide under these isocratic conditions was monitored at 260 nm using a Jasco MD-2010 Plus multi-wavelength detector fitted with a semi-micro flow cell. This multi-wavelength detector

also permitted the extraction of UV-absorbance spectra from the eluted nucleotide to confirm the identity of the nucleotide base. The column retention times of a set of nucleotide standards were determined to enable comparison with individual nucleotide species (i.e. to distinguish between AMP, ADP and ATP).

2.5.2. Reversed-phase complex composition assay

10 µg protein samples were dissolved in 20 mM Tris-HCl pH 8.0, 10 M urea, 2 mM EDTA, 5 mM TCEP, 1% acetic acid to give a final volume of 20 µl. The sample was applied to a Zorbax 300SB-C3 Narrow-Bore column (2.1 x 150 mm, 5 µm pore size, Agilent technologies) equilibrated in Buffer CC-A (see Table 2.15), mounted on a Jasco HPLC system controlled by Chrompass v1.8.6.1. software (Jasco). The column temperature was maintained at 60°C. The column was eluted using a gradient from Buffer CC-A to Buffer CC-B (see Table 2.16). The retention times of the eluted peaks were monitored at 220 nm using a Jasco MD-2010 Plus multi-wavelength detector fitted with a semi-micro flow cell.

Buffer CC-A	Buffer CC-B
HPLC- H_2O	Acetonitrile
0.05% TFA	0.05% TFA

Table 2.15. Reversed-phase complex composition assay buffer compositions.

Step	Time (min)	% A	% B
Pre-run	0	95	5
Start ramp	7	95	5
End ramp/start gradient	13	75	25
End gradient	40	40	60
Final elution	45	0	100
Final elution end	50	0	100
Re-equilibrate	55	95	5
End of experiment	65	95	5

Table 2.16. Complex composition assay gradient elution profile. Proteins in this assay did not elute before 25% B so a ramp step was introduced to save time. The final elution was a precaution to wash off any contaminants strongly bound to the column. The re-equilibration step ensured the column was ready for the start of another experiment.

2.5.3. Integration of UV absorbance peaks and stoichiometry calculation

Peak areas of chromatographic peaks were determined by integration of their UV absorbance and were used to quantify the amount of protein/nucleotide (in moles) contained within the peak using Equation 2.5.

$$N = \frac{A}{\varepsilon} \times \frac{f}{1000} \qquad \text{(Equation 2.5)}$$

Where N is the amount of protein/nucleotide in moles, A is the integrated peak absorbance at a given wavelength (220 nm or 280 nm for proteins, 260 nm for nucleotides) and f is the flow rate in ml per min. For proteins, when A is measured at 280 nm, ε is the molar extinction coefficient (ε_{280}). For nucleotides, measured at 260 nm, ε is the molar extinction coefficient of the nucleotide at 260 nm (ε_{260}). For data collected at 220 nm in RP-HPLC complex composition assays, assuming an average amino acid composition, ε is replaced by the number of residues of each chain (445 in Clp1, 107 in Pcf11(454-561)). The stoichiometry of Clp1-Pcf11 complexes was determined from the mole ratio of Pcf11 and Clp1 derived from 280 nm data or from the ratio of the residue-normalised Pcf11 and Clp1 absorbance at 220 nm.

2.6. Incorporation of radiolabelled ATP and scintillation counting

A stock solution of radiolabelled ATP was made in the following manner: 499 µl of ATP (concentration 1 mM) was spiked with 1 µl ATP [γ-^{32}P] (6000 Ci/mmol 10 mCi/ml, Perkin Elmer) to give a final ATP [γ-^{32}P] concentration of 20 µCi/ml. The 1 mM ATP [γ-^{32}P] stock was diluted into buffered protein samples to give molar ratios of 2:1, 5:1 and 10:1 (ATP [γ-^{32}P]: protein) in final volumes of 100 µl. Control reactions were prepared by dilution of 1 mM ATP [γ-^{32}P] into buffer containing no protein. Samples were incubated at room temperature for 1 h. After incubation, 1 µl of sample was removed for scintillation counting. Unincorporated ATP [γ-^{32}P] was then removed by centrifugation using PD spinTrap G25 desalting spin columns (GE healthcare) equilibrated in 25 mM Tris-HCl pH 8.0, 150 mM NaCl, 4.9 mM $MgCl_2$, 0.5 mM TCEP (the volume of sample after centrifugation was equal to the input volume), and 1 µl of the eluate was removed for scintillation counting. Samples were diluted into 500 µl dH_2O and 8 ml Liquiscint scintillation fluid (National Diagnostics), then mixed

thoroughly. Scintillation counting was performed on a Beckman LS 5000 CE scintillation counter using the ^{32}P channel. In samples containing protein, ATP exchange was measured by the difference between total ATP [γ-^{32}P] counts per million (CPM, before desalting) and eluted ATP [γ-^{32}P] CPM (after desalting) minus the background CPM (eluate from samples not containing protein).

2.7. Circular dichroism spectroscopy

Circular dichroism is the phenomenon that arises from the differential absorption of left- and right-handed circularly polarised light by chiral molecules. Circular dichroism spectroscopy (CD) exploits this effect to gain information about the structures of biomolecules. The peptide bonds of a protein are chromophores in the far-UV spectral region (190-250 nm) and are chiral. When in a regular, folded environment, secondary structure elements of proteins give rise to characteristic CD signals: α-helices display intense negative absorbance peaks at ~208 and 222 nm and an intense positive peak at ~198 nm. β-sheets give rise to a negative peak at ~215 nm and a positive peak at ~198 nm, and random coil has a weak positive absorbance peak at ~220 nm and a strong negative peak at 195-198 nm.

For CD experiments, protein samples were dialysed into phosphate buffer and concentrated to 0.15 mg/ml (as described in section 2.3.7). Far-UV CD spectra were recorded on a Jasco-J715 spectropolarimeter using a quartz cuvette with 1 mm pathlength. Spectra were taken with a 100 nm/min scan rate, 0.25 s time constant and 2 nm spectral bandwidth. Final spectra were the average of at least 20 scans. Data analysis was carried out with Dr S. Martin (NIMR) using the SpecPro program. The CD signal was measured in millidegrees and converted to the mean residue weight differential molar extinction ($\Delta\varepsilon_{mrw}$) using Equation 2.6.

$$\Delta\varepsilon_{mrw} = \frac{S \cdot (mrw)}{(3290) \cdot C \cdot l} \quad \text{(Equation 2.6)}$$

where S is the CD signal in millidegrees, mrw is the protein mean residue weight (a value close to 110 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, NIMR): the CONTIN, SELCON3 and CDSSTR methods of secondary structure estimation from CD data were used as recommended by Sreerama and Woody [197]. The final estimated secondary structure content was calculated by averaging the results provided by each method.

2.8. Yeast methods

2.8.1. Yeast transformations

A small amount of a glycerol stock of the desired yeast strain was inoculated into 20 ml yeast extract peptone dextrose medium (YPD, composition as described by Amberg *et al.* [198]) and incubated overnight at 30°C with shaking (180 RPM). The following morning, 1 ml of the overnight culture was inoculated into 50 ml fresh YPD and incubated for 5 h at 30°C in with shaking, and the yeast cells were harvested by centrifugation at 3,000*g* for 2 min. The cell pellet was washed in 30 ml dH_2O, then in 1 ml 100 mM lithium acetate, before being resuspended in 400 µl of 100 mM lithium acetate. A competent cell mix containing an aliquot of this cell suspension and the desired plasmid for transformation was prepared as described in Table 2.17. The competent cell mix containing the plasmid construct (transformation mix) was mixed thoroughly and incubated at 30°C for 30 min, followed by 42°C for at least 30 min and incubation on ice for 5 min. Transformed cells were then harvested by centrifugation at 6, 200*g* for 1 min, and resuspended in 600 µl dH_2O. 200 µl of transformed cells were plated on the appropriate selective medium (minimal medium minus leucine for *LEU2* plasmid constructs or minus uracil for *URA3* constructs, media compositions as described by Amberg *et al*. [198]) and plates incubated at 30°C. Colonies from successful transformants were typically visible after 2 days and at optimum size for selection after 3 days.

Reagent	Volume (μl)
Yeast cell suspension	50
50% PEG 3350 kDa	240
1 M lithium acetate	36
Carrier DNA 2 mg/ml * (single stranded salmon sperm, Sigma)	50
Desired plasmid construct (75 ng/μl)	2
dH_2O	32
Total	410

Table 2.17. Yeast competent cell mix. For multiple transformations, a master mix was made and divided between different tubes containing the desired plasmid constructs. * Salmon sperm DNA was denatured by incubation at 95°C for 10 min, then cooled on ice for 5 min prior to addition to the competent cell mix.

2.8.2. Construction of a haploid *pcf11Δ* yeast strain

A diploid *pcf11Δ/PCF11* yeast strain (see section 2.2.2) was transformed with the *URA3-PCF11* construct (see section 2.2.3) as described in section 2.8.1. Individual colonies were streaked onto minimal medium (minus uracil) and incubated at 30°C for 3 days until a thick patch of cells had grown. Cells were streaked from this plate onto minimal sporulation medium (composition as described by Amberg *et al.* [198]) and incubated at 23°C for 1-2 weeks. Cells were analysed under a light microscope for formation of tetrads (4 haploid spores). When sufficient numbers of tetrads were visible, the asci (capsules encasing the spores) were digested by mixing cells into 200 μl of 5% Glusulase solution (Perkin Elmer) and incubating at 30°C for 45 min. After this time, spores were settled at the base of the tube and the supernatant was decanted. The spores were resuspended in 500 μl dH_2O and then diluted 1 in 10. 50 μl aliquots of the diluted spore solution were deposited onto the edge of a YPD plate. 10-20 intact tetrads were dissected using a Singer MSM System dissection microscope. Individual spores from dissected tetrads were separated on the YPD plate using the microscope's computer grid system. The YPD plate containing dissected spores was incubated for 2-3 days to allow spores to germinate. Where 3 or 4 spores from a tetrad had germinated and multiplied by vegetative growth, the resulting haploid cells were streaked onto both YPD plates containing 100 μg/ml G418 antibiotic (Sigma) and 5-FOA plates (see section 2.8.3) and incubated at 30°C for 3 days. Yeast that were resistant to G418 but sensitive to 5-FOA were prospective haploid *pcf11Δ* strains: only yeast that is *pcf11Δ* could grow on G418 antibiotic as the cassette that disrupts *pcf11* confers kanamycin/G418 resistance, and sensitivity to 5-FOA confirmed the presence of the *URA3* plasmid construct containing the *PCF11* gene. To confirm the creation of a

haploid *pcf11Δ* strain, 5-FOA plasmid selection was then performed with haploid *pcf11Δ* transformed with the *LEU2-PCF11* construct (see section 2.8.3).

2.8.3. 5-FOA Plasmid shuffle

There are various methods by which a yeast strain can be manipulated so that the effects of mutations in target genes can be observed. A common approach to assay the functionality of a *S. cerevisiae* mutant is to delete the gene of interest from the chromosome and complement the deletion with a plasmid-borne copy of the gene containing the desired mutation. However, when genes are essential this poses a problem, as their deletion is lethal. To overcome this problem, a plasmid shuffle system has been developed based upon the principle that URA^+ yeast strains are sensitive to 5-fluoroorotic acid (5-FOA). This system allows the eventual phenotype of a gene mutant to be observed [199, 200].

The principles of the 5-FOA plasmid shuffle system are summarised in Figure 2.4. The first step in the system is to create a haploid strain in which the chromosomal copies of the target, *URA3* and *LEU2* genes have been deleted. Functional copies of the target and *URA3* genes are present at this stage on a *URA3* plasmid (see section 2.8.2 for haploid strain construction). This strain is subsequently transformed with a *LEU2* plasmid containing a functional copy of *LEU2* and a mutagenised copy of the target gene. Plating on minimal medium containing 5-FOA then allows selection against $URA3^+$ yeast.

5-FOA selection depends upon the conversion of the non-toxic 5-FOA compound to highly-toxic 5-flurouracil by the orotidine 5'-phosphate decarboxylase enzyme (OMP decarboxylase) encoded by the *URA3* gene [199, 200]. The 5-flurouracil product is a potent and irreversible inhibitor of thymidine synthase. Inhibition of thymidine synthase blocks the production of thymidine monophosphate (dTMP), leading to a cessation in DNA replication and lethality. Therefore, 5-FOA selection leads to the death of yeast cells containing the *URA3* plasmid bearing the wildtype copy of the chromosomally deleted target gene. Where the *URA3* plasmid has been lost by asymmetric plasmid segregation during mitosis, the yeast can survive on 5-FOA if they bear the *LEU2* plasmid with the mutagenised gene only if the encoded protein is viable.

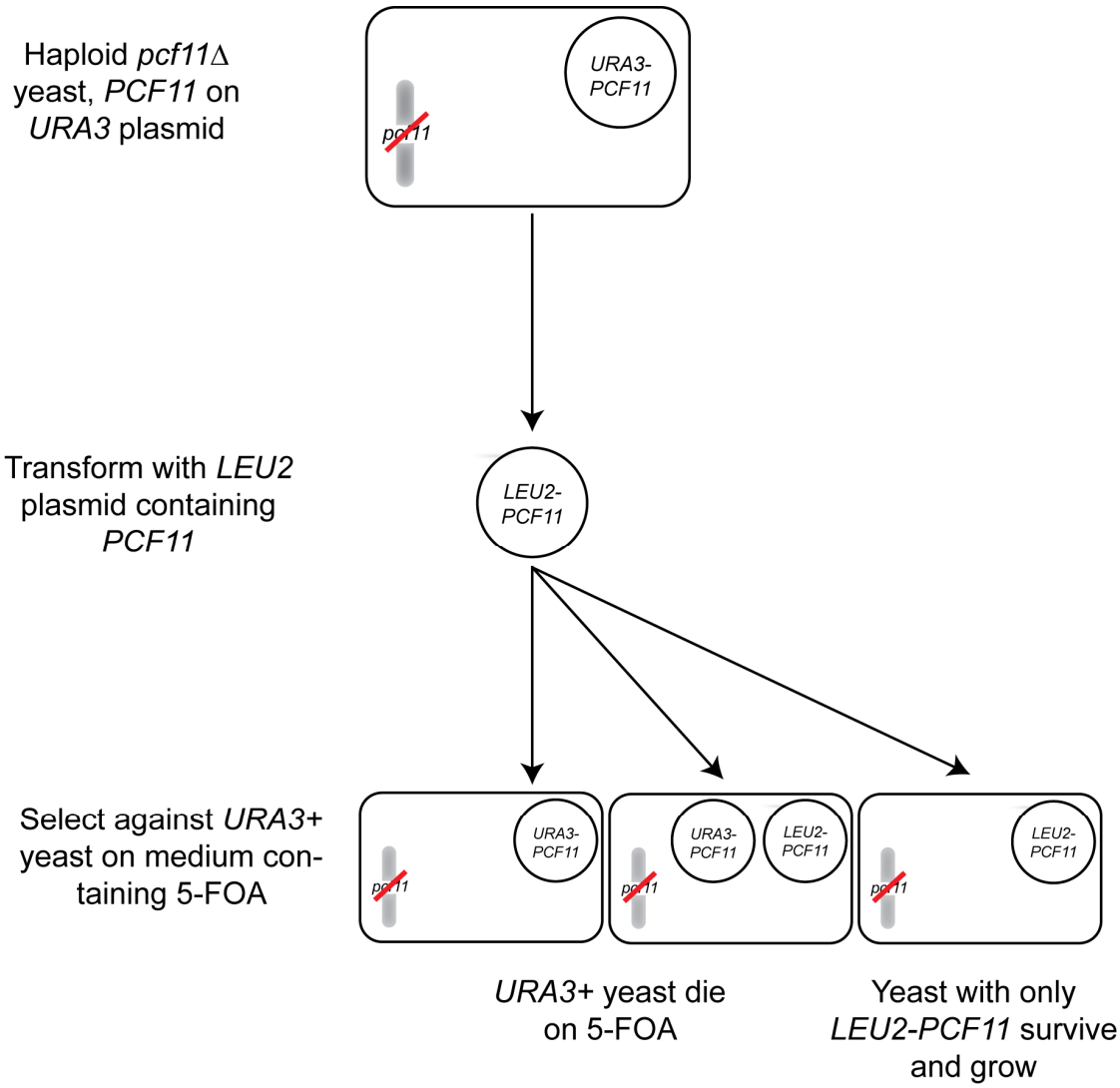


Figure 2.4. The 5-FOA plasmid shuffle system. The example given is for the haploid *pcf11Δ* strain. The wildtype copy of the target gene in the *LEU2* plasmid can be mutated directly by SDM. The 5-FOA plasmid shuffle can then be used to obtain a transformed yeast strain only expressing the mutant protein for further experiments. Where genes are essential, this is only possible if mutations within the gene do not completely abrogate essential functions.

The *clp1Δ* haploid yeast strain was supplied by Dr. M. Geymonat (formerly NIMR), and the *pcf11Δ* haploid yeast strain was created as described in section 2.8.3. These strains were transformed with *LEU2* plasmids containing copies of wildtype or mutant *clp1* or *pcf11* as described in section 2.8.1. An individual colony from *LEU2*$^+$ transformants was streaked onto minimal medium minus leucine and incubated at 30°C for 2-3 days. 5-FOA selection was carried out in the following manner: Cells were taken from the colony-derived cell patch and resuspended in 1 ml dH_2O. A further dilution of 30 µl into 1 ml was made to obtain the first point in a dilution series of cell suspensions. Three serial dilutions of 1:10 ratio (cells:dH_2O) were made from the first point in the series (total four points). 5 µl of each dilution was plated onto both minimal medium containing 5-FOA (1 mg/ml) and YPD. Plates were allowed to dry and incubated at 23°C for 5 days. To look for temperature sensitive phenotypes, a duplicate set of plates was incubated at 37°C (or other specified temperatures) for 3 days.

2.8.4. Growth rate assays

After 5-FOA selection, the desired yeast mutant was plated onto YPD and incubated overnight at 30°C. The following evening, cells from the plate were inoculated into 20 ml YPD broth and incubated overnight with shaking (180 RPM) at 30°C. The following morning, the cell density of the culture was measured. Briefly, 50 µl of cell culture was mixed with 10 ml 1X CASYton cell culture dilution solution (Schärfe system) in a sample vial. The mixture was then sonicated at 20% power for 3 s and the cell density measured using a CASY1 TT cell counter (Schärfe system). For growth rate assays, the overnight culture was diluted into fresh YPD broth to give a density of 1 X 10^6 cells/ml, and incubated with shaking at 30°C unless another temperature was specified. The cell density was measured every hour for eight hours as described above.

For assays with mutant yeast under control of the *GAL1* promoter, growth rate assays were performed as above except using YPD with 2% galactose and 2% sucrose (YPD gal/suc) as carbon sources (instead of glucose). To monitor the repression of the *GAL1* promoter, the desired yeast mutant was initially cultured in YPD gal/suc broth and diluted into fresh YPD glucose (YPD glu) broth before the start of the 8 h time course.

2.8.5. Western Blotting

A *LEU2* construct containing the desired gene mutant was transformed into the appropriate haploid deletion strain as described in section 2.8.1. A single colony from the transformation plate was streaked onto minimal media and incubated at 30°C for 3 days. Using a sterile loop, cells were deposited into 20 ml liquid minimal medium (minus leucine), and cultured overnight at 30°C with shaking (180 RPM). The following morning, 2 ml of the culture was diluted into 50 ml fresh liquid minimal medium (minus leucine) and incubated for 5 h at 30°C with shaking. The yeast cells were then harvested by centrifugation at 3,000*g*, washed once with 1 ml PBS then snap frozen in liquid nitrogen and stored at -80°C.

Whole cell extract was prepared by the following procedure. Defrosted yeast cell pellets were resuspended in 500 µl of yeast lysis buffer (50 mM Tris pH 7.4, 250 mM NaCl, 0.5 mM EDTA). To assist with lysis, Nonidet P-40 was added to a final concentration of 0.2% and 5 µl of 100X Halt Protease Inhibitor Cocktail (Thermo Scientific) was also added to inhibit protein degradation. Yeast cells were lysed with 500 µm diameter acid-washed glass beads (Sigma) using a FastPrep-24 Ribolyzer (MP Biomedicals). Typically, six 10 s bursts with 2 min rests on ice between bursts was sufficient for cell lysis. The whole-cell extract lysate was collected by centrifugation at 3,000*g* for 2 min, and clarified by centrifugation twice at 17,000*g* for 15 min at 4°C. The protein concentration of whole cell extract was determined by the Bio-Rad Protein assay following the manufacturer's protocol (Bio-Rad), using an Eppendorf BioPhotometer (Eppendorf).

Samples for Western blotting were composed of an equal mixture of yeast whole cell extract, 3X SDS loading buffer and 100 mM DTT. The final protein concentration of loading samples was 2 mg/ml. Before loading, samples were heated to 100°C for 5 min. SDS-PAGE was performed as described in section 2.3.6. After SDS-PAGE, the gel was soaked in 40 ml Western transfer buffer (25 mM Tris-HCl pH 8.3, 194 mM glycine, 10% methanol) for 20 min and the proteins transferred to an Immobilon-P Transfer membrane (Millipore) in a BioRad Trans-Blot Cell run at 100 V for 1 h. After protein transfer, the membrane was blocked overnight with 10 ml 1X Western Blocking Reagent (Roche) at 4°C with gentle agitation. The following morning the membrane was placed into 10 ml fresh 0.5X Western Blocking Reagent. The membrane was

probed by addition of a 1:10,000 dilution of mouse Anti-HA antibody (clone 12CA5, Roche) and incubation at room temperature for 1 h at room temperature with gentle agitation. To remove non-specifically associated primary antibody, the membrane was washed three times for 10 min with 20 ml Tris-buffered saline plus 0.1% Tween-20 (TBST). For detection, the membrane was placed into 20 ml 0.5X Western blocking buffer and a 1:20,000 dilution of peroxidise-conjugated goat anti-mouse secondary antibody added (Thermo Scientific). The membrane was incubated at room temperature for 1 h with gentle agitation and then washed three times as described above. SuperSignal West chemiluminescent substrate was applied in accordance with the manufacturer's protocol (Thermo Scientific) and protein bands visualised using X-ray film (Amersham Hyperfilm MP, GE healthcare). Exposures were 10-30 s, and the films were developed using a FujiFilm FPM-3800AD film processor.

2.8.6. Coimmunoprecipitation experiment

The desired yeast mutant was cultivated as described in section 2.8.5, except the culture volume was increased to 100 ml. Whole cell extract was prepared immediately as described in section 2.8.5 with the following modifications: the lysis buffer was composed of 25 mM Tris-HCl pH 7.4, 150 mM NaCl, 5 mM $MgCl_2$ and 5% glucose, with no addition of EDTA or Nonidet-P40. Protease inhibitors were used as before. In the absence of detergent, the lysate was clarified by centrifugation twice at 17,000*g* for 15 min at 4°C, and finally at 17,000*g* for 30 min at 4°C.

CoIP of HA-tagged proteins with associated binding partners was performed with the Pierce HA Tag IP/Co-IP Kit (Thermo Scientific) using the manufacturer's protocol, except that Tween-20 was omitted from buffers. 20 µl of the CoIP elution volume was resolved by SDS-PAGE as described in section 2.3.6 and the proteins were transferred to an Immobilon-P Transfer membrane as described in section 2.8.5. The membrane was blocked and probed with primary antibody as described in section 2.8.5 except a 1:1,000 dilution of rabbit anti-Clp1 antibody (produced for us by BioServ UK, Sheffield, UK) was used. The membrane was washed as described in section 2.8.5, and detection and visualisation with a 1:20,000 dilution of peroxidase conjugated goat anti-rabbit secondary antibody (Thermo Scientific) was performed as described in 2.8.5. For corresponding anti-HA detection, the membrane was stripped by incubation at room temperature for 15 min in 20 ml Restore Western Blot Stripping buffer (Thermo Scientific). The membrane was briefly washed in TBST before blocking and detection

of anti-HA tagged protein was performed (using the same protocol as described in 2.8.5).

2.8.7. Determination of protein half-life

The desired HA-tagged construct was transformed into the appropriate haploid yeast strain and selected by growth on 5-FOA as described in sections 2.8.1 and 2.8.3 respectively. Cells from the 5-FOA plate were streaked onto YPD and incubated at 30°C for 2 days. Cells from the YPD plate were inoculated into 20 ml YPD broth and cultured with shaking (180 RPM) overnight at 30°C. The following morning, the cell density of the culture was measured as described in section 2.8.4. The appropriate volume of culture was inoculated into 100 ml fresh YPD broth to give a cell density of 2 X 10^6 cells/ml, and the culture grown for 8 h with shaking at 30°C. After this time, cycloheximide (Sigma) was added to the culture to a final concentration of 50 μg/ml to inhibit protein synthesis, and 5 ml samples withdrawn from the culture at 0, 30, 60, 90 and 120 min time points post cycloheximide treatment. Cells from these samples were harvested immediately by centrifugation at 3,000*g*, and the cell pellets flash frozen in liquid nitrogen and stored at -20°C. Yeast whole cell extract was prepared as described in section 2.8.5, except the volume of lysis buffer used was reduced to 100 μl. Western blotting was performed as described in section 2.8.5. The developed film was scanned using an Epson Perfection V700 duplex scanner and the ImageJ v1.45s software (National Institutes of Health) was used to quantify the intensity of individual protein bands. The apparent half-life of proteins was derived from plots of band intensity (A) against time after cycloheximide treatment (t). The data were fitted directly to Equation 2.7 that describes the exponential decaying signal in terms of the initial amplitude (Ao) and the protein half-life ($t^{1/2}$).

$$A = A_0 e\,((-(ln2)/t^{1/2})t) \qquad \text{(Equation 2.7)}$$

2.8.8. Preparation of total RNA

The desired mutant HA-tagged construct was transformed into the appropriate haploid yeast strain and selected by growth on 5-FOA as described in sections 2.8.1 and 2.8.3 respectively. Cells from the 5-FOA plate were streaked onto YPD and incubated at 30°C for 2 days. Cells from the YPD plate were inoculated into 20 ml YPD broth and cultured overnight at 30°C with shaking (180 RPM). The following morning, the cell density of the culture was measured as described in section 2.8.4. The appropriate

volume of culture was inoculated into 50 ml fresh YPD broth to give a cell density of 2 X 10^6 cells/ml, and the culture grown for 8 h with shaking at 30°C. Cells were harvested by centrifugation at 3,000*g,* and the cell pellet flash frozen in liquid nitrogen and stored at -20°C overnight. The following day, total RNA was extracted from the cell pellet using the RiboPure Yeast Kit (Ambion) according to the manufacturer's protocol. Purified RNA was divided into 5 µl aliquots and stored at -80°C. The purity of RNA samples was assessed by UV absorbance using the A_{260}:A_{280} ratio (a high purity sample has a A_{260}:A_{280} ratio of 1.8 to 2.1). The quality of RNA samples was assessed by comparing the intensities of the 18S and 28S ribosomal RNA (rRNA bands) that were visualised by denaturing agarose gel electrophoresis (following the manufacturer's instructions, Ambion, see Appendix 8.7).

2.9. Poly(A) tail-length assay

The poly(A) tail-length assay is a technique used to assess the ability of specific yeast strains to polyadenylate pre-mRNA. The assay involves extracting and radiolabelling total RNA samples from the desired yeast strain. Total RNA is then digested using specific RNases that do not cleave at adenine bases, leaving polyadenylate tails as the only RNA fragments of significant length. Poly(A) tail-length is then analysed after denaturing polyacrylamide gel electrophoresis, and differences in polyadenylation are determined by changes in the size distribution of poly(A) tails between wildtype and the desired mutant yeast strains. Various adaptations of a similar poly(A) tail-length assay protocol have been cited in the literature in the past 25 years [91, 94, 104, 201, 202]. These protocols are very concise, and had to be considered as templates to formulate our own protocols, described below.

2.9.1. Poly(A) tail 3'-end labelling

Total RNA was radiolabelled using poly(A) polymerase and the ATP analogue, cordycepin 5'-triphosphate [α-^{32}P] (Perkin Elmer) that lacks a 3'-hydroxyl moiety and causes chain termination during RNA synthesis (hence each RNA chain is radiolabelled with one molecule of cordycepin). The components of the 3'-end labelling reaction used in the first method for poly(A) tail-length analysis are listed in Table 2.18. 3'-end labelling was performed at 37°C for 30 min, and then poly(A) polymerase heat inactivated by incubation at 65°C for 15 min. Unincorporated cordycepin was removed

by centrifugation using an Illustra MicroSpin G25 spin column (GE Healthcare) that was pre-equilibrated in 1X PAP buffer (Ambion).

Reagent	Original Volume (μl)	Modified Volume (μl)
Total RNA sample (1 μg/μl)	2.5	0.5
Poly(A) polymerase buffer (5X)	6	6
$MnCl_2$ (12.5 mM)	6	6
E. coli poly(A) polymerase 2U/μl	2	2
Nuclease-free water	13	15
[α-^{32}P] Cordycepin 5'-triphosphate (see legend)	0.5	0.5
Total	30	30

Table 2.18. Original and modified reaction setups for 3'-end labelling of total RNA samples. Poly(A) polymerase and associated buffers were purchased from Ambion. Cordycepin 5'-triphosphate: 3'-deoxyadenosine 5'-triphosphate [α-^{32}P], 250 μCi (9.25 MBq), 10 mM tricine pH 7.6, concentration 10 mCi/ml (Perkin Elmer).

2.9.2. Poly(A) tail digestion and purification

The volume of the radiolabelled RNA solution retrieved after spin column purification (described in section 2.9.1) was increased to 95 μl by addition of 65 μl 1X PAP buffer. RNA was then digested at 37°C for 2 h with a combination of RNase A and RNase T1 enzymes (Ambion), which hydrolyse RNA chains at C/U ribonucleotides and G ribonucleotides respectively. The reaction setup for RNA digests is presented in Table 2.19. After digestion, reactions were stopped and RNases removed by addition of 100 μl of termination buffer (2.5% SDS, 130 mM EDTA, 2 mg/ml proteinase K- Ambion) and incubation at 37°C for 30 minutes. 200μl of phenol:chloroform:IAA (25:24:1) pH 6.7 (Ambion) was then added to the mixture, and after vigorous mixing and centrifugation at 17,000g for 2 min, the aqueous phase containing the RNase A and RNase T1 resistant poly(A) tails was removed (~180 μl). The extracted RNA was precipitated overnight at -20°C following the addition of 20 μl 3M ammonium acetate (Ambion), 5 μl yeast tRNA (10 mg/ml, Ambion) and 2.5X volumes of 100% ethanol (500 μl). The following morning, RNA was recovered by centrifugation at 17,000*g* for 30 min at 4°C. The RNA pellet was washed once with 500 μl of ice-cold 70% ethanol to remove residual salt, and the pellet was recovered by centrifugation for 10 min at 4°C. Ethanol was removed by aspiration and the pellet was allowed to dry at room temperature. The RNA pellet containing poly(A) tails was dissolved in 20 μl of Gel loading buffer II (a formamide-based loading dye, Ambion).

Reagent	Original Volume (µl)	Modified Volume (µl)
Radio-labelled RNA sample	30	30
poly(A) polymerase buffer (1X)	65	-
100mM Tris-HCl pH 8.0 3M NaCl, 30 mM EDTA	-	4
Yeast tRNA (10 mg/ml)	2	-
RNase A (1 mg/ml)	2	3
RNase T1 (1000 U/µl)	1	3
Total	100	40

Table 2.19. Original and modified reaction setups for RNA digestion. Yeast tRNA and RNases were purchased from Ambion.

2.9.3. Visualisation of poly(A) tails

Prior to electrophoresis, poly(A) tail samples were heated at 95°C for 10 min. RNA poly(A) tail samples were resolved on 8.3 M urea 10% acrylamide/Bis (19:1, Ambion) polyacrylamide gels by electrophoresis at a constant power of 35 watts for 90 min in an EV500 DNA sequencing gel tank (Cambridge electrophoresis). After drying, gels were exposed to a storage-phosphor screen (GE Healthcare) and visualised on a Storm 860 phosphor-imager (GE Healthcare).

2.9.4. Modifications to poly(A) tail length assay protocols

Different modifications were made to the protocols described in sections 2.9.1-2.9.3 in an attempt to improve the resolution of poly(A) tails on acrylamide gels (discussed in section 4.3.1). Changes in the components of 3'-end labelling and RNA digestion reactions are listed in Tables 2.18 and 2.19 respectively. The modifications to the protocol that resulted in the largest improvements in resolution are as follows.

1) The volume of the RNA digest reaction was reduced to 40 µl (see Table 2.19).
2) RNA digests were terminated by adding 2.5 µl 10% SDS, 4 µl proteinase K and 3.5 µl dH_2O with incubation at 37°C for 30 min (total volume of terminated reaction is reduced from 200 µl to 50 µl).
3) RNA was precipitated overnight at -20°C after addition of either 5 µl 3 M ammonium acetate or 5 µl 3 M sodium acetate with 5 µl yeast tRNA (10 mg/ml, Ambion), 5 µl glycogen (5 mg/ml, Ambion) and 2.5X volumes of 100% ethanol (150 µl). RNA pellets were washed either once or twice with 70% ethanol.
4) The percentage of acrylamide/Bis in gels was increased to 15% to allow a more complete separation of poly(A) tails of different sizes.

2.9.5. DNA ladder preparation

A DNA oligonucleotide size standard ladder was created by mixing single-stranded DNA oligonucleotides of the following sizes (2 µl each, concentration 10 µM, oligos synthesised by Eurofins MWG): 12, 15, 20, 25, 30, 35 40, 46 50, 57, 65, 77, 96 and 125 nucleotides. The 5'-ends of oligonucleotides in this mixture were radiolabelled with ATP [γ-^{32}P] (Perkin Elmer) using T4 PNK (NEB). The components of the 5'-end labelling reaction are listed in Table 2.20. The labelling mixture was incubated at 37°C for 30 min and then T4 PNK was heat inactivated at 65°C for 15 min. Unincorporated ATP [γ-^{32}P] was removed by centrifugation using an illustra MicroSpin G25 spin column (GE Healthcare) and the oligonucleotide size standard mixture was purified by phenol-chloroform isoamyl alcohol extraction and ethanol precipitation as described in section 2.9.2. The DNA pellet was dissolved in 20 µl Gel loading buffer II (Ambion) and 2 µl was loaded onto polyacrylamide gels along with radiolabelled poly(A) tail RNA samples.

Reagent	Volume (µl)
Oligonucleotide mixture	5
T4 PNK buffer (10X)	4
ATP [γ-^{32}P] (see legend)	2
T4 PNK (10 U/µl)	2
dH_2O	27
Total	40

Table 2.20. Reaction setup for oligonucleotide radiolabelling. T4 PNK was purchased from NEB. ATP [γ-^{32}P] 250 µCi (9.25 MBq), 6000 Ci/mmol, concentration 10 mCi/ml (Perkin Elmer).

2.10. Quantitative real-time PCR

Quantitative real-time PCR (qPCR) is a technique that allows quantification of the amount of a target DNA sequence in a sample. Quantification is achieved by monitoring the amplification of target sequences through fluorescence reporter systems (reviewed in [203]). qPCR is also commonly used to quantify gene expression using reverse-transcribed RNA as template (RT-qPCR). It is also a useful tool in other areas of life sciences research. For example, qPCR of ChIP purified DNA has allowed quantification of genomic DNA sequences associated with nuclear proteins [204-206]. qPCR has also been used for single nucleotide polymorphism (SNP) genotyping [207], which has proved important for the detection of disease-associated SNPs in molecular diagnostics.

The RT-qPCR experiments used in this study employed the 5' nuclease (TaqMan) assay rather than the alternative SYBR-green based detection [208]. The 5' nuclease assay uses hydrolysis probe technology and takes advantage of the 5'->3' exonuclease activity of *Thermus aquaticus* DNA polymerase (Figure 2.5) [209]. During the annealing step of PCR, a dual-labelled hydrolysis probe hybridises to its complementary sequence within the target sequence. At this stage, the fluorescent reporter covalently attached to the 5'-end of the probe is in close proximity to a non-fluorescent quencher attached to the 3'-end of the probe. As a result, the fluorescence emitted by the reporter when excited by light of the appropriate wavelength is quenched. As the polymerase extends from 5' to 3' from the forward primer, the 5' exonuclease activity cleaves the probe, freeing the reporter and allowing detection of the fluorescent signal.

The most useful measurement for quantitative means is the determination of the threshold cycle (C_T), the PCR cycle at which the fluorescent signal of the reporter crosses a threshold where there is first a detectable significant increase in fluorescence (in the exponential phase of PCR). Different methods can be applied to derive the absolute or relative quantities of target from C_T values (see Applied Biosystems User Bulletin #2 [210]). The RT-qPCR experiments described in sections 2.9.3 and 2.9.4 were analysed using the relative standard curve method [210]. In this method, target is quantified in a dilution series of an RNA or DNA standard (of known quantity) to generate a standard curve. Comparison of the C_T value of a sample with the C_T values of the standard curve is used to determine the quantity of target (given as a number without units). Relative quantification is achieved by dividing the quantity of the sample through the quantity of a calibrator to give a final target quantity expressed as an *n*-fold difference to the calibrator (for example gene expression in a mutant relative to wildtype gene expression). For a worked example refer to Figure 2.6.

2.10.1. Reaction setup

RT-qPCR experiments were carried out in an Applied Biosystems 7500 Fast Real-Time PCR System. All primers and probe sequences were designed in accordance with parameters defined in the Primer Express 2.0 software (Applied Biosystems) and purchased from Applied Biosystems. Primers and probes for RT-qPCR are listed in Appendix 8.4. All target amplicons were detected by the 5' nuclease (TaqMan) assay using 5'-6-carboxyfluorescein (6-FAM) 3'-minor groove binder (MGB) labelled probes

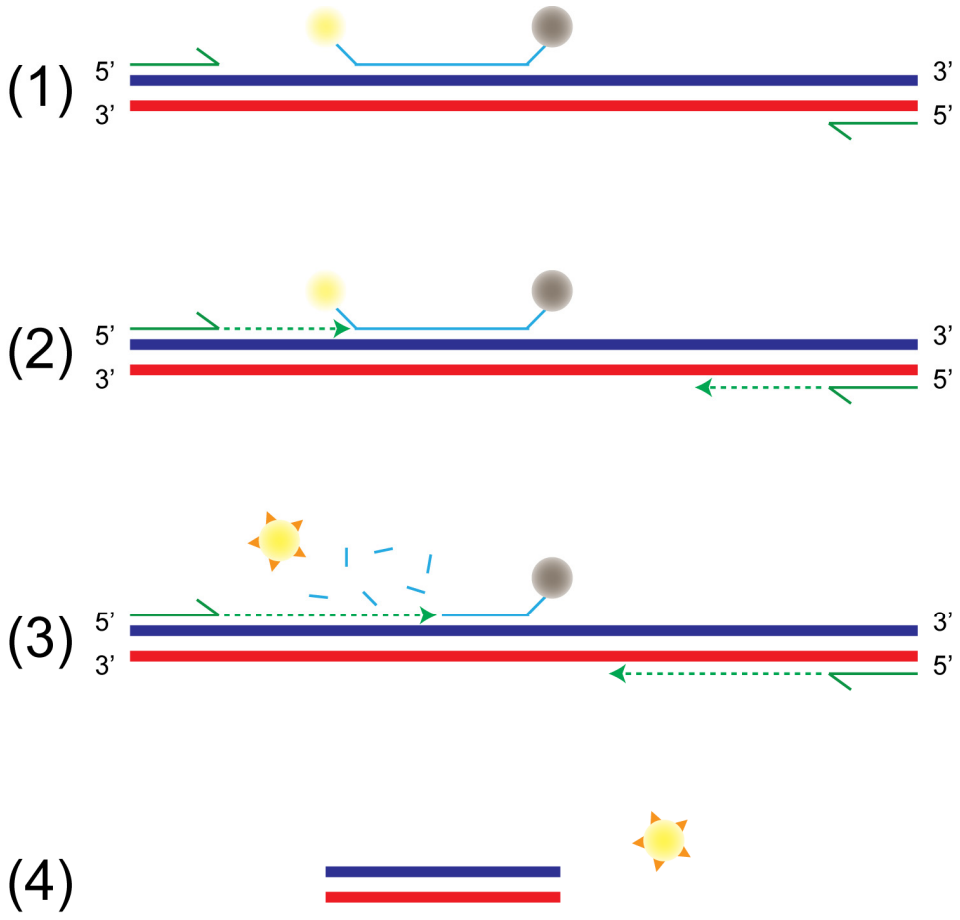


Figure 2.5. Hydrolysis (TaqMan) probes for qPCR. During the PCR cycle, (1) Primers (green) and probe (blue) bind to their complementary sequences (sense strand, dark blue) and (2) Taq polymerase extends from the primers (dashed green line) to transcribe the amplicon. (3) Upon reaching the probe (annealed to the sense strand here), Taq polymerase displaces and cleaves the probe using its 5'->3' exonuclease activity. This frees the fluorophore (yellow sphere) from the quencher (grey sphere) allowing emission of light from the fluorophore when excited by light of the correct wavelength in the thermal cycler. (4) The thermal cycler can quantify the fluorescent signal for further analysis, and PCR product is generated.

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 set up used for quantitation of all targets is shown in Table 2.21 and cycling parameters in Table 2.22. Each target was detected in triplicate within a single reaction plate and each reaction plate was duplicated (each target was detected a total of six times). Eight-point standard curves were generated for targets and endogenous controls from two-fold serial dilutions of *S. cerevisiae* genomic DNA (Stratagene) of concentration 100 μg/ml. Reaction thresholds were determined automatically by the 7500 v2.0.5 analysis software (Applied Biosystems).

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

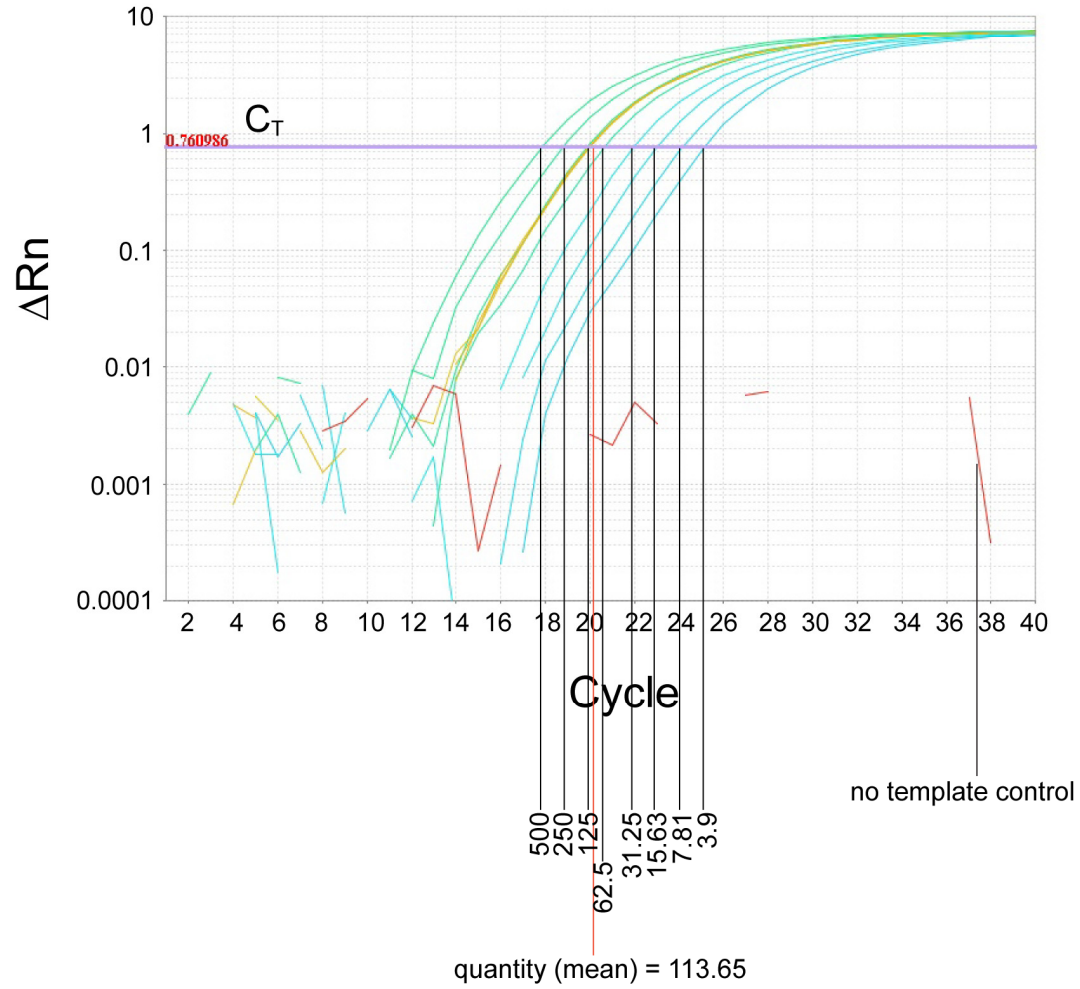
Table 2.21. Reaction setup for RT-qPCR experiments. The concentration of RNA was determined by UV absorbance at 260 nm. The A_{260}:A_{280} ratios for RNA samples was always between 1.8 and 2.1. For each target, a master mix was made and 20 μl aliquoted into individual wells of a PCR plate containing 5 μl sample RNA or genomic DNA standard.

Step	Temperature	Time
1	48°C (reverse transcription)	15 min
2	95°C (DNA polymerase activation)	10 min
3	95°C (melting)	15 s
4	60°C (annealing/extension)	1 min
5	Repeat 3-4, 40 cycles	

Table 2.22. Thermal cycling conditions for RT-qPCR experiments. 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.

2.10.2. *ACT1* Poly(A) site selection assay

The *ACT1* poly(A) site selection assay was designed to determine changes in *ACT1* cleavage/poly(A) site usage in cDNA derived from wildtype and mutant yeast strains. The reaction setup was as described in 2.10.1. Primer and probe choice was influenced by the previous identification of cleavage/poly(A) sites in the *S. cerevisiae ACT1* 3'-UTR using an RNase H cleavage/Northern blotting assay [153]. Primers and probes for quantification of target cDNA were designed to amplify only if cleavage/poly(A) sites



quantity (mean) *ACT1* 3' UTR normalised to *TDH2* =

113.65 ÷ 131.59* = 0.86

quantity (mean) relative to wildtype =

0.86 ÷ 0.92† = **0.93 or 93 % of wildtype expression**

* quantity (mean) *TDH2*

† quantity (mean) *ACT1 3' UTR* of wildtype normalised to *TDH2*

Figure 2.6. Worked example of the relative standard curve method. The fluorescence from the reporter is normalised (Rn) to the signal of a passive reference. The baseline-corrected normalised reporter (ΔRn) is calculated for each PCR cycle as follows: ΔRn (cycle) = Rn (cycle) – Rn (baseline). The quantity of target in the sample (yellow curves) is calculated by comparing the point at which the reporter signal in the sample crosses the C_T (purple line), with the C_T values of the genomic DNA standards (cyan curves) of known concentration. The calculations show how to derive the relative quantity of target in a mutant sample compared to a wildtype sample.

were intact (see target design in Figure 2.7A). The mean quantities of target cDNA were normalised to the mean quantity of *TDH2* cDNA, and the normalised quantities of each target were used to calculate the frequency at which each poly(A) site had been processed (this calculation is summarised in a worked example presented in Figure 2.7B).

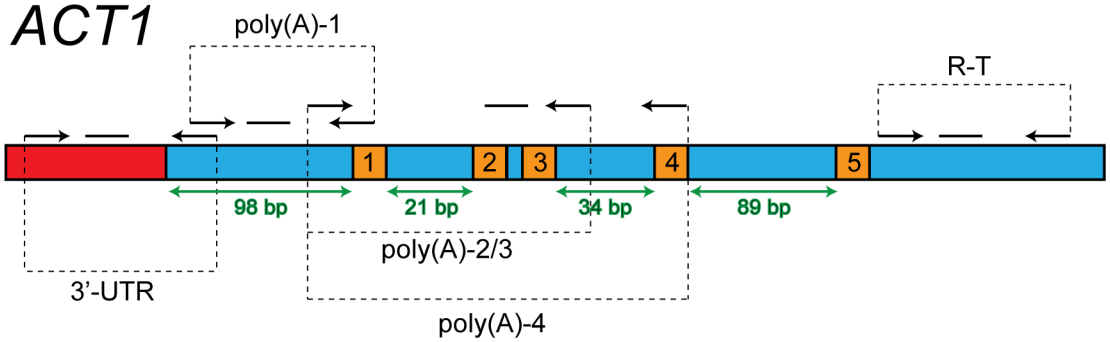
2.10.3. Analysis of *TDH2, ACT1, ADH1, CYC1* and *YPT1* transcription

Changes in the quantities of *TDH2, ACT1, ADH1, CYC1* and *YPT1* cDNAs between mutant and wildtype yeast strains (and therefore changes in the expression of these genes) were determined by RT-qPCR. Primers and probes for quantification of target cDNA were designed to amplify sequences within the ORFs or overlapping with 3'-UTRs of cDNA (but before poly(A) sites, see Appendix 8.4 for primer and probe sequences). The reaction setup was as described in section 2.10.1. Data was analysed by the relative standard curve method: quantities of mutant cDNA target were expressed relative to wildtype cDNA target.

2.10.4. *ACT1* Transcription readthrough assay

Changes in the levels of readthrough transcription (and therefore transcription termination efficiency) on the *ACT1* gene between mutant and wildtype yeast strains were determined by RT-qPCR. The locations of primers and probes for quantification of total *ACT1* cDNA and *ACT1* readthrough cDNA are shown in Figure 2.7A. The reaction setup was as described in 2.10.1. Data was normalised to the mean quantity of *TDH2* cDNA and analysed by the relative standard curve method: within a sample, the quantity of readthrough cDNA (R-T target) was expressed relative to the quantity of total *ACT1* cDNA (3'-UTR target).

A



B

normalised quantity mean 3'-UTR = 0.669
normalised quantity mean poly(A)-1 = 0.114
normalised quantity mean poly(A)-2/3 = 0.098
normalisedquantity mean poly(A)-4 = 0095

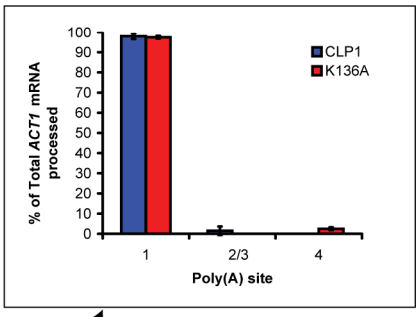
Amount of mRNA processed at each site obtained from difference in quantities of cDNA at adjacent sites

e.g. Amount of processing at site 1 =

quantity 3'-UTR - quantity poly(A)-1 =

0.669-0.114 = 0.555

Expressed as a percentage of Total processed *ACT1* mRNA =



Poly(A) site	Amount of processing	% of Total processed ACT1 mRNA
1	0.555	97
2/3	0.016	3
4	0.003	0
Total	0.574	

Figure 2.7. (A) Design of targets for RT-qPCR poly(A) site selection and transcription readthrough experiments. In the *ACT1* poly(A) site selection assay, targets were designed to overlap specific poly(A) sites: amplification was only possible where poly(A) sites were intact. Targets were designed to quantify the amounts of three intact poly(A) sites: 1, 2/3 and 4 (sites 2 and 3 are in such close proximity they were treated as a single site). In the *ACT1* transcription readthrough assay, the detected quantities of R-T and 3'-UTR amplicons were compared to quantify the level of readthrough transcription. All targets are labelled and indicated by dashed black lines. 3'-UTR, target to quantify gene expression; poly(A)-1/2/3/4, target to quantify intact poly(A) sites; R-T, target to quantify readthrough transcription. Primers are displayed as arrows, probes as solid black lines. Orange boxes represent poly(A) sites, blue boxes 3'-UTR sequence and red boxes the 3'-ends of genes. Distances between poly(A) sites are indicated in green. (**B**) Worked example of calculating poly(A) site usage from RT-qPCR data generated from a wildtype *CLP1* yeast-derived total RNA sample. Mean cDNA quantities in the left box are normalised to the mean cDNA quantity of *TDH2*. Poly(A) site selection can be compared between wildtype and mutants using histograms (right box).

3. *in vitro* analysis of recombinant Clp1-ATP binding mutants

3.1. Introduction and overview

Clp1 is a subunit of the cleavage factor IA complex (CFIA), an essential multi-subunit protein complex required for the cleavage and polyadenylation of pre-mRNAs in *S. cerevisiae* (see sections 1.3 and 1.3.4). Clp1 is associated with the CFIA complex only through its interaction with Pcf11 [99], but has also been shown to bind to several subunits of Cleavage and Polyadenylation Factor (CPF) including Ysh1, the proposed endonuclease that effects pre-mRNA cleavage [103]. The crystal structure of a Clp1-Pcf11 heterodimer is available, and shows that Clp1 is a three-domain protein that binds ATP at a nucleotide-binding site located within its large central domain (CD, see section 1.3.4 and Figure 1.5) [116]. The Clp1 CD is a member of the SIMIBI class of NTPases and contains three characteristic features: a P-loop/Walker A motif together with Switch I and Switch II motifs. These motifs are arranged around the ATP molecule with the majority of the Clp1-ATP interaction occurring between the Clp1 P-loop residues and the phosphate groups of ATP (see section 1.3.4 and Figure 1.6), although the interaction of Clp1 with ATP is also mediated by coordination of a magnesium ion by Switch I residues and a water molecule. The arrangement of the nucleotide-binding pocket observed in the crystal structure suggests that the putative catalytic residue is of insufficient proximity to initiate catalysis of ATP. In addition, biochemical experiments have failed to detect that any other nucleotide species are bound by recombinant Clp1, implying that ATP bound by Clp1 is stable to hydrolysis [116]. Despite these findings, it has been proposed that Clp1 possesses catalytic activity that is promoted by a conformational change induced by the interaction of Clp1 with an unidentified binding partner [116]. It has also been proposed that the orientation of ATP in the nucleotide-binding pocket may modulate the interaction of Clp1 with Pcf11 and that this is a method of communication between 3'-end processing and transcription machineries [121, 122]. Despite these theories, the role of the Clp1-ATP interaction remains enigmatic.

In order to understand the importance of the Clp1-ATP interaction, conserved residues in the Clp1-ATP binding pocket were mutated and the properties of the resultant recombinant mutant proteins were analysed in a series of hydrodynamic, biochemical

and biophysical experiments. These aimed to determine which mutations uncoupled ATP from Clp1, the effects of Clp1-ATP uncoupling on protein stability and secondary structure, and whether Clp1-ATP binding was also important for the Clp1-Pcf11 interaction.

3.2 Mutation of the Clp1-ATP binding pocket affects protein stability

3.2.1. Design of Clp1 mutants

To study the importance of Clp1-ATP binding, it was necessary to weaken ATP-binding or remove bound ATP completely and characterise any effects this had on purified recombinant Clp1. To achieve this, point mutations were introduced into the Clp1/Pcf11-EXP construct by SDM. Mutations were designed based upon interactions between Clp1 residues and ATP observed in the Clp1-Pcf11 crystal structure [116] (see section 1.3.4 and Figure 1.6), and from information about Clp1 residue conservation (see Figure 1.5C and Appendix 8.1). A full list of mutations is provided in Table 3.1. Mutation of P-loop residues and Lys321 in the CTD aimed to break the interaction of Clp1 with the triphosphate moiety of ATP. Switch II Asp251 was mutated to remove a hydrogen bond it shares with P-loop Thr137, a conserved interaction required for stability of the ATP-binding pocket. The crystal structure of Clp1 suggests Switch I Asp161 is not involved in ATP-binding or magnesium ion coordination. However, the equivalent Switch I residue in the NifH nitrogenase of *Azobacter vinelandii* is at the centre of the active site [116, 120]. It was therefore necessary to include Asp161 in the mutational analysis in the event that it makes an important contact with an extended water/magnesium ion network that may have been missed due to the limited resolution of the crystal structure. Clp1-ATP binding pocket mutants are hereafter referred to as Clp1-ATP mutants.

Residue	Location	Function	Mutated to	Mutant name
Lys136	P-loop	Binds ATP γ and β-phosphates	Glu	K136E
Thr137	P-loop	Binds ATP β-phosphate, ATP-binding pocket stability	Ala	T137A
Ser138	P-loop	Binds α-phosphate	Ala	S138A
Asp161	Switch I	Putative catalytic residue	Ala	D161A
Asp251	Switch II	ATP-binding pocket stability	Ala	D251A
Lys321	CTD	Binds α-phosphate	Ala	K321A

Table 3.1. Mutations made in the Clp1-ATP binding pocket in the Clp1/Pcf11-EXP construct.

3.2.2. Size-exclusion chromatography

SEC is a technique used to separate proteins by particle size and shape. For folded globular proteins, particle size is likely to increase in proportion with molar mass. Therefore SEC can be used as a relatively low-resolution method to distinguish proteins of different molar mass, oligomeric states and the formation of protein complexes such as heterodimers.

SEC was used as a purification step for recombinant Clp1-Pcf11(454-563) (hereafter Pcf11 refers to Pcf11(454-563)). When comparing the SEC profiles from different protein preparations it became evident that the hydrodynamic properties of some mutants had changed significantly compared to wildtype. As a result, the final step for purification of recombinant proteins (by SEC) also served as a first step in analysing the effects of mutations in the Clp1-ATP binding pocket.

The SEC profile of Clp1-Pcf11 wildtype showed a single peak with a retention volume of ~74 ml from a Superdex 200 (16/60) column (Figure 3.1A). Analysis of the protein composition of this peak by reversed-phase high-performance liquid chromatography (RP-HPLC) revealed two proteins with column retention times of 17 and 32 minutes (Figure 3.1B, both panels). SDS-PAGE of the RP-HPLC-eluted proteins showed the 17-minute peak contains a 15 kDa protein, slightly larger than the expected size of 12 kDa for Pcf11(454-563) (Figure 3.1C). The 32-minute peak contains a 50 kDa protein which corresponds to the expected size of 51.7 kDa for Clp1 (Figure 3.1C). The molecular weights of both proteins were verified by mass spectrometry (performed by Dr. S. Howell (NIMR)) and mass spectrum deconvolution reports are presented in Appendix 8.5). This confirmed that the 17 minute RP-HPLC peak contained protein with a molar mass of 12017.5 Da, which is equal to the calculated molar mass of a Pcf11 polypeptide containing residues 454-561. This is two residues shorter than the Pcf11 454-563 polypeptide encoded by the Clp1/Pcf11-EXP construct, and is probably the result of a C-terminal cleavage event in *E. coli* that shortens the polypeptide by two residues. Mass spectrometry of the 17 minute peak also revealed two smaller proteins of 5794.5 kDa and 6239.6 kDa (see Appendix 8.5). These proteins are Pcf11 residues 454-508 and 509-561 respectively. The amino acids at positions 508 and 509 are Asp and Pro respectively. The peptide bond between these amino acids is particularly acid labile, and it is likely that a significant amount of Pcf11 454-561 was

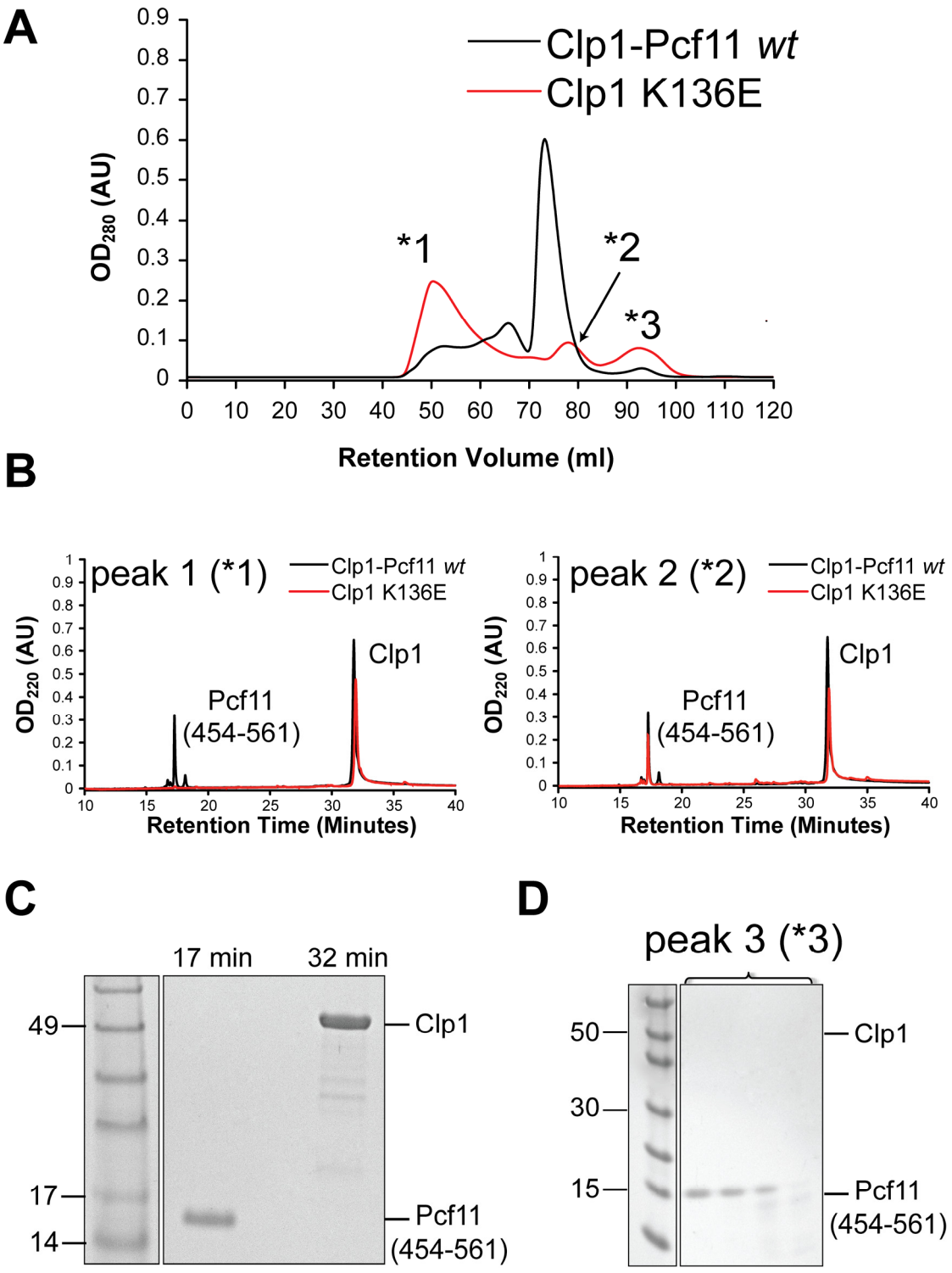


Figure 3.1. SEC of Clp1 and Clp1-ATP mutants. (**A**) Comparison of Clp1 wildtype (black line) and Clp1 K136E (red line) SEC profiles. The three separate peaks from K136E SEC are labelled with asterisks and numbered. (**B**) Comparison of SEC peak compositions determined by RP-HPLC for Clp1 wildtype (black lines), K136E peak 1 (red line, left panel) and K136E peak 2 (red line, right panel). (**C**) SDS-PAGE characterisation of protein peaks from the Clp1-wildtype RP-HPLC experiment. The retention times of the collected RP-HPLC peaks are indicated above the gel lanes. (**D**) SDS-PAGE of fractions eluted in K136E peak 3.

cleaved between these residues through exposure to RP-HPLC conditions (60 ºC, 0.05% TFA) for a significant period prior to mass spectrometry. Mass spectrometry of the 32 minute RP-HPLC peak showed protein within it had a molar mass of 51777.3 Da (Appendix 8.5), which is equal to the calculated molar mass of full-length Clp1 with an intact C-terminal hexahistidine sequence (His-tag).

In contrast to wildtype, SEC of Clp1 K136E gave rise to three distinct peaks (Figure 3.1A). The first and largest peak eluted at ~45 ml, which is approaching the void volume of the SEC column (where particle size is larger than the pore size of the gel matrix, leading to particle exclusion). The second peak had a retention volume of ~76 ml, which corresponds well with the single wildtype Clp1-Pcf11 peak. The third peak eluted at ~92 ml, indicating the presence of a protein with a smaller molar mass than wildtype. RP-HPLC of a sample from Clp1 K136E peak 1 revealed only the 32 minute Clp1 peak (Figure 3.1B, left panel), whereas RP-HPLC of K136E peak 2 displayed both the 17 minute Pcf11 peak and 32 minute Clp1 peak, indicating the Clp1 K136E peak 2 sample is a Clp1-Pcf11 complex (Figure 3.1B, right panel). The protein concentration of Clp1 K136E peak 3 was insufficient for RP-HPLC analysis, however an SDS-PAGE experiment revealed this peak contained only Pcf11 (Figure 3.1D). Similar to Clp1 K136E, SEC of Clp1 T137A also gave rise to three resolved peaks with retention times of ~45 ml, ~76 ml and ~92 ml (peaks 1-3 respectively, Figure 3.2), showing the T137A mutation also generates a different SEC profile of protein peaks compared to wildtype. The SEC profile of Clp1 D251A was also distinct from that of wildtype Clp1-Pcf11, with peaks at ~45 ml and ~92 ml, corresponding to peaks 1 and 3 observed in SEC experiments with the Clp1 K136E and T137A mutants (Figure 3.2). However, no peak corresponding to a Clp1-Pcf11 complex was observed with this mutant. The SEC profiles of S138A and K321A closely resembled the Clp1-Pcf11 wildtype SEC profile (Figure 3.2), with only single peaks (retention times of ~75 ml) revealing these protein samples contain only Clp1-Pcf11 complex.

SEC revealed that point mutation of Lys136, Thr137 and Asp251 in the Clp1-ATP binding pocket disrupted the Clp1-Pcf11 complex, but mutation of Ser138 and Lys321 had little or no effect. The disruption of the Clp1-Pcf11 complex by mutation of residues in the Clp1 P-loop was intriguing, as the Clp1-ATP binding pocket is located on the opposite face of the protein to the Clp1-Pcf11 interface (see Figure 1.5A),

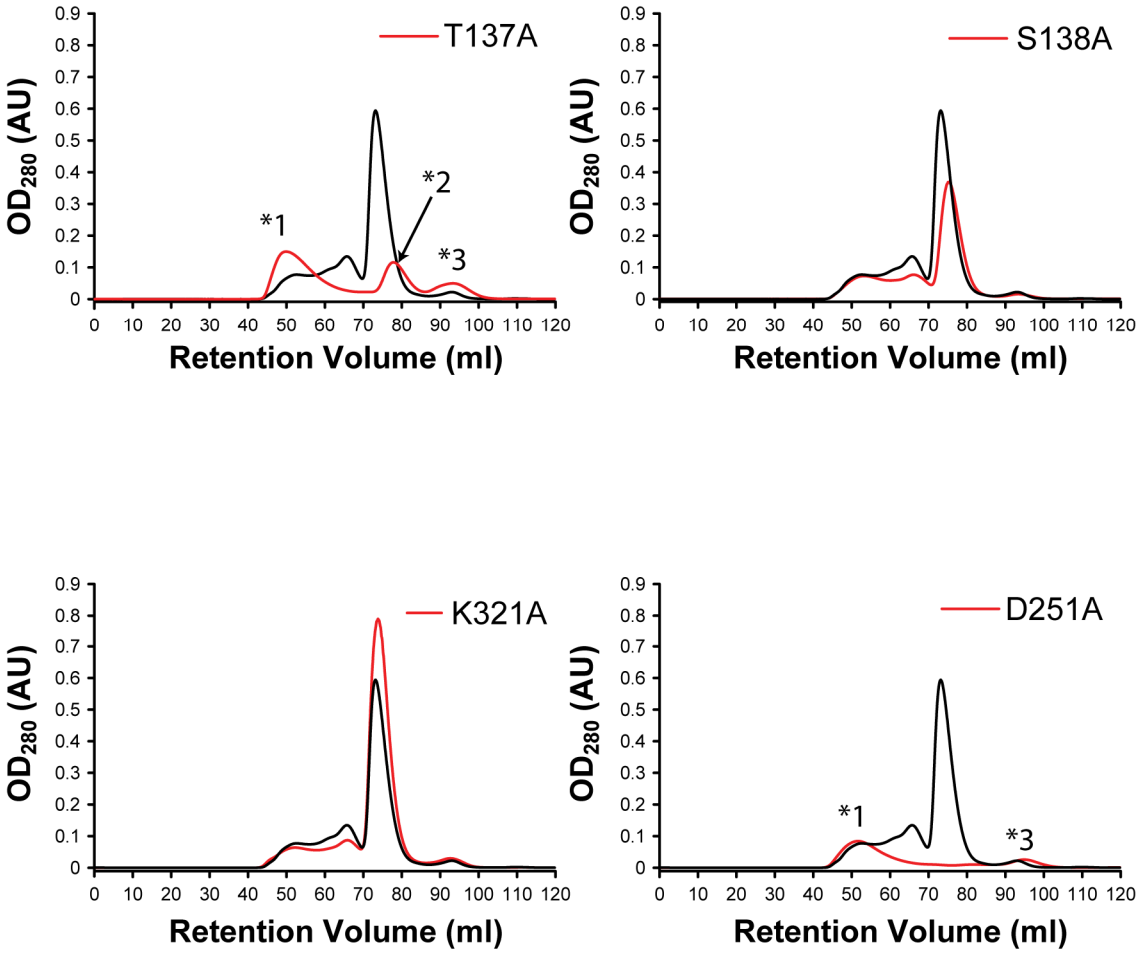


Figure 3.2. SEC profiles of Clp1-ATP mutants. Each mutant (red line) is labelled and the wildtype SEC profile (black line) is shown for comparison. Resolved peaks in SEC of T137A and D251A are labelled with asterisks and numbered.

providing no obvious reason as to why a mutation in one site should affect the function of the other. The fact that Clp1 eluted close to the void volume in the K136E and T137A SEC experiments suggested that these mutations render the Clp1-Pcf11 complex unstable and lead to the aggregation of Clp1.

3.2.3. Multiangle laser light scattering

MALLS is a technique that combines static light scattering and protein concentration measurements to determine the absolute molar mass of molecules in solution. It can be used to determine the molar masses and oligomerisation states of proteins, and when combined with size-exclusion chromatography (SEC-MALLS), to assess the distribution of different molar mass species within chromatographic peaks. If the protein within a chromatographic peak is of a single molar mass the distribution is said to be "uniform" or narrow. However, if a peak contains multiple protein species of different molar masses the distribution is "non-uniform" and can be very broad [211-213]. Thus a non-uniform mass distribution of a single protein is indicative of protein self-association or aggregation. SEC experiments revealed that mutation of the Clp1-ATP binding pocket likely affects protein stability. To further probe the effects of mutations on protein stability, SEC-MALLS experiments were conducted to investigate whether mutations affected the oligomerisation state, or caused aggregation of Clp1-Pcf11.

In a SEC-MALLS experiment with SEC-purified wildtype Clp1-Pcf11, the signal from the differential refractive index (dRI) detector showed a peak eluted from a Superdex 200 (10/30) SEC column with a retention time of ~26 minutes (Figure 3.3). Combining the dRI and light scattering (LS) data from this peak enabled calculation of the molar mass of protein within it and an assessment of the distribution of different molar mass species. Protein in this peak showed a narrow distribution of molar mass (0.3% of the weight-averaged molar mass indicative of a uniformly-distributed protein species) and analysis of Debye plots gave a weight-averaged molar mass of 60.2 kDa, close to the expected mass of 63.8 kDa for a Clp1-Pcf11 heterodimer (Figure 3.3). SEC-MALLS analysis of the Clp1 S138A and K321A mutants also revealed single chromatographic peaks that eluted with retention times of ~26 minutes. These peaks both contained a narrow molar mass distribution and gave derived molar masses of 62.4 kDa and 61.6 kDa respectively, consistent with the molar mass of a Clp1-Pcf11 heterodimer. This

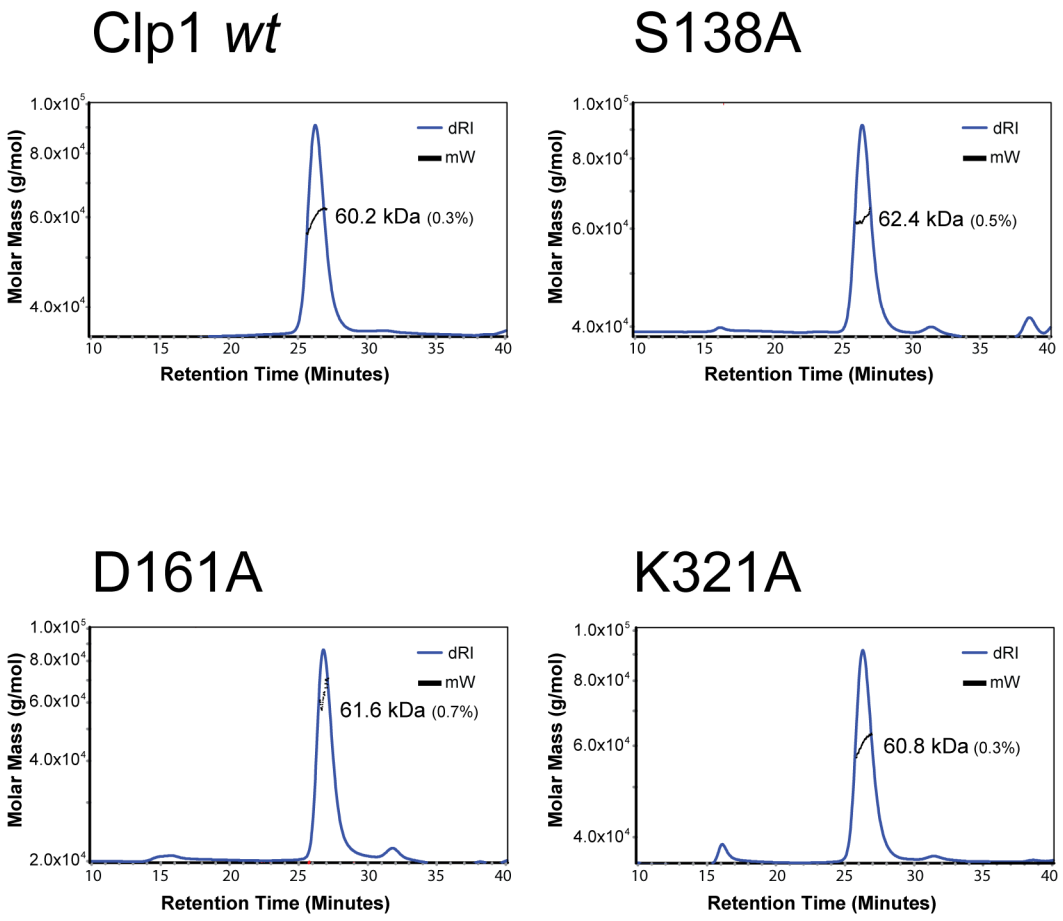


Figure 3.3. MALLS analysis of Clp-ATP mutants. The S138A, D161A and K321A mutants eluted from SEC with similar profiles to Clp1 wildtype (see Figure 3.2). The chromatograms recorded by the differential refractometer (dRI, blue line) for each sample are shown. Black dots are individual measurements of the molar mass of material recorded at 1 s intervals throughout elution of chromatographic peaks. Each dot represents the molar mass determined from an individual Debye plot and the combination and averaging of these measurements was used to determine the overall weight-averaged molar mass of protein in a peak. The derived molar masses are displayed next to peaks along with a percentage error associated with the averaging that provides a measure of the mass spread within each peak.

was in agreement with the findings of SEC experiments that the retention times of S138A and K321A protein peaks corresponded with the retention time of a Clp1-Pcf11 complex (see section 3.2.2 and Figure 3.2). A further Clp1 mutant, Clp1 D161A (Switch I point mutation) was also included in the SEC-MALLS analysis. This D161A sample also eluted in a single peak with a retention time of ~26 minutes, had a uniform distribution and a derived molar mass of 60.8 kDa, close to that of a Clp1-Pcf11 complex (Figure 3.3). In contrast to these results, the first SEC-purified peaks (peak 1) from K136E, T137A and D251A mutants eluted with retention times of ~16 minutes, which is around the void volume of the SEC column (Figure 3.4). In these samples the protein was excluded from the column because of the high molar mass, greater than 1.0 x 10^6 Da. Moreover, individual Debye plots generated at short intervals throughout the elution of the excluded peaks in K136E (peak 1), T137A (peak 1) and D251A (peak 1) samples showed molar mass distributions that varied by up to an order of magnitude, signifying Clp1 in these peaks was highly heterogeneous and in a self-associated or aggregated state. SEC-MALLS experiments with the second SEC-purified peaks (peak 2) of Clp1 K136E and T137A revealed these samples also contained a significant proportion of material excluded from the SEC column with high molar mass and large non-uniform molar mass distribution. However, the K136E (peak 2) and T137A (peak 2) protein samples also contained four additional poorly resolved chromatographic peaks not observed in SEC-MALLS experiments with wildtype Clp1-Pcf11 and other mutants (Figure 3.4, peaks labelled A-D). The dRI signals showed the amount of protein eluting in these peaks was far greater than that in the excluded protein peaks (~16 minute peaks), indicating that the majority of Clp1 in the K136E (peak 2) and T137A (peak 2) samples was not aggregated. This is in stark contrast to the distribution and concentration of protein peaks in the K136E (peak 1) and T137A (peak 1 samples), in which Clp1 eluted solely in the void volume of the SEC column. The LS signals of the additional A-D peaks observed in the K136E (peak 2) and T137A (peak 2) samples were decreased compared to that of the excluded aggregated peaks, which suggested the A-D peaks contained protein of lower particle size that permeated the SEC matrix. The calculated molar masses of the 'B' peaks in K136E (peak 2) and T137A (peak 2) samples were 59.2 and 62.5 kDa respectively, revealing that these samples still contain intact Clp1-Pcf11 complex (Figure 3.4). The calculated molar masses of the 'A' peaks in these samples were 95.1 and 98.1 kDa respectively (Figure 3.4), close to the expected molar mass of two Clp1 molecules (a Clp1 homodimer, 102.4 kDa), and could be

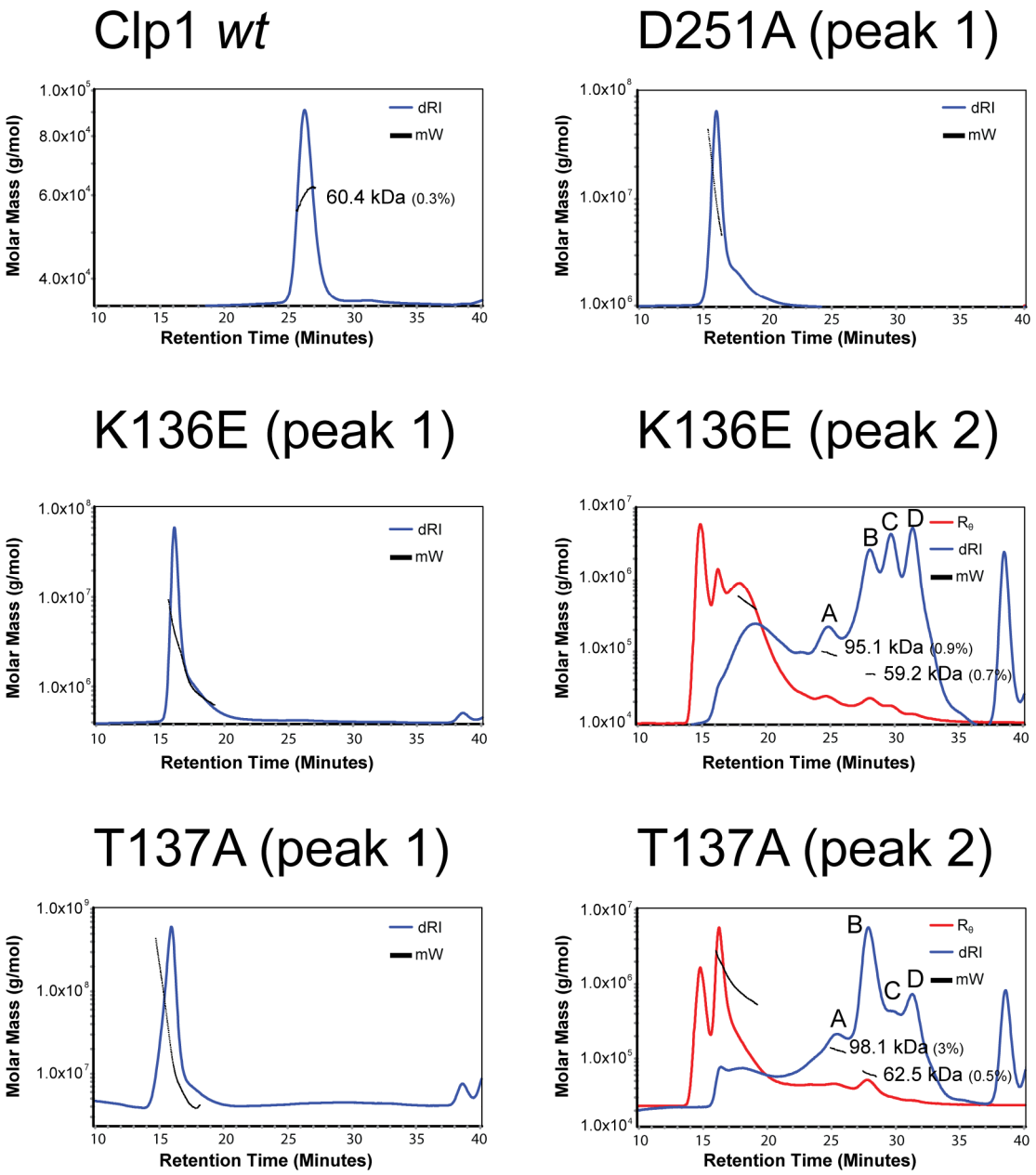


Figure 3.4. MALLS analysis of Clp-ATP mutants. The K136E, T137A and D251A mutants eluted from SEC with profiles distinct from Clp1 wildtype (see Figures 3.1 and 3.2). The chromatograms recorded by the light scattering (R_θ) and differential refractive index detectors (dRI) are shown in red and blue respectively. Peaks A-D that are referred to in the text are labelled. Black dots are individual measurements of the molar mass. The overall weight averaged molar mass and percentage errors are derived and displayed as in Figure 3.3.

representative of the beginning of Clp1 aggregation that may only occur after loss of the Pcf11 fragment. The molar masses of proteins in the 'C' and 'D' peaks observed in SEC-MALLS experiments with the Clp1 K136E (peak 2) and T137A (peak 2) samples could not be determined due to limitations of low protein concentration and low scattered light intensities. If the aggregation of Clp1 is preceded by loss of Clp1-Pcf11 binding, it is likely that one of the 'C' or 'D' peaks could contain released Pcf11. SEC-MALLS analysis of the protein content of Clp1 K136E (peak 3), T137A (peak 3) and D251A (peak 3) was also not possible due to low protein concentrations and low scattered light intensities, although previous SDS-PAGE gel analysis showed these samples are comprised only of free Pcf11 (Figure 3.1D).

SEC-MALLS experiments showed that mutation of Lys136, Thr137 and Asp251 residues within the Clp1-ATP binding pocket causes instability of Clp1-Pcf11 complexes and results in the aggregation of recombinant Clp1. The possible identification of an intermediate Clp1 species of ~100 kDa, points to the formation of a Clp1 homodimer. This is likely a precursor to larger Clp1 aggregates and also suggests that the loss of Clp1-Pcf11 binding precludes Clp1 instability. Misfolding of protein into an incorrect conformation often causes aggregation. In the case of the Clp1 K136E, T137A and D251A, it appears that mutations in the Clp1-ATP binding pocket that reduce Clp1-ATP binding have dramatic effects on Pcf11-binding and overall protein stability. One possibility is that the Clp1-ATP pocket mutations directly affect Clp1 secondary structure leading to misfolding. Another possibility is that ATP-binding is intimately linked to Pcf11-binding and that loss of Pcf11 causes the instability of Clp1. These ideas are examined in the proceeding sections

3.2.4. Far-UV circular dichroism spectroscopy

Circular dichroism (CD) spectroscopy measures the differential absorption of left- and right-handed circularly polarised light by chiral molecules. In the far-UV region, CD can give information about the secondary structure of proteins. Far-UV CD spectra of Clp1 mutants were recorded to determine whether protein instability observed by SEC-MALLS was associated with changes in Clp1 secondary structure.

The wildtype Clp1-Pcf11 CD spectrum displayed negative CD peaks at ~208 nm and ~222 nm (Figure 3.5A), characteristic of a mixed α/β protein and in agreement with the

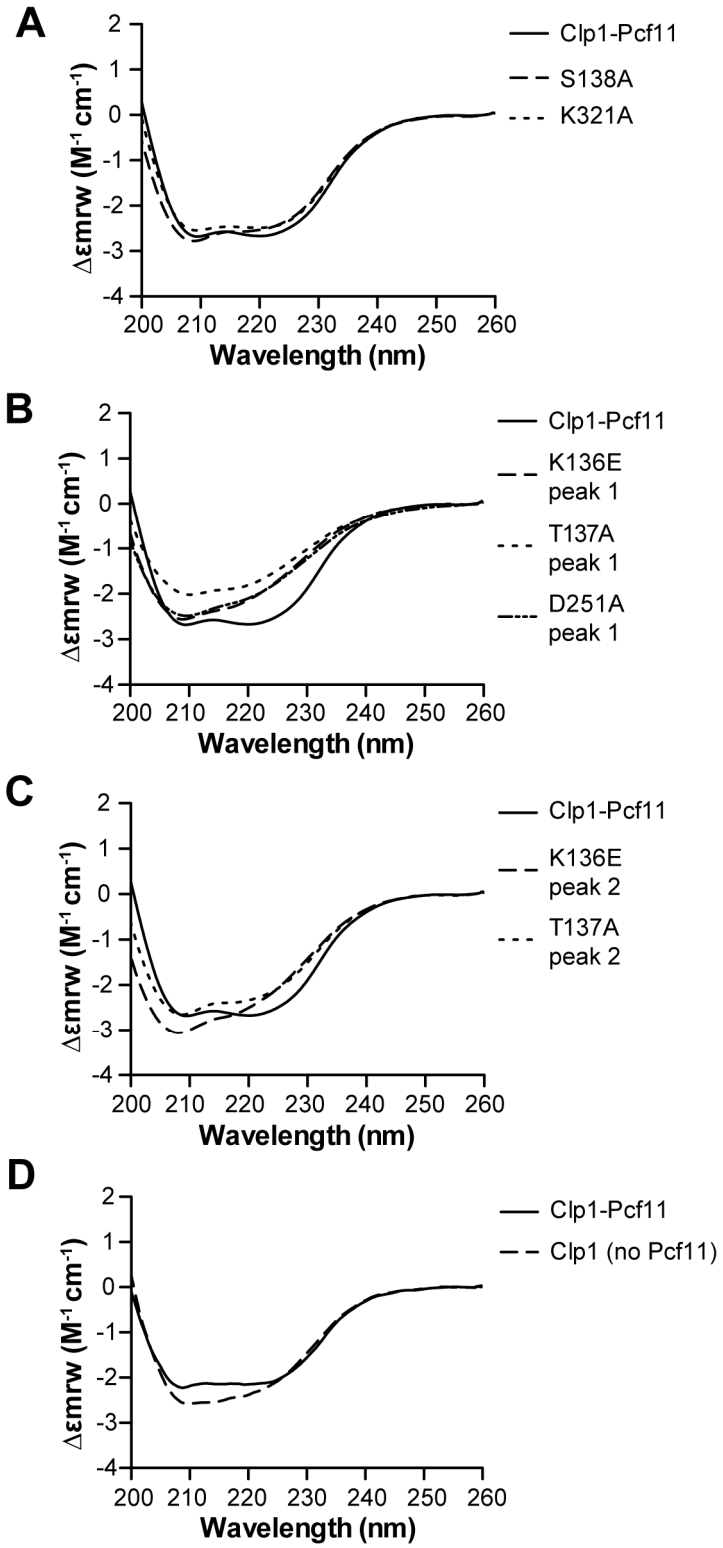


Figure 3.5. Far-UV CD of Clp1-ATP mutants. (**A**) CD spectra of mutants that have similar hydrodynamic properties to wildtype, (**B**) CD spectra of samples from mutants containing aggregated Clp1 and (**C**) CD spectra of samples from mutants that contain Clp1-Pcf11 heterodimers, but with wide molar mass distributions. (**D**) Comparison of the CD spectra recorded from Clp1 and the Clp1-Pcf11 complex. The spectra are expressed in mean residue weight differential molar extinction and lines corresponding to individual CD spectra are identified in the key on the right hand side of each set of spectra.

Clp1-Pcf11 crystal structure (see Figure 1.5A). The S138A and K321A mutants had CD spectra indistinguishable from wildtype, indicating they are also mixed α/β proteins with identical secondary structures to wildtype Clp1-Pcf11 (Figure 3.5A). In contrast, the CD spectra of K136E (peak 1), T137A (peak 1) and D251A (peak 1) overall had diminished amplitude compared to Clp1-Pcf11 and a pronounced reduction in negative CD at ~222 nm (Figure 3.5B). This likely reflects a partial loss in secondary structure and was also observed in the CD spectra of K136E (peak 2) and T137A (peak 2) (Figure 3.5C). One possibility is that the diminished amplitude of K136E, T137A and D251A CD spectra compared to wildtype were due to the loss of the Pcf11 fragment, as Clp1 and a Clp1-Pcf11 complex may give rise to significantly different CD spectra. To rule out the changes in CD spectra were not due to loss of the Pcf11 fragment, the CD spectra of a Clp1-Pcf11 complex and Clp1, shown to be monomeric by SEC-MALLS, were compared (Figure 3.5D). These data demonstrate that the CD spectrum of Clp1 actually displays a small increase in amplitude at 208-225 nm compared to the Clp1-Pcf11 spectrum, implying Clp1 has an increased secondary structure content compared to a Clp1-Pcf11 complex. To investigate the differences in the CD spectra, the CDpro program [197] was used to analyse and compare the Clp1-Pcf11 complex and free Clp1 secondary structure element compositions, the results of which are presented in Table 3.2. Secondary structure prediction revealed that the percentages of β-sheet and turn do not change significantly between Clp1-Pcf11 complex and free Clp1. However, The α-helix content of free Clp1 is higher than a Clp1-Pcf11 complex, at 37% compared to 27% (Table 3.2). This corresponds with an increased random coil content in the Clp1-Pcf11 complex over free Clp1, with 37% compared with 29%. These changes in secondary structure composition show it is likely that Pcf11 is largely unstructured, and that the differences in the bound and free CD spectra can be explained by the averaging that is used to obtain the mean residue weight CD. Thus, the CD of Clp1-Pcf11 at 208 and 225 nm is reduced because it is averaged to include the unstructured Pcf11 fragment. This finding is also consistent with results presented in chapter 5 that show a larger Pcf11 fragment containing two zinc-finger motifs is still largely unstructured (see section 5.2.3).

CD spectroscopy has demonstrated that residues in the Clp1-ATP binding pocket are important for the maintenance of Clp1 secondary structure. Combining these findings with the hydrodynamic characterisation of Clp1-ATP mutants, it is evident that

mutations cause loss of secondary structure that results in Clp1 instability and leads to protein aggregation. It is possible that loss of Clp1 secondary structure could be the direct result of amino acid substitutions, or alternatively it could be indirectly caused through alteration of Clp1 interaction with ATP.

	α-helix (%)	β-sheet (%)	Turn (%)	Random coil (%)
Clp1-Pcf11 complex	26.87 (1.44)	17.63 (1.32)	18.10 (2.13)	37.4 (3.67)
Clp1	36.75 (2.48)	15.90 (1.84)	18.28 (1.62)	29.07 (2.71)

Table 3.2. Prediction of Clp1-Pcf11 and Clp1 secondary structure element compositions (using the CDpro program [197].

3.3. ATP is a structural component of Clp1

3.3.1. Mutations in Clp1 P-loop and Switch II abrogate Clp1-ATP binding

The hydrodynamic and biophysical characterisation of Clp1-ATP mutants revealed that some mutations cause loss of secondary structure and protein instability. To determine whether protein instability correlated with a loss of Clp1-ATP binding, the nucleotide contents of Clp1-ATP mutants were analysed by RP-HPLC nucleotide assays.

A previous study demonstrated that ATP was the only nucleotide extracted from recombinant Clp1 expressed and purified from *E. coli* [116]. In agreement with this, the extracted nucleotide from wildtype Clp1-Pcf11 had a column retention time of ~8 minutes, which was consistent with the retention time of an ATP nucleotide standard (Figure 3.6A). A UV absorbance spectrum recorded from the eluted nucleotide showed a maximal absorbance at 260 nm, confirming the nucleotide had an adenine base (Figure 3.6A, inset). Importantly, no ADP was detected suggesting that bound ATP is stable to hydrolysis. Nucleotide assays with samples extracted from Clp1 S138A, D161A and K321A mutants showed these proteins also copurify from *E. coli* with a single nucleotide species that has a retention time of ~8 minutes consistent with ATP. Contrary to this, nucleotide assays with samples extracted from Clp1 K136E (peaks 1 and 2), T137A (peaks 1 and 2) and D251A (peak 1) mutants did not detect any bound nucleotides (Figure 3.7), signifying these mutants do not copurify with bound ATP.

The results of nucleotide assays show that mutation of Clp1 Lys136, Thr137 and Asp251 abrogates Clp1-ATP binding, but Ser138, Asp161 and Lys321 mutations do not

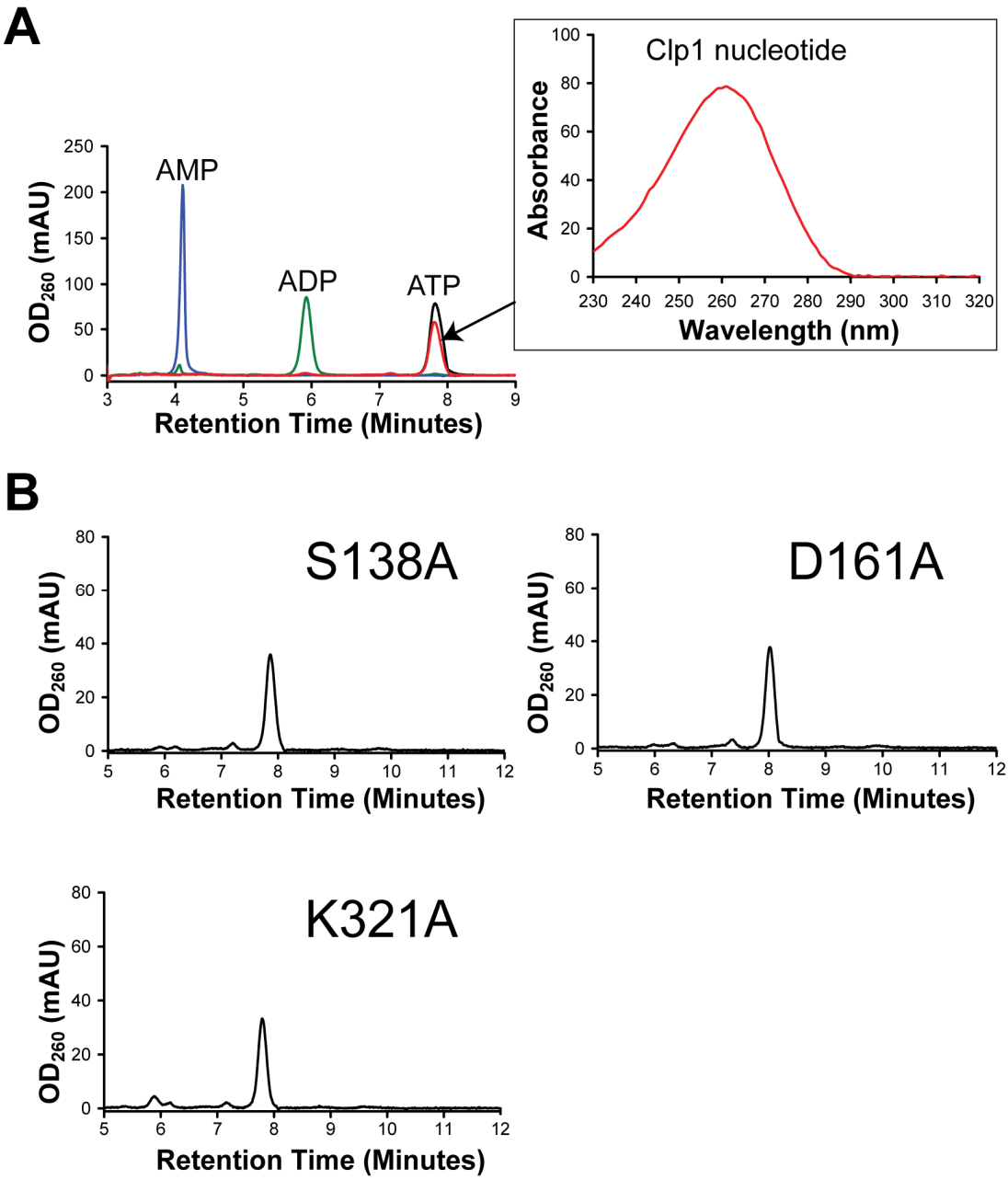


Figure 3.6. Analysis of the nucleotide content of Clp1-Pcf11 wildtype and stable Clp1-ATP binding mutants. (**A**) Comparison of the retention time of the Clp1-extracted nucleotide (red line) with nucleotide standards AMP (blue), ADP (green) and ATP (black). A UV-absorbance spectrum recorded at the time of elution of the Clp1-extracted nucleotide peak is shown (inset). (**B**) Chromatograms from nucleotide assays with samples extracted from Clp1 S138A, D161A and K321A mutants (labelled above chromatograms).

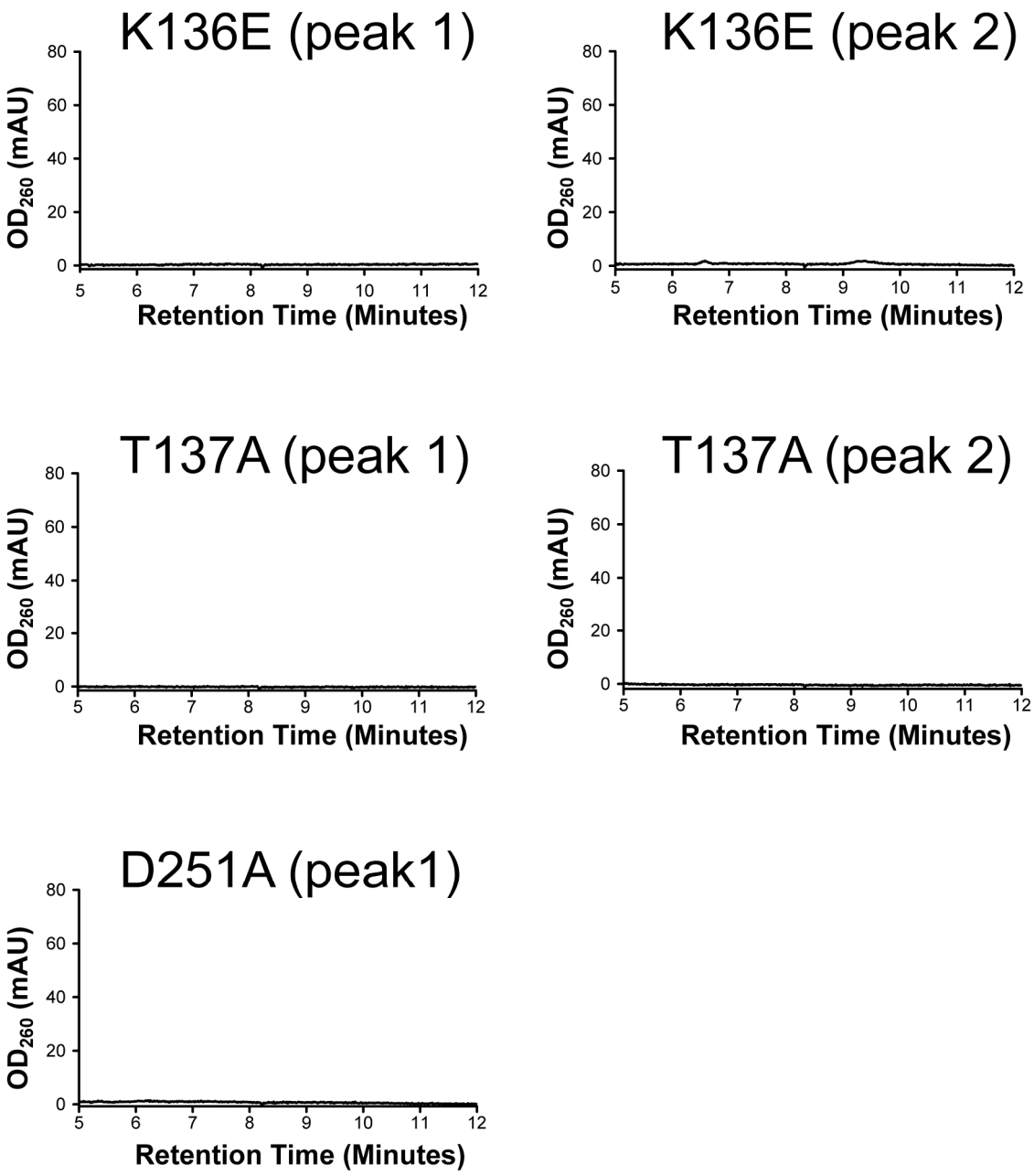


Figure 3.7. Nucleotide assays with unstable Clp1-ATP binding mutants. Each mutant sample is labelled above the relevant chromatogram. Compare with the retention times of nucleotide standards in Figure 3.6A.

affect Clp1-ATP binding. These findings demonstrate that the instability and aggregation of Clp1 is linked directly to the abolishment of Clp1-ATP binding. In light of these findings, it is probable that the Clp1-Pcf11 interaction is dependent on Clp1-ATP binding, as without ATP Clp1 cannot maintain the necessary conformation for its interaction with Pcf11. On the other hand it still remains possible that the K136E, T137A and D251A mutations are the cause of structural changes in Clp1, with loss of ATP a consequence.

3.3.2. ATP is essential for Clp1 stability

Mutation of Clp1 at Lys136 and Thr137 in the P-loop and Asp251 in the Switch II motif abrogates Clp1-ATP binding. This loss of ATP corresponds strongly with a loss of the Clp1-Pcf11 interaction, a reduction in Clp1 secondary structure and protein instability. To ascertain that these effects were a result of uncoupling the Clp1-ATP interaction and not merely a consequence of mutation, it was necessary to characterise the effects of ATP removal on wildtype Clp1-Pcf11.

To remove bound ATP, the wildtype Clp1-Pcf11 complex was incubated with 20 mM EDTA for 12 h. EDTA is a chelating agent that is commonly used to remove metal ions from the active sites of enzymes. In the Clp1-Pcf11 crystal structure, there is a Mg^{2+} ion present in the ATP-binding pocket coordinated by the phosphate groups of the bound ATP molecule, a water molecule and the side-chains of surrounding amino acids (see Figure 1.6). Thus, disruption of this network of interactions by chelation of the Mg^{2+} ion should release ATP from the Clp1 nucleotide-binding site. In a SEC-MALLS experiment with untreated Clp1-Pcf11, two chromatographic peaks eluted from the SEC column with retention times of ~16 minutes and ~26 minutes respectively (Figure 3.8A). Protein in the ~16 minute peak was shown to be of high molar mass with a wide molar mass distribution, consistent with aggregated Clp1 observed in previous SEC-MALLS experiments (see section 3.2.3 and Figure 3.4). The ~26 minute peak contained a protein species with a calculated molar mass of 60.6 kDa, and a narrow molar mass distribution, characteristic of a Clp1-Pcf11 heterodimer. Integration of the UV absorbance peak areas at 280 nm allowed quantification of the amounts of protein in the different chromatographic peaks. The 26 minute peak contained 2.74 nmoles (+/- 0.3) of Clp1-Pcf11 complex in the untreated sample, whereas after EDTA treatment the amount of Clp1-Pcf11 complex was much lower at 1.29 nmoles (+/- 0.08) (this is also

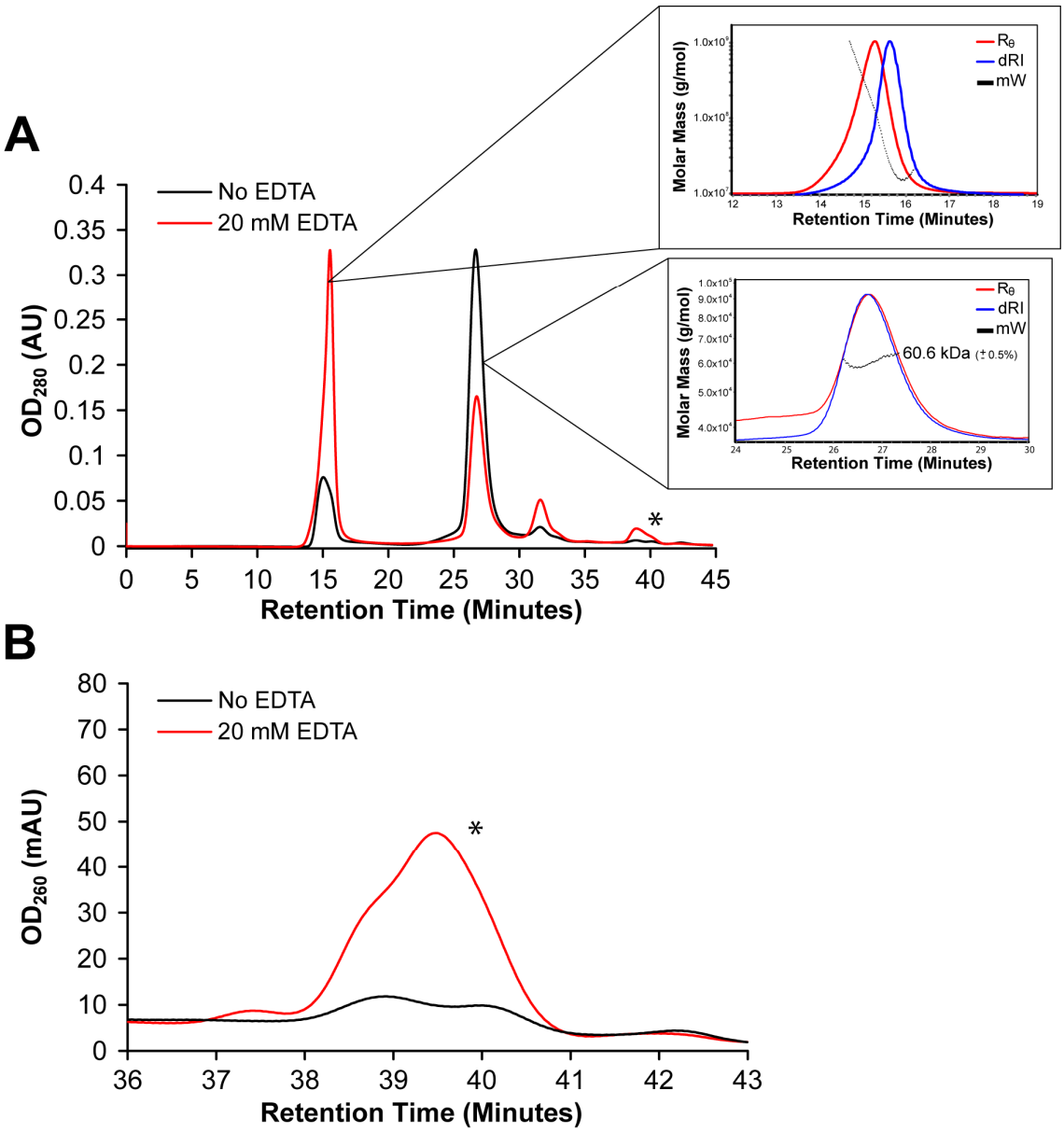


Figure 3.8. MALLS analysis of Clp1-Pcf11 stability after EDTA chelation of the Mg^{2+} ion in the Clp1-ATP binding pocket. (**A**) SEC profiles from MALLS experiments with Clp1-Pcf11 (black line) and a sample treated with 20 mM EDTA (red line). The MALLS chromatograms (inset) demonstrate protein aggregation in the 16 minute peak and the Clp1-Pcf11 heterodimer in the 26 minute peak. The asterisk indicates an additional peak observed at 39 minutes after EDTA treatment (**B**) The increased absorbance at 260 nm of the 39 minute peak after treatment with 20 mM EDTA (red line and asterisk), demonstrating the release of ATP from Clp1 after chelation of Mg^{2+} from the nucleotide binding site.

evident from the height of absorbance peaks at 280 nm, Figure 3.8A). The decrease in the amount of Clp1-Pcf11 complex after EDTA treatment also corresponded with a large increase in the amount of Clp1 in the 16 minute peak from 0.78 nmoles (+/-0.06) to 2.6 nmoles (+/- 0.2) (Figure 3.8A). The change in proportion of these peaks after EDTA treatment shows that EDTA chelation causes an increase Clp1 instability that is characterised by increased Clp1 aggregation. This was accompanied by the appearance of a peak with a retention time of ~39 minutes and a maximal UV absorbance at 260 nm, indicative of released ATP (Figure 3.8B). Furthermore, another small peak in UV absorbance at 280 nm was present in the SEC-MALLS experiment with untreated Clp1-Pcf11 (retention time ~32 minutes, Figure 3.8A), although the low protein concentration and scattered light intensity of this peak prohibited derivation of the molar mass of protein within it. After EDTA treatment, the UV absorbance peak area of this peak increased four-fold (illustrated well by the increase in peak height, Figure 3.8A). It is likely that the increase in the area of this peak denotes a rise in the amount of free Pcf11 that follows destabilisation of the Clp1-Pcf11 complex by abrogation of Clp1-ATP binding.

SEC-MALLS analysis of Clp1-Pcf11 demonstrated that chelation of Mg^{2+} by EDTA releases ATP from Clp1, causing Pcf11 dissociation, Clp1 instability and subsequent aggregation. The tendency for apo-Clp1 (ATP free Clp1) to aggregate made it impossible to characterise the affinity of the Clp1-ATP interaction by direct titration. However, because Clp1 stability depends on ATP, it was hypothesised that perhaps Clp1 is constitutively ATP-bound. To test this hypothesis, attempts were made to exchange the ATP bound by Clp1 (in a Clp1-Pcf11 complex) with radiolabelled ATP (ATP [γ-^{32}P]). Here, Clp1-Pcf11 was incubated with increasing amounts of ATP [γ-^{32}P], and then the free-ATP removed by desalting prior to scintillation counting. The results of these exchange experiments are presented in Table 3.3 and details of the experimental method are presented in section 2.6. Upon incubation of ATP [γ-^{32}P] with Clp1-Pcf11 at a 2:1 (ATP [γ-^{32}P]: protein) molar ratio, the radioactive signal detected by scintillation counting of the protein fraction after desalting was 3,670 counts per minute (CPM). This corresponds to the background level of ATP [γ-^{32}P] observed in control experiments with no Clp-Pcf11, far less than the expected radioactive signal of 16,001 CPM for a sample in which 100% of ATP in Clp1 is exchanged. This result showed that a 2:1 (ATP [γ-^{32}P]:protein) molar ratio was not sufficient to promote the exchange

of ATP bound in Clp1 with ATP [γ-^{32}P]. At the higher molar ratio of 5:1 (ATP [γ-^{32}P]:protein), ATP exchange was also not observed as the radioactive signal of 6,062 CPM was actually lower than the background level and far below the expected signal of 20,697 CPM for 100% ATP exchange. A final attempt to exchange ATP bound by Clp1 was made by incubating ATP [γ-^{32}P] with Clp1-Pcf11 at a 10:1 (ATP [γ-^{32}P]:protein) molar ratio. After the incubation period and removal of free-ATP, the radioactive signal was 10,263 CPM. This was lower than the background signal and not close to the expected signal of 33,055 CPM for 100% ATP exchange indicating no ATP had been exchanged.

The results of this assay designed to exchange the ATP bound in Clp1 with radiolabelled ATP shows that ATP is "non exchangeable" under the conditions tested and that ATP is likely very tightly bound in the Clp1 nucleotide-binding pocket. The fact that ATP was not exchanged can be explained by details provided by the Clp1-Pcf11 crystal structure (see section 1.3.4 and Figure 1.5B) [116]. In the structure the Clp1 nucleotide-binding pocket is completely occluded by the packing of the Clp1 NTD against the Clp1 CD, and this lack of access makes it highly unlikely that free ATP can exchange with that already bound. Taken with the results of the SEC-MALLS experiment with EDTA-chelated Clp1-Pcf11, it is likely that Clp1 is constitutively bound to ATP, which maintains the stability of Clp1 and facilitates the Clp1-Pcf11 interaction.

ATP [γ-^{32}P]:protein molar ratio	**Total ATP [γ-^{32}P] CPM**	**Background ATP [γ-^{32}P] CPM**	**Expected ATP [γ-^{32}P] CPM with 50 μM Clp1-Pcf11**	**Actual ATP [γ-^{32}P] CPM with 50 μM Clp1-Pcf11**	**% ATP exchanged**
2:1	24,852	3,575	16,001	3,670	0
5:1	65,750	8,271	20,697	6,062	0
10:1	122,552	20,629	33,055	10,263	0

Table 3.3. The ATP molecule bound by Clp1 is non-exchangeable. To try to exchange bound ATP, Clp1-Pcf11 samples were incubated with increasing molar ratios of ATP [γ-^{32}P]. Expected ATP [γ-^{32}P] CPM represents the expected radioactive signal if 100% of the cold ATP bound by Clp1-Pcf11 is exchanged with radioactive ATP [γ-^{32}P]. The results showed ATP [γ-^{32}P] was never detected above background levels, indicating no ATP was exchanged (highlighted red).

3.4. Summary and conclusions

Hydrodynamic characterisation of recombinant Clp1-ATP mutants by SEC and SEC-MALLS showed that in the Clp1 K136E and T137A P-loop mutants, and Clp1 D251A Switch II mutant, the Clp1-Pcf11 interaction is uncoupled causing protein instability and subsequent aggregation of Clp1. Differences in the far-UV CD spectra of these mutants compared to wildtype Clp1-Pcf11 revealed a reduction in secondary structure was also associated with Pcf11 dissociation. Reversed-phase nucleotide assays demonstrated that the Clp1 K136E, T137A and D251A mutants do not contain bound ATP, providing a link between loss of Clp1-ATP binding and abolishment of the Clp1-Pcf11 interaction. This link was confirmed when it was shown that EDTA chelation of the Mg^{2+} ion and subsequent loss of ATP from the Clp1-nucleotide binding pocket directly caused dissociation of Pcf11 and Clp1 instability. The fact that ATP bound by Clp1 could not be exchanged with radiolabelled ATP even at high molar ratios suggests that the Clp1-ATP interaction is very tight and that Clp1 may be constitutively ATP-bound. Considering the results presented in this chapter, it is hypothesised that Clp1 requires ATP as a co-factor to maintain a stable folded conformation necessary for interaction with Pcf11. Loss of Clp1-ATP binding is associated with changes in Clp1 secondary structure and the release of Pcf11, which is likely to disrupt CFIA formation *in vivo*.

4. Functional analysis of the Clp1-ATP interaction in *S. cerevisiae*

4.1 Introduction and overview

The functions of the Rna14, Rna15 and Pcf11 subunits of CFIA in pre-mRNA 3'-end processing and transcription have been investigated in a number of different studies using a combination of *in vitro* and *in vivo* techniques (described in sections 1.3.4 and 1.7). Until recently, there was little information available with regards to Clp1 function. However, during the completion of this work, two studies were published that have shed some light upon the role of Clp1 in transcription and 3'-end processing. Firstly, Haddad *et al*. uncovered two *clp1* temperature sensitive mutants (*clp1-5* and *clp1*-12) that were unable to cleave and polyadenylate pre-mRNA *in vitro* [121]. They also found that *in vivo* depletion of Clp1 led to defective transcription termination on small nucleolar RNA (snoRNA) genes. The second study by Holbein *et al.* showed that cellular depletion of Clp1 interfered with both pre-mRNA 3'-end formation and PolII transcription termination [180]. In terms of the function of Clp1-ATP binding, there are now contrasting results in the literature. Ghazy *et al.* recently claimed that mutation of Clp1 P-loop residues did not abrogate Clp1-ATP binding, but was lethal in *S. cerevisiae* because it trapped Clp1 in a conformation that prohibited its association with Pcf11 [122]. On the other hand, Holbein *et al.* recognised that mutation of P-loop residues removes bound ATP [180]. However, they argued that Clp1-ATP binding was not necessary for transcription, transcription termination, 3'-end processing and indeed *S. cerevisiae* viability.

A thorough characterisation of recombinant Clp1 mutants using hydrodynamic, biophysical and biochemical methods showed that preventing Clp1 from binding ATP resulted in the loss of Pcf11-binding, protein instability and aggregation of Clp1 (see chapter 3). As Clp1 is an essential protein in *S. cerevisiae*, it was hypothesised that the effects of mutations that abrogate Clp1-ATP binding could be lethal. Alternatively, Clp1-ATP mutants may reveal new phenotypes that may illuminate the role of Clp1 in transcription termination or pre-mRNA 3'-end processing.

In order to investigate whether Clp1-ATP binding is obligatory for the function of Clp1 in pre-mRNA 3'end processing and transcription termination, a comprehensive *in vivo*

mutational analysis was employed to probe the phenotypes of *clp1* ATP-binding mutants in *S. cerevisiae*.

4.2. Clp1-ATP binding is essential in *S. cerevisiae*

4.2.1. Mutations that abrogate Clp1-ATP binding are lethal

To investigate the effects of Clp1-ATP mutations in *S. cerevisiae*, the 5-FOA plasmid shuffle system was used to determine if any of the *clp1*-ATP mutants characterised *in vitro* could rescue the lethal phenotype of a chromosomal *clp1* deletion (*clp1*Δ).

The haploid *clp1*Δ yeast strain (which contains the *URA3-CLP1* construct) does not grow on 5-FOA medium (Figure 4.1A), consistent with the sensitivity of *URA3*+ yeast to 5-FOA and that deletion of the only chromosomal copy of *clp1* is lethal. Haploid *clp1*Δ yeast formed colonies on 5-FOA medium after transformation with the *LEU2-CLP1* construct, indicating that a plasmid-borne copy of the *CLP1* gene is able to reverse the lethal phenotype (Figure 4.1A, upper panel). Yeast transformed with constructs harbouring the *clp1* S138A and *clp1* D161A alleles also formed colonies after 5-FOA selection, and the density of yeast growth in each serial dilution "spot" revealed there is no obvious difference between the growth of these mutants compared with wildtype *CLP1* yeast. When the growth temperature was shifted to 37°C, *clp1* S138A and D161A mutants still grew comparably with wildtype yeast, confirming the gene products are stable and that these alleles are not temperature sensitive at 37 ºC (Figure 4.1A, lower panel). In contrast, yeast transformed with the *clp1* K136E, T137A and D251A alleles displayed no growth after 5-FOA selection at either 23 ºC or 37 ºC (Figure 4.1A, both panels). However, each mutant grew comparably with wildtype *CLP1* yeast on the pre-5-FOA selection control plate (transformed strain plated on YPD medium, Figure 4.1A, left panels). This ruled out the possibility that the inability of these yeast mutants to grow on 5-FOA was due to unrelated growth defects that could have been introduced into the haploid *clp1*Δ yeast strain. Western blotting experiments confirmed that the level of protein expression of all mutants was comparable to wildtype (Figure 4.1B), excluding the possibility that the lethality of *clp1* K136E, T137A and D251A alleles was due to abnormalities in protein expression from these genes. The *clp1* K136E, T137A and D251A alleles are therefore lethal in *S. cerevisiae*. It was hypothesised that the K136E mutation may significantly disrupt the conformation of the Clp1 nucleotide-binding pocket (as the charge of the amino acid side chain at this

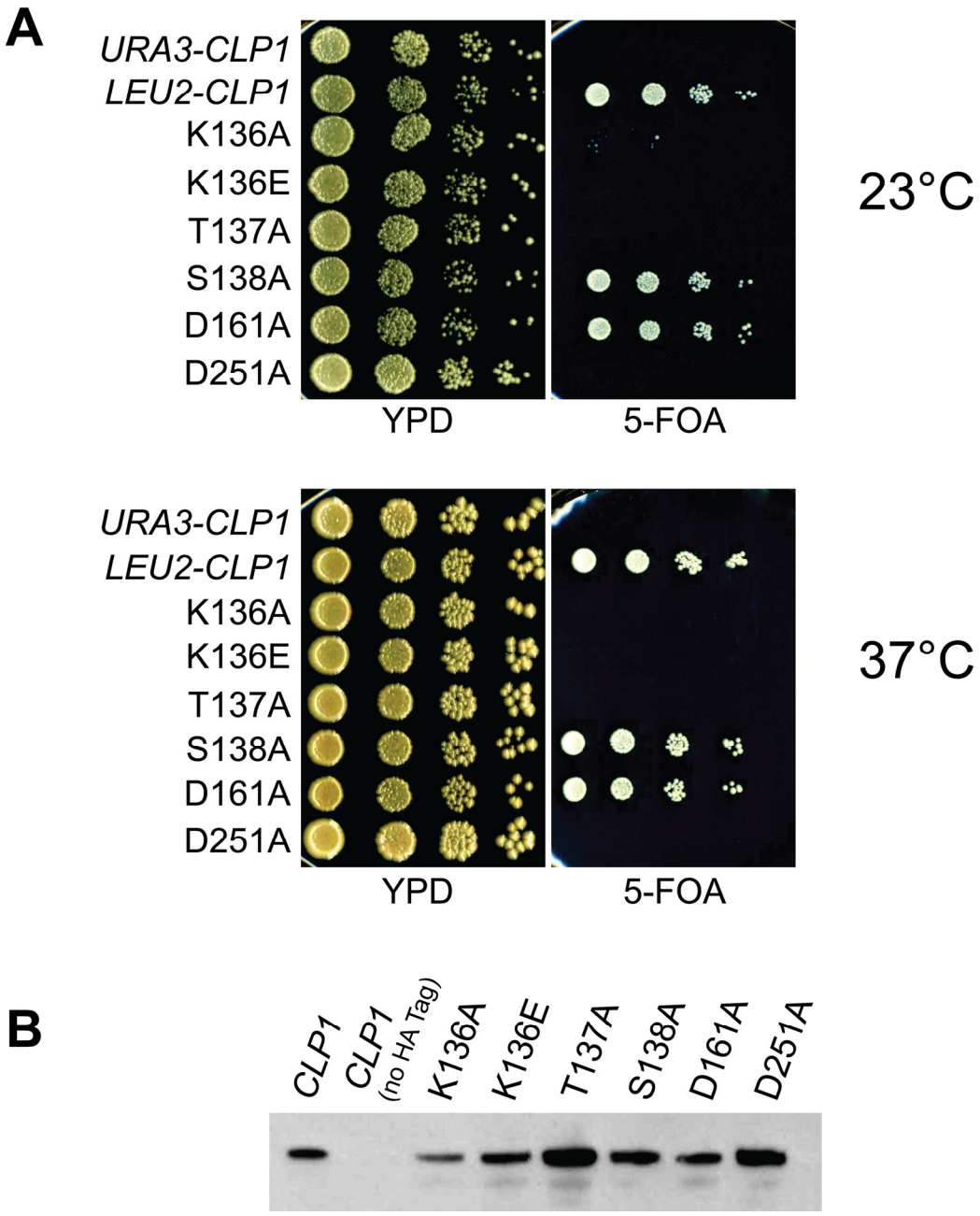


Figure 4.1. 5-FOA plasmid shuffle assay to determine the effects of mutations in the *clp1* gene on *S. cerevisiae* viability. (**A**) "Spot test" assays displaying the results of *clp1*Δ complementation with *clp1*-ATP mutant alleles performed at 23°C (top panels) and 37°C (bottom panels). Each spot represents a serial dilution (left to right: 10^4-10^1). The right panels show yeast transformants plated on 5-FOA medium and the left panels show transformants plated on YPD medium (control for growth defects in the *clp1*Δ strain). *URA3-CLP1* (top lanes) were not transformed with *LEU2* constructs containing *CLP1/clp1* alleles (control to check the efficacy of 5-FOA against *URA3*$^+$ yeast). *LEU2-CLP1* (second lanes from top) denotes the *clp1*Δ strain transformed with the *LEU2-CLP1* construct (wildtype gene). The lower lanes are the *clp1*Δ strain transformed with *LEU2* constructs containing each mutant allele indicated on the left hand side. (**B**) Evaluation of Clp1 expression by western blotting of protein extracts derived from *CLP1* wildtype and *clp1* mutants. Detection is with an antibody directed against the HA-tag at the C-terminus of Clp1 wildtype and mutants. Each mutant is labelled above the relevant lane of the blot.

position is swapped from positive to negative by this mutation), and that instead the K136A mutation may be more conservative in its effects on Clp1-ATP binding. When the *clp1*Δ strain was transformed with a construct bearing the *clp1* K136A allele, no growth was observed after 5-FOA selection at either 23°C or 37°C (Figure 4.1A, both panels). The pre-5-FOA selection control showed this was not due to growth defects in the *clp1*Δ strain (Figure 4.1A, left panels), or decreased expression of the Clp1 K136A protein (Figure 4.1B), indicating that the *clp1* K136A allele is also lethal in *S. cerevisiae*.

Although the 5-FOA plasmid shuffle assay showed that *clp1* S138A and D161A yeast mutants grew comparably with wildtype *CLP1* yeast, it could not be ruled out that there were minor growth defects associated with these mutants that were not observed due to the relatively insensitive spot test assays. To investigate the possibility that the *clp1* S138A and D161A mutations might cause minor growth defects in *S. cerevisiae*, growth rate assays were performed to allow a more quantitative comparison of the growth characteristics of these mutants with wildtype *CLP1* yeast. Figure 4.2 shows that the growth of the *clp1* S138A and D161A yeast mutants is indistinguishable from wildtype *CLP1* yeast over an 8 hour time period, with both reaching the exponential growth phase (doubling time of 1.5-2 hours in YPD medium) after 2 hours when cultured in YPD at 30°C. This demonstrates that the growth characteristics of *clp1* S138A and D161A mutants are identical to wildtype and as a result these mutants were considered poor candidates for analysis of pre-mRNA processing abnormalities.

Analysis of the effects of *clp1*-ATP mutations in *S. cerevisiae* using the 5-FOA plasmid shuffle system revealed that the K136A, K136E, T137A and D251A mutations (which disrupt Clp1-ATP binding *in vitro*, see section 3.3.1) in the *clp1* gene were lethal. However, the *clp1* S138A and D161A mutant alleles (these mutations did not affect Clp1-ATP binding *in vitro*, see section 3.3.1) were able to restore growth to a *clp1*Δ strain to wildtype levels. These results demonstrate that Clp1-ATP binding is likely to be essential in *S. cerevisiae.* The fact that the *clp1* D161A allele (the putative catalytic residue mutant) is viable suggests that either there is no enzymatic activity associated with Clp1, or that Clp1 enzymatic activity is not essential. Because the *clp1* K136A, K136E, T137A and D251A mutants were not viable, it was not possible to extract RNA in order

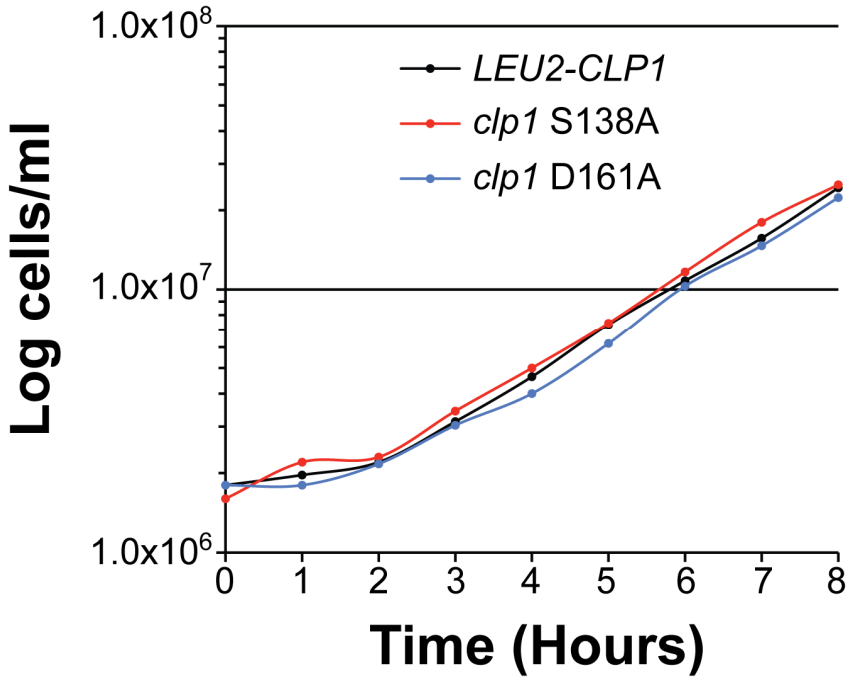


Figure 4.2. Analysis of growth rate of *CLP1* wildtype yeast (black), *clp1* S138A (red) and D161A (blue) mutants. The assay was conducted in YPD broth at 30°C.

to investigate the consequences of breaking the Clp1-ATP interaction on pre-mRNA 3'-end processing.

4.2.2. Overexpression of lethal *clp1*-ATP mutant alleles restores growth

When the deletion of a gene of interest is lethal (as is the case for deletion of *clp1*), the classical yeast genetics approach to investigate gene function is to characterise the effects of "conditional" mutations. Conditional mutations affect the gene product so that it can function in a "permissive" condition allowing propagation of yeast cells, but a shift to a "restrictive" condition renders the gene product inactive allowing experiments to test the effects of gene function loss to be performed. The most common example of conditional mutants are temperature sensitive mutants. A temperature sensitive mutant is able to grow at the permissive temperature (usually 25-30°C) but not at a temperature sufficiently high enough to induce the yeast heat shock response (35-40°C), but that is still permissive to the growth of wildtype cells [214].

None of the *clp1* mutants from the *in vivo* experiment described in section 4.2.1 displayed temperature sensitive phenotypes, meaning that mRNA could only be extracted and analysed from the viable S138A and D161A *clp1* mutants where the Clp1-ATP interaction was shown to be intact *in vitro* (see section 3.3). In an attempt to rescue lethal phenotypes and procure mRNA for later experiments, *clp1* mutants were overexpressed using a *GAL1* promoter and the 5-FOA plasmid shuffle assay was repeated to identify viable yeast mutants. The *GAL1* promoter is constitutively active in the presence of galactose and repressed in the presence of glucose [215]. It was reasoned that if Clp1 K136E, K136A, T137A and D251A still possessed any affinity for ATP, overexpression of these mutants could generate a small fraction of ATP-bound Clp1 sufficient enough to support yeast cell propagation.

The haploid *clp1*Δ *URA3*$^{+}$ yeast strain did not grow on 5-FOA in the presence of either galactose or glucose, consistent with the sensitivity of *URA3*$^{+}$ yeast to 5-FOA and the lethality of the chromosomal *clp1* deletion (Figure 4.3, central and right panels). Haploid yeast transformed with the *GAL1-CLP1* construct formed colonies on 5-FOA containing galactose (Figure, 4.3, central panel). This demonstrates that constitutive transcription of *CLP1* from the *GAL1* promoter in the presence of galactose is able to reverse the lethal phenotype of the chromosomal *clp1* deletion. However, transcription

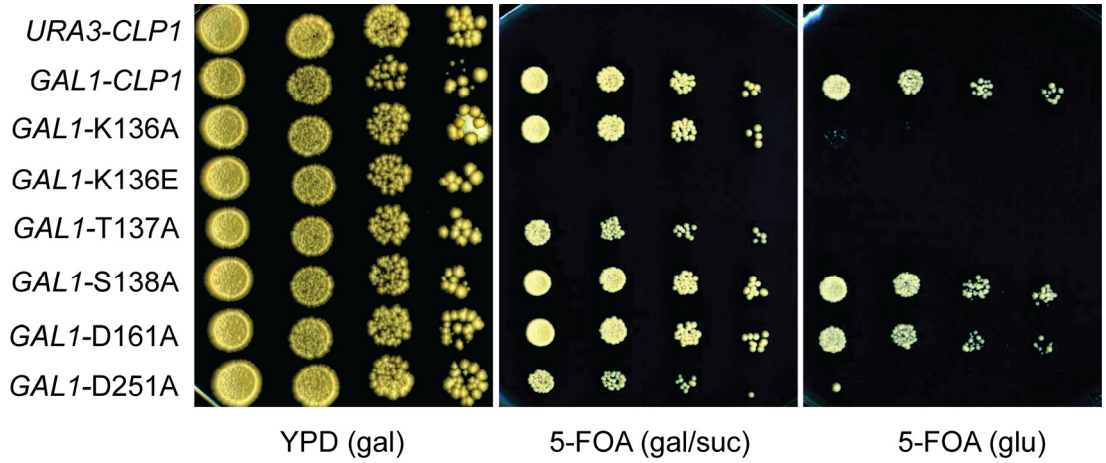


Figure 4.3 Spot test assays displaying the results of *clp1*Δ complementation with *clp1*-ATP mutant alleles expressed from the *GAL1* promoter. Each spot represents a serial dilution (left to right 10^4-10^1). The left panel shows yeast transformants plated on YPD medium (control for growth defects in the *clp1*Δ strain). The central panel shows yeast plated on 5-FOA medium containing galactose and sucrose (galactose activation of the *GAL1* promoter) and the right panel shows yeast plated on 5-FOA medium containing glucose (glucose repression of the *GAL1* promoter). The *clp1*Δ strain was transformed with *GAL1* constructs containing each mutant allele (labelled on the left hand side).

of *CLP1* from the *GAL1* promoter should have been shut off in the presence of glucose (glucose repression) resulting in lethality and the absence of yeast colonies from the 5-FOA glucose plate. This was not the case, as yeast colonies formed on 5-FOA containing glucose in all serial dilution spots (Figure 4.3, right panel). This indicates that glucose repression of the *GAL1* promoter was not completely effective, and that there was enough transcription of *CLP1* to support yeast growth. Similarly, the *GAL1-clp1* S138A and D161A alleles also restored growth to the *clp1*Δ strain after 5-FOA selection on both galactose and glucose media, reaffirming the inefficient glucose repression of *GAL1* promoter activity (Figure 4.3, central and right panels respectively). Yeast transformed with the *GAL1-clp1* K136E construct was unable to grow on 5-FOA containing either galactose or glucose (Figure 4.3, central and right panels respectively), demonstrating that overexpression of K136E was unable to complement the lethal *clp1* deletion. Growth of *GAL1-clp1* K136E yeast on the pre-5-FOA selection control was comparable to wildtype yeast, showing that the absence of growth on 5-FOA media was not due to defects in the *clp1*Δ strain (Figure 4.3, left panel). In contrast to this, when the *clp1*Δ yeast strain was transformed with *GAL1-clp1* K136A, T137A or D251A constructs, growth was observed in all serial dilution spots after 5-FOA selection on 5-FOA plates containing galactose, but no growth was observed on 5-FOA glucose plates (Figure 4.3, compare central and right panels). This demonstrates that overexpression of *clp1* K136A, T137A and D251A alleles from the *GAL1* promoter is able to rescue the lethal phenotypes of the same *clp1* alleles expressed at natural levels (see section 4.2.1 and Figure 4.1A). The absence of growth on 5-FOA glucose suggests that glucose repression of the transcription of these *clp1* mutants is sufficient to cause lethality to yeast.

Because 5-FOA is a toxic compound, it was deemed necessary to rule out that its presence in growth media did not contribute to the glucose repression-induced lethality of the *GAL1-clp1* K136A, T137A and D251A yeast mutants. The growth characteristics of GAL1-*clp1* yeast mutants were determined at 30°C on minimal agar medium containing galactose or glucose without 5-FOA. Figure 4.4A shows that *GAL1-CLP1* (*LEU2*$^{+}$) yeast grew equally well on both minimal galactose and minimal glucose media, reaffirming that glucose repression of *CLP1* transcription from the *GAL1* promoter was not sufficient to cause lethality. Contrary to the results of the 5-FOA plasmid shuffle assay (see Figure 4.3), the growth of the *GAL1-clp1* D251A yeast

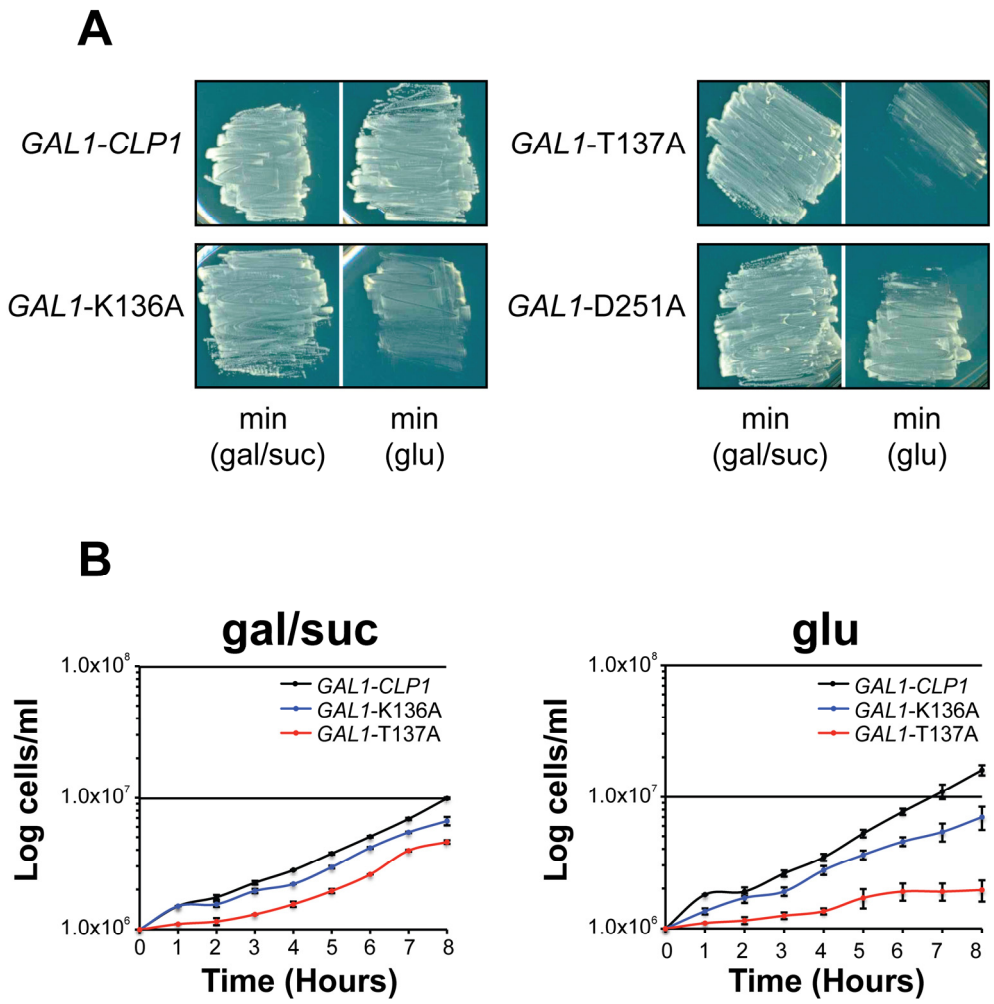


Figure 4.4. Glucose repression of *CLP1* wildtype and *clp1* mutant transcription. (**A**) Plate-based growth assay. Mutants are labelled to the left of each panel and the composition of the growth medium indicated below the panels. (**B**) Growth rate assays comparing the growth of *GAL1-clp1* K136A and T137A yeast mutants with *GAL1-CLP1* wildtype yeast (identified by inset key) at 30°C before and after glucose repression. The composition of the growth media are indicated above the graphs. Error bars represent the standard deviation from the mean of a set of three replicate assays.

mutant was not significantly different on medium containing galactose to growth on medium containing glucose (Figure 4.4A). This suggests that in the initial experiment, both the toxicity of 5-FOA and transcriptional repression of *clp1* D251A contributed towards lethality on the 5-FOA glucose plate. However, in keeping with the results of the 5-FOA plasmid shuffle assay, growth of the *GAL1-clp1* K136A and T137A yeast mutants was diminished on minimal media containing glucose compared to growth of these mutants on minimal media with galactose (Figure 4.4A). Growth rate assays were conducted to further characterise the growth phenotypes of the *GAL1-clp1* K136A and T137A mutants. In the presence of galactose, the growth of the *GAL1-clp1* K136A and T137A mutants were reduced 1.5-fold and 2-fold respectively compared with *GAL1-CLP1* wildtype yeast growth after 8 hours (Figure 4.4B, left panel). Differences in growth rates were also observed when strains were initially grown in galactose then shifted to glucose to repress *GAL1* promoter activity (Figure 4.4B, right panel). Growth of *GAL1-clp1* K136A yeast was 2.3-fold reduced compared to wildtype yeast after 8 hours, but with no evidence of a gradual cessation in growth that would be indicative of successful glucose repression. On the other hand, after 8 hours the cell density of the *GAL1-clp1* T137A yeast mutant was 8.2-fold less than that of *GAL1-CLP1* wildtype yeast. Furthermore, *GAL1-clp1* T137A growth was static between the 5 and 8 hour time points, suggesting the combination of transcriptional repression by glucose and the effect of the T137A mutation on Clp1-ATP binding acted together to cause lethality.

The restoration of growth to the *clp1*Δ yeast strain by overexpression of *clp1* K136A, T137A and D251A mutants is probably the result of an increase in the amount of ATP-bound Clp1. Yeast containing the *clp1* T137A allele under control of the *GAL1* promoter is similar to a conditional mutant where galactose conditions are permissive due to a high level of induction of T137A expression, and the lethal phenotype is observed upon a shift to the restrictive conditions of glucose repression. The glucose repression observed in these experiments on the whole is poor and it is possible that the cessation of T137A growth in glucose is due to a combination of 5-FOA toxicity, reduced T137A transcription and the consequences of disrupting Clp1-ATP binding. However, the experiments described above point to T137A as being a promising candidate for the investigation of 3'-end processing defects in mRNA associated with disruption of the Clp1-ATP interaction.

4.3. Analysis of pre-mRNA 3'-end processing in *clp1*-ATP mutants

Previous studies have shown that *S. cerevisiae rna14, rna15* and *pcf11* temperature sensitive mutants exhibit defects in pre-mRNA 3'-end cleavage and polyadenylation [88-94]. More recently, the same effects have been demonstrated with *clp1* temperature sensitive mutants, although no link between Clp1-ATP binding and 3'-end processing has been uncovered [121]. Clp1 is an essential component of the CFIA complex that is obligatory for both cleavage and polyadenylation of pre-mRNAs in *S. cerevisiae*. It was hypothesised that mutations that disrupt Clp1-ATP binding might result in aberrant pre-mRNA 3'-end processing and are lethal for this reason.

4.3.1. Analysis of poly(A) tail-lengths

Within the pre-mRNA processing field, the cleavage and polyadenylation characteristics of a yeast strain are typically investigated separately using *in vitro* assays. Cleavage assays involve testing the ability of partially-purified extracts from the desired yeast strains to cleave *in vitro*-transcribed RNA precursors (such as *GAL7* and *CYC1* precursor RNAs). The most commonly utilised polyadenylation assays either test the ability of extracts from the desired yeast strains to polyadenylate a pre-cleaved RNA precursor, or assess the length distribution of poly(A) tails in total RNA samples extracted from the desired yeast strains. Poly(A) "tail-length assays" can be used to compare the length distribution of poly(A) tails generated in wildtype and mutant yeast strains in order to reveal if a yeast mutant displays defects in transcript polyadenylation. The method involves extraction and purification of total RNA from the desired yeast strain, followed by digestion of non-poly(A) RNA by RNases and radiolabelling at the 3'-ends of poly(A) tails [201] (see section 2.9). Poly(A) tails are then visualised using autoradiography after their separation (by size) by denaturing polyacrylamide gel electrophoresis. To determine if *clp1* mutations that abrogate Clp1-ATP binding affect polyadenylation of pre-mRNAs, poly(A) tail-length assays were conducted using total RNA isolated from *GAL1-CLP1* wildtype yeast and *GAL1-clp1* yeast mutants.

Poly(A) tails isolated from a wildtype *GAL1-CLP1* total RNA sample were detected by autoradiography (Figure 4.5A). Comparison of the position of these poly(A) tails with the position of DNA size standards showed that poly(A) tail size distribution ranged from ~20 to ~80 nucleotides. This is consistent with the range of *S. cerevisiae* poly(A)

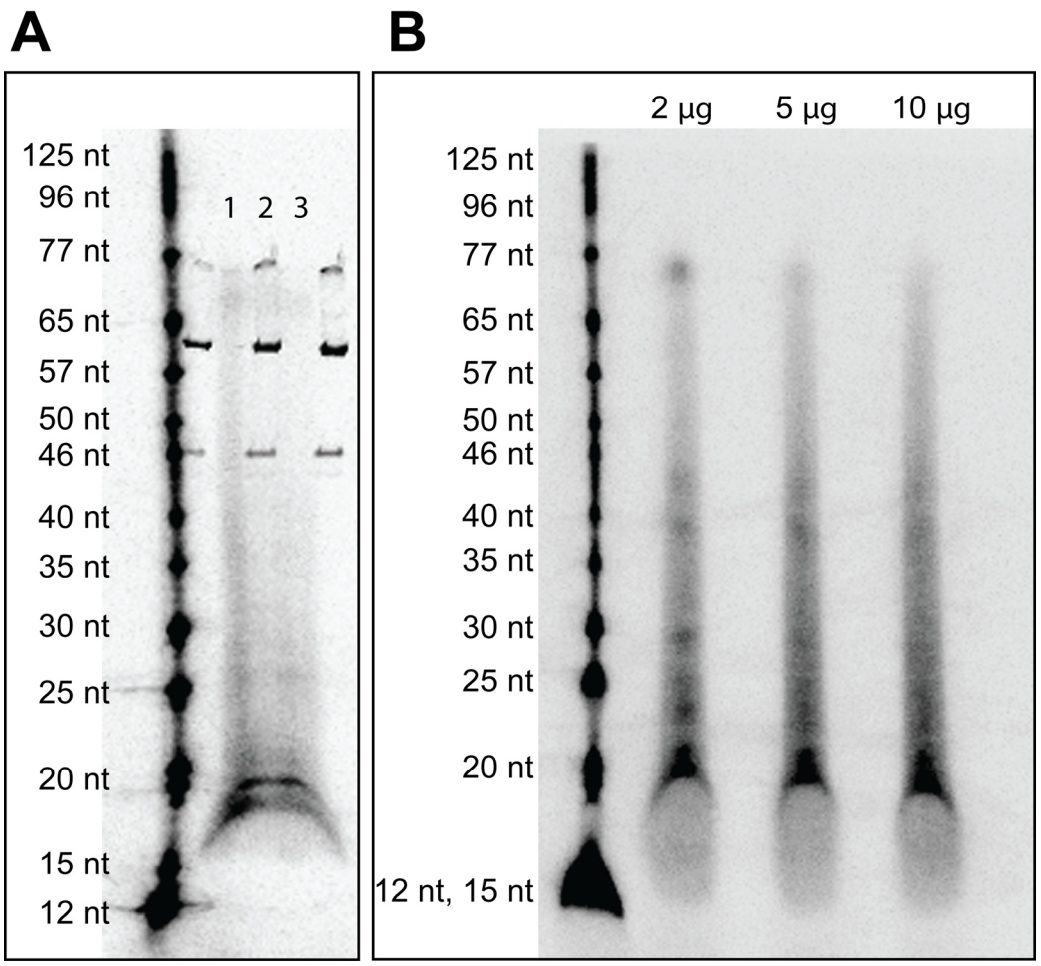


Figure 4.5. Poly(A) tail-length analysis. (**A**) poly(A) tails purified from *GAL1-CLP1* wildtype yeast cultures that were separated by electrophoresis on 10% polyacrylamide gels and visualised by autoradiography. Lane 1: digestion and purification protocol as cited in section 2.9; Lane 2: modified conditions, once ethanol precipitated (see section 2.9); Lane 3: modified conditions, twice ethanol precipitated (see section 2.9). The far left lane contains radiolabelled oligonucleotide size standards (sizes marked on left-hand side). The intense bands at 77, 65 and 46 nt are contaminants on the phosphor-screen from a previous experiment with the ^{14}C radioisotope. (**B**) Poly(A) tails purified from *GAL1-CLP1* wildtype yeast cultures that were separated by electrophoresis on 15% polyacrylamide gels and visualised by autoradiography. The modified digestion and purification conditions used are described in section 2.9. The amount of total RNA radiolabelled before RNase digestion is indicated above each lane. The far left lane contains radiolabelled oligonucleotide size standards (sizes marked on left-hand side).

tail-lengths observed in other studies that have used poly(A) tail-length assays [91, 94, 202]. However, the bands observed were poorly resolved. This indicated that the isolation of poly(A) tails was not optimal or that the conditions of the polyacrylamide gel electrophoresis were not appropriate for the level of resolution required. Modifications were made to the digestion and electrophoresis conditions according to published protocols (described in section 2.9) and Figure 4.5B shows the best results that were achieved using a modified tail-length assay. Intense poly(A) tail bands of 25, 30 and 40 nucleotides in length were visible that were not previously observed (compare Figure 4.5A and B). However, the overall resolution was still poor and not sensitive enough to confidently infer changes in pre-mRNA polyadenylation. Resolution did not improve when increasing the amount of RNA that was labelled (Figure 4.5B), indicating that the signal amplitude itself was not a problem and that sub-optimal digest and gel electrophoresis conditions were the likely cause of poor poly(A) tail resolution.

Poly(A) tail-length assays demonstrated that poly(A) tails of a normal length distribution were generated on mRNAs transcribed by wildtype *GAL1-CLP1* yeast. However, high-resolution, defined by visualisation of individual bands of different poly(A) tail-lengths, is required to infer differences in polyadenylation between wildtype yeast and mutants. Because the resolution of this technique could not be improved it was deemed insufficient to continue its use as a method for uncovering pre-mRNA processing phenotypes.

4.3.2. The Clp1-ATP interaction does not affect poly(A) site selection on *ACT1* mRNA

After the discontinuation of poly(A) tail-length assays it was decided to utilise reverse transcription quantitative real-time PCR (RT-qPCR) as a "readout" for pre-mRNA processing phenotypes. The RT-qPCR "poly(A) site selection assay" was designed to quantify the utilisation of different processing sites in the 3'-UTRs of mRNAs transcribed from the yeast actin gene (*ACT1*), allowing a comparison of the site usage in *CLP1* wildtype yeast and the *clp1* ATP-binding yeast mutants. Using wildtype *CLP1*, the relative usage of each *ACT1* processing site was determined in order to formulate the pattern of cleavage/poly(A) site usage on the *ACT1* gene (see section 2.9.2 and Figure 2.6 for experimental design). Comparison of cleavage/poly(A) site (hereafter

termed poly(A) site) usage between wildtype yeast and mutants was therefore able to determine whether Clp1-ATP binding is important for poly(A) site selection. To enable the conditional phenotype of T137A to be exploited, total RNA was extracted from yeast where the expression of *CLP1* wildtype or *clp1* mutants had been subjected to glucose repression (see section 2.8.8).

The RT-qPCR experiment using total RNA from a *GAL1-CLP1* wildtype yeast RNA sample revealed 98% of *ACT1* mRNA is processed at the first poly(A) site, demonstrating that the first site proximal to the 3' exon is the major poly(A) site of this gene (Figure 4.6B, location of cleavage/poly(A) sites depicted in Figure 4.6A). This is consistent with the findings of the study which first characterised the *ACT1* 3'-UTR using northern blotting experiments [153]. The second and third sites are situated close together in the *ACT1* 3'-UTR and were assayed as one poly(A) site due to limitations in primer and probe design. Transcripts processed at the second/third poly(A) sites comprised just 2% of total *ACT1* mRNA and *ACT1* mRNA processed at the fourth poly(A) site was not detected at all. *ACT1* poly(A) site selection assays with total RNA from the *GAL1-clp1* K136A mutant revealed poly(A) site usage on the *ACT1* gene was comparable to wildtype, with 97% processing occurring at the first poly(A) site, but with no processing at the second/third sites and 3% usage of the fourth site (Figure 4.6B). *ACT1* poly(A) site selection was also similar to wildtype in both *GAL1-clp1* S138A and D161A yeast mutants, with the respective frequencies of site usage being 95% and 96% for the first site, 4% and 0% for the second/third sites and 1% and 4% for the fourth site (Figure 4.6B). The *GAL1-clp1* T137A yeast mutant was considered promising for the analysis of pre-mRNA processing defects because of its conditional phenotype. Here, the *ACT1* poly(A) site selection assay showed that *ACT1* mRNA processed at the first site comprised 90% of total mRNA, with the frequencies of second/third and fourth site selection determined to be 4% and 6% respectively (Figure 4.6B). However, whilst this represents a 10% decrease in the usage of the first poly(A) site compared with wildtype, no radical redistribution of poly(A) site usage is apparent, suggesting that the loss of Clp1-ATP binding results in only modest effects on poly(A) site selection on *ACT1* pre-mRNAs.

Examination of the raw data generated from RT-qPCR amplification plots revealed that although site selection was only moderately affected, it was apparent that the mean quantity of *ACT1* mRNA in *GAL1-clp1* T137A RNA samples was ~2-fold reduced

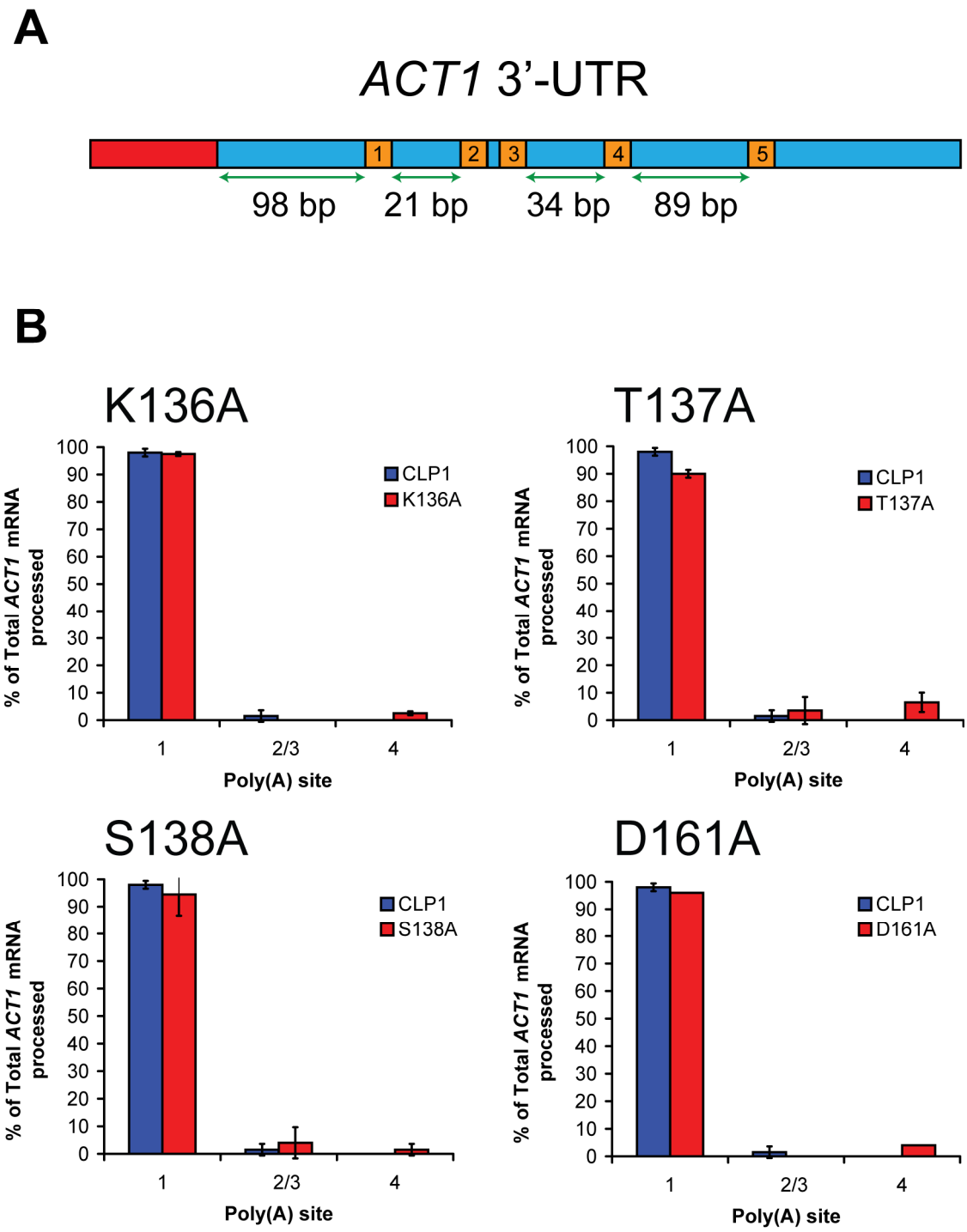


Figure 4.6. Poly(A) site selection on *ACT1* mRNA in *GAL1-clp1* ATP-binding mutants. (**A**) A schematic of the *ACT1* 3'-UTR. The 3'-end of the *ACT1* ORF is coloured red, 3'-UTR sequence is coloured blue and poly(A) sites are coloured orange and are numbered. Distances between poly(A) sites in bp are indicated as defined by Mandart and Parker [153]. (**B**) Results of the *ACT1* poly(A) site selection RT-qPCR experiment. The histograms show the percentages of total *ACT1* mRNA that were processed at each of three poly(A) sites in wildtype and mutant yeast. In each panel, the degree of processing at poly(A) sites in mutants compared to wildtype is displayed. The panels are labelled above histograms with the identity of each mutant, wildtype histograms are depicted in blue, mutants in red. Error bars represent the standard deviation from the mean of two experiments and in each experiment targets were detected in triplicate.

compared to the mean quantity found in *GAL1-CLP1* wildtype RNA samples (a mean mRNA quantity of 80.3 compared to 136.5, see Table 4.1). Similarly, the mean quantity of *TDH2* mRNA (the normalisation control used in this experiment) was 42.4 in *GAL1-clp1* T137A RNA samples compared to 211.75 in *GAL1-CLP1* wildtype RNA samples, which represents a 5-fold decrease in *TDH2* transcription. A reduction in the mRNA levels of these genes was not observed in any of the mutant samples tested

The results of the RT-qPCR poly(A) site selection assay demonstrated that Clp1-ATP binding does not contribute significantly to poly(A) site choice on the *ACT1* gene in *S. cerevisiae*. However, the quantities of *ACT1* and *TDH2* mRNAs were significantly lower in *clp1* T137A total RNA samples compared to wildtype *CLP1* and other *clp1* mutant RNA samples. These genes are normally highly-expressed, so it is likely that loss of Clp1-ATP binding caused by the T137A mutation results in transcription defects. A reduction in *ACT1* and *TDH2* mRNA levels was not observed in the *clp1* K136A mutant. However, this mutant also did not display a conditional phenotype when its transcription was repressed by glucose (see section 4.4.2), and so it is likely transcription is far less affected in this case.

	CLP1	**K136A**	**T137A**	**S138A**	**D161A**
Mean Quantity *TDH2* (SD)	211.75 (2.1)	351.55 (17.75)	42.4 (7.3)	225.95 (168.95)	309 (22.95)
Mean Quantity *ACT1* (SD)	136.5 (5.85)	202.1 (3.85)	80.3 (19.05)	137.9 (88.95)	195.75 (14.5)

Table 4.1. *TDH2* and *ACT1* mRNA quantities determined by RT-qPCR of total RNA extracted from *CLP1* wildtype and *clp1* mutants. *TDH2* and *ACT1* mRNA quantities in the T137A RNA sample are highlighted red to illustrate their reduction compared to wildtype and other mutant quantities. The quantities displayed were determined by comparison to the relevant standard curves and are not normalised to or expressed as relative to the expression of any other gene. The figures in brackets are the standard error derived from evaluation of the mean. C_T and quantity values generated in all RT-qPCR experiments are located in Appendix 8.6.

4.3.3. Clp1-ATP binding plays an important role in transcription

To investigate the effects of disrupting the Clp1-ATP interaction on transcription, the quantities of mRNAs transcribed from *ACT1*, *TDH2* and three other genes were analysed by RT-qPCR of total RNA samples extracted from wildtype *GAL1-CLP1* and *GAL1-clp1* yeast mutants. This enabled a comparison of the amounts of transcription of these genes in *GAL1-clp1* yeast mutants relative to the amounts of transcription in wildtype (*GAL1-CLP1)* yeast. The three additional genes chosen for transcript quantification were *ADH1* (alcohol metabolism), *YPT1* (membrane trafficking) and

CYC1 (respiration). These genes were chosen because they are involved in different cellular pathways, are highly-expressed and their expression does not fluctuate significantly during the yeast cell cycle (gene expression information obtained from *Saccharomyces* Genome Database, http://www.yeastgenome.org/).

The mean quantity of *TDH2* mRNA detected in total RNA samples extracted from the *GAL1-clp1* K136A yeast mutant represented 147% of the quantity detected in *GAL1-CLP1* wildtype total RNA samples (Figure 4.7A). Similarly, the amount of *ACT1* transcription was also increased in the *GAL1-clp1* K136A mutant at 149% of the amount of wildtype transcription. The quantities of *YPT1*, *ADH1* and *CYC1* mRNAs detected relative to wildtype were 80%, 100% and 80% respectively, demonstrating that transcription of these genes is not significantly affected by the *clp1* K136A mutation. In the *GAL1-clp1* S138A yeast mutant, the quantities of *ACT1, YPT1, ADH1* and *CYC1* mRNAs were between 70-100% of the *GAL1-CLP1* wildtype mRNA quantities (Figure 4.7A), showing that transcription of these genes in this yeast mutant is similar to wildtype. In the *GAL1-clp1* D161A mutant, the quantities of transcription of *TDH2, ACT1, YPT1, ADH1* and *CYC1* genes compared to wildtype was more variable than in the K136A and S138A mutants. Transcription of *TDH2* was decreased nearly 2-fold (61% of wildtype) and transcription of *ACT1* was increased ~1.5-fold (145% of wildtype) compared to wildtype (Figure 4.7A). On the other hand, transcription of *YPT1, ADH1* and *CYC1* genes was relatively similar in *GAL1-clp1* D161A to wildtype transcription (relative quantities of 74%, 95% and 99% respectively). The largest differences in transcription were found in the *GAL1-clp1* T137A mutant. The quantity of *TDH2* mRNA detected in this yeast mutant was just 10% of the quantity of *TDH2* detected in *GAL1-CLP1* wildtype yeast, indicating a severe reduction in the amount of transcription of this gene (Figure 4.7, A). This was around 2-fold lower than was detected in the *ACT1* poly(A) site selection assay (*TDH2* transcription in T137A was 20% relative to wildtype here). *ADH1* transcription in this mutant was also diminished to 10%, and *ACT1* transcription reduced to 40% of wildtype transcription. The quantities of *CYC1* and *YPT1* mRNAs detected in *GAL1-clp1* T137A RNA samples were ~80% of the mRNA quantities detected in wildtype yeast RNA samples, indicating transcription of these genes is not significantly affected by the *clp1* T137A mutation.

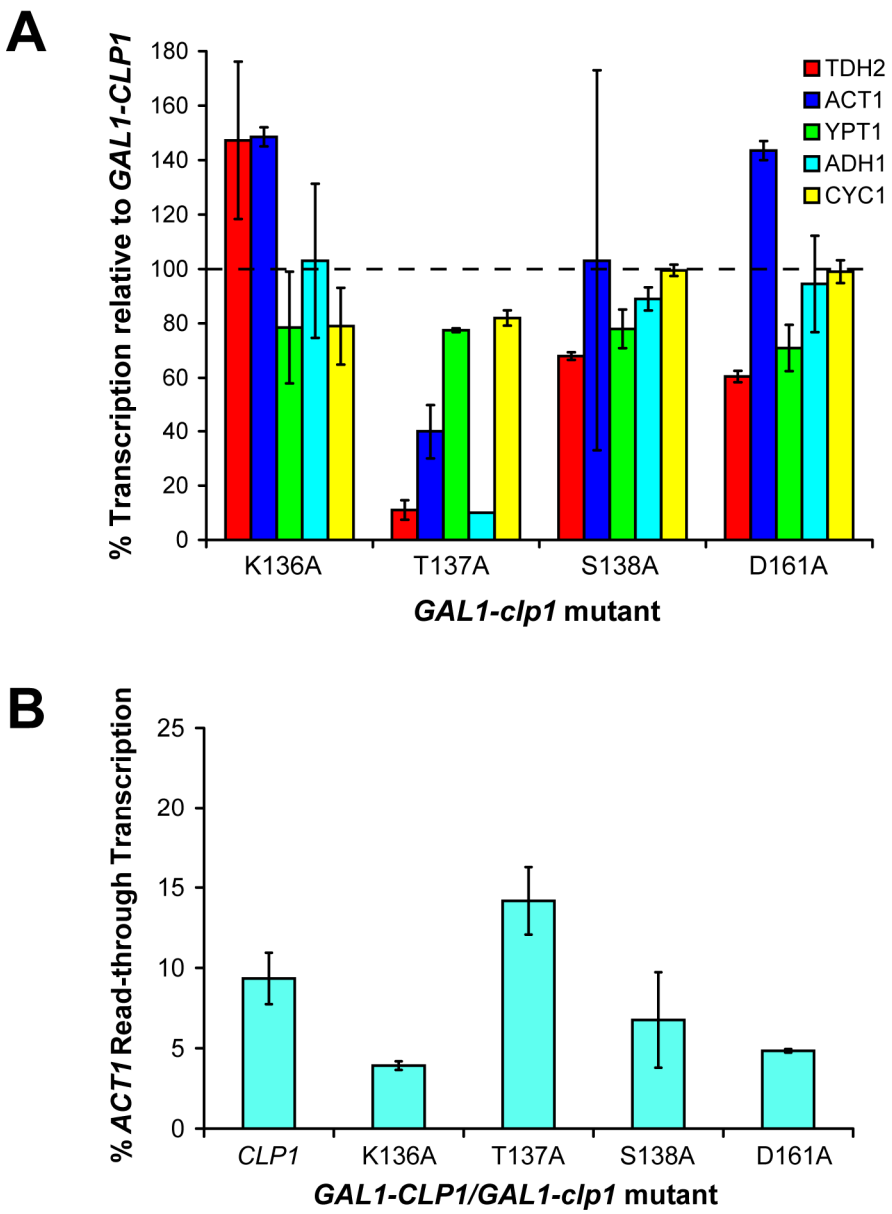


Figure 4.7. Gene expression and transcription termination in *GAL1-clp1* yeast mutants compared to wildtype yeast. (**A**) Histogram to show the amounts of transcription from *TDH2, ACT1, YPT1, ADH1* and *CYC1* genes in *GAL1-clp1* ATP-binding mutants compared to wildtype. The amount of mRNA transcribed from each gene was determined by RT-qPCR of total RNA samples extracted from *GAL1-clp1* ATP-binding mutants. Each gene is identified in the inset key. (**B**) Histogram to show quantification of transcription termination on the *ACT1* gene (determined by percentage of *ACT1* readthrough transcription) in *GAL1-CLP1* wildtype yeast and *GAL1-clp1* ATP-binding mutants (each mutant is labelled). Error bars represent the standard deviation from the mean of two experiments and in each experiment targets were detected in triplicate.

Analysis of mRNA quantities by RT-qPCR revealed the *clp1* T137A mutation, causes reduced transcription of the *TDH2*, *ACT1* and *ADH1* genes. The differences in the levels of transcription of these genes between the *GAL1-clp1* T137A yeast mutant and *GAL1-CLP1* wildtype yeast were far more prominent than the small increases/decreases in transcription observed between wildtype yeast and the other *GAL1-clp1* yeast mutants. Taking these results into consideration, it is possible that the growth defects displayed by the *GAL1-clp1* T137A mutant are the result of severe defects in transcription caused by loss of Clp1-ATP binding, and that defective transcription results in lethality (when *clp1* T137A is not overexpressed). A wider profile of the *GAL1-clp1* T137A yeast mutant transcriptome is required to determine the extent of transcriptional abnormalities associated with this mutant. The *clp1* K136A mutation seemingly does not affect transcription although it has also been shown to abolish the Clp1-ATP interaction (see section 3.3.1). It is probable that reduced transcription was not observed with the *GAL1-clp1* K136A mutant as it did not display the same conditional phenotype as *GAL1-clp1* T137A after glucose repression of the *GAL1* promoter (see section 4.2.2).

4.3.4. Clp1-ATP binding may be important for transcription termination on the *ACT1* gene

Two recent studies have demonstrated that cellular depletion of Clp1 affects PolII transcription termination in *S. cerevisiae* [121, 180]. The most recent research into transcription termination in *S. cerevisiae* points to a unified allosteric/torpedo model in which both transcription of the cleavage/poly(A) site and cleavage of the pre-mRNA are essential in halting transcription (see section 1.7.1) [176]. Aberrant transcription termination is defined by "transcription run-on" or "readthrough transcription", where PolII continues to transcribe after passing the most distal poly(A) site from the gene being transcribed. To determine whether Clp1-ATP binding is important for transcription termination, an RT-qPCR assay was used to compare the quantities of readthrough transcription on the *ACT1* gene in wildtype yeast and the *GAL1-clp1* K136A, T137A, S138A and D161A yeast mutants.

RT-qPCR experiments using total RNA extracted from *GAL1-CLP1* wildtype yeast showed that *ACT1* readthrough transcripts comprised ~9% of total *ACT1* mRNA (Figure 4.7B). This amount was higher than expected and is indicative of inefficient

transcription termination. The amount of *ACT1* readthrough transcription in *GAL1-CLP1* yeast is noticeably higher than that observed in experiments with *PCF11* wildtype yeast described in chapter 6 (less than 1% readthrough transcription on the *ACT1* gene), and may reflect decreased levels of Clp1 protein caused by reduced *CLP1* gene transcription (after glucose repression of the *GAL1* promoter). To confirm this, it would be necessary to perform further experiments that use more efficient methods to deplete Clp1 in *S. cerevisiae* (for example the degron approach). The amount of *ACT1* readthrough transcription was lower in the *GAL1-clp1* K136A, S138A and D161A mutants compared to wildtype, with *ACT1* readthrough transcripts comprising 4%, 7% and 5% of total *ACT1* mRNA respectively (Figure 4.7B). In contrast, the proportion of *ACT1* readthrough transcripts to total *ACT1* mRNA was slightly elevated in the *GAL1-clp1* T137A mutant at 15%, revealing a small increase in readthrough compared to wildtype (Figure 4.7B).

The *GAL1-clp1* T137A yeast mutant displayed a small increase in *ACT1* readthrough transcription compared to wildtype *GAL1-CLP1* yeast, suggesting Clp1-ATP binding may be important for transcription termination. However, because this increase was only small (~5%) it is difficult to ascertain as to whether it is significant. Results presented in chapter 6 show that the Pcf11-Clp1 interaction is necessary for transcription termination on the *ACT1* gene. When considering the fact that the T137A mutation causes loss of the Clp1-Pcf11 interaction *in vitro* (see section 3.2), it is possible that the increase in *ACT1* readthrough transcription in the *GAL1-clp1* T137A mutant is significant. Clarification of the role of Clp1-ATP binding in PolII transcription termination will require a more in-depth analysis of yeast transcripts from a larger selection of genes.

4.4. Summary and conclusions

An *in vivo* analysis in *S. cerevisiae* of the effects of mutations in the *CLP1* gene showed that *clp1* ATP-binding yeast mutants that abrogate Clp1-ATP binding *in vitro* are not viable (*clp1* K136A, K136E, T137A and D251A), whereas *clp1* ATP-binding mutants that do not affect Clp1-ATP binding *in vitro* displayed no growth abnormalities compared to *CLP1* wildtype yeast (*clp1* S138A and D161A). The upregulation of mutant *clp1* gene transcription by galactose induction of the *GAL1* promoter was able to reverse the lethal phenotypes of the *clp1* K136A, T137A and D251A mutations. This

was likely due to the fact that these mutations may not completely abolish Clp1-ATP binding and that increased protein levels might produce a small fraction of ATP-bound Clp1 that is sufficient to maintain the essential function of Clp1 to *S. cerevisiae.* The *clp1* T137A yeast mutant showed a conditional phenotype where growth in medium containing glucose provides a restrictive condition which is possibly caused by a combination of repression of *GAL1* promoter activity, 5-FOA toxicity and the loss of Clp1-ATP binding caused by this mutation. Analysis of mRNA extracted from *GAL1-clp1* yeast mutants by RT-qPCR showed that the frequency of poly(A) site usage on *ACT1* mRNA was only modestly affected in the *GAL1-clp1* T137A yeast mutant compared to wildtype, indicating that Clp1-ATP binding may not be a major effector of poly(A) site selection. However, the quantities of mRNA transcribed from five different genes were significantly lower in the *GAL-clp1* T137A mutant compared to wildtype and other mutants, positing a role for the Clp1-ATP interaction in the maintenance of transcription levels in *S. cerevisiae*. Results of RT-qPCR experiments also suggested that the Clp1-ATP interaction may be involved in effecting transcription termination, although a deeper investigation is required to provide conclusive evidence of this.

The fact that mutation of the Clp1 putative catalytic residue, Asp161, had no effect on Clp1-ATP binding *in vitro* and caused no growth abnormalities in *S. cerevisiae* suggests that Clp1-ATP binding fulfils no catalytic function, or that the catalytic capability of Clp1 is redundant. This gives further substance to the idea that ATP is a structural component of Clp1 that is essential for Clp1 stability and the adoption of a conformation that is conducive to CFIA assembly. It is likely that this is necessary for the maintenance of appropriate levels of transcription from a variety of essential genes and for efficient PolII transcription termination.

5. *in vitro* analysis of recombinant Pcf11-Clp1 binding mutants

5.1. Introduction and overview

Pcf11 is a subunit of the cleavage factor IA complex (CFIA), an essential multi-subunit protein complex required for the cleavage and polyadenylation of pre-mRNAs in *S. cerevisiae* (see sections 1.3 and 1.3.3). Pcf11 interacts with all three other CFIA subunits (Rna14, Rna15 and Clp1) and also core subunits of the cleavage and polyadenylation factor complex (CPF) [89, 99, 103]. The amino-terminal domain (NTD) of Pcf11 interacts with PolII CTD, which has been shown to be crucial for transcription termination at a number of gene loci [106, 110, 115], and cellular depletion of Pcf11 has also been associated with a decrease in transcription [89, 176, 186]. The crystal structure of a Clp1-Pcf11 heterodimer shows an intimate association between Clp1 and Pcf11 [116]. In particular, in Pcf11 three highly conserved residues in the Pcf11-Clp1 interaction domain (Clp1 ID) are extensively associated with residues in the Clp1 CD and in a cleft between the Clp1 CD and CTD via hydrogen bonding and hydrophobic interactions. The results of a study published during the completion of the work presented here showed that the *pcf11* RW-A temperature sensitive allele (which was identified independently in this study) causes a weakened Pcf11-Clp1 interaction that results in defects in pre-mRNA 3'-end processing and transcription termination [122]. However, despite the obvious importance of the Pcf11-Clp1 interaction in both of these processes, the mechanism by which the interaction of these proteins influence 3'-end processing and transcription termination is still unclear.

To study the significance of the Pcf11-Clp1 interaction, it was necessary to first assess the properties of the Pcf11 CTD that contains the Clp1 ID. Following this, conserved residues in the Clp1 ID were mutated and properties of the resultant recombinant mutant proteins were characterised using hydrodynamic and biochemical techniques. The experiments aimed to determine which mutations uncoupled Clp1 from Pcf11, the effects of Pcf11-Clp1 uncoupling on Clp1 stability, and whether the Pcf11-Clp1 interaction was important for Clp1-ATP binding.

5.2. Characterisation of the Pcf11 CTD

Analysis of the Pcf11 amino acid sequence together with findings from previous studies has shed some light onto the properties of the Pcf11 CTD. The crystal structure of a Clp1-Pcf11 heterodimer revealed that the Pcf11-Clp1 ID is comprised of Pcf11 CTD residues 475-499 (see section 1.3.3 and Figure 1.4) [116]. Furthermore, it has been shown that the Pcf11-Yra1 binding site overlaps this region (for a domain map of Pcf11 see Figure 1.3A) [170]. Other prominent features of the Pcf11 CTD are two zinc finger motifs that flank the Clp1/Yra1 ID. Zinc finger motifs are structural motifs that use zinc ions to stabilise their folds, and commonly serve as interaction modules that bind to other proteins, DNA or RNA. However, no functions of Pcf11 have yet been associated with these motifs. Apart from residues 475-499, no structural information is available for the Pcf11 CTD, yet it is known to be the interaction site for at least two protein factors.

5.2.1. Purification of Pcf11 456-626

Characterisation of the Pcf11 CTD required the production of recombinant protein containing this domain. A pET-22b construct containing DNA sequence encoding Pcf11 residues 456-626 fused to a C-terminal hexahistidine sequence tag had been generated previously and was made available by Dr I. Taylor (NIMR). This fragment of Pcf11 contains the Clp1/Yra1 ID and the second zinc finger motif of Pcf11. Pcf11 456-626 was overexpressed from the pET-22b construct in *E. coli* BL21(DE3) cells and purified by nickel-affinity chromatography and size-exclusion chromatography. SDS-PAGE gel analysis of protein fractions from the nickel-affinity chromatography purification showed a dense band of overexpressed protein was visible in the crude lysate but not in the insoluble pellet fraction (Figure 5.1A). Comparison of this protein band with protein size standards showed it was between 20-30 kDa in size, which is slightly larger than the estimated molecular weight of 20.3 kDa for the Pcf11 456-626 protein. The fact that Pcf11 456-626 was not visible in the fraction extracted from the nickel column flow-through demonstrated that the C-terminal His-tag was present and bound the nickel resin, although some protein was lost in two subsequent wash steps along with some higher molecular weight contaminants that bound non-specifically to the nickel column (Figure 5.1A). Protein with the correct SDS-PAGE migration was observed in fractions 1-11 after elution of the nickel column with an imidazole buffer,

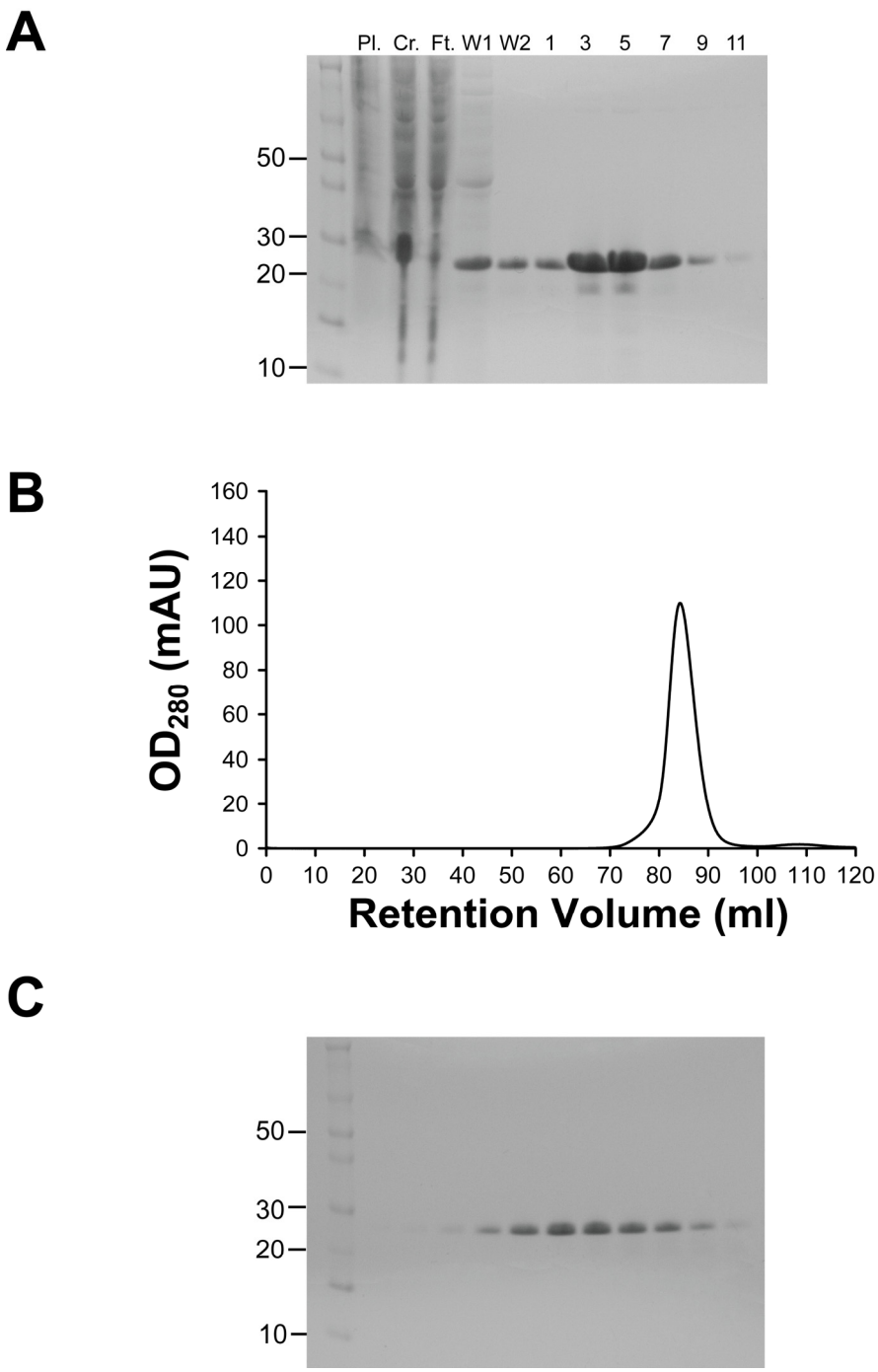


Figure 5.1. Purification of the Pcf11 CTD. (**A**) SDS-PAGE of nickel-affinity purification of Pcf11 456-626 (6-His Tagged) after overexpression of the protein in *E. coli*. Each stage of the purification was analysed: Pl., cell pellet (insoluble protein); Cr., crude clarified lysate; Ft., flow-through from nickel column; W1, first wash of nickel column; W2, second wash of nickel column; 1-11 (ml), eluted fractions from nickel column. Protein size standards are labelled on the left (in kDa). (**B**) SEC purification profile of Pcf11 456-626 (black line) following nickel-affinity purification. (**C**) SDS-PAGE of fractions eluted by SEC. Protein size standards are labelled on the left (in kDa).

along with small amounts of a lower molecular weight protein contaminant present in fractions 3-5 (Figure 5.1A). As a further purification step, SEC was performed to separate any remaining contaminants and protein oligomers by size using a HiLoad 16/60 Superdex 75 SEC column. The SEC profile of Pcf11 456-626 displayed a single peak that eluted from the column with a retention volume of ~85 ml (Figure 5.1B). This suggested that the protein was pure and that its oligomeric state was likely to be homogeneous. SDS-PAGE analysis of sample fractions taken throughout the elution of the chromatographic peak showed that it contained Pcf11 456-626 with no other visible contaminants (Figure 5.1C), and so was suitable for use in subsequent biophysical experiments.

5.2.2. Multiangle laser light scattering analysis of Pcf11 456-626

A study published during the completion of this work determined the stoichiometry of the CFIA complex subunits to be 2:2:1:1 (Rna14:Rna15:Clp1:Pcf11) [92]. SEC-MALLS experiments described here aimed to determine the oligomerisation state and solution molar mass of Pcf11 456-626.

In a SEC-MALLS experiment with Pcf11 456-626, the signal from the dRI detector showed a peak eluted from the SEC column (Superdex 75 10/30) with a retention time of ~30 minutes (Figure 5.2A). The combination of dRI and light scattering measurements taken from this peak allowed calculation of molar mass and therefore the oligomeric state of the protein within it. Pcf11 456-626 displayed a narrow molar mass distribution (0.4% of the weight-averaged molar mass) with a weight-averaged molar mass of 19.4 kDa. This experimentally obtained value is very close to the expected mass of a Pcf11 456-626 monomer (20.3 kDa with the His-tag fusion), meaning its migration by SDS-PAGE (between the 20-30 kDa size standards, see Figure 5.1A, C) is anomalous rather than a result of an incorrect sequence.

The fact that Pcf11 456-626 is monomeric in solution is in agreement with the recently determined stoichiometry of Pcf11 in the CFIA complex (1 molecule of Pcf11 per CFIA complex) [92], although the analysis provided here does not rule out the possibility that the oligomeric state of Pcf11 is regulated by its N-terminal or central domains.

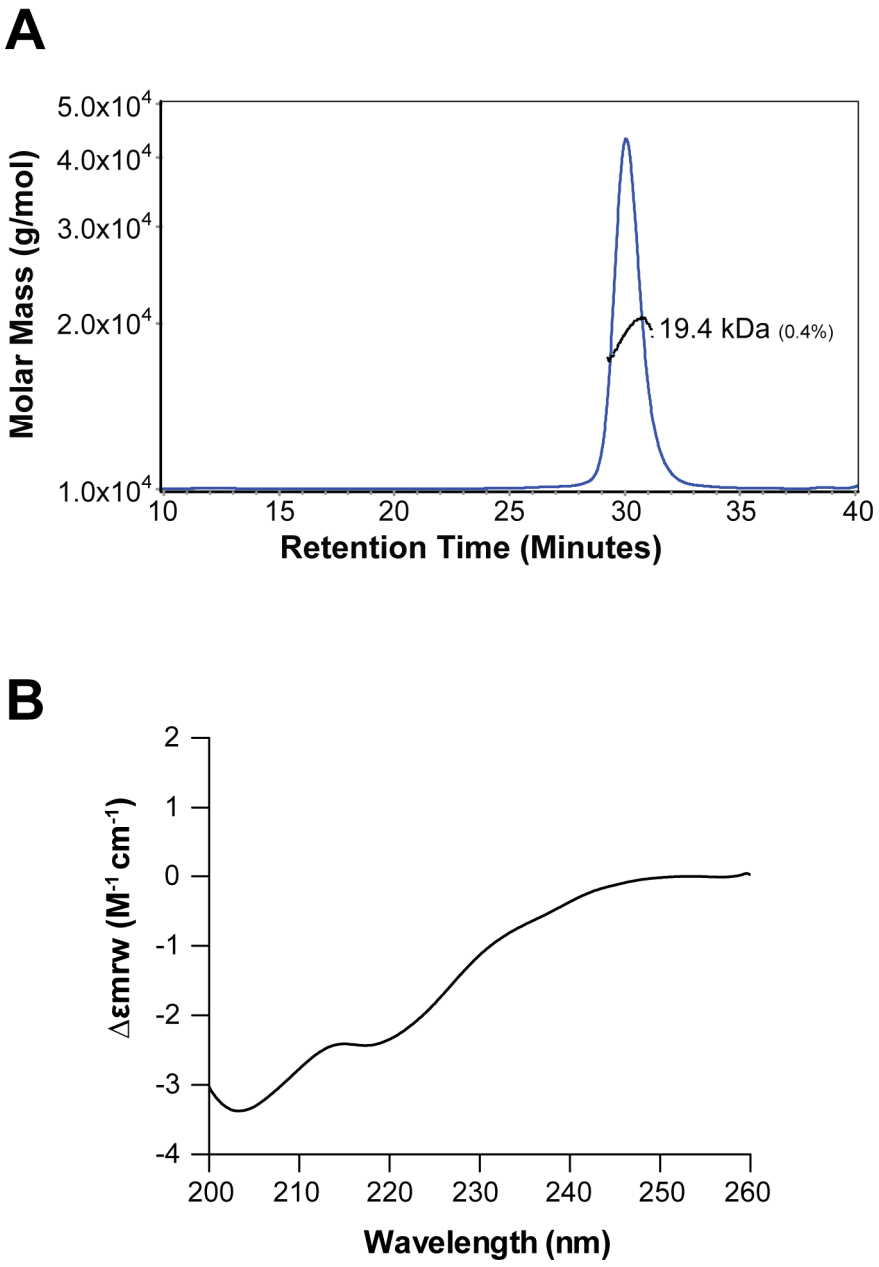


Figure 5.2. SEC-MALLS and CD characterisation of the Pcf11 CTD. (**A**) Chromatogram from SEC-MALLS analysis of Pcf11 456-626. The differential refractive index (dRI, blue line) is plotted to show the location of chromatographic peaks. The points in black are individual measurements taken from Debye plots of the weight-averaged molar mass of material eluted in the chromatographic peaks. (**B**) Far-UV CD spectrum (260-200 nm) of Pcf11 456-626 expressed as the differential mean residue weigh molar extinction ($\Delta\varepsilon_{mrw}$).

5.2.3. Far-UV circular dichroism of Pcf11 456-626

Residues 475-499 of Pcf11 comprise the protein interface in the crystal structure of a Clp1-Pcf11 heterodimer [116]. However, with regards to the structure of the remainder of the Pcf11 CTD, no data is available. Pcf11 associates with many other proteins including CFIA and CPF subunits, and PolII. In order to dynamically associate with its many different binding partners, Pcf11 might require a degree of flexibility, accomplished by adopting an unstructured and highly adaptable conformation. Alternatively, the Pcf11 CTD could adopt a highly-ordered folded conformation, with a number of interaction surfaces presented to facilitate the binding of protein partners. To gain insight into the secondary structure of the Pcf11 CTD, far-UV CD spectroscopy was performed with recombinant Pcf11 456-626.

The far-UV CD spectrum of Pcf11 456-626 displays negative CD peaks at ~222 nm and 205 nm (Figure 5.2B), indicative that Pcf11 456-626 contains both α and β secondary structure elements. However, the intense negative CD in the ~205 nm region and below indicates that a high proportion of the protein is composed of irregular or random coil, and is therefore disordered. The CDpro program [197] was used to predict the composition of Pcf11 456-626 secondary structure elements from far-UV CD data. The predicted secondary structure content of Pcf11 456-626 is 16.7% α-helix, 14.5% β-sheet, 15.9% turn and 52.9% random coil. This analysis has revealed that the Pcf11 456-626 is likely to be considerably unstructured.

It is probable that the predicted α and β secondary structure elements of Pcf11 456-626 are confined to the second zinc finger motif of the Pcf11 CTD, and that few or none of the remaining residues are structured. The Clp1 and Yra1 binding sites have previously been mapped to the Pcf11 CTD, although it may contain the interaction domains for other protein factors. Considering this, the disordered nature of the Pcf11 CTD may be necessary for it to act as a flexible binding platform for multiple interaction partners.

5.2.4. Pcf11 CTD deletion analysis

Although the Pcf11-Clp1 ID has already been characterised by the determination of a high-resolution crystal structure, it is possible that the structure only shows a small segment of the Pcf11-Clp1 ID. To date, the Pcf11-Clp1 interaction has not been investigated in great detail and it is possible that the Pcf11-Clp1 ID is more extensive

than is currently thought, or that there are multiple sites that mediate the Pcf11-Clp1 interaction. To determine if other parts of the Pcf11 CTD are involved in Clp1-binding, biophysical experiments were planned to map the Pcf11-Clp1 ID in greater detail. In order to carry out these experiments, several different Pcf11 CTD deletion mutants were cloned, overexpressed and purified from *E. coli* using a pGEX-6P-1 expression system, glutathione-S-transferase-affinity chromatography and SEC (see sections 2.3.3 and 2.3.5). Figure 5.3A shows a schematic of the Pcf11 CTD from which a variety of constructs were designed. Pcf11 400-626 encompasses the entire Pcf11 CTD, Pcf11 416-626 excludes the residues N-terminal to the first zinc finger motif, Pcf11 400-499, 416-499 and 420-499 are constructs that all contain the first zinc finger and Clp1 ID observed in the crystal structure, but lack the second zinc finger and C-terminal residues.

The consecutive steps of the glutathione-S-transferase (GST) purification were analysed by SDS-PAGE. Analysis of the Pcf11 400-499 purification showed three intensely-stained protein bands were visible in a sample of the protein that bound to the glutathione sepharose beads (Figure 5.3B, lane labelled 'B.'). Comparison of these bands with protein size standards showed that two species had molar masses of approximately 30 and 35 kDa respectively. The larger of these bands was likely to be the GST-Pcf11 400-499 fusion (expected molar mass of 38 kDa), and the second band possibly a degradation product of this protein. A third band was present with a lower molar mass of around 25 kDa. The identity of this band was also unclear, and it is likely that this was also a degradation product of the GST-Pcf11 400-499 fusion. After cleavage of the GST-tag with 3C protease, a further intensely-stained band was visible at around 20 kDa (Figure 5.3B, lane labelled 'C.B.'). Although the apparent molar mass of this band was slightly lower than that expected for the cleaved GST-tag (26 kDa), its intensity and the fact it appears in the other Pcf11-GST purifications suggests it can only be cleaved GST. Furthermore, the appearance of this band corresponded with a reduction in the intensity of the 35 kDa band, signifying that the GST-tag had been cleaved from Pcf11 400-499. However, the post cleavage eluted fraction (labelled 'C.E.' in Figure 5.3B) did not contain a 12 kDa protein equivalent to Pcf11 400-499, with only low levels of protein contaminants and some carry-over of the GST-tag visible. Similarly, SDS-PAGE analysis of the Pcf11 416-499 and Pcf11 420-499 GST-purifications showed that these GST-fusion proteins were also expressed, but no proteins of the correct size (10 and 9.5 kDa respectively) were recovered after

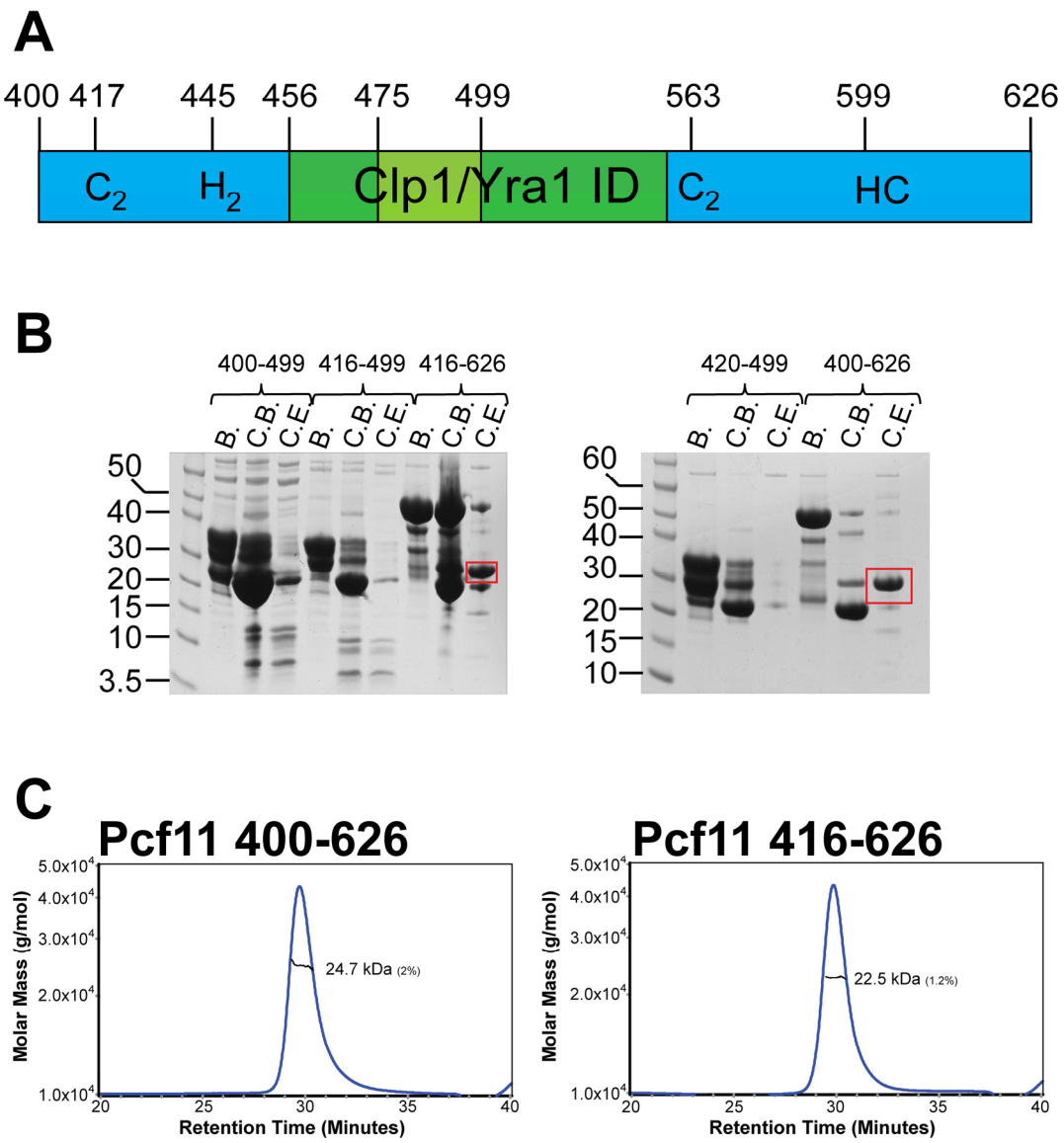


Figure 5.3. Pcf11 CTD deletion analysis. (**A**) Schematic of the Pcf11 CTD. The zinc finger regions and the Clp1/Yra1 ID are coloured blue and green respectively and residue positions indicated above. The light green box represents Pcf11 residues observed in the Clp1-Pcf11 crystal structure. (**B**) SDS-PAGE of Pcf11 CTD construct GST purifications. The identity of each construct is shown above the relevant section of the gel and lanes are labelled: B., protein bound to glutathione sepharose beads; C.B., sample of resin after cleavage of the GST tag from target protein with 3C protease; E.B., sample of the protein eluted into buffer after 3C protease digestion. The red boxes indicate soluble target protein. Sizes of molecular weight standards are shown to the left of each gel (sizes in kDa) (**C**) Chromatograms from SEC-MALLS experiments with Pcf11 400-626 and Pcf11 416-626. The differential refractive index (dRI, blue line) of each protein sample (labelled) is plotted to illustrate the location of chromatographic peaks. The black points are measurements of the weight-averaged molar mass recorded throughout the elution of the chromatographic peak.

cleavage of the GST-tags (Figure 5.3B). These results demonstrate that Pcf11 400-499, 416-499 and 420-499 are unstable in the absence of a fused GST-tag. On the other hand, GST purification of Pcf11 400-626 and 416-626 was successful. SDS-PAGE analysis showed in the Pcf11 400-626 purification, a ~50 kDa protein was bound by the glutathione Sepharose beads, and in the Pcf11 416-626 purification an overexpressed ~45 kDa protein was visible (Figure 5.3B, lanes labelled 'B.'). These molar masses were comparable to the expected combined masses of each protein fused with a GST-tag (51 kDa and 48 kDa respectively). After cleavage, intensely-stained bands corresponding to cleaved GST-tags were bound to Sepharose beads in both purifications (Figure 5.3B, lanes labelled 'C.B.'). SDS-PAGE of the eluted fractions from Pcf11 400-626 and Pcf11 416-626 GST purifications revealed both of these fractions contained intensely-stained protein species that migrated between 20 kDa and 30 kDa relative to the protein size standards (Figure 5.3B, lanes labelled 'C.E.'). These proteins were consistent with the expected molar masses of Pcf11 400-626 (26.2 kDa) and Pcf11 416-626 (24.2 kDa).

After a SEC purification step, the molar masses and oligomeric states of these proteins were analysed by SEC-MALLS. Both proteins eluted from the SEC column with a retention time of ~30 minutes (Figure 5.3C). The molar mass determined by Debye analysis of the light scattering data for Pcf11 400-626 was 24.7 kDa, close to the expected molar mass of 26.2 kDa. The molar mass determined for Pcf11 416-626 was 22.5 kDa, close to the expected molar mass of 24.2 kDa. The uniformity of molar masses calculated from the individual Debye plots generated throughout both of the protein peaks showed the proteins were monomeric with no tendency for self-association.

Purification of recombinant Pcf11 CTD deletion mutants showed that when protein constructs were terminated at residue 499, recombinant protein was unstable/insoluble after cleavage of GST-tags. The fact that the crystal structure of Clp1-Pcf11 shows that this residue is located in an unstructured loop suggests that it is unlikely that termination of the construct here affects the fold of the protein. However, the two protein constructs that were soluble (Pcf11 400-626 and 416-626) both contain the second zinc finger. Therefore, an alternative explanation is that the second zinc finger is important for the stability of the Pcf11 CTD, and that its deletion results in the protein misfolding. The

initial objective of purifying recombinant Pcf11 CTD deletion mutants was to use them in biophysical experiments aimed at probing the full extent of the Pcf11-Clp1 interaction. However, the fact that Clp1 is unstable and aggregates in the absence of bound Pcf11 made it impossible to perform these experiments (see sections 3.3 and 5.3).

5.3. Mutation of conserved Pcf11 residues in the Clp1-ID affects the stability of Clp1

5.3.1. Design of Pcf11 mutants

It has been shown that disrupting ATP from the Clp1-ATP binding pocket results in Clp1 instability and loss of the Pcf11-Clp1 interaction (see sections 3.2 and 3.3). Therefore, in addition to ATP-binding it is possible that the Pcf11-Clp1 interaction may be important for Clp1 stability. To further assess the nature and function of the Pcf11-Clp1 interaction, point mutations designed to weaken and/or disrupt the binding interface were introduced into the Clp1/Pcf11-EXP construct by SDM. These mutations were designed based upon interactions between Pcf11 and Clp1 residues observed in the Clp1-Pcf11 crystal structure [116] (see section 1.3.3. and Figure 1.4B), and the conservation of Pcf11 residues in the Clp1 ID of Pcf11 (see Figure 1.6C). A full list of mutations is presented in Table 5.1. Gln478 makes multiple contacts with other polar residues in Clp1. Arg480, Trp 482 and Trp489 are all involved in extensive hydrogen-bonding networks with Clp1 residues, with Trp482 and Trp489 also mediating hydrophobic interactions. These three residues are also highly-conserved across a wide range of organisms, implying that they may contribute to an essential function. Recombinant mutant proteins were all expressed from the Clp1/Pcf11(454-563)-EXP construct in *E. coli* (see section 2.2.3) and are hereafter referred to as Pcf11-Clp1 mutants.

Residue	Location	Function	Mutated to	Mutant name
Gln478	Clp1 ID	Binds Clp1 CD-CTD cleft residues Ser342, Ala197, Gly196, Thr194	Asn	Q478N
Arg480	Clp1 ID	Binds Clp1 CD-CTD cleft residues Gln191, Ser192, Glu331, Tyr332	Ala	R480A
Trp482	Clp1 ID	Binds Clp1 CD-CTD cleft residues Leu406, Pro426, Val427	Ala	W482A
Trp489	Clp1 ID	Binds Clp1 CD residues Gln204, Met206, Arg236	Ala	W489A

Table 5.1. Mutations made in the Pcf11-Clp1 interface in the Clp1/Pcf11-EXP construct.

5.3.2. Size-exclusion chromatography

SEC experiments with recombinant Clp1-ATP binding mutant proteins revealed that some mutations caused the Clp1-Pcf11 complex to be separated into distinct chromatographic peaks. Subsequent experiments confirmed this was due to the instability of Clp1 after loss of ATP (see sections 3.2 and 3.3). To determine if mutations in the Pcf11-Clp1 interface also affected the stability of Clp1, SEC experiments with Pcf11-Clp1 mutants were also performed.

The SEC profile of recombinant wildtype Clp1-Pcf11 displayed a single peak with a retention volume of ~75 ml (Figure 5.4). The retention volume of this peak is consistent with a peak that was previously shown to contain Clp1-Pcf11 complex (see section 3.2.2). SEC of the Pcf11 Q478N mutant also produced a single peak with a retention volume of ~75 ml, consistent with a Clp1-Pcf11 heterodimer (Figure 5.4). In contrast, the SEC profile of the R480A mutant showed that protein eluted in three separate peaks (Figure 5.4). The first peak had a retention volume of ~45 ml, which is close to the void volume of the SEC column. This indicated the presence of a protein of high molar mass, and it is likely that this was aggregated Clp1 that had previously been observed in SEC experiments with Clp1-ATP mutants (see section 3.2.2). The second peak displayed a retention volume of ~82 ml, which is ~7 ml later than the wildtype Clp1-Pcf11 complex. This shift in retention volume suggested that this peak contained protein of a slightly lower molar mass than a Clp1-Pcf11 heterodimer, and was likely to be monomeric Clp1. The third peak eluted with a retention volume of ~92 ml, which was consistent with the retention volume of Pcf11 454-561 (previously characterised by SEC and SDS-PAGE, see section 3.2.2 and Figure 3.1). Similar to the R480A mutant, the SEC profiles of the W482A and W489A mutants displayed three resolved peaks, with retention volumes of ~45 ml, ~82 ml and ~92 ml (peaks 1-3 respectively, Figure 5.4), showing that these samples are also composed of a mixture of aggregated and monomeric Clp1.

SEC revealed that point mutation of Arg480, Trp482 and Trp489 in the Pcf11-Clp1 ID resulted in Clp1 instability that is characterised by a chromatographic peak containing aggregated Clp1. The resolution of three protein peaks was also observed in the SEC profiles of Clp1-ATP mutants, and it was shown that this was due to loss of Clp1-Pcf11 binding leading to instability and aggregation of Clp1 (see sections 3.2 and 3.3). It is

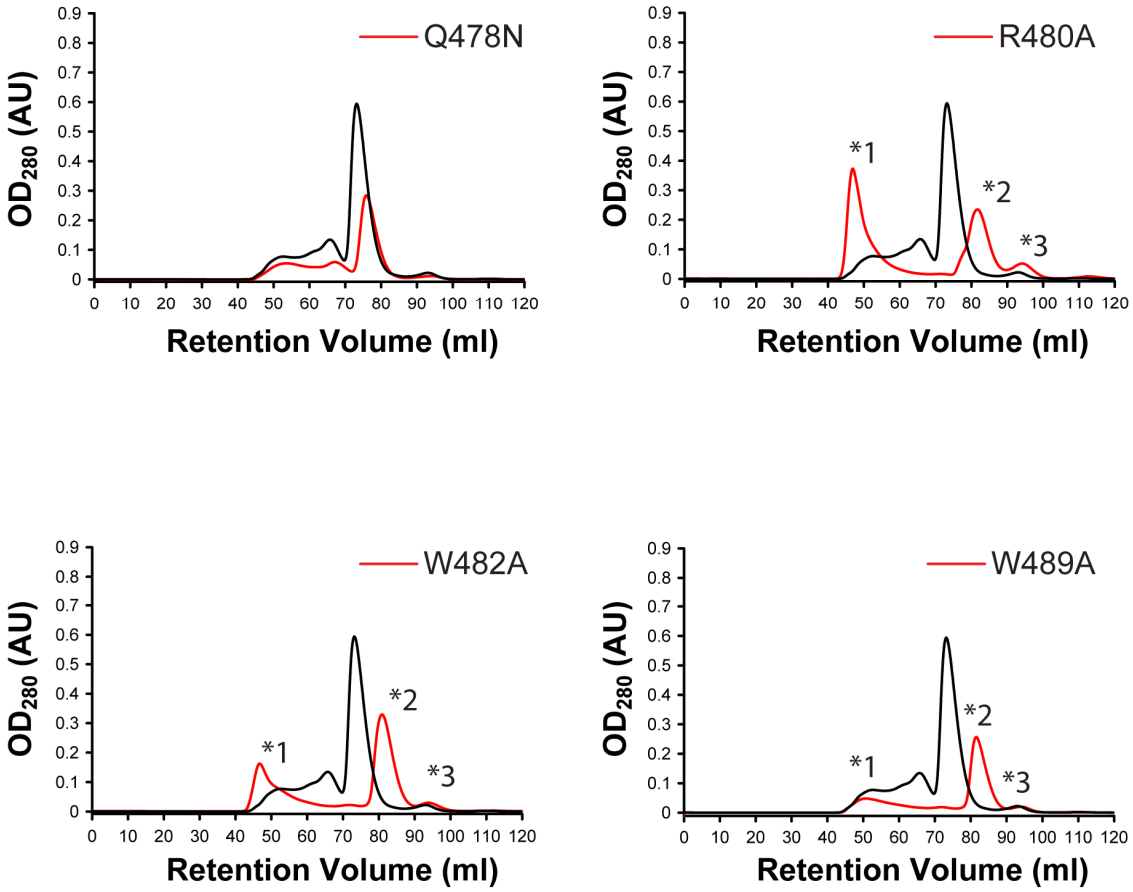


Figure 5.4. SEC analysis of Pcf11-Clp1 mutants. Chromatograms from each mutant (red line) are displayed and compared to the wildtype SEC profile (black line). Resolved peaks in the R480A, W482A and W489A chromatograms are labelled with asterisks and numbered.

therefore likely that Clp1 instability observed in SEC experiments with Pcf11-Clp1 mutants is also caused by direct abrogation of the Pcf11-Clp1 interaction.

5.3.3. Multiangle laser light scattering

SEC experiments revealed that mutation of the Pcf11-Clp1 interface may affect the stability of Clp1. To ascertain the effects of Pcf11-Clp1 mutations on Clp1 stability, SEC-MALLS experiments were conducted to determine the composition of protein species in samples purified by SEC and to look for evidence of Clp1 instability and aggregation.

In a SEC-MALLS experiment with SEC-purified wildtype Clp1-Pcf11, the signal from the differential refractive index (dRI) detector showed a peak eluted from the SEC column with a retention time of ~26 minutes (Figure 5.5). The weight-averaged molar mass of protein in this peak was determined to be 60.2 kDa, close to the expected mass of 63.8 kDa for a Clp1-Pcf11 heterodimer. Analysis of individual Debye plots throughout the eluted peak demonstrated that the molar mass distribution was narrow, consistent with a single protein species with no indication of self-association. SEC-MALLS analysis of the Q478N mutant also revealed a single chromatographic peak containing protein with a narrow molar mass distribution and a weight average of 57.4 kDa (Figure 5.5). This lies between the expected values of Clp1 (51.7 kDa) and a Clp1-Pcf11 complex (63.7 kDa). However, the column retention volume/time of Q478N in both the SEC and SEC-MALLS experiments was identical to that of a wildtype Clp1-Pcf11 complex (see section 5.3.2, Figure 5.4 and Figure 5.5). Furthermore, the distribution of molar mass derived from individual Debye plots throughout the eluted peak suggested the Clp1-Pcf11 stoichiometry was not affected in the Q478N mutant. In contrast to Clp1-Pcf11 wildtype and Q478N, the first SEC-purified peak of the R480A mutant (peak 1) eluted with a retention time of ~16 minutes, which is close to the void volume of the SEC column (Figure 5.5). The average molar mass of protein within this peak was very high (above 1.0×10^7 Da) and the molar mass distribution throughout the peak was also very broad (1.0×10^7 to 4.0×10^8 Da), similar to what was observed in SEC-MALLS experiments with recombinant Clp1-ATP binding mutants (see section 3.2.3 and Figure 3.4). A SEC-MALLS experiment with the second SEC-purified peak of the R480A mutant (peak 2) revealed two chromatographic peaks (Figure 5.6). This was somewhat unexpected as the protein purified previously as a single

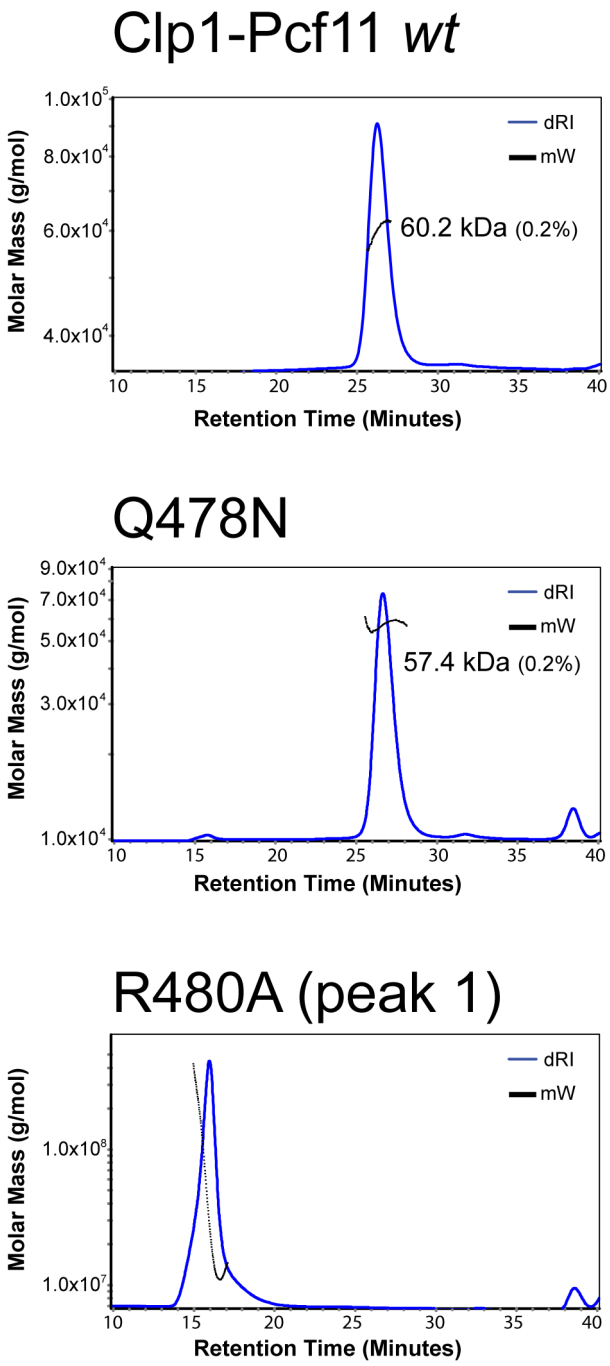


Figure 5.5. SEC-MALLS analysis of Pcf11-Clp1 mutants. Chromatograms (blue line) recorded by the differential refractometer for wildtype and mutants are shown in each panel. The black points are individual measurements of the molar mass from Debye plots generated from the LS and dRI data recorded throughout the elution of the chromatographic peak. The overall weight-averaged molar mass derived from averaging the individual measurements is displayed next to each peak along with an associated error giving an indication of spread of molar mass.

chromatographic peak using SEC. The first peak had a retention time of ~16 minutes, and contained protein of high molar mass and wide molar mass distribution ($1.0x10^7$-$1.0x10^8$ kDa), consistent with the heterogeneous aggregates of Clp1 observed previously. The second peak had a retention time of ~29 minutes, slightly greater than that of Clp1-Pcf11, suggesting the protein in this peak was likely to be of a lower molar mass than the Clp1-Pcf11 complex. Subsequent analysis of Debye plots yielded a narrow molar mass distribution and showed that the weight-averaged molar mass of protein in this peak was 53.8 kDa, close to the expected mass of 51.8 kDa for monomeric Clp1. SEC-MALLS analysis of W482A (peak 2) and W489A (peak 2) mutants revealed the characteristics of these samples were similar to those of R480A (peak 2). Both of these samples gave rise to two chromatographic peaks with retention times of ~16 minutes and ~29 minutes (Figure 5.6). The ~16 minute peaks contained wide distributions of high molar mass and the ~29 minute peaks contained proteins with weight-averaged molar masses of 50 kDa (W482A) and 52 kDa (W489A).

The molar masses of the ~29 minute protein peaks observed in the R480A (peak 2), W482A (peak 2) and W489A (peak 2) samples are close to that of monomeric Clp1. This suggests that in these cases Pcf11(453-561) no longer copurifies with His-tagged Clp1, and that mutation of Pcf11 Arg480, Trp482 and Trp489 abolishes the interaction with Clp1. By contrast, introduction of the conservative mutation Q478N into Pcf11 still supports Pcf11-Clp1 interaction and copurification of an intact Clp-Pcf11 complex. Unexpectedly, samples from R480A (peak 2), W482A (peak 2) and W489A (peak 2) mutants also contained protein of high molar mass despite the fact that they had already been purified by SEC to separate protein aggregates (Figure 5.6). The dRI of chromatographic peaks in these samples showed that the amount of protein eluting as high molar mass aggregates was similar to the ~50 kDa peaks, indicating a large proportion of Clp1 had aggregated subsequent to SEC purification. This reveals that disruption of the Pcf11-Clp1 interaction has a negative impact on Clp1 stability.

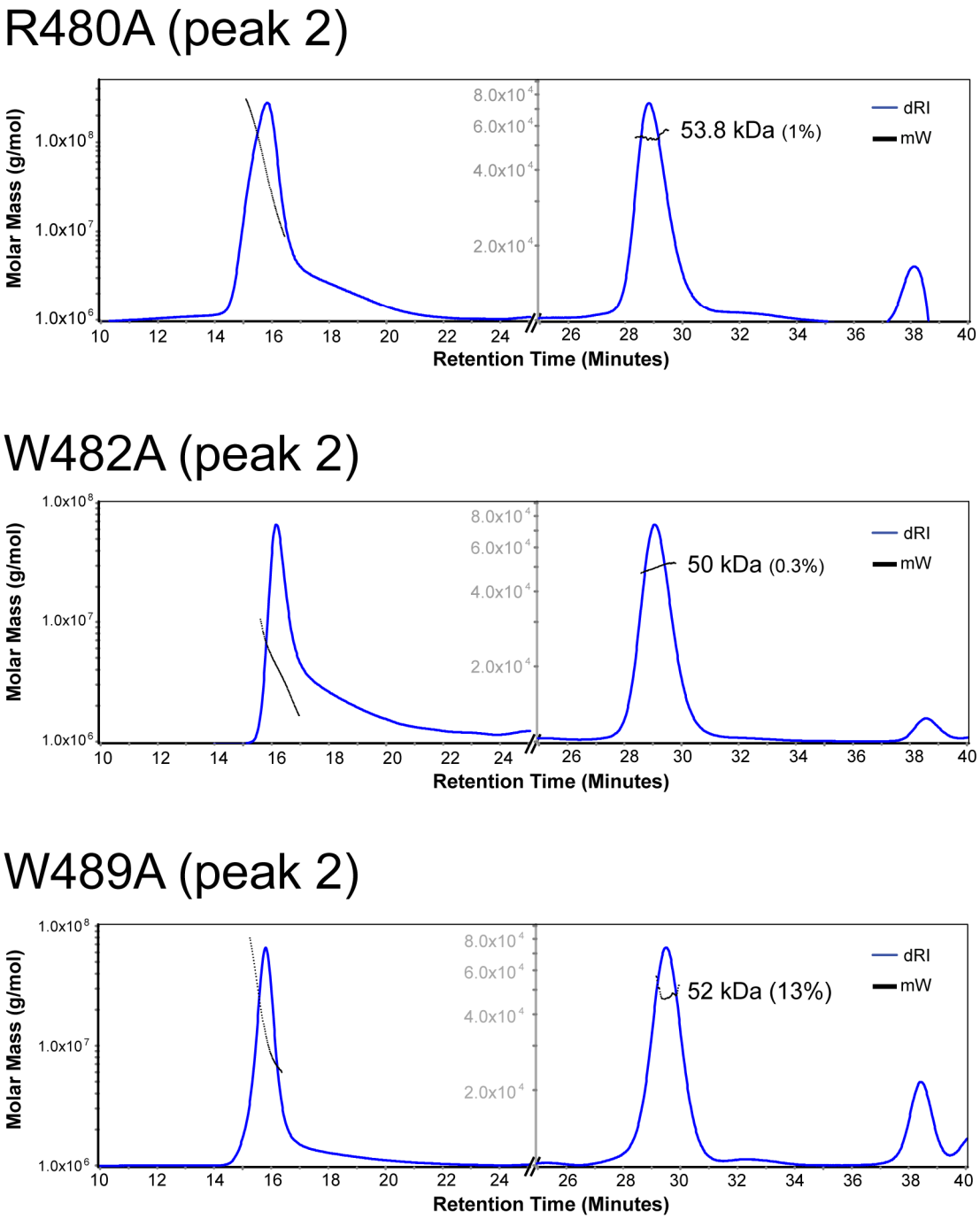


Figure 5.6. SEC-MALLS analysis of Pcf11-Clp1 mutants. Chromatograms (blue line) recorded by the differential refractometer for each mutant are shown in each panel. The *x*-axes of chromatograms are broken at 25 minutes to allow re-scaling of the *y*-axes (in grey) for better illustration of the molar mass of peaks that eluted after 25 minutes. The black points are individual measurements of the molar mass from Debye plots generated from the LS and dRI data recorded throughout the elution of the chromatographic peak. The overall weight-averaged molar mass derived from averaging the individual measurements is displayed next to each peak along with an associated error giving an indication of spread of molar mass.

5.4. The Pcf11-Clp1 interaction promotes Clp1-ATP binding and Clp1 stability

5.4.1. Mutations in the Pcf11-Clp1 ID disrupt the Pcf11-Clp1 interaction

SEC-MALLS experiments with recombinant Pcf11-Clp1 binding mutants strongly suggested that mutation of the Pcf11-Clp1 ID at Arg480, Trp482 or Trp489 abrogated the Pcf11-Clp1 interaction. However, to prove conclusively that these residues are essential for Pcf11-Clp1 binding, a reversed-phase high-performance liquid chromatography (RP-HPLC) complex composition assay was used to determine whether Pcf11-Clp1 mutations had indeed disrupted the interaction and prohibited the copurification of Pcf11 with His-tagged Clp1 from *E. coli*.

Analysis of the protein complex composition of a Clp1-Pcf11 wildtype sample showed two protein peaks eluted from the reversed-phase column with retention times of 17 and 32 minutes respectively (Figure 5.7). SDS-PAGE of the RP-HPLC-eluted proteins revealed the 17 minute peak contained Pcf11(454-561) and the 32 minute peak contained Clp1 (see section 3.2.2 and Figure 3.1C). The integration of the areas of both the 220 nm and 280 nm UV absorbance peaks corresponding to these proteins showed that the Clp1-Pcf11 complex has a stoichiometry of 1:1 (Clp1:Pcf11, for peak integration and stoichiometry calculation method see section 2.5.3). A complex composition assay with the Q478N mutant also revealed two peaks with retention times of 17 and 32 minutes (Figure 5.7). Integration of the UV absorbance peak areas demonstrated that the ratio of Clp1 to Pcf11 was 1:1, confirming the findings of SEC and SEC-MALLS experiments that mutation of Gln478 does not affect Pcf11-Clp1 binding or the stoichiometry of the complex. In contrast, analysis of R480A (peak 1) showed that only the 32 minute peak corresponding to Clp1 was present (Figure 5.7). Complex composition assays with R480A (peak 2), W482A (peak 2) and W489A (peak 2) protein samples also showed these samples only contained Clp1 (Figure 5.7).

The complex composition assays with Pcf11-Clp1 binding mutants demonstrate that mutation of Arg480, Trp482 and Trp489 results in loss of the Pcf11-Clp1 interaction, confirming that Clp1 aggregation is preceded by loss of its interaction with Pcf11. Moreover, previous SEC experiments with R480A and W482A mutants revealed a ~92 ml peak that was shown to contain isolated Pcf11 (Figure 5.4 and Figure 3.1). This suggests the mutant Pcf11 fragments do interact with Clp1 when initially expressed in

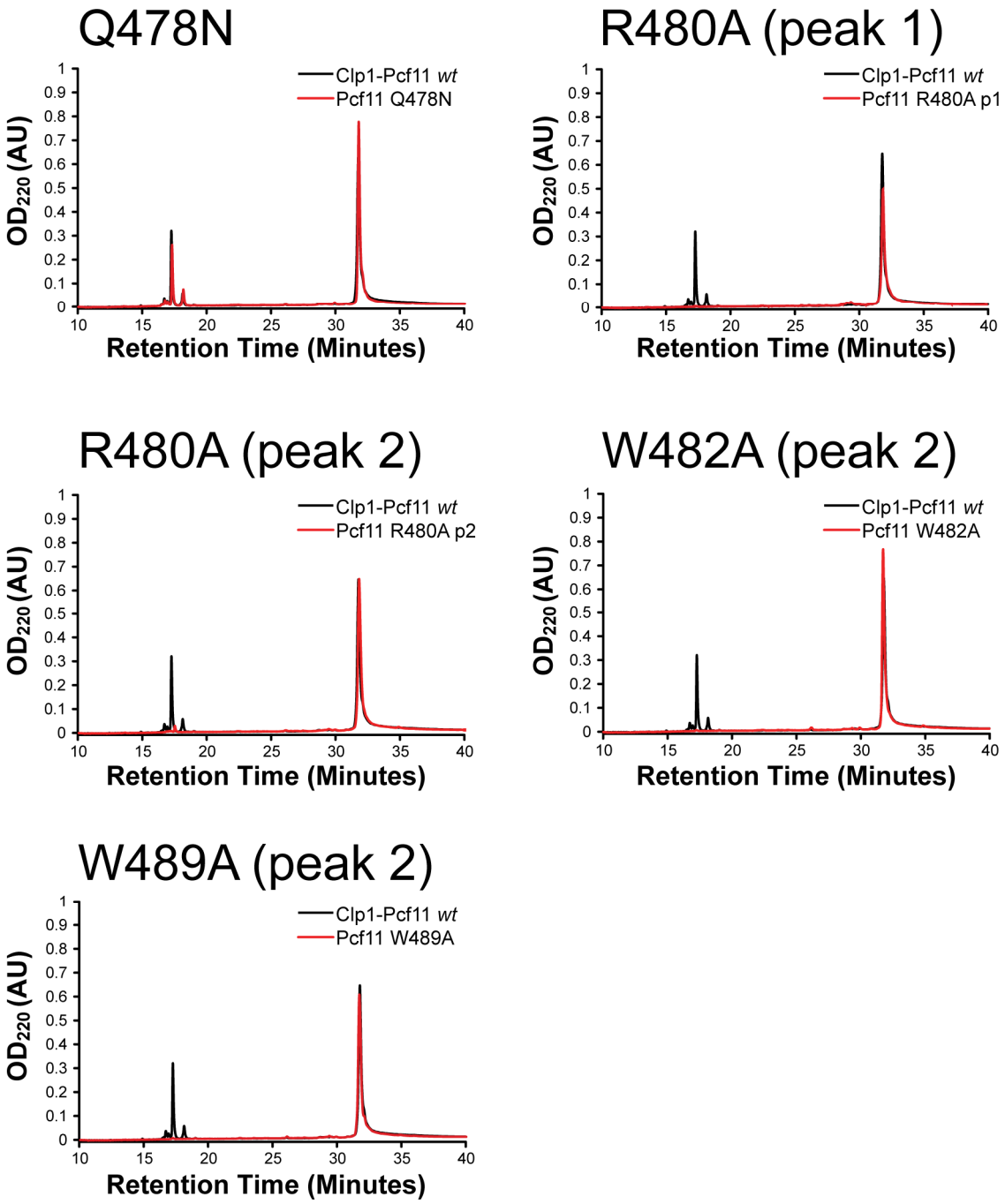


Figure 5.7. Clp1-Pcf11 complex composition assays. Chromatograms from RP-HPLC analysis of Clp1-Pcf11 wildtype and Pcf11-Clp1 binding mutants are presented. In each panel a wild type chromatogram (black) is overlaid with that from the mutant (red) indicated.

E. coli. However, it is evident that the interaction is substantially weaker and throughout purification dissociation of the complex occurs, notably prior to SEC. When considering the results of RP-HPLC complex composition assays and SEC experiments together, it is evident that mutation of Arg480, Trp482 and Trp489 weakens, but not abolishes the Pcf11-Clp1 interaction. These "weakened" complexes were actually utilised to purify isolated Clp1 with some success and enabled the raising of a Clp1 antibody (see sections 2.8.6 and 6.3). However, it was not possible to use Clp1 in direct binding experiments aimed at quantifying the Pcf11-Clp1 affinity due to the aggregation of Clp1 in the absence of Pcf11. Nevertheless, these results now posit a role for Pcf11 in the maintenance of Clp1 structure and stability, although it is unclear whether this is linked with Clp1-ATP binding.

5.4.2. The Pcf11-Clp1 interaction promotes Clp1-ATP binding

Characterisation of Clp1-ATP mutant proteins by hydrodynamic, biophysical and biochemical approaches showed that bound ATP is essential for Clp1 stability, and that the aggregation of Clp1 is preceded by dissociation of Pcf11 (see sections 3.2 and 3.3). The results of SEC-MALLS experiments and RP-HPLC complex composition assays with Pcf11-Clp1 mutant proteins demonstrated that the Pcf11-Clp1 interaction is also important for Clp1 stability, suggesting that Clp1-ATP binding and the Pcf11-Clp1 interaction may be linked. To test whether disruption of the Pcf11-Clp1 interaction affected Clp1-ATP binding, RP-HPLC nucleotide assays were performed with Pcf11-Clp1 mutant protein samples.

An RP-HPLC nucleotide assay with a sample extracted from Clp1-Pcf11 wildtype showed that only a single nucleotide species copurifies with Clp1-Pcf11 from *E. coli* (Figure 5.8A). This nucleotide had a retention time of ~8 minutes, which was consistent with the retention time of an ATP nucleotide standard. A UV absorbance spectrum taken from the eluted nucleotide showed a peak in absorbance at 260 nm, confirming the nucleotide had an adenine base (Figure 5.8A, inset). Nucleotide with a retention time identical to the ATP standard was also extracted from Clp1 in the Pcf11 Q478N mutant sample (Figure 5.8B, compare nucleotide retention time to ATP in Figure 5.8A). However, nucleotide was absent from the Clp1 in the Pcf11-Clp1 R480A (peak 1) mutant sample (Figure 5.8B), and similarly to when ATP is lost through direct mutation of Clp1-ATP binding residues, the result is protein instability (see section 3.3,

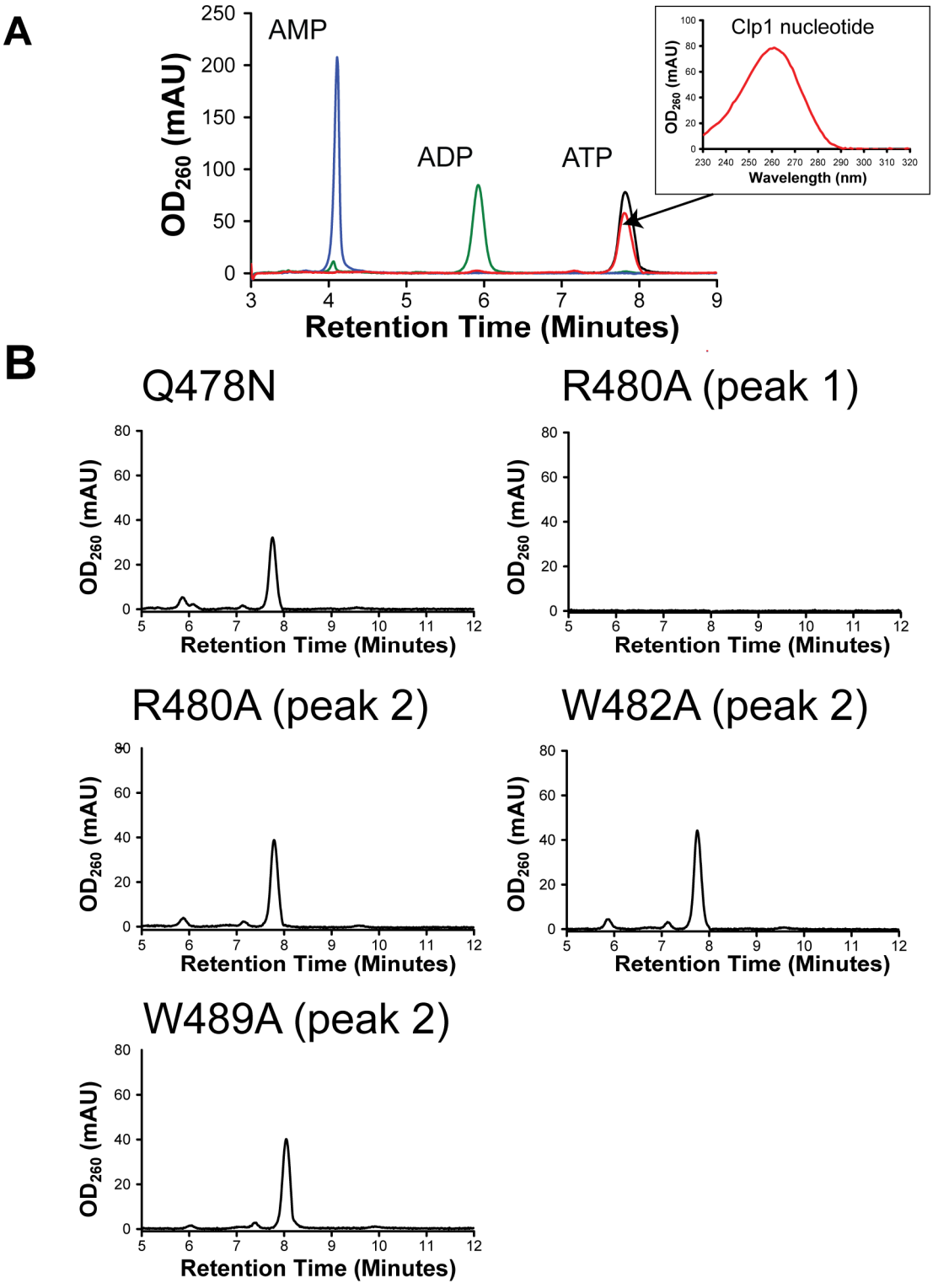


Figure 5.8. RP-HPLC analysis of the nucleotide content of Pcf11-Clp1 binding mutants. (**A**) Comparison of the retention time of the Clp1-extracted nucleotide (red line) with nucleotide standards AMP (blue), ADP (green) and ATP (black). A UV-absorbance spectrum recorded at the time of elution of the Clp1-extracted nucleotide peak is shown (inset). (**B**) Chromatograms from nucleotide assays with samples extracted from the Pcf11-Clp1 binding mutants (identities of mutants are labelled).

Figures 3.7 and 3.8). Nucleotide assays with samples extracted from peak 2 mutants (where only Clp1 is present but not mutant Pcf11 fragments) showed that these proteins do contain bound ATP (Figure 5.8B). The fact that ATP is bound in these samples suggests that the Pcf11-Clp1 interaction is not totally required for Clp1-ATP binding. However, SEC-MALLS experiments with these samples showed that large proportions of Clp1 had subsequently aggregated after SEC purification (see 5.3.3, Figures 5.5 and 5.6). Therefore, as Clp1 instability and aggregation has been shown to be a direct consequence of ATP removal (see section 3.3.2 and Figure 3.8), it is likely that some of the ATP observed in RP-HPLC nucleotide assays of peak 2 samples may have already been released from Clp1.

The findings of RP-HPLC nucleotide assays with Pcf11-Clp1 mutants show that the Pcf11-Clp1 interaction is required for the tight association of ATP with Clp1 and that bound ATP is an absolute requirement for Clp1 stability. Comparison of the nucleotide content and stability of Pcf11-Clp1 mutants also suggests that free Clp1 displays a reduced affinity for ATP compared to a Clp1-Pcf11 complex, which leads to ATP dissociation and subsequent Clp1 instability. Drawing from these results, it is evident that Clp1-ATP binding and the Pcf11-Clp1 interaction are both equally important for Clp1 stability because they are dependent upon each other. Clp1-ATP binding was found to be essential in *S. cerevisiae,* which could be due to an important role in transcription (see sections 4.2 and 4.3). Because the Pcf11-Clp1 interaction and Clp1-ATP binding are likely to be linked, it is probable that direct disruption of the Pcf11-Clp1 interaction will also be lethal in *S. cerevisiae*.

5.5. Summary and conclusions

Biophysical studies of the Pcf11 CTD reveal it is a largely disordered domain, with CD data showing it contains a small number of α and β structural elements likely confined to the zinc fingers that flank the Pcf11-Clp1 ID. Based on these observations, it is proposed that the unstructured nature of the Pcf11 CTD might serve to create the flexibility required for the interaction of Pcf11 with multiple protein partners. The biophysical characterisation of recombinant Pcf11-Clp1 binding mutants showed that mutation of Arg480, Trp482 and Trp489 in the Pcf11-Clp1 ID weakened the Clp1-Pcf11 interaction, which in turn caused ejection of bound ATP and resulted in an

unstable Apo form of Clp1. Moreover, the findings of reversed-phase nucleotide assays inferred that although Clp1 does bind ATP in the absence of Pcf11, the nucleotide was rapidly lost when Pcf11 is not bound. These results suggest that the binding of Pcf11 and ATP to Clp1 are linked events and demonstrate that that removal of either interaction destabilises the other. As both ATP and Pcf11 are critical for the maintenance of Clp1 structure and stability, it is likely that nucleotide-bound holo-Clp1 and associated Pcf11 are obligatory for formation of functional CFIA.

6. Functional analysis of the Pcf11-Clp1 interaction in *S. cerevisiae*

6.1. Introduction and overview

Numerous studies have identified that Pcf11 is essential for pre-mRNA 3'-end formation and efficient PolII transcription termination (described in sections 1.3.3 and 1.7.1). Despite this, information about the functional relevance of the Pcf11-Clp1 interaction has only been provided very recently in a single study that was published during the completion of this work. Although it was shown that the Pcf11-Clp1 interaction was required for both 3'-end processing and transcription termination, the method by which this interaction modulates these processes was not determined. Now, characterisation of recombinant Pcf11-Clp1 mutants in Chapter 5 of this thesis clearly demonstrates that disrupting the Pcf11-Clp1 affects the Clp1-ATP interaction causing nucleotide loss and resulting in Apo-Clp1 instability. Based on these *in vitro* observations, it was hypothesised that the weakening of the Pcf11-Clp1 interaction would be lethal or produce growth-restricted phenotypes in *S. cerevisiae*, similar to that caused by the *GAL1-clp1* T137A allele (see section 4.2). Therefore, in order to investigate whether the Pcf11-Clp1 interaction is important for pre-mRNA 3'-end processing and transcription termination, an *in vivo* mutational analysis was conducted to probe the phenotypes of *pcf11* Clp1-binding mutants in *S. cerevisiae*.

6.2. The Pcf11-Clp1 interaction is essential in *S. cerevisiae*

6.2.1. Previously characterised Pcf11-Clp1 mutants do not disrupt the Pcf11-Clp1 interaction in *S. cerevisiae*

To investigate the effects of Pcf11-Clp1 ID mutations in *S. cerevisiae*, the 5-FOA plasmid shuffle system was used to determine if any of the *pcf11* Clp1-ID mutants that were characterised *in vitro* could rescue the lethal phenotype of a chromosomal *pcf11* deletion (*pcf11*Δ).

The haploid *pcf11*Δ yeast strain (which contains the *URA3-PCF11* construct) does not grow on 5-FOA medium (Figure 6.1A), consistent with the lethality of the chromosomal *pcf11* deletion and the sensitivity of *URA3*+ yeast to 5-FOA. Haploid *pcf11*Δ yeast formed colonies on 5-FOA medium after transformation with the *LEU2-PCF11* construct, indicating a plasmid-borne copy of the *PCF11* gene is able to reverse

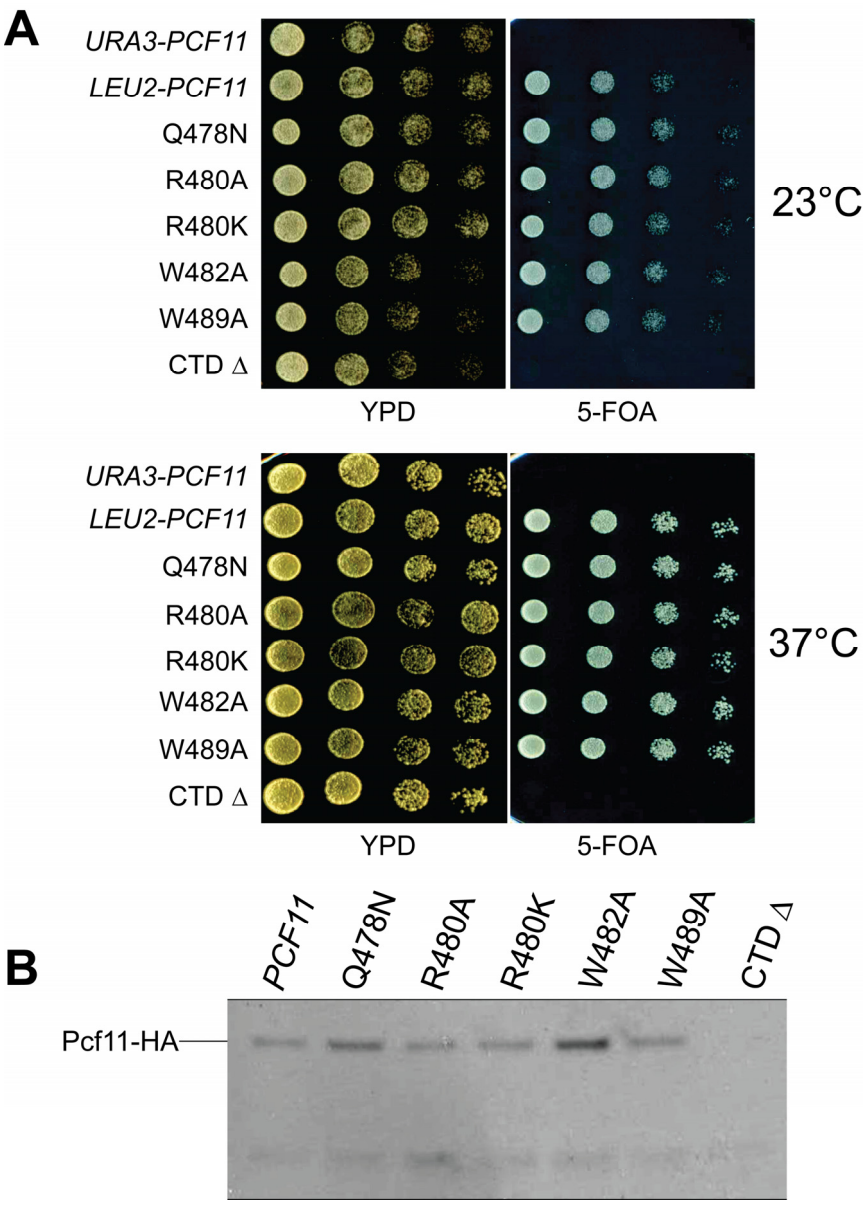


Figure 6.1. 5-FOA plasmid shuffle assay to determine the effects of single mutations in the *pcf11* gene on *S. cerevisiae* growth. (**A**) "Spot test" assays displaying the results of *pcf11*Δ complementation with *pcf11*-Clp1 ID mutant alleles performed at 23°C (top panels) and 37°C (bottom panels). Each spot represents a serial dilution (left to right: 10^4-10^1). The right panels show yeast transformants plated on 5-FOA medium and the left panels show transformants plated on YPD medium (control for growth defects in the *pcf11*Δ strain). *URA3-PCF11* (top lanes) were not transformed with *LEU2* constructs containing *PCF11/pcf11* alleles (control to check the efficacy of 5-FOA against *URA3*$^+$ yeast). *LEU2-PCF11* (second lanes from top) denotes the *pcf11*Δ strain transformed with the *LEU2-PCF11* construct (wildtype gene). The *pcf11*Δ strain was transformed with *LEU2* constructs containing each mutant allele (labelled on the left hand side). (**B**) Western blotting of protein extracts derived from *PCF11* wildtype and *pcf11* mutants to evaluate Pcf11 protein expression levels using an antibody directed against the HA-tag attached to the C-terminus of Pcf11 wildtype and mutants. Each mutant is labelled above the relevant section of the blot. The position of the expected Pcf11 band is indicated on the left-hand side of the blot.

the lethal phenotype (Figure 6.1A, upper panel). Yeast transformed with constructs harbouring the mutant *pcf11* Q478N, R480A, R480K, W482A and W489A alleles also formed colonies after 5-FOA selection (Figure 6.1A, upper panel). The density of yeast growth in each serial dilution spot revealed there is no obvious difference between the growth of these mutants and wildtype *PCF11* yeast. Western blotting was performed to determine if Pcf11 protein expression in *pcf11* yeast mutants was similar to the expression of Pcf11 in wildtype yeast. Probing *PCF11* wildtype yeast extract with an anti-HA antibody (proteins were C-terminally HA-tagged) produced two bands (Figure 6.1B). The larger band corresponded well with the molar mass of Pcf11-HA (76 kDa). The smaller band is likely a degradation product of Pcf11, or an uncharacterised isoform of the protein. Probing extracts from *pcf11* Q478N, R480A, R480K, W482A and W489A yeast mutants with anti-HA antibody showed that the intensities of these two bands were similar in all mutants compared to wildtype. This clearly demonstrates that there are no significant differences in the expression of Pcf11 mutant proteins compared to wildtype (Figure 6.1B).

The fact that the R480A, W482A and W489A mutant alleles were capable of complementing the lethal chromosomal deletion of *pcf11* was surprising given the results of *in vitro* experiments that showed these mutations weakened the Pcf11-Clp1 interaction (see section 5.3). In an effort to obtain reduced growth or lethal phenotypes, the growth temperature of the 5-FOA plasmid shuffle assay was increased to 37°C. Figure 6.1A (lower panel) shows that at 37°C, the *pcf11* R480A, R480K, W482A and W489A yeast mutants grew in all dilution spots after 5-FOA selection, demonstrating that growth of these mutants is unchanged at 37°C compared to 23°C and suggesting the gene products are stable. One possibility is that the *pcf11* R480A, W482A and W489A yeast mutants are viable due to the fact that the Pcf11-Clp1 interaction is not essential in *S. cerevisiae*. However, it is more likely that these single mutations are not sufficient to disrupt the Pcf11-Clp1 interaction *in vivo*.

To determine whether the mutant proteins expressed from *pcf11* alleles were able to interact with Clp1, coimmunoprecipitation (CoIP) of Clp1 with Pcf11 from yeast extracts derived from *pcf11* mutant yeast strains was undertaken. After immunoprecipitation (IP) of proteins associated with Pcf11 in wildtype *PCF11* yeast, Western blotting analysis with an anti-Clp1 antibody revealed a single Clp1 band

migrating at a slightly lower molar mass than the Clp1-HA positive control (Figure 6.2, upper panel). This finding affirms that wildtype Pcf11 and Clp1 are interaction partners in *S. cerevisiae*. CoIP of proteins associated with Pcf11 Q478N followed by anti-Clp1 detection revealed a band of the same size and intensity to that observed in the experiment with wildtype yeast, reaffirming the finding from *in vitro* experiments that this mutation does not affect Pcf11-Clp1 binding (Figure 6.2, upper panel). Clp1 was also detected in CoIP experiments with *pcf11* R480A, W482A and W489A mutant yeast extracts (Figure 6.2, upper panel). Furthermore, the intensity of each Clp1 band was similar to the intensity of the Clp1 band detected in the experiment with wildtype *PCF11* yeast extract. These results clearly demonstrate that mutation of Pcf11 Arg480, Trp482 and Trp489 individually is not sufficient to abolish the Pcf11-Clp1 interaction *in vivo*. CoIP experiments with extracts from the *pcf11* Q478N, R480A and W482A mutants also showed a second band of slightly lower molecular weight than Clp1 (Figure 6.2, upper panel). The identity of this band has not been confirmed, although it may represent a Clp1 degradation product. To check that the level of Pcf11-HA captured by the anti-HA agarose used in CoIP experiments was comparable in mutants and wildtype, the CoIP Western blot was stripped and re-probed with an anti-HA antibody. Figure 6.2 (lower panel) shows that similar amounts of two intense bands corresponding to Pcf11-HA and a possible Pcf11 degradation product or shorter isoform, are detected in CoIP experiments with *PCF11* wildtype yeast and all of the tested mutant yeast extracts.

In a further effort to disrupt the Pcf11-Clp1 interaction, a construct harbouring a *pcf11* CTD deletion allele (*pcf11* CTDΔ) was made (see Figure 1.3A and Appendix 8.2 for construct details). This construct was unable to restore growth to the *pcf11*Δ yeast strain after 5-FOA selection (Figure 6.1A). However, Western blotting revealed a lack of protein expression (Figure 6.1B), indicating the encoded protein is either not expressed or it is rapidly degraded.

The findings of the 5-FOA plasmid shuffle assay and CoIP experiments showed that mutation of Arg480, Trp482 and Trp489 did not affect *S. cerevisiae* viability and did not sufficiently diminish or abolish the Pcf11-Clp1 interaction *in vivo*. Based on these observations, it was apparent that in order to determine if Pcf11-Clp1 binding is essential in *S. cerevisiae* a more comprehensive mutational analysis was required.

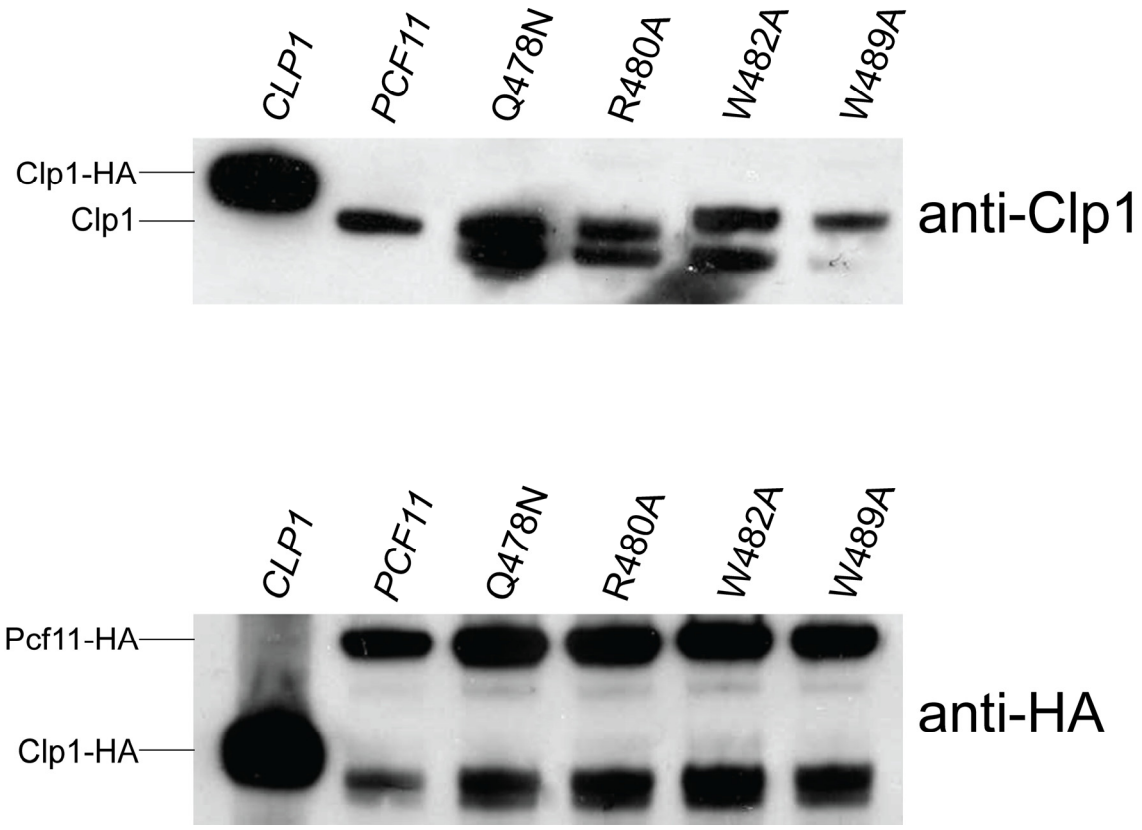


Figure 6.2. Coimmunoprecipitation experiment with yeast extracts from *pcf11* single mutants using anti-HA antibody agarose (see section 2.8.6). The top panel shows the CoIP Western blot probed with anti-Clp1 antibody and the bottom panel shows the same blot probed for Pcf11 (with the anti-HA antibody). Yeast extracts that were used in this experiment are indicated above the blots. *CLP1* is a positive control for the anti-Clp1 and anti-HA antibodies. The identity of protein bands is shown on the right-hand side of the blots.

6.2.2. Mutation of multiple conserved residues uncouples the Pcf11-Clp1 interaction and is lethal

Point mutation of the conserved Pcf11 Arg480, Trp482 and Trp489 residues did not substantially affect the Pcf11-Clp1 interaction in *S. cerevisiae*. Therefore, in a further effort to uncouple the Pcf11-Clp1 interaction in *S. cerevisiae*, *LEU2-pcf11* constructs that express double and triple Pcf11-Clp1 ID mutants were created (see Table 6.1 and Appendix 8.2). The 5-FOA plasmid shuffle system was used to determine whether these mutants could restore growth to the *pcf11*Δ strain.

Residue	Location	Function	Mutated to	Mutant name
Arg480	Clp1 ID	See Table 5.1	Ala	RW-A
Trp482			Ala	
Trp482	Clp1 ID	See Table 5.1	Ala	WW-A
Trp489			Ala	
Arg480			Ala	
Trp482	Clp1 ID	See Table 5.1	Ala	RWW-A
Trp489			Ala	

Table 6.1. Mutations made in the *LEU2-pcf11* constructs for expression of Pcf11-Clp1 ID double and triple mutants in *S. cerevisiae*.

As observed previously, the lethal phenotype of the chromosomal *pcf11* deletion was reversed by complementation with a wildtype copy of the *PCF11* gene or the *pcf11* W482A allele (Figure 6.3A). Haploid *pcf11*Δ yeast transformed with a construct containing the *pcf11* RW-A double mutant also formed colonies after 5-FOA selection. At 23°C, this yeast mutant displayed growth in all serial dilution spots that is indistinguishable from wildtype growth (Figure 6.3A, upper panel). However, when the growth temperature was shifted to 37°C, only small colonies were formed by *pcf11* RW-A yeast at the first point in the cell dilution series, indicating that the growth of the *pcf11* RW-A mutant was severely restricted (Figure 6.3A, lower panel). The inability of this mutant to grow at the higher temperature condition raised the possibility that the *pcf11* RW-A allele is temperature sensitive (further investigated in section 6.2.3). When the *pcf11*Δ strain was transformed with the *pcf11* WW-A and *pcf11* RWW-A alleles, no growth was observed after 5-FOA selection at either 23°C or 37°C (Figure 6.3A, both panels). Transformants of these alleles displayed no growth defects on the pre-5-FOA selection control (transformed *pcf11*Δ strain plated on YPD medium). To rule out that the lethal phenotypes of *pcf11* WW-A and RWW-A mutants were due to

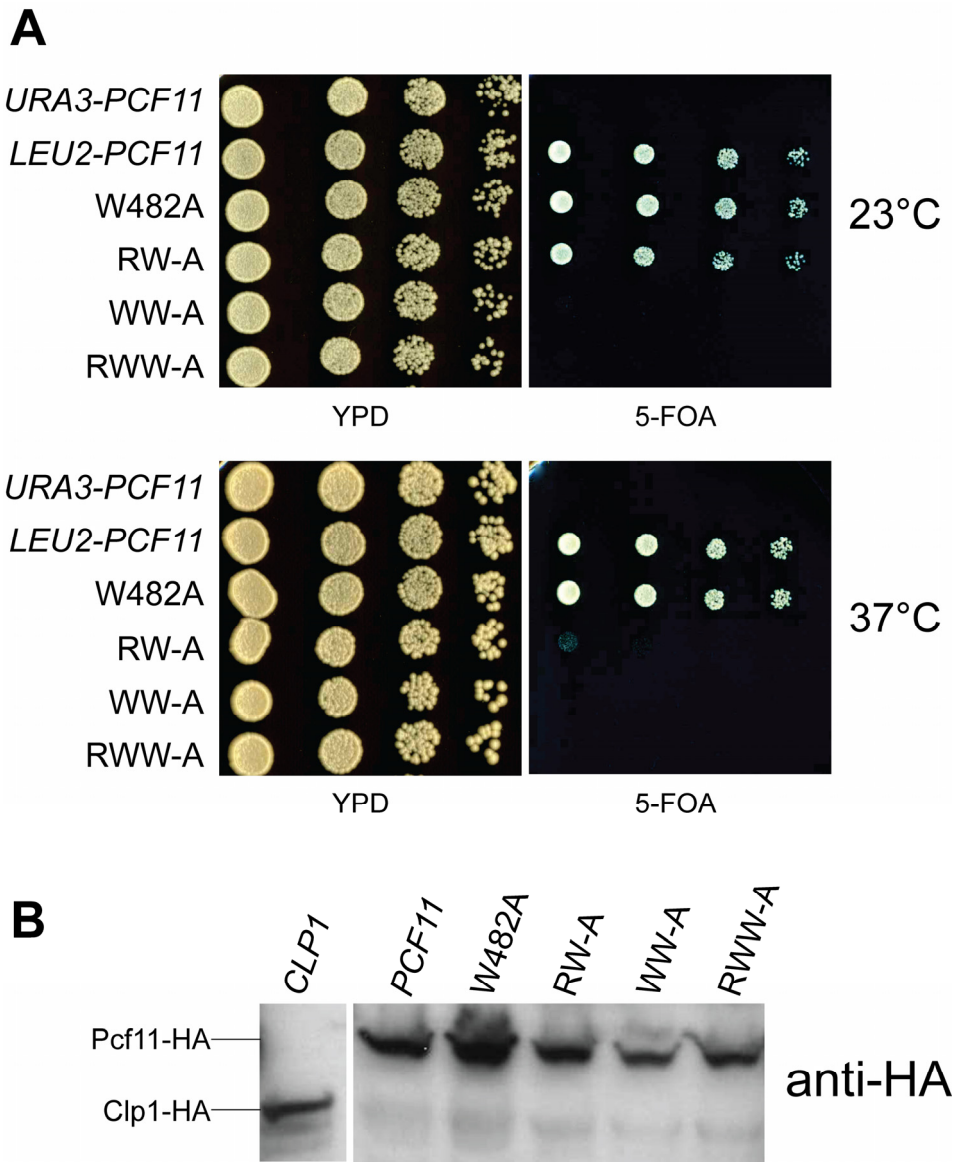


Figure 6.3. 5-FOA plasmid shuffle assay to determine the effects of double/triple mutations in the *pcf11* gene on *S. cerevisiae* growth. (**A**) "Spot test" assays displaying the results of *pcf11*Δ complementation with *pcf11*-Clp1 ID mutant alleles performed at 23°C (top panels) and 37°C (bottom panels). Each spot represents a serial dilution (left to right: 10^4-10^1). The right panels show yeast transformants plated on 5-FOA medium and the left panels show transformants plated on YPD medium (control for growth defects in the *pcf11*Δ strain). *URA3-PCF11* (top lanes) were not transformed with *LEU2* constructs containing *PCF11/pcf11* alleles (control to check the efficacy of 5-FOA against *URA3*$^+$ yeast). *LEU2-PCF11* (second lanes from top) denotes the *pcf11*Δ strain transformed with the *LEU2-PCF11* construct (wildtype gene). The *pcf11*Δ strain was transformed with *LEU2* constructs containing each mutant allele (labelled on the left hand side). (**B**) Western blotting of protein extracts derived from *PCF11* wildtype and *pcf11* mutants to evaluate Pcf11 expression levels using an anti-HA antibody. Each mutant is labelled above the relevant section of the blot, *CLP1* is a positive control for the anti-HA antibody. The position of expected bands are indicated on the left-hand side of the blot.

abnormalities in protein expression, Western blotting was performed. Detection of HA-tagged Pcf11 (with anti-HA antibody) in protein extract from wildtype *PCF11* yeast showed an intense band that correlated with the expected position of Pcf11-HA, and a less intense band of a lower molar mass (Figure 6.3B). The positions of these bands on the Western blot were consistent with those detected in a previous Western blotting experiment (see section 6.2.1 and Figure 6.1B). Western blotting of *pcf11* W482A, RW-A, WW-A and RWW-A yeast extracts revealed identical bands of similar intensities to those observed in the experiment with wildtype yeast extract (Figure 6.3B), indicating that protein expression was not significantly different in any of these mutants compared with wildtype yeast. Therefore, the inability of *pcf11* WW-A and *pcf11* RWW-A alleles to complement the *pcf11*Δ strain is due to the fact that the *pcf11* WW-A and *pcf11* RWW-A mutations are lethal in *S. cerevisiae*.

To determine if lethality was due to the uncoupling of the Pcf11-Clp1 interaction, attempts were made to CoIP Clp1 with Pcf11 from yeast extracts derived from the double/triple *pcf11* yeast mutants (CoIP performed before 5-FOA selection, see section 2.8.6). As seen previously, CoIP experiments with wildtype *PCF11* and *pcf11* W482A mutant yeast extracts demonstrated intact Pcf11-Clp1 interactions. This is evident from intense single bands observed after probing the CoIP Western blot with anti-Clp1 antibody (Figure 6.4, upper panel). Clp1 was also coimmunoprecipitated with Pcf11 RW-A, although the diminished intensity of the Clp1 band revealed that the amount of Clp1 was severely reduced (Figure 6.4, upper panel). Clp1 was not detected in CoIP experiments with extracts from *pcf11* WW-A and RWW-A yeast mutants (Figure 6.4, upper panel). To determine that diminished levels of Clp1 or absence of Clp1 was not due to poor expression of Pcf11 (and hence reduced amounts of Clp1), the CoIP blot was stripped and re-probed using an anti-HA antibody against the Pcf11 HA-tag. Bands corresponding to Pcf11-HA and a possible Pcf11 degradation product were clearly visible in extracts from *PCF11* wildtype and the *pcf11* W482A and RW-A mutants (Figure 6.4, lower panel). The intensities of the Pcf11 bands were also similar between wildtype and these mutants indicating there were no abnormalities in protein expression. Therefore, the reduced intensity of the Clp1 band in the *pcf11* RW-A CoIP experiment was likely a consequence of loss of Clp1 through the weakening of the Pcf11-Clp1 interaction. Bands corresponding to Pcf11 were also visible in *pcf11* WW-A and RWW-A yeast extracts, confirming that Clp1 was not detected in CoIP

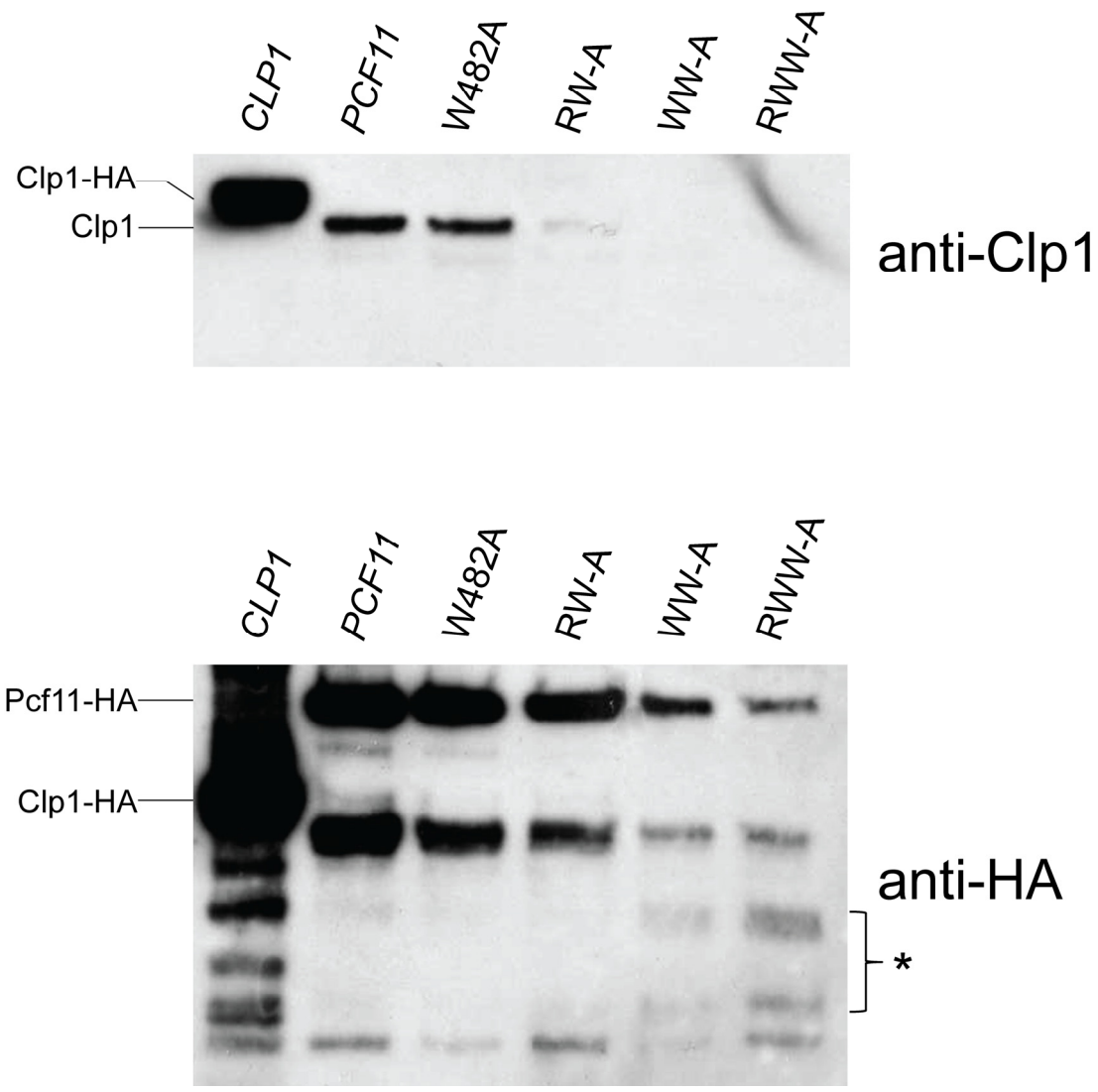


Figure 6.4. Coimmunoprecipitation experiment with yeast extracts from *pcf11* double/triple mutants using anti-HA agarose. The top panel shows the CoIP Western blot probed for bound Clp1 (with anti-Clp1 antibody) and the bottom panel shows the same blot probed for Pcf11 (with the anti-HA antibody). Yeast extracts that were used in this experiment are labelled above the blots, *CLP1* is a positive control for the anti-Clp1 and anti-HA antibodies. The identities of protein bands are indicated with the labels on the right-hand side of the blots. The asterisk indicates bands that may be Pcf11 degradation products.

experiments because the WW-A and RWW-A mutations abrogate Pcf11-Clp1 binding. However, although Pcf11 WW-A and RWW-A were detected on the CoIP Western blot, the intensities of the detected bands were significantly reduced compared to the wildtype levels (Figure 6.4, lower panel). This implies that the expression of these mutant proteins is reduced compared to wildtype Pcf11 expression. However, significant quantities of bands of lower molar mass were also present, suggesting that Pcf11 WW-A and Pcf11 RWW-A mutants are more prone to degradation than wildtype.

The characterisation of double and triple *pcf11* mutants in *S. cerevisiae* revealed that the Pcf11-Clp1 interaction is essential. The findings of CoIP experiments also suggested that Pcf11 mutants that were unable to interact with Clp1 were also prone to degradation. As it is known that these Pcf11 mutations are in an unstructured loop, it is unlikely that they cause misfolding of Pcf11 and protein instability. Instead, it is possible that Clp1-binding is required for the maintenance of Pcf11 stability. Growth of the *pcf11* RW-A mutant was significantly reduced in elevated temperature conditions, suggesting that *pcf11* RW-A could be a temperature sensitive allele. The identification of a condition that completely restricts the growth of the *pcf11* RW-A yeast mutant would be beneficial for investigation of pre-mRNA 3'-end processing and transcription defects associated with the disruption of the Pcf11-Clp1 interaction.

6.2.3. *pcf11* RW-A is a temperature sensitive allele

The findings of the 5-FOA plasmid shuffle assay described in section 6.2.2 revealed that growth could be restored to the *pcf11*Δ yeast strain after complementation with the *pcf11* RW-A allele. However, an increase in temperature to 37°C severely inhibited the growth of this mutant, which implied *pcf11* RW-A is a temperature sensitive allele (see section 4.2.2 for an explanation of temperature sensitive alleles). Despite this, because the presence of 5-FOA in growth media had enhanced negative growth effects observed in *clp1* ATP-binding yeast mutants (see section 4.2.2), it was necessary to further investigate the growth characteristics of the *pcf11* RW-A yeast mutant. To ascertain whether *pcf11* RW-A is a temperature sensitive allele, growth of the *pcf11* RW-A yeast mutant was monitored in the absence of 5-FOA (after 5-FOA selection) on both minimal and YPD media at 30°C and 39°C.

The *pcf11*Δ strain complemented with a construct containing a wildtype copy of the *PCF11* gene grew densely on YPD and minimal agar plates at both 30°C and 39°C (Figure 6.5A, upper panels). The growth of the *pcf11* RW-A yeast mutant on both YPD and minimal media was indistinguishable from wildtype at 30°C (Figure 6.5A, lower left panels). However, at 39°C growth of the *pcf11* RW-A mutant was abolished on both YPD and minimal media (Figure 6.5A, lower right panels). Although encouraging, a quantitative comparison of the growth of *pcf11* RW-A and wildtype yeast at 39°C was required, and to achieve this a growth rate assay was performed (Figure 6.5B). The assay demonstrated that the cell density of *PCF11* wildtype yeast cultured in YPD broth at 39°C increases steadily over an 8 hour period, although the exponential growth phase is not reached (Figure 6.5B). The fact that the exponential growth phase was not reached after 8 hours is likely due to reduced growth rate caused by induction of the yeast thermal stress response at 39°C [214]. Steady growth was also observed when *PCF11* wildtype yeast was cultured in minimal broth (Figure 6.5B), although the cell density reached after 8 hours was not as high as in rich media. In contrast, the cell densities of *pcf11* RW-A yeast cultures (YPD and minimal) did not change over the 8 hour time period, with cell counts after 8 hours that were ~3-fold reduced compared to wildtype cultures (Figure 6.5B).

The fact that growth of the *pcf11* RW-A yeast mutant was wildtype-like at 30°C, but completely restricted at 39°C confirmed that *pcf11* RW-A is a conditional yeast mutant in which the function of the gene product is inactivated at 39°C. The identification of the *pcf11* RW-A temperature sensitive allele provides a useful tool for experiments designed to investigate 3'-end processing and transcription abnormalities that may be associated with the uncoupling of the Pcf11-Clp1 interaction.

6.2.4. Overexpression of lethal *pcf11*-Clp1 mutant alleles does not restore growth

The work presented in chapter 4 of this thesis demonstrated that mutations that abrogate Clp1-ATP binding are lethal in *S. cerevisiae*, but lethal phenotypes can be overcome by up-regulating *clp1*-ATP mutant gene transcription (see sections 4.2.1 and 4.2.2). In an attempt to rescue the lethal phenotypes of the *pcf11* WW-A and RWW-A mutants, these alleles were overexpressed from the *GAL1* promoter (after galactose induction), and the 5-FOA plasmid shuffle assay was repeated to identify viable yeast mutants.

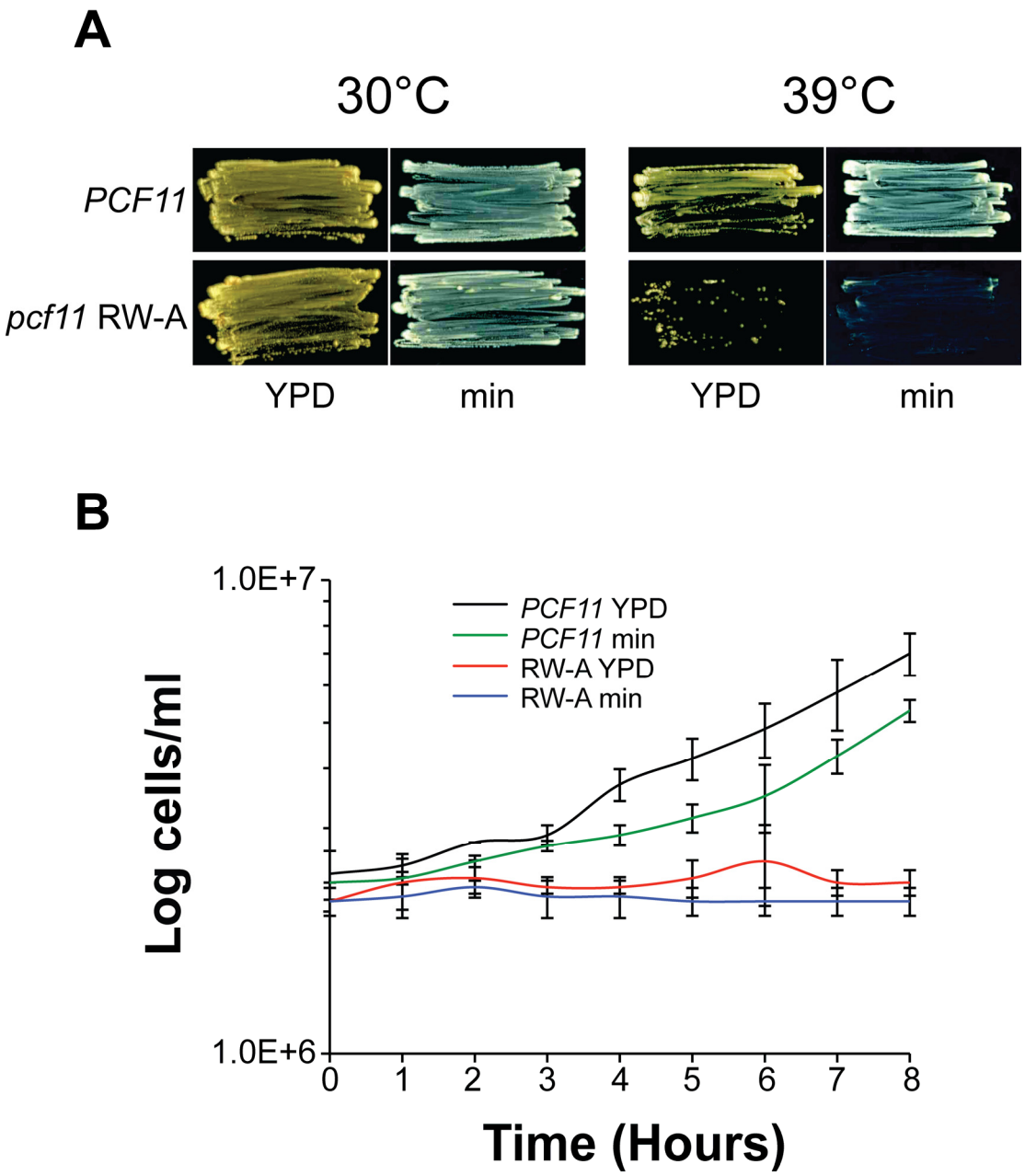


Figure 6.5. Growth characteristics of the *pcf11* RW-A yeast mutant. (**A**) Viability of *PCF11* wildtype yeast and the *pcf11* RW-A mutant on YPD agar and minimal media agar at 30°C (permissive temperature) and 39°C (restrictive temperature), growth temperatures and media are indicated in each panel. (**B**) Growth rate assay comparing the culture density of the *pcf11* RW-A mutant with wildtype yeast in both YPD and minimal media broths at 39°C. The inset key identifies each condition and error bars represent the standard deviation from the mean of a set of three replicate assays.

The haploid *pcf11*Δ yeast strain did not grow on 5-FOA in the presence of either galactose or glucose, which is consistent with the sensitivity of *URA3*$^{+}$ yeast to 5-FOA and the lethality of the chromosomal *pcf11* deletion (Figure 6.6A, central and right panels). Haploid yeast transformed with the *GAL1-PCF11* construct formed colonies on 5-FOA containing galactose (Figure 6.6A, central panel). This demonstrates that constitutive transcription of the *PCF11* gene after activation of the *GAL1* promoter by galactose reverses the lethal phenotype of the chromosomal *pcf11* deletion and that overexpression is not deletrious to yeast viability. *GAL1-PCF11* wildtype yeast also formed colonies on 5-FOA containing glucose (Figure 6.6A, right panel). However, although *GAL1-PCF11* wildtype yeast colonies were visible in all serial dilution spots on 5-FOA glucose medium, the size of individual colonies was visibly reduced compared to the size of colonies on galactose medium (see last serial dilution spot in series, Figure 6.6A, right panel). This observation suggests that glucose repression of the *GAL1* promoter did not completely suppress *PCF11* transcription, but gene expression was reduced sufficiently to produce a mild growth defect. Similarly, the *GAL1-pcf11* W482A allele also restored growth to the *pcf11Δ* strain after 5-FOA selection on both galactose and glucose media (Figure 6.6A, central and right panels). The small colonies on the glucose plate reaffirmed that there is probably a mild growth defect caused by glucose repression of *GAL1* promoter activity. The growth of *GAL1-pcf11* RW-A yeast was indistinguishable from wildtype *GAL1-PCF11* yeast on 5-FOA galactose (Figure 6.6A, central panel). However, on the 5-FOA glucose plate, growth of *GAL1-pcf11* RW-A yeast was limited to a few small colonies in the first serial dilution spot. This indicates that the reduced transcription of this mutant combined with the weakened Pcf11-Clp1 interaction caused by the *pcf11* RW-A mutation (see section 6.2.2) results in a severe growth defect. No growth was observed on galactose or glucose media after 5-FOA selection when the *pcf11*Δ strain was complemented with *pcf11* WW-A or *pcf11* RWW-A alleles (Figure 6.6A, central and right panel). However, growth of *GAL1-pcf11* WW-A/RWW-A yeast on the pre-5-FOA selection control was comparable to wildtype *GAL1-PCF11* yeast, showing that the absence of growth on 5-FOA media was not due to defects in the *pcf11*Δ strain (Figure 6.6A, left panel). To rule out that growth was not observed after complementation with these alleles because of aberrant expression of the mutant proteins, Western blotting was performed. Detection of wildtype Pcf11 by Western blotting of *GAL1-PCF11* wildtype yeast extract produced two intense bands (Figure 6.6B). The larger molar mass

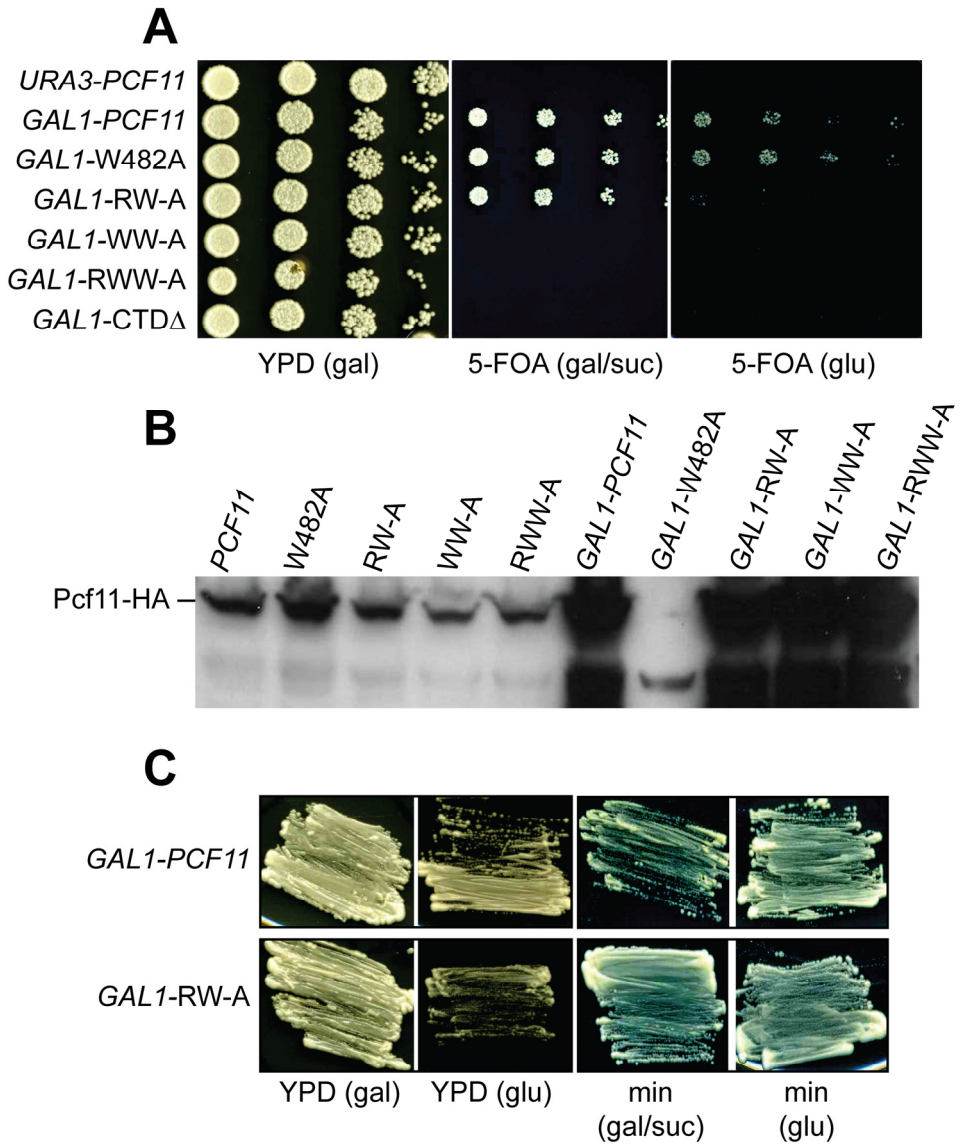


Figure 6.6. Overexpression of *pcf11* double/triple mutants from the *GAL1* promoter. (**A**) Spot test assays displaying the results of *pcf11*Δ complementation with *pcf11* mutant alleles expressed from the *GAL1* promoter. Each spot represents a serial dilution (left to right 10^4-10^1). The left panel shows yeast transformants plated on YPD medium (control for growth defects in the *pcf11*Δ strain). The central panel shows yeast plated on 5-FOA medium containing galactose and sucrose (galactose activation of the *GAL1* promoter) and the right panel shows yeast plated on 5-FOA medium containing glucose (glucose repression of the *GAL1* promoter). The *pcf11*Δ strain was transformed with *GAL1* constructs containing each mutant allele (labelled on the left hand side). (**B**) Western blotting of protein extracts to evaluate Pcf11/mutant Pcf11 protein expression levels (using anti-HA antibody). The amount of Pcf11 detected was compared (from equal quantities of protein extract) when *PCF11/pcf11* genes were transcribed from the native *PCF11* promoter (first five lanes) and from the *GAL1* promoter after galactose induction (last five lanes). Each mutant is labelled above the relevant section of the blot and the expected size of Pcf11-HA is indicated. (**C**) Glucose repression of *PCF11* wildtype and *pcf11* RW-A transcription (labelled to left of panels). The composition of the growth medium is indicated below the panels.

band was consistent with the molar mass of Pcf11-HA, and the lower molar mass band was consistent with a possible Pcf11 degradation product or smaller isoform (also observed in previous Western blotting experiments). The Pcf11 band intensities were far greater than those detected in the same amount of yeast extract derived from *PCF11* wildtype yeast, where *PCF11* transcription was controlled by the native *PCF11* promoter (Figure 6.6B). This demonstrated that the Pcf11 protein was successfully overexpressed after upregulation of *PCF11* gene transcription from the *GAL1* promoter. Western blotting of extracts derived from the *GAL1-pcf11* RW-A, WW-A and RWW-A yeast mutants also revealed two intense bands that were indicative of overexpressed protein (Figure 6.6B). Therefore, the reason that the *pcf11* WW-A and RWW-A yeast mutants did not grow after 5-FOA selection was because these alleles are lethal in *S. cerevisiae* and not because of defective protein expression. Western blotting of the *GAL1-pcf11* W482A extract failed to detect any Pcf11, and although this suggested that Pcf11 W482A expression was abnormal, it is more likely due to experimental error (incorrect loading of the SDS-PAGE gel) considering the fact that the *GAL1-pcf11* W482A allele was able to restore growth to the *pcf11*Δ strain after 5-FOA selection (Figure 6.6A).

The fact that the *GAL1-pcf11* RW-A yeast mutant did not grow well on 5-FOA glucose medium suggested that the introduction of glucose into growth media could provide a restrictive condition that may be advantageous for the extraction of mRNA and analysis of mRNA processing defects. However, because 5-FOA is a toxic compound, it was deemed necessary to rule out that its presence in growth media did not contribute to the glucose repression-induced lethality. The growth characteristics of the *GAL1-pcf11* RW-A yeast mutant were determined at 30°C on minimal agar medium containing galactose or glucose without 5-FOA. Figure 6.6C (top panels) shows that *GAL1-PCF11* wildtype yeast grew equally well on YPD and minimal media that contained either galactose or glucose. This reiterated the finding that glucose repression of *PCF11* transcription from the *GAL1* promoter was not sufficient to cause lethality. The growth of the *GAL1-pcf11* RW-A yeast mutant was comparable to growth of *GAL1-PCF11* wildtype yeast on YPD and minimal agar plates that contained either galactose or glucose (Figure 6.6C, bottom panels), demonstrating that the *GAL1-pcf11* RW-A yeast mutant does not display a conditional phenotype after glucose repression of the *GAL1*

promoter, and that 5-FOA toxicity in the initial assay was partly responsible for the observed growth defect.

GAL1-overexpression of the lethal *pcf11* WW-A and RWW-A alleles did not restore yeast viability. This result was not surprising, as CoIP experiments had previously revealed that these mutations completely abolish the Pcf11-Clp1 interaction, which is essential in *S. cerevisiae* (see section 6.2.2). The *GAL1-pcf11* RW-A yeast mutant did not display a conditional phenotype after glucose repression of the *GAL1* promoter, although this allele was already confirmed to be temperature sensitive (see section 6.2.3).

6.3. Abolition of the Pcf11-Clp1 interaction causes increased turnover of Pcf11

CoIP experiments designed to probe the ability of Pcf11 double/triple mutants to interact with Clp1 *in vivo* showed that Pcf11 WW-A and Pcf11 RWW-A appeared to be degraded and that cellular extracts contained much lower amounts of full-length Pcf11 (see section 6.2.2 and Figure 6.4). Based on this observation, it was hypothesised that the lower amounts of Pcf11 WW-A and RWW-A mutants were not due to their poor expression, but possibly the result of an increase in Pcf11 degradation directly associated with the loss of the Pcf11-Clp1 interaction. To test this hypothesis, the half-life of Pcf11 wildtype protein and the Pcf11 RWW-A mutant protein were determined using Western blotting of cellular extracts as a function of time following the inhibition of protein synthesis by cycloheximide.

As expected, upon the introduction of cycloheximide to block nascent protein synthesis, the intensity of the wildtype Pcf11 band decreased slowly over the time course, consistent with slow protein turnover (Figure 6.7A). Similarly, the Pcf11 W482A band intensity also decreased at a rate comparable to wildtype. Although the 30 minute time point has increased intensity relative to the zero time point, this is likely a loading error as in the remainder of the time course the band intensity decreases in a similar fashion to wildtype Pcf11 band intensity. Western blotting of extracts derived from the *pcf11* RWW-A showed that the intensity of the Pcf11 RWW-A band detected at the 0 minute time point was comparable to that of Pcf11 and Pcf11 W482A bands (Figure 6.7A). However, at subsequent time points the intensities of Pcf11 RWW-A bands were diminished compared to wildtype Pcf11 and Pcf11 W482A bands. Furthermore, after

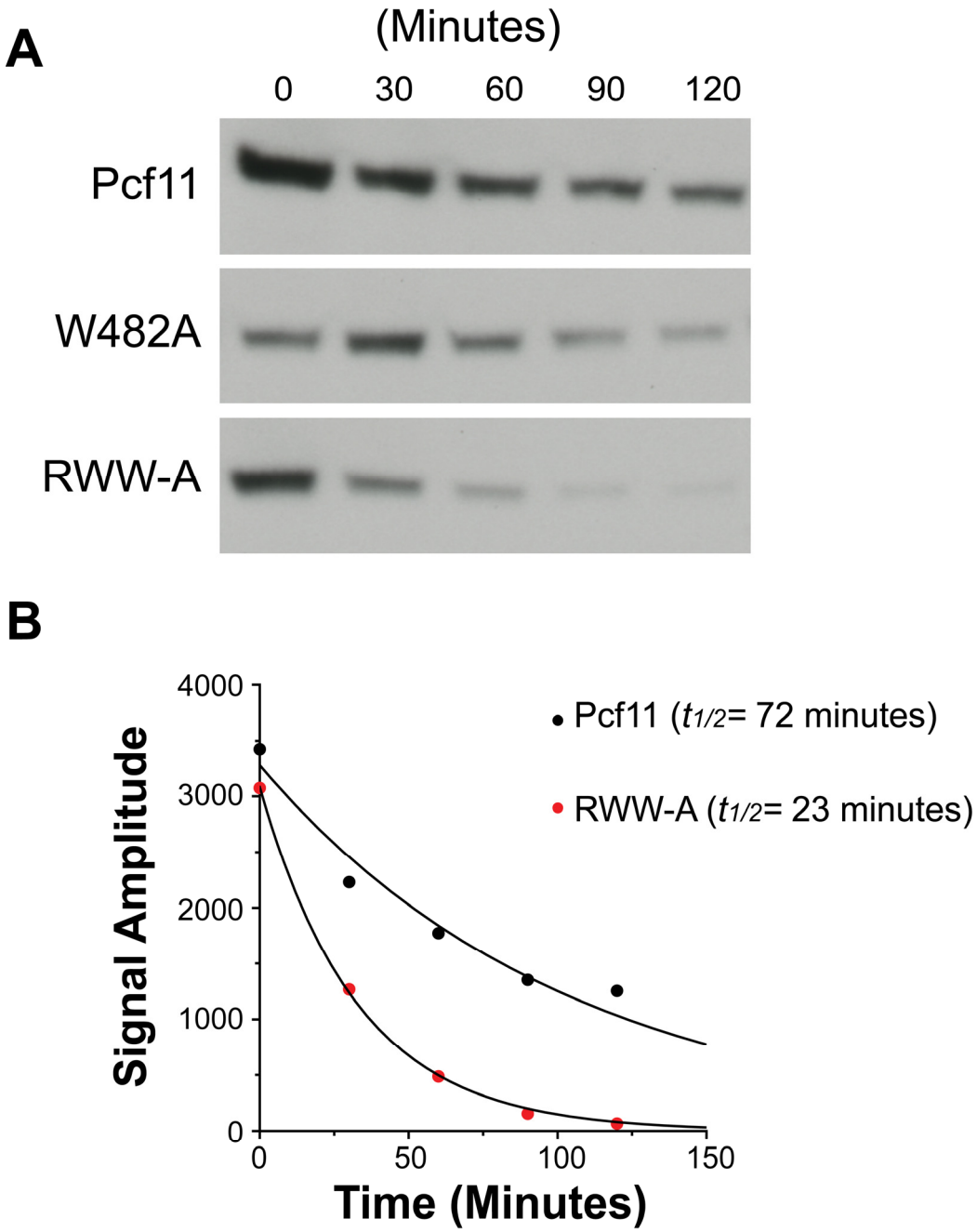


Figure 6.7. Turnover of Pcf11 and Pcf11 RWW-A after cycloheximide inhibition of protein synthesis. (**A**) Western blot analysis of Pcf11, Pcf11-W482A and Pcf11-RWW-A in yeast extracts taken at the indicated time points after cycloheximide treatment. Pcf11 and mutants (labelled on left) were detected with anti-HA antibody. (**B**) Intensities of Pcf11 and Pcf11 RWW-A protein bands detected by Western blotting (bands denoted as individual points, see inset key) were fit to an exponential decay function (black lines), allowing calculation of each protein's apparent half life (see inset key).

120 minutes only a very faint band corresponding to Pcf11 RWW-A was visible, indicating the protein was almost entirely turned over.

In order to quantify Pcf11 turnover in terms of a cellular protein half-life, the signal intensities of Pcf11 wildtype and Pcf11 RWW-A protein bands were integrated and the fall in signal intensities fit to an exponential decay function (see section 2.8.7). Plotting these data shows that the signal amplitude of the Pcf11 RWW-A band drops at a more rapid rate than the signal amplitude of wildtype Pcf11 bands (Figure 6.7B). Analysis of these data gives a calculated half-life of Pcf11 wildtype of 72 minutes. By comparison, the calculated half-life of Pcf11 RWW-A is only 23 minutes, a 3.1-fold reduction compared to wildtype Pcf11 half-life. These results confirm that the turnover of Pcf11 RWW-A is increased compared to wildtype Pcf11. Intriguingly, the reduction in the Pcf11 RWW-A half-life is coincident with the loss of the Pcf11-Clp1 interaction caused by the RWW-A mutation (determined by the CoIP experiment, see section 6.2.2 and Figure 6.4). Although there is still a remote possibility that Pcf11 RWW-A is rapidly turned over because the mutations cause instability, it is more likely these findings demonstrate that the interaction of Clp1 with Pcf11 is essential for maintaining Pcf11 levels in the cell.

6.4. The Pcf11-Clp1 interaction does not influence poly(A) site choice on *ACT1* mRNA

Investigation of the growth characteristics of the *pcf11* RW-A yeast mutant revealed it has a conditional phenotype that is affected by growth temperature (temperature sensitive mutant/allele, see section 6.2.3). The conditional phenotype enables function of the gene product supporting growth of yeast in "permissive" conditions, and inactivation of the gene product by moving to "restrictive conditions", allowing experiments that test the effects of losing gene function to be performed. To investigate the effect of uncoupling the Pcf11-Clp1 interaction on pre-mRNA 3'-end processing, total RNA samples were extracted from the *pcf11* RW-A yeast mutant after growth at permissive (30°C) and restrictive (39°C) temperatures. Total RNA samples were also extracted from *PCF11* wildtype yeast as a control. These total RNA samples were used as templates in reverse transcription quantitative real-time PCR (RT-qPCR) experiments designed to detect the quantities of specific *ACT1* mRNA sequences. The quantities of

each *ACT1* mRNA sequence were compared to formulate the pattern of cleavage/poly(A) site usage on the *ACT1* gene (see section 2.9.2 and Figure 2.7 for experimental design). Comparison of cleavage/poly(A) site (hereafter termed poly(A) site) usage between wildtype *PCF11* and *pcf11* RW-A yeast at permissive and restrictive temperatures was therefore able to determine whether the Pcf11-Clp1 interaction is important for poly(A) site selection.

When total RNA samples extracted from *PCF11* yeast grown at 30°C were used as template, RT-qPCR poly(A) site selection assays showed that 97% of *ACT1* mRNA had been processed at the first poly(A) site, indicating the first site is the major poly(A) site of this gene (Figure 6.8). *ACT1* mRNA processed at the second/third and fourth poly(A) sites comprised just 1% and 2% of total *ACT1* mRNA respectively. These results were mirrored when total RNA samples extracted from *PCF11* yeast grown at 39°C were used as template (Figure 6.8), with 95% of processing occurring at the first poly(A) site, 2% of processing at the second/third site and 3% processing at the fourth site. When total RNA samples extracted from the *pcf11* RW-A yeast mutant grown at 30°C were used as template, the poly(A) site selection assay revealed an *ACT1* mRNA poly(A) site selection pattern similar to that observed in experiments with *PCF11* wildtype RNA (Figure 6.8). The first site was selected for processing in 95% of *ACT1* mRNA and the second/third and fourth sites were used far less frequently, comprising just 1% and 4% of *ACT1* mRNA respectively. The *ACT1* poly(A) site selection assay using total RNA template derived from *pcf11* RW-A yeast that had been subjected to a period of growth at the restrictive high temperature of 39°C revealed a pattern of site selection that was slightly altered compared with wildtype (Figure 6.8). The first site was processed in 92% of *ACT1* mRNA and no *ACT1* mRNA that had been processed at the second/third poly(A) site was detected at all. Processing of the fourth site was slightly elevated at 8% of total processing. However, these differences are very small and it is still clear that the first poly(A) site on *ACT1* mRNA remains the dominant site for 3'-end processing compared to the other three sites tested in the assay.

When examining the raw data generated from RT-qPCR amplification plots, it was evident that the mean quantity of *ACT1* mRNA in *pcf11* RW-A 39°C total RNA samples was nearly 4-fold less than that observed in *pcf11* RW-A 30°C total RNA samples (see Table 6.2). Similarly, the mean quantity of *TDH2* mRNA was 2.6-fold reduced in *pcf11* RW-A 39°C samples compared to wildtype 39°C samples. However,

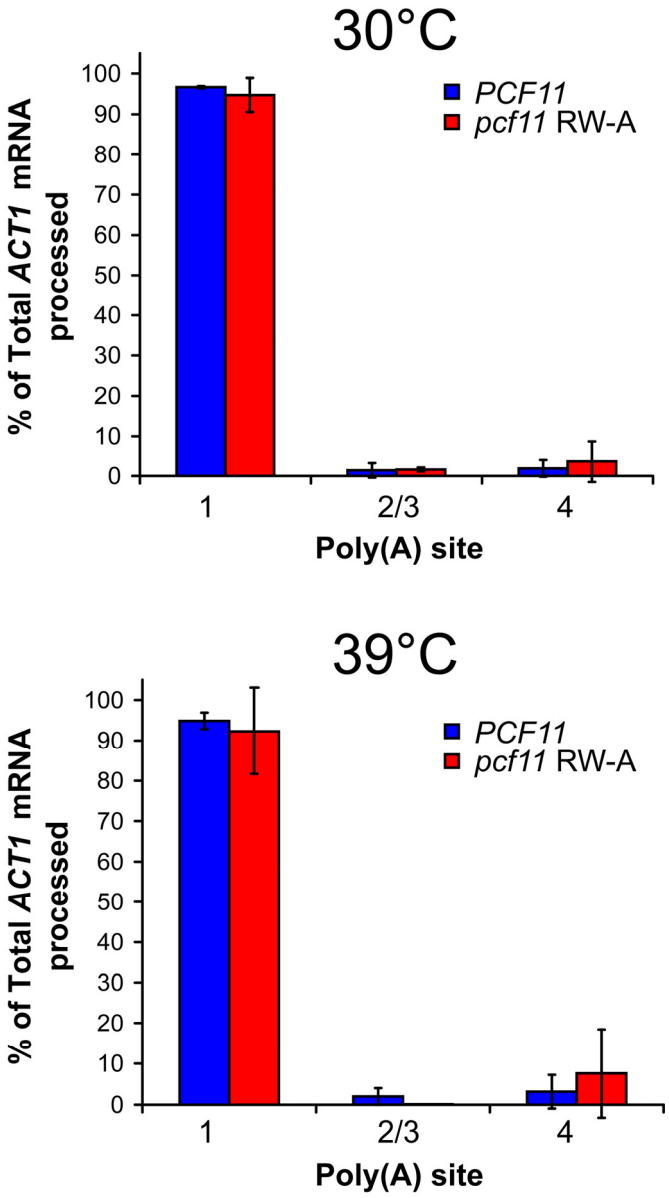


Figure 6.8. Results of the RT-qPCR *ACT1* poly(A) site selection assay with wildtype and *pcf11* RW-A yeast. The histograms show the percentages of total *ACT1* mRNA that were processed at each of three poly(A) site locations in wildtype (*PCF11*) and mutant (*pcf11* RW-A) yeast (see colour scheme inset) at 30°C (left histogram) and 39°C (right histogram). Error bars represent the standard deviation from the mean of two experiments and in each experiment targets were detected in triplicate.

the quantities of these transcripts were similar in wildtype and mutant RNA samples when yeast had been cultured at 30°C (the mean quantity of *TDH2* mRNA was 149.6 in *pcf11* RW-A samples compared to 137.15 in wildtype samples, and the mean quantity of *ACT1* mRNA was 113.15 in *pcf11* RW-A samples compared to 145.5 in wildtype samples).

The results of the RT-qPCR poly(A) site selection assay demonstrated that *ACT1* mRNA poly(A) site selection is not significantly different in the *pcf11* RW-A yeast mutant compared to *PCF11* wildtype yeast even during growth at a temperature at which *pcf11* RW-A is inviable (39°C) due to loss of the Pcf11-Clp1 interaction (see sections 6.2.2 and 6.2.3). This demonstrates that the Pcf11-Clp1 interaction is not required for poly(A) site selection on *ACT1* pre-mRNA. However, the lower quantities of *ACT1* and *TDH2* mRNAs in total RNA samples extracted from the *pcf11* RW-A yeast mutant grown at 39°C may reflect defects in transcription caused by loss of the Pcf11-Clp1 interaction.

	***PCF11* 30°C**	***PCF11* 39°C**	**RW-A 30°C**	**RW-A 39°C**
Mean Quantity *TDH2* (SE)	137.15 (10.9)	114.7 (11.45)	149.6 (15.8)	43.4 (3.9))
Mean Quantity *ACT1* (SE)	145.5 (14.25)	64.8 (5.95)	113.15 (6.15)	17.1 (0.6)

Table 6.2. *TDH2* and *ACT1* mRNA quantities determined by RT-qPCR of total RNA extracted from *PCF11* wildtype and the *pcf11* RW-A mutant. *TDH2* and *ACT1* mRNA quantities in the RW-A 39°C RNA sample are highlighted red to illustrate their reduction compared to wildtype and the RW-A 30°C quantities. The quantities displayed were determined by comparison to the relevant standard curves and are not normalised to or expressed as relative to the expression of any other gene. The figures in brackets are the standard error derived from evaluation of the mean. C_T and quantity values generated in all RT-qPCR experiments are presented in Appendix 8.6.

6.5. The Pcf11-Clp1 interaction is important for transcription

Analysis of the raw data from RT-qPCR poly(A) site selection assays revealed diminished transcription of the *ACT1* and *TDH2* genes in the *pcf11* RW-A yeast mutant at the restrictive temperature. To further investigate defects in transcription associated with the *pcf11* RW-A mutation, the quantities of mRNAs transcribed from *ACT1, TDH2, ADH1, YPT1* and *CYC1* genes were determined by RT-qPCR as a measure of the amount of the transcription of these genes. Total RNA samples extracted from

PCF11 wildtype and *pcf11* RW-A yeast grown in both permissive (30°C) and restrictive (39 C) conditions were used as template.

The mean quantity of *TDH2* mRNA detected in total RNA samples extracted from the *pcf11* RW-A yeast mutant grown at 30°C represented 88% of that detected in *PCF11* wildtype total RNA samples (Figure 6.9A). This indicates the amount of transcription of the *TDH2* gene in the *pcf11* RW-A yeast mutant is not significantly altered compared to wildtype at the permissive growth temperature of 30°C. The amounts of *ACT1* and *YPT1* transcription in the *pcf11* RW-A mutant relative to wildtype at 30°C were 113% and 85% respectively, demonstrating that transcription of these genes is also not significantly affected by the *pcf11* RW-A mutation at the permissive temperature (Figure 6.9A). The amount of *ADH1* transcription in the *pcf11* RW-A yeast mutant was slightly decreased at 72% of the wildtype amount, whereas *CYC1* transcription was increased ~1.5-fold in *pcf11* RW-A compared to *PCF11* wildtype at 30°C (Figure 6.9A). These are relatively small changes in transcriptional level that may be caused by the *pcf11* RW-A mutation. However, it was expected that any transcription phenotypes generated by the *pcf11* RW-A mutation would be more clearly visible after yeast underwent a period of growth at the restrictive temperature of 39°C. After a shift to the restrictive temperature (39°C), the mean quantity of *TDH2* mRNA detected in RNA extracted from the *pcf11* RW-A yeast mutant was just 40% of that detected in *PCF11* wildtype RNA samples (Figure 6.9A). There was also a significant reduction in mutant *TDH2* transcription relative to wildtype compared to that observed at 30°C (88% mutant relative to wildtype). The amounts of *ACT1* and *YPT1* transcription in the *pcf11* RW-A mutant relative to wildtype at 39°C were just 31% (3.3-fold reduction) and 22% (4.5-fold reduction) respectively (Figure 6.9A), showing that transcription of these genes in the *pcf11* RW-A mutant was significantly reduced at the restrictive temperature. At 30°C, *ADH1* transcription in the *pcf11* RW-A mutant was 72% of the amount observed in *PCF11* wildtype yeast. At 39°C, the amount of *ADH1* transcription in *pcf11* RW-A yeast fell further to just 27% (3.7-fold reduction) of wildtype *ADH1* transcription (Figure 6.9A). Transcription of *CYC1* was 1.5-fold increased in the *pcf11* RW-A mutant compared to wildtype at 30°C. However, at the restrictive temperature of 39°C the amount of *CYC1* transcription in the *pcf11* RW-A mutant was reduced to 38% relative to wildtype *CYC1* transcription (2.7-fold reduction, Figure 6.9A).

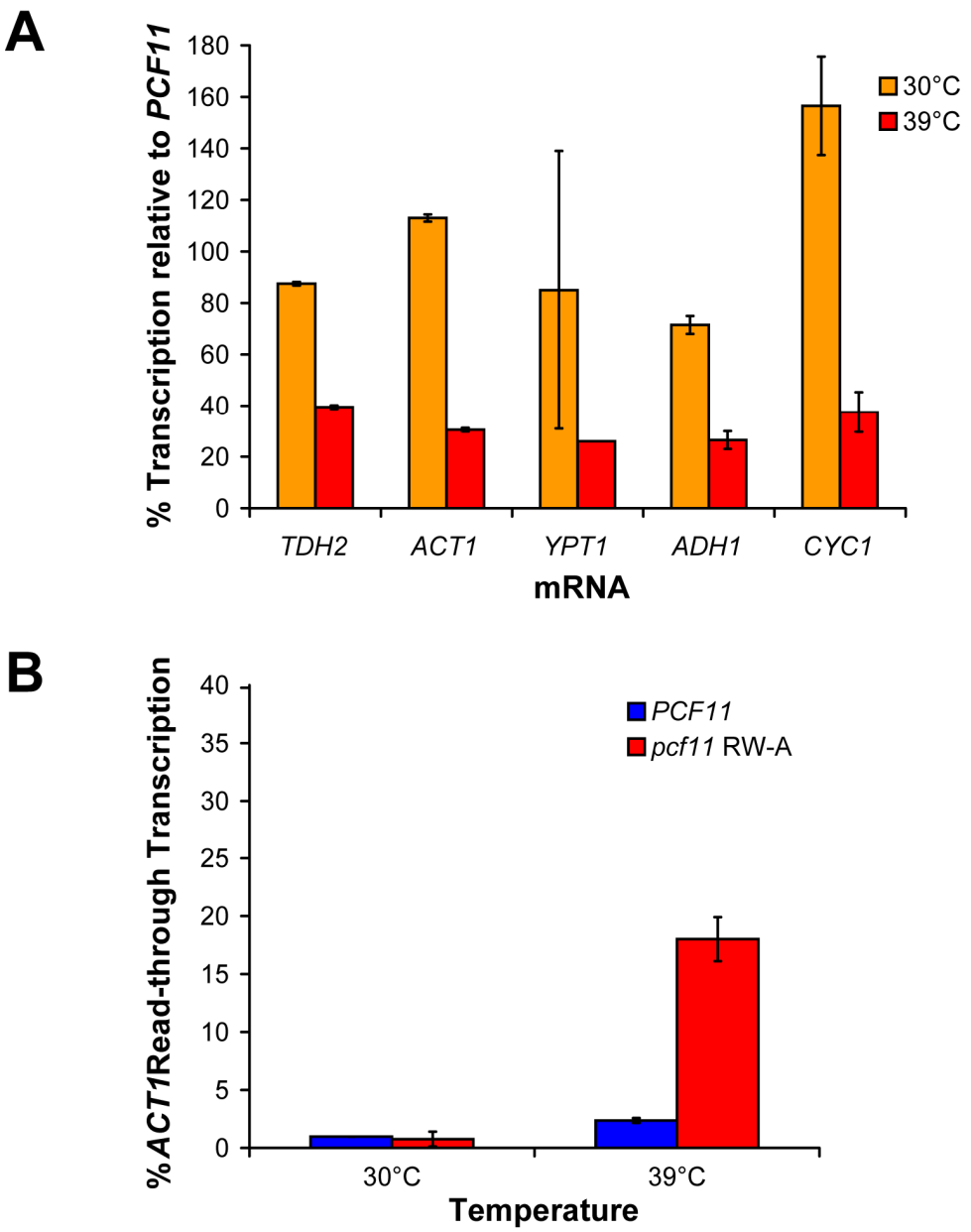


Figure 6.9. RT-qPCR analysis of gene expression and transcription termination in *pcf11* RW-A mutant yeast. (**A**) Histogram showing the level of transcription of *TDH2, ACT1, YPT1, ADH1* and *CYC1* genes in the *pcf11* RW-A mutant compared to wildtype yeast at 30°C (orange) and 39°C (red). (**B**) Histogram showing quantification of transcription termination on the *ACT1* gene (determined by the percentage of *ACT1* readthrough transcription) in wildtype yeast (blue) and the *pcf11* RW-A mutant (red) at both 30°C and 39°C. Error bars represent the standard deviation from the mean of two experiments and in each experiment targets were detected in triplicate.

The amount of transcription of the *TDH2, ACT1, YPT1, ADH1* and *CYC1* genes was unaffected by the temperature sensitive *pcf11* RW-A mutation at the permissive temperature of 30°C. However, at the restrictive temperature of 39°C, transcription of these genes is severely reduced in the *pcf11* RW-A mutant compared to wildtype. It has previously been shown that the Pcf11 RW-A mutation diminishes the Pcf11-Clp1 interaction at 30°C (see section 6.2.2 and Figure 6.4) and is lethal at 39°C (see section 6.2.2 and Figure 6.3). Considering these findings with the RT-qPCR transcription data discussed above, it is probable that the maintenance of transcription levels in *S. cerevisiae* is dependent on the Pcf11-Clp1 interaction, and that abrogation of this interaction results in severe transcription defects that are lethal to *S. cerevisiae*. A wider profile of the transcriptome of the conditional *pcf11* RW-A mutant would be advantageous to determine the proportion and classes of genes in which transcription is most and least compromised.

6.6. The Pcf11-Clp1 interaction is required for transcription termination on the *ACT1* gene

Past studies have shown that mutations in *pcf11* cause transcription termination defects in *S. cerevisiae* [106, 110, 122] (see section 1.7.1). Defective transcription termination is characterised by "transcription run-on" or "readthrough transcription", where PolII continues to transcribe after passing the most distal poly(A) site of the gene being transcribed. To determine whether the Pcf11-Clp1 interaction is important for transcription termination, a RT-qPCR assay was used to compare the amounts of readthrough transcription on the *ACT1* gene in wildtype yeast and the *pcf11* RW-A mutant. Readthrough transcription was calculated by investigating what proportion of total *ACT1* mRNA contained sequences downstream of the most distal (from the end of the *ACT1* ORF) *ACT1* poly(A) site (see section 2.9.4 and Figure 2.7 for experimental design).

RT-qPCR transcription readthrough experiments using total RNA extracted from wildtype yeast cultured at 30°C revealed that *ACT1* readthrough mRNA transcripts comprised less than 1% of total *ACT1* mRNA (Figure 6.9B). This is indicative of highly-efficient transcription termination. When wildtype yeast was cultured at 39°C, *ACT1* readthrough transcripts made up ~2% of total *ACT1* mRNA, suggesting the

increase in temperature had not dramatically affected the efficiency of transcription termination (Figure 6.9B). Total RNA extracted from the *pcf11* RW-A yeast mutant cultured at 30°C also contained a very low proportion of *ACT1* readthrough mRNA transcripts at only 1% of total *ACT1* mRNA (Figure 6.9B). However, when the *pcf11* RW-A mutant was grown in restrictive conditions (39°C), *ACT1* readthrough transcripts accounted for 18% of total *ACT1* mRNA. This represents a large increase in *ACT1* readthrough transcription compared to that observed in the *pcf11* RW-A mutant at 30°C and wildtype yeast at both temperatures, and denotes an increase in failed transcription termination under restrictive conditions.

Transcription termination on the *ACT1* gene was unaffected by growth temperature in yeast with a wildtype copy of the *PCF11* gene. However, the higher growth temperature of 39°C that is restrictive to the growth of yeast containing the *pcf11* RW-A allele caused failed transcription termination in nearly one-fifth of total *ACT1* transcription. These findings highlight the necessity of the Pcf11-Clp1 interaction for correct transcription termination on the *ACT1* gene and raise the possibility that lethality caused by loss of the Pcf11-Clp1 interaction may be the result of a combination of reduced transcription and impaired transcription termination.

6.7. Summary and conclusions

An *in vivo* analysis in *S. cerevisiae* of the effects of mutations in the *PCF11* gene showed that the R480A, W482A and W489A *pcf11*-Clp1 ID yeast mutants that abolished the Pcf11-Clp1 interaction *in vitro* did not display any growth abnormalities compared to wildtype yeast. A CoIP experiment showed that mutation of single residues in the Pcf11-Clp1 ID was not sufficient to abolish the Pcf11-Clp1 interaction *in vivo*. A more comprehensive mutational analysis demonstrated that mutation of both conserved tryptophans in the Pcf11-Clp1 ID (*pcf11* WW-A), or mutation of all three highly-conserved residues (*pcf11* RWW-A) abolished the Pcf11-Clp1 interaction *in vivo* and was lethal in *S. cerevisiae*. Determination of protein half-lives also revealed that Pcf11 RWW-A is turned over rapidly *in vivo* in comparison to wildtype Pcf11, demonstrating that the Pcf11-Clp1 interaction is essential in maintaining Pcf11 stability. A temperature sensitive allele, *pcf11* RW-A, was also uncovered, which permitted the investigation of pre-mRNA processing defects associated with loss of Pcf11-Clp1 binding. A RT-qPCR assay designed to map the pattern of poly(A) site selection on the

ACT1 gene showed that *ACT1* poly(A) site selection was not significantly different in the *pcf11* RW-A yeast mutant compared to wildtype, raising the possibility that the Pcf11-Clp1 interaction may not be required for pre-mRNA 3'-end processing. However, when the *pcf11* RW-A mutant was subjected to restrictive growth conditions, the quantities of mRNA transcribed from five different genes was significantly lower compared to wildtype yeast. This suggests that the Pcf11-Clp1 interaction might have a key role in transcription in *S. cerevisiae*. Furthermore, a RT-qPCR experiment with *PCF11* wildtype and *pcf11* RW-A RNA samples showed that transcription termination on the *ACT1* gene was severely disrupted when the *pcf11* RW-A mutant was exposed to the restrictive temperature, demonstrating that the Pcf11-Clp1 interaction is probably vital for transcription termination in *S. cerevisiae*.

These results advocate that the Pcf11-Clp1 interaction is obligatory for the maintenance of the stability of both proteins, which is likely important for the integrity of CFIA. Evidence points towards the Pcf11-Clp1 interaction as essential for PolII transcription and transcription termination in *S. cerevisiae*.

7. Discussion

7.1. Clp1, ATP and the enzyme question

Since the elucidation of the Clp1-Pcf11-ATP crystal structure, research groups working in the pre-mRNA 3'-end processing field have postulated that Clp1 might be capable of hydrolysing its bound ATP. Specifically, it has been suggested that Clp1 may be an ATPase or a protein kinase that is responsible for the conformational rearrangements that are necessary for 3'-end cleavage [116, 121]. The findings presented in this thesis are contrary to these theories. A reversed-phase nucleotide assay demonstrated that a recombinant Clp1-Pcf11 complex contained bound ATP at a high stoichiometric ratio, whereas if Clp1 was capable of ATP hydrolysis it would be expected to release the nucleotide and/or that traces of ADP or even AMP might be present. The ATP bound in Clp1 was also shown to be non-exchangeable with free-ATP, suggesting the Clp1-ATP interaction is tight and that a continuous cycle of ATP hydrolysis is unlikely. Furthermore, mutation of the putative catalytic residue in the Clp1 Switch I motif (Asp161) did not appear to affect transcription or pre-mRNA 3'-end processing in *S. cerevisiae*. Although this does not prove conclusively that Clp1 is not catalytically competent (any enzymatic function of Clp1 could be redundant and hence viability of the yeast mutant may not be affected), taken with the aforementioned *in vitro* results it seems the most probable scenario. In support of this, *in vitro* experiments with recombinant Clp1 that were carried out in past studies have failed to detect any ATPase activity (I. Taylor, unpublished data) or DNA/RNA/protein kinase activity [189].

It is curious that Clp1 should contain a domain with conserved structural elements that secure its membership in a family of ATPases (SIMIBI-class NTPases [119]), yet possess no catalytic potential itself. Indeed, the two other SIMIBI-class members that are closest in structural relation to Clp1 are the active ATPases NifH (*Azobacter vinelandii*) and ArsA (*E. coli*) [216-218]. These proteins utilise ATP hydrolysis to transmit protein conformational changes. Recent studies that have aimed to characterise the human homologue of Clp1 (hClp1) may shed light onto why Clp1 has no catalytic function. The P-loop, Switch I and Switch II motifs are all highly-conserved between Clp1 and hClp1 (see Figure 1.5C and Appendix 8.1), and these proteins share sequence identity of 23% overall [117]. hClp1 is a mammalian pre-mRNA 3'-end processing

factor that is essential for transcript cleavage but not polyadenylation [117]. However, an intriguing difference between Clp1 and hClp1 is that hClp1 kinase activity is well documented. hClp1 has been identified as the kinase that phosphorylates the 5'-hydroxyl of synthetic short interfering RNAs (siRNAs) used for RNA interference (RNAi) [188]. However, the natural kinase function of hClp1 is to phosphorylate the 5'-hydroxyl of 3'-tRNA exons in the yeast-like tRNA splicing pathway in vertebrates [189]. Ramirez *et al.* have found that hClp1 can rescue both growth-defective and lethal phenotypes in *S. cerevisiae* caused by mutations or deletion of the gene that encodes the yeast tRNA ligase (*trl1*Δ) [189]. This was shown to be dependent upon the kinase activity of hClp1, as the *hclp1* Asp151 mutant (equivalent residue to Switch I Asp161 in yeast Clp1) was not able to restore growth to wildtype levels. These findings led the authors to postulate that despite their similarity, the yeast and human Clp1 proteins are not functional orthologues. Although the function of hClp1 as a tRNA splicing enzyme has been made clear, studies investigating other mammalian homologues of yeast-type tRNA splicing enzymes have suggested that the yeast-type tRNA splicing pathway may be redundant in mammalian cells [219, 220].

Considering the information available on hClp1 with our own findings, we postulate that yeast Clp1 was also a kinase involved in the yeast tRNA splicing pathway at some point in the past. However, kinase activity has since been lost been lost through a lack of evolutionary selective pressure due to its functional redundancy. There are a number of points to reinforce this argument: (1) the stringent conservation of the motifs necessary for kinase/ATPase between Clp1 and hClp1 reflecting a conserved function; (2) the absence of kinase or ATPase activities in recombinant Clp1; (3) the existence of an enzyme in yeast (Trl1) that performs all steps of the tRNA splicing pathway, which could have led to the redundancy of Clp1 as a tRNA splicing enzyme; (4) the possible redundancy of the yeast-type tRNA splicing pathway in humans; (5) evidence that phoClp1, an "ancestral" Clp1 protein from the archaeal species *Pyrococcus horikoshii* is an RNA kinase that can act in lieu of yeast Trl1, suggesting that the common ancestor of both Clp1 and hClp1 is a kinase [221] .

In terms of pre-mRNA 3'-end processing, it is highly likely that the role of Clp1-ATP binding is purely to maintain Clp1 structure, which is important for the integrity of CFIA. This will be discussed in section 7.2. It has been posited that hClp1 may utilise its kinase activity to maintain the 5'-phosphate group of the 3' nascent mRNA chain

(after pre-mRNA cleavage) to allow entry of the Xrn1 nuclease and subsequent transcription termination (torpedo model of transcription termination, see section 1.7.1). However, there is no evidence to support this, and instead, we hypothesise that hClp1 also plays a structural role that bridges the 3'-end processing complexes as has been previously theorised [117].

7.2. Clp1-ATP binding and the Pcf11-Clp1 interaction are linked

The comprehensive analysis of recombinant mutant proteins presented in this thesis demonstrates that both ATP and Pcf11 are crucial for the stability of Clp1. Figure 7.1 summarises the events that lead to Clp1 instability, showing that disruption of the interaction of Clp1 with either ATP or Pcf11 also results in the loss of the other interaction. This suggests that Clp1 contains a structural element that links the nucleotide binding pocket and the Pcf11 ID. However, in the Clp1-Pcf11 crystal structure no obvious link is apparent. It is instead possible that the binding of either ATP or Pcf11 results in a small conformational change in Clp1 that presents or extends the binding interface for the other molecule. Another alternative is that loss of the Clp1-Pcf11 interaction (which is always a prerequisite to Clp instability) brings about a change in Clp1 that promotes instability. Examination of the Clp1-Pcf11 crystal structure shows that two of the highly-conserved Pcf11 residues (Trp482 and Trp489) make hydrophobic contacts with Clp1 residues in a cleft between the Clp1 CD and CTD (see Figure 1.4B). This could be significant, as abrogation of the Pcf11-Clp1 interaction potentially exposes a hydrophobic region of Clp1 promoting instability, self-association and aggregation *in vitro*. To ascertain if this was a viable theory, hydrophobic areas of the Clp1 surface were mapped to the structure of Clp1 using a hydrophobicity index described by Fasman [222], the results of which are presented in Figure 7.1B. This analysis revealed that the Clp1 CD-CTD cleft is significantly hydrophobic, and that in the absence of Pcf11 this would become exposed. Moreover, the fact that the Clp1 CD-CTD cleft is large and that Pcf11 is so intimately associated within it suggests that loss of Pcf11 would likely result in substantial changes in Clp1 structure, most likely the collapse of the walls of the CD-CTD cleft. Therefore, based upon the results of our experiments, details of the Clp1-Pcf11 structure and the hydrophobic map, we postulate that the abrogation of the Clp1-Pcf11 interaction causes both a loss of Clp1 structure and the presentation of a hydrophobic surface in Clp1, both of which are detrimental to Clp1 tertiary structure and stability.

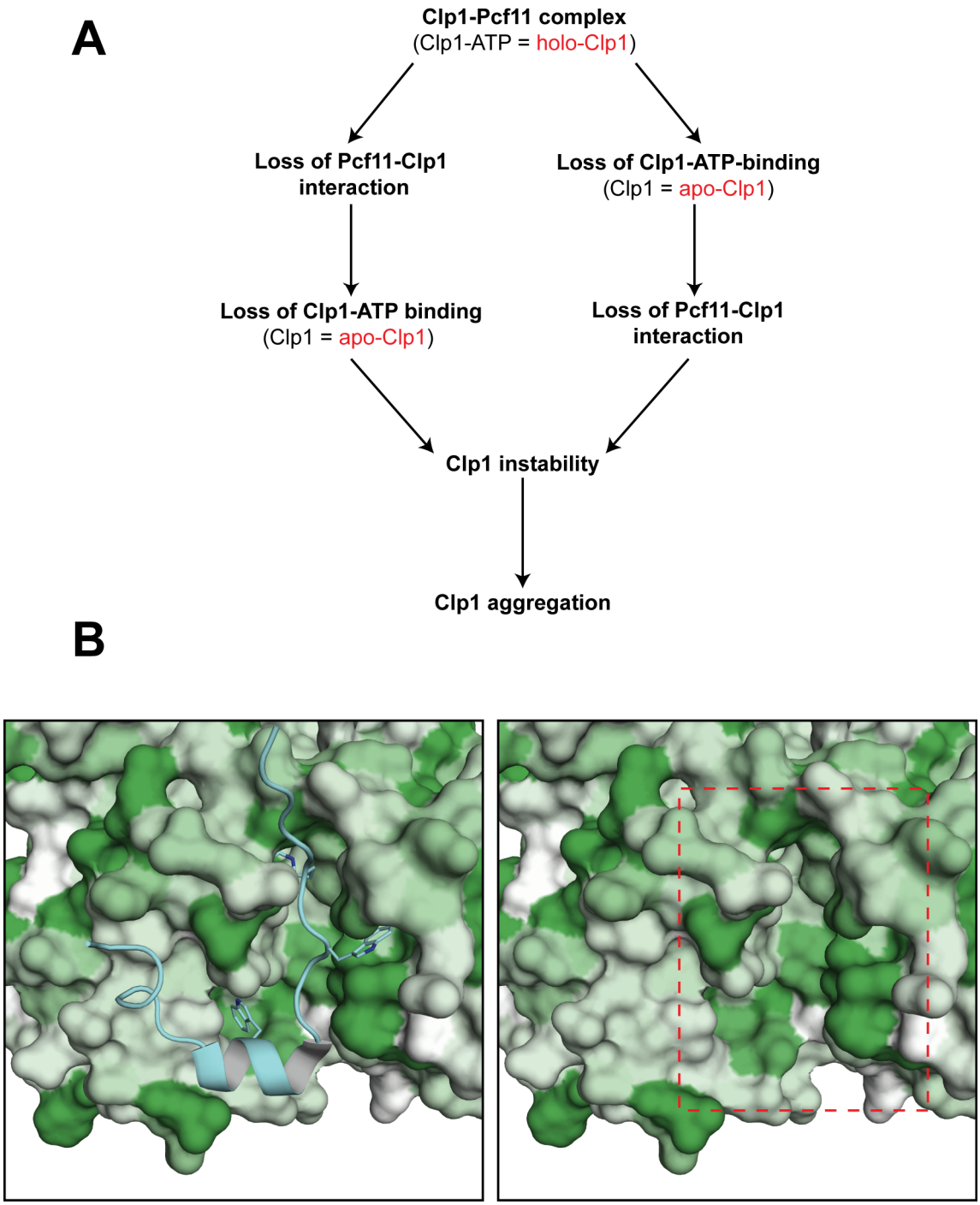


Figure 7.1. The events that lead to Clp1 instability and aggregation. (**A**) Schematic to show that Clp1-ATP binding and the Pcf11-Clp1 interaction are linked events and removal of one interaction destabilises the other resulting in Clp1 instability and aggregation. (**B**) Map of hydrophobic areas observed in the Clp1-Pcf11 crystal structure focusing on the Clp1 central domain-C-terminal domain cleft. A hydrophobic residue index described by Fasman [222] was applied to a surface representation of Clp1 using deepening shades of green to illustrate increasing hydrophobicity. The structure is shown with (left panel) and without (right panel) the bound Pcf11 fragment (cyan cartoon with side chains of Arg480, Trp482 and Trp489 depicted as sticks). The dashed red box indicates the Clp1-Pcf11 interaction site. *PDB code: 2NPI.*

A recent study has downplayed the importance of ATP as a structural element of Clp1. As was observed in this study, Holbein *et al.* found that recombinant Clp1 mutants carrying the K136A or T137A mutations were extremely unstable and resulted in protein aggregation [180]. However, they also claimed that these mutant *clp1* alleles were viable in *S. cerevisiae* and produced no obvious transcription or 3'-end processing abnormalities. An explanation for these seemingly paradoxical findings that calls into question their validity, is that the two alleles were introduced into the *clp1*Δ yeast strain on 2-μ plasmid shuttle vectors that replicate to high copy number in yeast cells (50-100 per cell). In experiments described in this thesis, Clp1 mutants were introduced on the centromeric plasmid pRS315, which is maintained at low copy-number (1-2 per cell). Consequently, it is likely that the *clp1* K136A and T137A yeast mutants were deemed viable due to a copy-number induced overexpression of mutant Clp1 proteins from these alleles, which increases a very small portion of ATP-bound functional Clp1. A similar effect was observed in experiments presented in chapter 4 of this thesis, when the K136A and T137A mutants were overexpressed from the *GAL1* promoter. Therefore, we proffer that the true effect of abrogating Clp1-ATP binding is lethality. In another study, Ghazy *et al* demonstrate that *clp1* K136E and T137E mutants are unable to interact with Pcf11 and are lethal in *S. cerevisiae*, but they do not associate this with a loss of Clp1-ATP binding [122]. Rather, they postulate that these mutations trap Clp1 in a conformation in which ATP is repositioned in the nucleotide-binding pocket. They speculate that the nucleotide-binding pocket has "extra-space" that could accommodate the differential orientation of ATP as a molecular switch that modulates the Pcf11-Clp1 interaction. By contrast, work presented in this thesis clearly shows that mutation of Lys136 or Thr137 in the Clp1 P-loop abrogates Clp1-ATP binding *in vitro*.

We therefore conclude that both the Clp1-ATP and Pcf11-Clp1 interactions are required to maintain stable Clp1 for incorporation into the CFIA complex, and that the lethality observed upon removal of the Clp1-ATP or Pcf11-Clp1 interactions is due to a structural collapse in Clp1 that prevents the formation and essential functions of CFIA.

7.3. Subunit composition of CFIA during transcription

Previously, it was thought that CFIA was recruited as an intact unit to the PolII elongation complex through the interaction of PolII CTD with the Pcf11 CID [50].

Crucially, the recruitment of CFIA was found to be dependent on ongoing transcription and the maintenance of PolII CTD Ser2 phosphorylation, which increases throughout transcription elongation as PolII reaches the 3'-ends of genes [173, 178]. From evidence presented here and from recently published data [169, 170], we now propose that rather than recruitment of intact CFIA, the subunit composition of CFIA changes throughout transcription.

This idea is based upon our finding that disruption of the Pcf11-Clp1 interaction causes instability and aggregation of Clp1, and that we hypothesise that this results in the abolition of CFIA assembly causing lethality in yeast mutants. This hypothesis was upheld by the findings of another study that was published during the collation of the work presented here. Haddad *et al*. found that mutations in Clp1 that negated Clp1-Pcf11 binding also prevented the interaction of Pcf11 with Rna14 and Rna15 (or in other terms a complete disintegration of CFIA) [121]. Furthermore, the same mutations correlated with loss of transcription termination *in vivo* and 3'-end formation *in vitro*, reiterating the fact that Clp1 and intact CFIA are obligatory for both processes. Despite the importance of Clp1 to CFIA, there is evidence to suggest that CFIA does not always contain Clp1. The mRNA export factor Yra1 has been shown to interact directly with Pcf11 in a manner that is mutually exclusive with Clp1 [169, 170] (see section 1.6.3). Consistent with this, the ChIP profiles of Clp1, Yra1 and Pcf11 across a number of genes also suggest the subunit composition of CFIA may be flexible. The peaks of Clp1 and Pcf11 recruitment were shown to coincide near the poly(A) sites in the 3'-UTRs of some genes, whereas Yra1 recruitment peaks near the 3' ends of ORFs and subsequently declines. In addition, Yra1 and Pcf11 are recruited steadily throughout ORFs whereas Clp1 seems to be specifically recruited at 3'-ends [170].

In light of the observations from the studies outlined above, we postualte that the subunit composition of CFIA changes during PolII-mediated gene transcription and that CFIA is recruited progressively throughout transcription of a gene. It is likely that Clp1 is not initially incorporated, and that instead Yra1 maintains the interaction of Pcf11 with Rna14 and Rna15. At a later stage of transcription (possibly after transcription of the poly(A) site), Clp1 replaces Yra1 in CFIA. It is possible that at this point Yra1 is "handed off" to the pre-mRNA to assist in its essential role as an mRNA export factor. An analysis of the half-lives of Pcf11 and the Pcf11-Clp1 binding defective Pcf11 RWW-A mutant presented in section 6.3 demonstrated that Clp1 is essential in

protecting Pcf11 from protein degradation. Yra1 may also act to maintain Pcf11 stability during transcription elongation, where Clp1 is absent from CFIA. The significance of CFIA formation in view of pre-mRNA 3'-end processing and transcription termination is discussed below.

7.4. The function of CFIA in transcription, termination and 3'-end processing

Through the exploitation of conditional *clp1* and *pcf11* yeast mutants and combined RT-qPCR experiments, we have shown that Clp1 and Pcf11 play a crucial part in transcription and transcription termination on the *ACT1* gene. Mutation of the Clp1 P-loop at Thr137 was shown to abrogate Clp1-ATP binding and was associated with a small increase in readthrough transcription at the *ACT1* gene locus. Although this represents only a small reduction in transcription termination, it is likely that this is a genuine effect of the Clp1 T137A mutation. This is because removal of ATP from Clp1 was shown to abrogate the Pcf11-Clp1 interaction, and we have shown that disruption of the Pcf11-Clp1 interaction caused by the *pcf11* RW-A allele causes a large increase in failed *ACT1* transcription termination. The *pcf11* RW-A allele was found to negate Pcf11-Clp1 binding at the non-permissive temperature and was identified separately in both the work described here and by Ghazy *et al.* [122]. In support of our findings that this allele has a negative effect on *ACT1* transcription termination, Ghazy *et al.* demonstrated that abrogation of the Pcf11-Clp1 interaction through this allele causes increases in readthrough transcription at the 3'-ends of the *ADH2* and *GAL7* genes. Although transcription termination defects are well documented with the *pcf11-2, pcf11-9* and *pcf11-13* temperature sensitive mutants (that remove the CID-PolII CTD interaction) [106, 176], the findings of this study and work presented by Ghazy *et al.* are the first reports that the Pcf11-Clp1 interaction is required for efficient PolII transcription termination. Considering these findings with our discovery that Pcf11 is rapidly degraded in the absence of Clp1, we suggest that a major function of Clp1 is as a chaperone of Pcf11 essential for its functions in transcription termination. It may also be that Clp1 is essential in maintaining a conformation of CFIA that is necessary for the recruitment of termination factors including the Rat1 exonuclease.

RT-qPCR experiments also showed that abolition of the Pcf11-Clp1 interaction causes significant decreases in the transcription of the *ACT1*, *ADH1, CYC1, TDH2* and *YPT1* genes. Although such abnormalities in transcription levels have been observed before

in Pcf11-depleted yeast or with yeast expressing the *pcf11-2* or *pcf11-9* temperature sensitive alleles [89, 176, 186], our work is the first report that the Pcf11-Clp1 interaction may be essential in the maintenance of PolII transcription levels. Similarly, the *clp1* T137A allele that expresses ATP-deficient Clp1 was associated with significant reductions in *ACT1, ADH1* and *TDH2* transcription. Holbein *et al.* have recently presented contrasting results that suggest that the Clp1 P-loop and Clp1-ATP binding are dispensable for transcription and transcription termination [180]. However, we question these findings, and again draw attention to the fact that their use of high-copy number plasmids for mutant *clp1* allele expression likely results in overexpression of Clp1 mutants and masks the true effects of P-loop mutations. Given that Clp1-ATP binding and the Pcf11-Clp1 interaction are linked (and both are obligatory for CFIA assembly), it is probable that the transcriptional abnormalities associated with *pcf11* RW-A and *clp1* T137A mutants are due to the aberrant or non-existent formation of CFIA. This suggests that CFIA may have an additional function during transcription itself. Support for this comes from the recent finding that CFIA interacts with the transcription factor TFIIB in a manner that is required for gene looping and the recycling of transcription factors back to the promoter (see section 1.7.2). In addition to this, the hindrance of gene loop formation has been shown to result in decreases in transcription levels [181, 182, 184, 185]. If the results of our experiments are interpreted in view of the information available on gene looping in yeast, the malformation of CFIA caused by abrogation of the Pcf11-Clp1 interaction is likely to prevent the formation of gene loops. The effect of this is that transcription factors are no longer recycled from gene terminator regions back to promoters, making transcription re-initiation and a continuous transcription cycle impossible. The end result is diminished transcription kinetics and a marked reduction in gene expression.

RT-qPCR experiments using total RNA extracted from the conditional *clp1* T137A and *pcf11* RW-A yeast mutants revealed that Clp1-ATP binding and the Pcf11-Clp1 interaction had little influence over poly(A) site selection on *ACT1* mRNA. However, our RT-qPCR assays do not rule out that mutations affected the efficacy of the cleavage and polyadenylation reactions. Indeed, a study published during the completion of this work showed that the *pcf11* RW-A yeast mutant is unable to cleave and polyadenylate an mRNA precursor both *in vitro* and *in vivo* [122]. This fits well with our findings that this mutant caused defective transcription termination and that a unified allosteric/torpedo model of transcription termination dictates that pre-mRNA cleavage

is obligatory for transcription termination (see section 1.7.1) [176]. Another recent study uncovered two *clp1* temperature sensitive mutants, *clp1-5* and *clp1-12*, that were unable to both cleave and polyadenylate *CYC1* pre-mRNA *in vitro* [121]. However, the authors did not link these deficiencies with a loss of Clp1-ATP binding. The true effect of Clp1-ATP binding on the cleavage and polyadenylation reactions remains unknown, although we speculate both activities would be lost through abrogation of the Pcf11-Clp1 interaction and the dismantling of CFIA. To our knowledge, no information was previously available regarding the roles of Clp1 and Pcf11 in poly(A) site selection. The fact that Clp1-ATP binding and the Pcf11-Clp1 interaction were seemingly dispensable in this process was surprising. However, a fresh characterisation of the export factor Yra1 may provide an explanation for the redundancy of Clp1 and Pcf11 in poly(A) site selection. Johnson *et al.* found that Yra1 competes for Pcf11 binding with Clp1 [170]. As discussed in section 7.3, the ChIP profiles of Clp1 were found to be distinct to those of Pcf11 and Yra1 across a number of genes, suggesting that the composition of CFIA changes during transcription. Furthermore, excess Yra1 was shown to inhibit cleavage and polyadenylation *in vitro*, suggesting CFIA containing Yra1 is not competent for the catalytic steps of 3'-end formation. To determine if Yra1 had other influences on 3'-end processing, the authors used an RNA sequencing approach to survey the 3'-end transcriptome of wildtype and Yra1-depleted yeast cells. They found that Yra1 modulated poly(A) site choice on mRNA transcribed from 1,430 genes, but had no effect on poly(A) site choice at 2,782 other genes. An in-depth analysis of the 1,430 "Yra1-sensitive" genes revealed that a large proportion of these genes (including the *ACT1* gene) contain divergent or non-optimally positioned efficiency elements. The authors postulate that one function of Yra1 may be to assist in conferring specificity in poly(A) site choice on genes with such sequence elements [170]. Interpreting these findings in the light of our own, we speculate that Clp1 and the Pcf11-Clp1 interaction is dispensable for poly(A) site selection and that poly(A) site choice is likely to be modulated by the Rna14 and Rna15 subunits of CFIA. The assistance of Hrp1 (CFIB) and Yra1 may be required to locate poly(A) sites where pre-mRNA sequence elements are poorly positioned or degenerate. In accordance with the ChIP profile of Clp1, recruitment of Clp1 may occur at a later stage in transcription near to the poly(A) site, to enable the assembly of catalytically competent 3'-end processing machineries.

An amalgamation of the results presented here with the findings of past (and recent) studies has provided enough information to present a hypothesis for how CFIA might regulate transcription, pre-mRNA 3'-end processing and transcription termination. In Figure 7.2, a simplified model for CFIA function in these processes is presented. Firstly, following transcription initiation, phosphorylation of PolII CTD at Ser2 increases to facilitate transcription elongation and enable recruitment of Pcf11. During transcription elongation, Pcf11 acts as a scaffold for the recruitment of Rna14, Rna15 and Yra1, which constitutes a CFIA complex that is required for poly(A) site selection. At this stage it is possible that part of the function of Yra1 is to stabilise Pcf11 by protecting it from protein degradation (similar to the role we have postulated for Clp1), which allows the maintenance of CFIA architecture. Upon transcription into the gene 3'-UTR, the Rna14 and Rna15 subunits are essential for poly(A) site selection (two copies of Rna15 bind to G/U rich sequences flanking the cleavage site). Hrp1 and Yra1 may also assist in poly(A) site selection in 3'-UTRs with divergent or degenerate sequence elements through their direct associations with RNA. Following poly(A) site selection, Clp1 competes with Yra1 for Pcf11-binding, leading to an alteration in CFIA composition to include Clp1. The inclusion of Clp1 is required to remodel the conformation of CFIA so that it promotes pre-mRNA 3'-end cleavage. This may directly involve interactions between Clp1 and CPF subunits, namely the Ysh1 endonuclease that is likely to be the effector of cleavage. After cleavage, polyadenylation is initiated by Pap1. Clp1 acts to bolster the stability of Pcf11 and promotes the Pcf11 conformation required for its essential function in transcription termination. Part of this function may be to facilitate the recruitment of Rat1 (exonuclease) to the free 5'-end of the 3'-RNA cleavage product. This model also accommodates the unified allosteric/torpedo model for transcription termination in which transcription of the poly(A) site, recruitment of factors that promote transcription termination and degradation of the 3'-RNA cleavage product are all required.

7.5. Completing the picture

In light of current thinking and information in the literature, the findings of work presented here can be condensed into three key theories: (1) The function of Clp1-ATP binding is purely to maintain Clp1 structure to allow CFIA assembly. (2) The Pcf11-Clp1 interaction is important for transcription and transcription termination. (3) The subunit composition of CFIA may be dynamic to facilitate its roles in transcription,

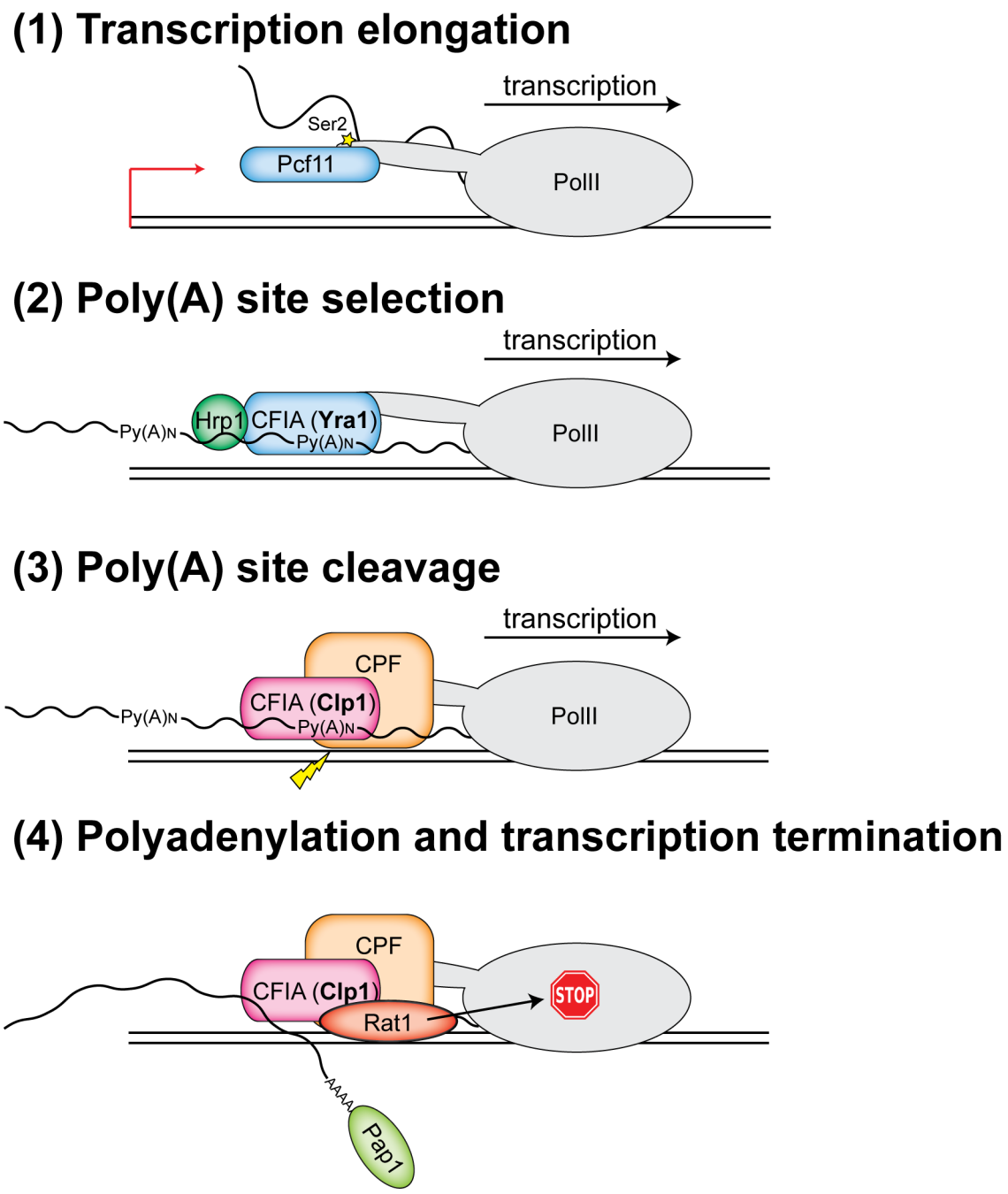


Figure 7.2. A model for CFIA function in transcription, transcription termination and pre-mRNA 3'-end processing (key events are outlined in the headings and are numbered). (1) Phosphorylation of PolII CTD Ser2 promotes transcription elongation and recruitment of Pcf11. (2) Pcf11 acts as a scaffold to recruit other CFIA subunits including Yra1 (and not Clp1) to facilitate poly(A) site selection. (3) Clp1 exchanges with Yra1 in the CFIA complex, conformational changes allow direct communication with the putative Ysh1 endonuclease of the CPF complex to effect 3'-end cleavage. (4) Pap1 polyadenylates the free 3'-end of the transcript, whilst the Pcf11-Clp1 interaction ensures that Pcf11 maintains a stable conformation required for recruitment of transcription termination factors including the Rat1 exonuclease. The red arrow indicates the promoter of a gene. The yellow star signifies phosphorylation of PolII CTD. CFIA is composed of Pcf11, Rna14 and Rna15 subunits and either Yra1 or Clp1 (denoted in brackets, bold text). $Py(A)_N$: cleavage/polyadenylation site in the 3'-UTR of a gene, the yellow lightning bolt signifies pre-mRNA cleavage.

pre-mRNA 3'-end processing and transcription termination. To provide substance to these theories, a range of different experiments would be required.

We have postulated that Clp1-ATP binding is required for the cleavage reaction through formation of a cleavage competent CFIA. Other groups have shown that temperature sensitive *clp1* alleles (*clp1-5* and *clp1-12*) or cellular depletion of Clp1 abolishes cleavage and polyadenylation [99, 121, 180]. However, these deficiencies in activity have never been linked with Clp1 P-loop mutations or loss of ATP. Although the *GAL1-clp1* T137A yeast mutant described here has been shown to be a conditional mutant (inactive in glucose), this is not as robust as finding a genuine temperature sensitive allele associated with a weakening of the Clp1-ATP interaction. It would be advantageous for future experiments looking at the role of Clp1-ATP binding to identify a Clp1-ATP binding mutant that is temperature sensitive. This would involve a more comprehensive mutational analysis of residues in or around the Clp1 P-loop motif. Following an *in vitro* characterisation of such a mutant using the methods already described here, an *in vitro* cleavage assay would be able to quickly determine whether loss of Clp1-ATP binding resulted in a loss of cleavage activity. In this assay, extracts from the relevant yeast mutant prepared at both permissive and restrictive temperatures would be assessed for the ability to cleave introduced radiolabelled pre-mRNAs, and the products visualised by denaturing gel electrophoresis.

It was hypothesised that during transcription elongation and poly(A) site selection, the CFIA complex may contain Yra1 in place of Clp1. To investigate this theory further, it would be useful to clone and express the Yra1 protein and determine whether it can be incorporated into a stable complex with Rna14, Rna15 and Pcf11. The Yra1-Pcf11 interaction has already been described [169, 170], therefore a good place to start would be to try to reconstitute this interaction *in vitro* and determine the stoichiometry of a Yra1-Pcf11 complex by SEC-MALLS.

To our knowledge, RT-qPCR was a novel approach by which to analyse changes in poly(A) site selection on yeast mRNAs. However, although this is a sensitive technique, there were also some drawbacks. To begin with, this approach requires prior knowledge of poly(A) signal locations in 3'-UTRs. The design of primers and probes is

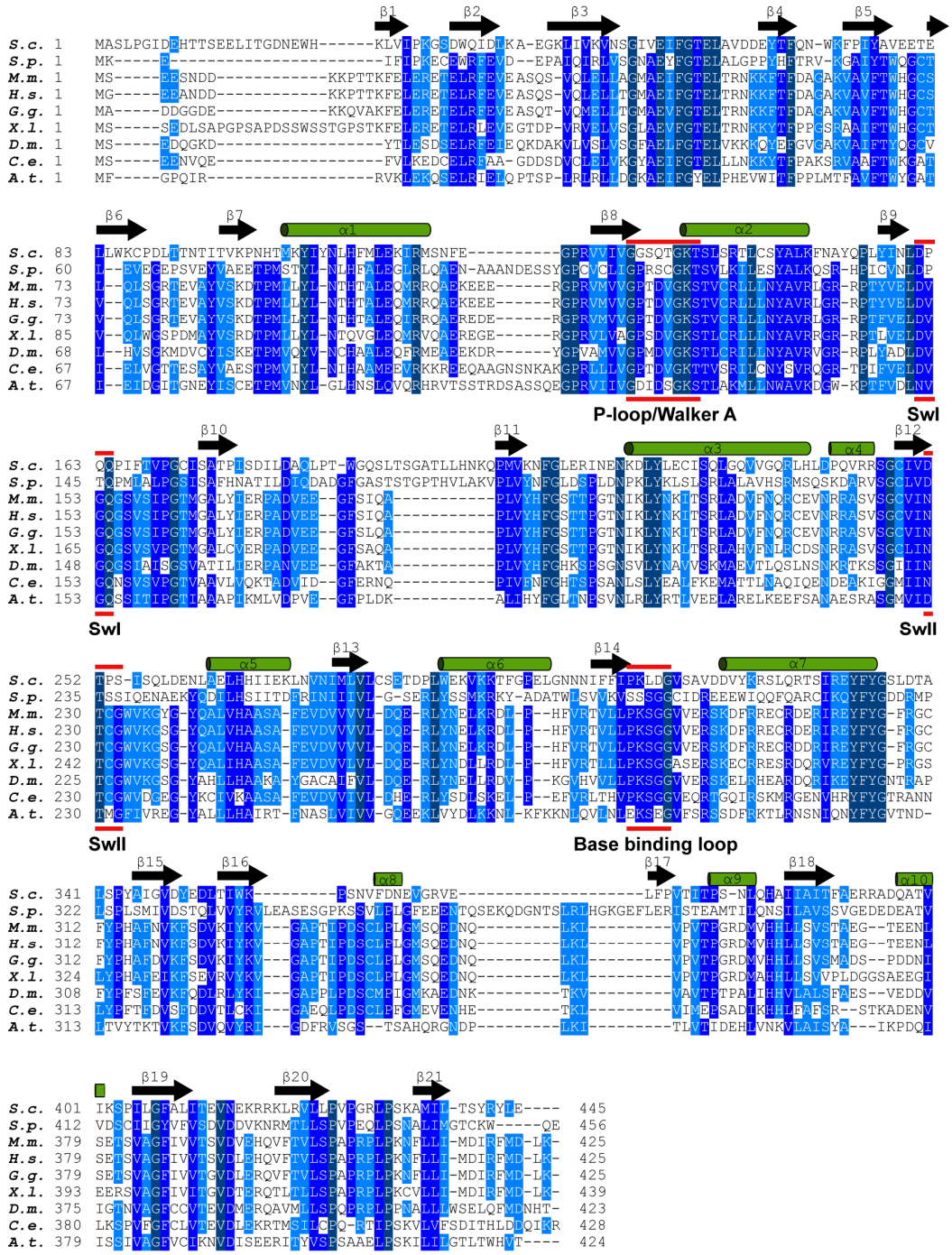
also limited to sequence content, and therefore the experiment may not be possible with some gene transcripts. In addition, as was observed here, when mutant phenotypes affect PolII transcription, it is likely that the steady-state levels of mRNA levels are all affected. This renders any endogenous control (usually a stably expressed gene) not fit for purpose. Indeed, because of this fact some of the RT-qPCR data presented here was not normalised to the expression of an endogenous control. Nevertheless, we are confident from our replicate datasets and by assaying RNA template quality and concentration thoroughly that reductions in transcription are genuine (see section 2.8.8 and Appendix 8.7). Moreover, defective transcription in *S. cerevisiae* as a result *pcf11* mutation is well described in the literature [89, 122, 176, 186]. The greatest shortcoming of this RT-qPCR approach is that the readout of the phenotype of a mutation is confined to only a single mRNA species. This renders it very trivial to conclude that the effects of a mutation (for example the removal of the Pcf11-Clp1 interaction) cause a global effect on a cellular process (for example the amount of transcription of a large set of yeast genes). Therefore, we propose that the greatest complement to the mutational analyses we have already carried out would be a global analysis of transcription and pre-mRNA 3'-end processing. In practice, this means a comparison of the entire transcriptomes of wildtype yeast with the *GAL1-clp1* T137A and *pcf11* RW-A yeast mutants. In recent years, experiments designed to achieve this have been made possible by the advent of next generation sequencing (massively parallel or high-throughout sequencing) technologies. RNA sequencing (RNA-seq) refers to the isolation and high-throughput sequencing of an entire transcriptome (reviewed in [223]). This is achieved by making cDNA libraries by reverse transcription of total RNA samples using an Oligo (dT) primer complementary to poly(A) tails (and followed by second strand synthesis). The libraries are then fragmented into shorter cDNA strands and these smaller strands are sequenced (shotgun sequencing) and mapped to the genome in question. Powerful data analysis techniques have been developed to enable characteristics such as transcript quantification to be determined. This would be extremely useful when applied to total RNA derived from the *pcf11* RW-A yeast mutant, as it would reveal the true extent to which the Pcf11-Clp1 interaction is involved in transcription. In terms of an approach to monitor global changes in 3'-end processing, RNA-seq has been adapted to analyse changes in poly(A) site selection across the entire transcriptome by surveying the junction between the 3'-UTRs and poly(A) tails of transcripts. This powerful approach has recently been used to characterise the use of poly(A) sites in *C. elegans* and *S. cerevisiae* [170, 224], and

would be a vastly superior alternative to our RT-qPCR assay. If applied to total RNA extracted from the *pcf11* RW-A mutant, this approach could be used to test our hypothesis that the Pcf11-Clp1 interaction (and Clp1) is not required for poly(A) site choice.

We have hypothesised that the composition of CFIA changes during transcription, with the effectual swapping of the Yra1 and Clp1 proteins in order to manage the requirement for poly(A) site choice or 3'-end cleavage and transcription termination during gene expression. This idea was formed from the results of ChIP-Chip experiments recently performed by Johnson *et al*, which showed that the peaks of Yra1 and Pcf11 occupancy across genes differed from the peaks of Clp1 occupancy [170]. ChIP experiments are used to identify the positions to which proteins or protein complexes bind on DNA. This is achieved by cross-linking genomic DNA to protein, shearing DNA into small fragments and isolating the protein of interest by immunoprecipitation. The cross-linking is then reversed, and the bound DNA sequences are often analysed by quantitative real-time PCR (qPCR). The ChIP-Chip (ChIP coupled with microarray analysis) and ChIP-seq (ChIP coupled with DNA-seq) methods are two further adaptations to ChIP that have allowed the genome-wide analysis of DNA sequences to which specific proteins associate (reviewed in [225]). The application of either of these approaches would greatly assist in testing our theories about CFIA formation and transcription termination. Although the *pcf11* RW-A mutant has been used to generate some ChIP data [122], this was far from a global analysis of transcription termination. If our hypotheses are correct then we would expect to see the occupancy of Pcf11 carrying the RW-A mutation increase on large numbers of genes after poly(A) signals, which would be representative of global defects in transcription termination.

8. Appendix

8.1. Clp1 multi-species sequence alignment



Species names: ***S.c.*** *Saccharomyces cerevisiae* (budding yeast); ***S.p.*** *Schizosaccharomyces pombe* (fission yeast); ***M.m.*** *Mus Musculus* (mouse); ***H.s.*** *Homo sapiens* (human); ***G.g.*** *Gallus gallus* (chicken); ***X. l.*** *Xenopus laevis* (African clawed frog); ***D.m.*** *Drosophila melanogaster* (fruit fly); ***C.e.*** *Caenorhabditis elegans* (nematode worm); ***A.t.*** *Arabidopsis thaliana* (mouse-ear cress).

Species names for Pcf11-Clp1 ID sequence alignment (Figure 1.4, C)**:** ***S.c.*** *Saccharomyces cerevisiae* (budding yeast); ***K.l.*** *Kluyveromyces lactis* (lactose yeast); ***S.p.*** *Schizosaccharomyces pombe* (fission yeast); ***M.m.*** *Mus musculus* (mouse); ***H.s.*** *Homo sapiens* (human); ***M.g.*** *Meleagris gallopavo* (wild turkey); ***X.t.*** *Xenopus tropicalis* (Western clawed frog); ***D.r.*** *Danio rerio* (Zebrafish); ***C.e.*** *Caenorhabditis elegans* (nematode worm).

8.2. Mutant plasmid constructs

Construct name	Vector	Organism	Marker	Function
Clp1 K136A-EXP	pACYCDUET1 (Novagen)	*E. coli*	camR	Expression of specified mutant Clp1-Pcf11(454-563) complex
Clp1 K136E-EXP	pACYCDUET1 (Novagen)	*E. coli*	camR	Expression of specified mutant Clp1-Pcf11(454-563) complex
Clp1 T137A- EXP	pACYCDUET1 (Novagen)	*E. coli*	camR	Expression of specified mutant Clp1-Pcf11(454-563) complex
Clp1 S138A- EXP	pACYCDUET1 (Novagen)	*E. coli*	camR	Expression of specified mutant Clp1-Pcf11(454-563) complex
Clp1 D161A- EXP	pACYCDUET1 (Novagen)	*E. coli*	camR	Expression of specified mutant Clp1-Pcf11(454-563) complex
Clp1 D251A-EXP	pACYCDUET1 (Novagen)	*E. coli*	camR	Expression of specified mutant Clp1-Pcf11(454-563) complex
Clp1 K321A-EXP	pACYCDUET1 (Novagen)	*E. coli*	camR	Expression of specified mutant Clp1-Pcf11(454-563) complex
Pcf11 Q478N-EXP	pACYCDUET1 (Novagen)	*E. coli*	camR	Expression of specified mutant Clp1-Pcf11(454-563) complex
Pcf11 R480A-EXP	pACYCDUET1 (Novagen)	*E. coli*	camR	Expression of specified mutant Clp1-Pcf11(454-563) complex
Pcf11 W482A-EXP	pACYCDUET1 (Novagen)	*E. coli*	camR	Expression of specified mutant Clp1-Pcf11(454-563) complex
Pcf11 W489A-EXP	pACYCDUET1 (Novagen)	*E. coli*	camR	Expression of specified mutant Clp1-Pcf11(454-563) complex
LEU2-clp1-K136A	pRS315 (ATCC)	*S. cerevisiae/ E. coli*	*LEU2* ampR	Testing viability of specified mutant allele in *S. cerevisiae*
LEU2-clp1-K136E	pRS315 (ATCC)	*S. cerevisiae/ E. coli*	*LEU2* ampR	Testing viability of specified mutant allele in *S. cerevisiae*
LEU2-clp1-T137A	pRS315 (ATCC)	*S. cerevisiae/ E. coli*	*LEU2* ampR	Testing viability of specified mutant allele in *S. cerevisiae*
LEU2-clp1-S138A	pRS315 (ATCC)	*S. cerevisiae/ E. coli*	*LEU2* ampR	Testing viability of specified mutant allele in *S. cerevisiae*
LEU2-clp1-D161A	pRS315 (ATCC)	*S. cerevisiae/ E. coli*	*LEU2* ampR	Testing viability of specified mutant allele in *S. cerevisiae*
LEU2-clp1-D251A	pRS315 (ATCC)	*S. cerevisiae/ E. coli*	*LEU2* ampR	Testing viability of specified mutant allele in *S. cerevisiae*

Construct name	Vector	Organism	Marker	Function
LEU2-clp1-K321A	pRS315 (ATCC)	*S. cerevisiae/ E. coli*	*LEU2* ampR	Testing viability of specified mutant allele in *S. cerevisiae*
LEU2-pcf11-Q478N	pRS315 (ATCC)	*S. cerevisiae/ E. coli*	*LEU2* ampR	Testing viability of specified mutant allele in *S. cerevisiae*
LEU2-pcf11-R480A	pRS315 (ATCC)	*S. cerevisiae/ E. coli*	*LEU2* ampR	Testing viability of specified mutant allele in *S. cerevisiae*
LEU2-pcf11-R480K	pRS315 (ATCC)	*S. cerevisiae/ E. coli*	*LEU2* ampR	Testing viability of specified mutant allele in *S. cerevisiae*
LEU2-pcf11-W482A	pRS315 (ATCC)	*S. cerevisiae/ E. coli*	*LEU2* ampR	Testing viability of specified mutant allele in *S. cerevisiae*
LEU2-pcf11-W489A	pRS315 (ATCC)	*S. cerevisiae/ E. coli*	*LEU2* ampR	Testing viability of specified mutant allele in *S. cerevisiae*
LEU2-pcf11-RW-A*	pRS315 (ATCC)	*S. cerevisiae/ E. coli*	*LEU2* ampR	Testing viability of specified mutant allele in *S. cerevisiae*
LEU2-pcf11-WW-A*	pRS315 (ATCC)	*S. cerevisiae/ E. coli*	*LEU2* ampR	Testing viability of specified mutant allele in *S. cerevisiae*
LEU2-pcf11-RWW-A*	pRS315 (ATCC)	*S. cerevisiae/ E. coli*	*LEU2* ampR	Testing viability of specified mutant allele in *S. cerevisiae*
LEU2-pcf11-CTDΔ*	pRS315 (ATCC)	*S. cerevisiae/ E. coli*	*LEU2* ampR	Testing viability of specified mutant allele in *S. cerevisiae*

*** Notes:** RW-A, R480A and W482A double mutant; WW-A, W482A and W489A double mutant; RWW-A, R480A, W482A and W489A triple mutant; CTDΔ, deletion of Pcf11 CTD (residues 393-626).

8.3. Oligonucleotide primers for PCR

Primers for cloning

Primer name + direction	Sequence (5' to 3')	Restriction site
Pcf11 400 F	GCTGGCG**GGATCC***CAAGAAAATCCCATCGTCC*	*Bam*HI
Pcf11 416 F	GCTGGCG**GGATCC***AAGCCAAATAAATGTAGC*	*Bam*HI
Pcf11 420 F	GCTGGCG**GGATCC***TGTAGCGTGTGTGGTAAG*	*Bam*HI
Pcf11 499 R	CGCCAGC**GTCGAC***GGAGGTAATTTCATCATC*	*Sal*I
Pcf11 626 R	CGCCAGC**GTCGAC***TTTTGTGACCAATTTCTTTAAG*	*Sal*I
CLP1 promoter F	CGAA**CCGCGG***TATGGAAGTTTACTTGATCCGGAG*	*Sac*II
CLP1 promoter R	CGCC**AAGCTTC**GGATCCA**CC*ATGG***TGGTT*TAATCGTGTTGGA*	*Hind*III/ *Nco*I
PCF11 promoter F	CGAA**CCGCGG***CAGCGGCGAAGAAGTCTAGCCTATTCC*	*Sac*II
PCF11 promoter R	CGCC**AAGCTTC**GGATCCA**CC*ATGG***TGAAA*AGAGAGTATTAC*	*Hind*III/ *Nco*I
GAL1 promoter F*	GCTGGCG**CCGCGG***CTCTTTGGAACTTTCAGTAATAC*	*Sac*II
GAL1 promoter R*	CGCCAGC**CC*ATGG***TTTTTTCTCCTTGACGTTAAAG*	*Nco*I
CLP1 F	CGCCAGC**CC*ATGG***CAAGTCTACCTGG*	*Nco*I
CLP1 R	CGTGGCG**TCTAGA***CTCTAAATATCTATAT*G	*Xba*I
CLP1-HA R	GCTGGCG**CTGCAG***CTCTAAATATCTATA*TG	*Pst*I
PCF11 F	CGCCAGC**CC*ATGG***ATCACGACACAGAAG*	*Nco*I
PCF11 R	GCTGGCG**TCTAGA***TTTTGTGACCAATTTCTTTAAG*	*Xba*I
PCF11-HA R	GCTGGCG**CTGCAG***TTTTGTGACCAATTTCTTTAAG*	*Pst*I
PCF11-HA R *Bgl*II	GCTGGCG**AGATCT***TTTTGTGACCAATTTCTTTAAG*	*Bgl*II
HA-tag F*	CGCC**AAGCTT**GGAATTCT**TCTAGA**TAA**CTGCAG***AGCGGCCGCATCTTTTACCCATAC*	*Hind*III/ *Xba*I/ *Pst*I
HA-tag R*	CGCC**CTCGAG***CCAACATAGATCAAAATTATAGCC*	*Xho*I
HA-tag F *Bgl*II *	CGCC**AAGCTT**GGAATTCT**TCTAGA**TAA**AGATCT***AGCGGCCGCATCTTTTACCCATAC*	*Hind*III/ *Xba*I/ *Bgl*II

***Notes:** The *GAL1* promoter was amplified by PCR using the pESC-*URA* plasmid vector (Stratagene) as template. This plasmid contains the divergent *GAL1/GAL10* promoters. The 3X haemagglutinin tag (HA-tag) plus terminator sequence was amplified by PCR using the pRS305-HA construct as template. This construct is a yeast shuttle vector (Addgene) that contains a triple repeat of amino acids 98-106 from the human influenza haemagglutinin protein (HA epitope) followed by a yeast 3'-UTR sequence. This construct was provided by Dr. M. Geymonat (formerly NIMR).

Primers for mutagenesis

Primer name + direction	Sequence (5' to 3')
Clp1 P-loop R	*GGTTTGCGAACCACCAACAATCACCACTCG*
Clp1 K136A F	*GGG**GCG**ACTTCCCTATCGAGGACACTATG*
Clp1 K136E F	*GGG**GAG**ACTTCCCTATCGAGGACACTATG*
Clp1 T137A F	*GGGAAG**GCT**TCCCTATCGAGGACACTATG*
Clp1 S138A F	*GGGAAGACT**GCC**CTATCGAGGACACTATG*
Clp1 D161A F	*CTATACATCAATCTT**GCC**CCTCAACAGCCC*
Clp1 D161A R	*GGGCTGTTGAGG**GGC**AAGATTGATGTATAG*
Clp1 D251A F	*CAGGGTGCATTGTC**GCT**ACGCCATCAATATCAC*
Clp1 D251A R	*GTGATATTGATGGCGT**AGC**GACAATGCACCCTG*
Clp1 Base loop R	*CATCAACTGCGGATACACCATCCAATTTAGGAATG*
Clp1 K321A F	*ATGTATAC**GCA**AGATCCTTGCAGAGGACATCC*
Clp1 K321R F	*ATGTATAC**CGC**AGATCCTTGCAGAGGACATCC*
Pcf11 Q478N F	*GGAAAAATATT**AAC**TCAAGAAACTGG*
Pcf11 Q478N R	*CCAGTTTCTTGA**GTT**AATATTTTTCC*
Pcf11 R480A F	*TATTCAATCA**GCC**AACTGGTATTTGAGTGACTCTC*
Pcf11 R480A R	*GAGAGTCACTCAAATACCAGTT**GGC**TGATTGAATA*
Pcf11 R480K F	*TATTCAATCA**AAG**AACTGGTATTTGAGTGACTCTC*
Pcf11 R480K R	*GAGAGTCACTCAAATACCAGTT**CTT**TGATTGAATA*
Pcf11 W482A F	*CAATCAAGAAAC**GCG**TATTTGAGTGACTC*
Pcf11 W482A R	*GAGTCACTCAAATA**CGC**GTTTCTTGATTG*
Pcf11 W489A F	*GAGTGACTCTCAA**GCG**GCAGCCTTCAAAG*
Pcf11 W489A R	*CTTTGAAGGCTGC**CGC**TTGAGAGTCACTC*
Pcf11 RW-A F	*ATTCAATCA**GCC**AAC**GCG**TATTTGAGTGACTC*
Pcf11 RW-A R	*GAGTCACTCAAATA**CGC**GTT**GGC**TGATTGAAT*
Pcf11 STOP 499 F	*GATGAAATTACCTCC**TGA**GACTCGAGCGGCCGC*
Pcf11 STOP 499 R	*GCGGCCGCTCGAGTC**TCA**GGAGGTAATTTCATC*
Pcf11 STOP 626 F	*GAAATTGGTCACAAAA**TGA**GACTCGAGCGGCCGC*
Pcf11 STOP 626 R	*GCGGCCGCTCGAGTC**TCA**TTTTGTGACCAATTTC*

8.4. Primers and probes for reverse-transcription quantitative real-time PCR experiments

RT-qPCR primers

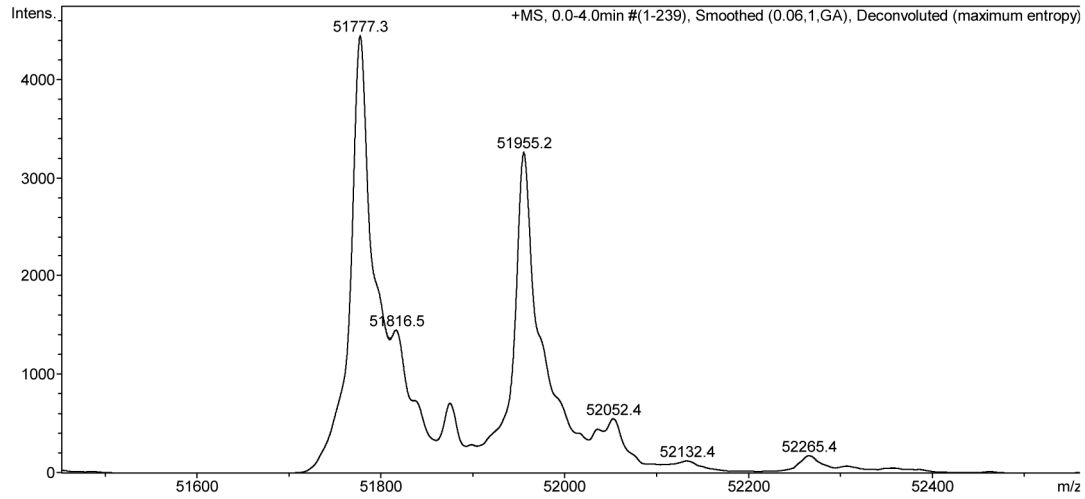
Primer name + direction	Sequence (5' to 3')	Tm (°C)
ACT1 3'-UTR F	*CAACAAATGTGGATCTCAAAACAAG*	58.1
ACT1 3'-UTR R	*CATAAACATACGCGCACAAAAGCA*	59.3
ACT1 poly(A)-1 F	*TTCACCACAAGTGTTTCTAATCTCT*	58.1
ACT1 poly(A)-1 R	*CAGTGCTTAAACACGTCTTTTCC*	58.9
ACT1 poly(A)-2/3 F	*CTATTTCTATTTTTAAACCA CCTCTCA*	59.1
ACT1 poly(A)-2/3 R	*GGAGAGACAAAACACATTATATCAATG*	58.9
ACT1 poly(A)-4 R	*GAAATTTTCGTAGAAAAGGGAGAGA*	58.1
ACT1 R-T F	*CGCTTTTCAGTTTTATAGGTTTATGC*	58.5
ACT1 R-T R	*GGTATCAAAACGCCGGACTCA*	59.8
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	*CACTCAGAAGAAGGAAGACAAAG*	57.1
YPT1 R	*AAGGTAGAGGCGCGCTTGTT*	57.3

RT-qPCR probes

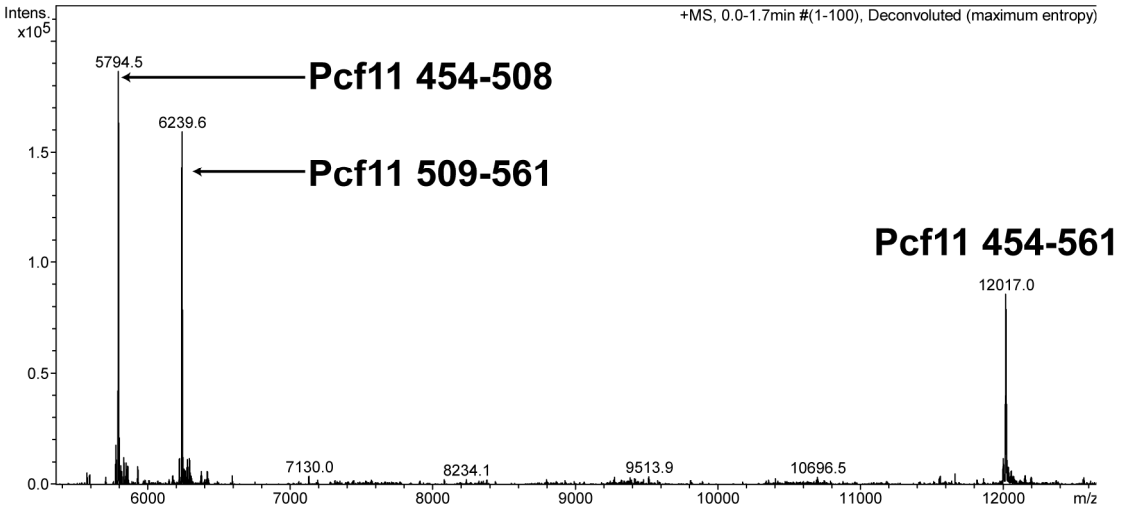
Probe name	Dye/quencher	Sequence (5' to 3')	Tm (°C)
ACT1 3'-UTR	6-FAM/MGB	*CGAAAGTGGTCCATCTATCGTTC ACCACAAGT*	68.2
ACT1 poly(A)-1	6-FAM/MGB	*TGTGCGCGTATGTTTATGTATGTA CCTCTCTCTCT*	68.3
ACT1 poly(A) 2/3	6-FAM/MGB	*AGACGTGTTTAAGCACTGACTTTA TCTACTTTTTGTACG*	67.4
ACT1 R-T	6-FAM/MGB	*CCTTCTTGGGAGGATCTTGTAGA ACCGCCATT*	69.5
TDH2	6-FAM/MGB	*CCAAGAAAGAGACCCAGCTAACT TGCCATG*	68
ADH1	6-FAM/MGB	*CTTCTTCGCCAGAGGTTTGGTCA AGTCTCCAA*	69.5
CYC1	6-FAM/MGB	*CATGTCAGAGTACTTGACTAACC CAAAGAAATATATTCC*	67.4
YPT1	6-FAM/MGB	*CAACCTGAAGGGACAGAGTTTAA CCAACACCG*	67.4

8.5. Clp1 and Pcf11(454-563) mass spectrum deconvolution reports

Clp1-6His mass spectrum deconvolution report



Pcf11 454-563 mass spectrum deconvolution report



8.6. C_T and quantity data from RT-qPCR experiments

GAL1-CLP1

Amplicon	C_T Mean (S.D.) Exp1	C_T Mean (S.D.) Exp2	Quantity Mean (S.D.) Exp1	Quantity Mean (S.D.) Exp2
***ACT1* 3'-UTR**	20.67 (0.07)	19.71 (0.04)	140.7 (6.35)	132.4 (3.4)
***ACT1* poly(A)-1**	22.77 (0.09)	22.24 (0.07)	23.95 (1.45)	31.35 (1.25)
***ACT1* poly(A)-2/3**	23.02 (0.02)	21.91 (0.01)	20.7 (0.3)	30.6 (0.15)
***ACT1* poly(A)- 4**	24.58 (0.13)	19.89 (0.38)	19.95 (1.45)	33.45 (7.3)
***ACT1* R-T**	22.29 (0.04)	22.62 (0.01)	15.7 (0.5)	11.1 (0.01)
***TDH2* (1)**	19.22 (0.05)	17.96 (0.01)	210.25 (8)	213.25 (1.9)
***TDH2* (2)**	17.95 (0.01)	17.74 (0.11)	223.75 (2.1)	273.6 (23.46)
ADH1	19.72 (0.05)	17.17 (0.04)	73.25 (2.75)	482.45 13.95)
CYC1	21.56 (0.06)	21.23 (0.11)	7.4 (0.15)	8.6 (0.75)
YPT1	22.43 (0.03)	21.99 (0.02)	10.23 (0.19)	13.05 (0.24)

GAL1-clp1 K136A

Amplicon	C_T Mean (S.D.) Exp1	C_T Mean (S.D.) Exp2	Quantity Mean (S.D.) Exp1	Quantity Mean (S.D.) Exp2
***ACT1* 3'-UTR**	18.92 (0.04)	18.94 (0.05)	199.35 (5.9)	204.85 (7.3)
***ACT1* poly(A)-1**	21.56 (0.08)	20.95 (0.08)	33.1 (1.6)	35 (1.9)
***ACT1* poly(A)-2/3**	20.89 (0.25)	21.57 (0.22)	33.9 (5.5)	35 (5)
***ACT1* poly(A)- 4**	22.52 (0.1)	22.56 (0.11)	29.35 (1.85)	29.8 (2)
***ACT1* R-T**	22.88 (0.04)	21.67 (0.13)	7.3 (0.18)	8.2 (0.65)
***TDH2* (1)**	17.42 (0.01)	17.3 (0.05)	335 (4.24)	364.15 (13.6)
***TDH2* (2)**	17.4 (0.07)	17.66 (0.02)	336.9 (17.6)	289.8 (4.95)
ADH1	16.91 (0.09)	17.09 (0.02)	563.95 (38.9)	511.8 (7.8)
CYC1	21.73 (0.01)	21.71 (0.09)	6.55 (0.01)	5.95 (0.45)
YPT1	22.53 (0.04)	22.58 (0.03)	9.45 (0.28)	8.4 (0.2)

GAL1-clp1 T137A

Amplicon	C_T Mean (S.D.) Exp1	C_T Mean (S.D.) Exp2	Quantity Mean (S.D.) Exp1	Quantity Mean (S.D.) Exp2
***ACT1* 3'-UTR**	21.8 (0.16)	20.22 (0.05)	66.85 (6.95)	93.75 (3.45)
***ACT1* poly(A)-1**	23.78 (0.13)	22.63 (0.11)	12.6 (1)	16.9 (1.2)
***ACT1* poly(A)-2/3**	24 (0.17)	21.71 (0.14)	8.7 (1.1)	24.55 (2.35)
***ACT1* poly(A)- 4**	26.74 (0.17)	23.55 (0.16)	5.85 (0.6)	17.25 (1.7)
***ACT1* R-T**	23.47 (0.02)	21.94 (0.03)	10.3 (0.15)	11.95 (0.2)
***TDH2* (1)**	21.77 (0.07)	20.24 (0.09)	37.25 (1.8)	47.5 (2.9)
***TDH2* (2)**	20.37 (0.04)	20.67 (0.01)	36.5 (0.95)	29.8 (0.25)
ADH1	20.07 (0.09)	20.31 (0.03)	56.5 (3.9)	49.3 (1)
CYC1	21.98 (0.05)	22 (0.11)	5.45 (0.2)	4.75 (0.4)
YPT1	22.85 (0.08)	22.92 (0.07)	7.45 (0.01)	6.45 (0.33)

***GAL1-clp1* S138A**

Amplicon	C_T Mean (S.D.) Exp1	C_T Mean (S.D.) Exp2	Quantity Mean (S.D.) Exp1	Quantity Mean (S.D.) Exp2
***ACT1* 3'-UTR**	20.49 (0.05)	18.9 (0.06)	75 (2.65)	200.8 (7.5)
***ACT1* poly(A)-1**	22.14 (0.06)	21.58 (0.12)	14.7 (0.65)	35.2 (2.9)
***ACT1* poly(A)-2/3**	22.67 (0.18)	21.61 (0.22)	9.6 (1.25)	36.4 (5.4)
***ACT1* poly(A)- 4**	24.88 (0.03)	22.49 (0.09)	7.35 (0.13)	34.35 (2.2)
***ACT1* R-T**	22.91 (0.01)	19.29 (0.21)	6.65 (0.01)	9.2 (1.3)
***TDH2* (1)**	18.97 (0.01)	14.38 (0.19)	106.5 (1.1)	345.4 (45.95)
***TDH2* (2)**	18.68 (0.05)	18.42 (0.03)	150.2 (5.95)	173.9 (4.5)
ADH1	13.89 (0.19)	17.6 (0.04)	206.8 (25.95)	413.45 (13.8)
CYC1	21.85 (0.09)	21.78 (0.06)	7.45 (0.45)	8.35 (0.39)
YPT1	22.73 (0.1)	22.49 (0.04)	8.45 (0.65)	9.45 (0.33)

***GAL1-clp1* D161A**

Amplicon	C_T Mean (S.D.) Exp1	C_T Mean (S.D.) Exp2	Quantity Mean (S.D.) Exp1	Quantity Mean (S.D.) Exp2
***ACT1* 3'-UTR**	18.95 (0.03)	18.91 (0.09)	206 (3.6)	185.45 (12.75)
***ACT1* poly(A)-1**	21.02 (0.04)	21.38 (0.01)	38.35 (0.95)	38.35 (0.15)
***ACT1* poly(A)-2/3**	20.98 (0.09)	21.2 (0.28)	40.3 (2.65)	38.55 (7.1)
***ACT1* poly(A)- 4**	22.56 (0.05)	22.67 (0.23)	32.8 (1.1)	32.15 (4.85)
***ACT1* R-T**	22.33 (0.39)	22.65 (0.05)	10 (0.25)	8.65 (0.3)
***TDH2* (1)**	17.46 (0.04)	17.58 (0.18)	325.65 (9.95)	293.15 (35.6)
***TDH2* (2)**	18.79 (0.02)	18.53 (0.04)	138 (2.15)	160.8 (4.8)
ADH1	13.68 (0.18)	17.66 (0.05)	239.35 (28.95)	397.15 (15)
CYC1	21.85 (0.08)	21.8 (0.09)	7.55 (0.5)	8.25 (0.55)
YPT1	22.83 (0.04)	22.64 (0.02)	7.85 (0.25)	8.45 (0.1)

***GAL1-clp1* D251A**

Amplicon	C_T Mean (S.D.) Exp1	C_T Mean (S.D.) Exp2	Quantity Mean (S.D.) Exp1	Quantity Mean (S.D.) Exp2
***ACT1* 3'-UTR**	19.19 (0.07)	18.93 (0.04)	193.15 (8.85)	198.45 (5.85)
***ACT1* poly(A)-1**	21.49 (0.07)	21.05 (0.08)	35.15 (1.6)	31.35 (1.6)
***ACT1* poly(A)-2/3**	21.24 (0.21)	21.06 (0.23)	40.15 (5.8)	37.2 (5.7)
***ACT1* poly(A)- 4**	22.46 (0.22)	20.46 (0.15)	30.1 (3.95)	31.7 (3.15)
***ACT1* R-T**	22.83 (0.09)	22.92 (0.02)	6.25 (0.4)	6.15 (0.09)
TDH2	17.54 (0.01)	17.56 (0.04)	296.35 (0.85)	305.3 (9.15)

PCF11 30°C

Amplicon	C_T Mean (S.D.) Exp1	C_T Mean (S.D.) Exp2	Quantity Mean (S.D.) Exp1	Quantity Mean (S.D.) Exp2
***ACT1* 3'-UTR**	19.95 (0.02)	16.58 (0.01)	113.65 (1.65)	177.35 (1.1)
***ACT1* poly(A)-1**	23.83 (0.1)	24.17 (0.04)	7.15 (0.45)	11.85 (0.3)
***ACT1* poly(A)-2/3**	24.11 (0.17)	24.85 (0.14)	6.95 (0.75)	6.95 (0.65)
***ACT1* poly(A)- 4**	26.4 (0.29)	26.89 (0.08)	3.35 (0.6)	6.15 (0.25)
***ACT1* R-T**	25.82 (0.03)	26.28 (0.03)	1.1 (0.01)	1.5 (0.01)
TDH2	18.89 (0.05)	15.28 (0.41)	131.6 (4.2)	142.7 (40.95)
ADH1	17.95 (0.02)	15.53 (0.04)	308.8 (5.06)	148.4 (4)
CYC1	22.58 (0.1)	22.76 (0.14)	3.55 (0.25)	5.2 (0.04)
YPT1	23.85 (0.06)	23.79 (0.05)	3.2 (0.15)	4.4 (0.15)

PCF11 39°C

Amplicon	C_T Mean (S.D.) Exp1	C_T Mean (S.D.) Exp2	Quantity Mean (S.D.) Exp1	Quantity Mean (S.D.) Exp2
***ACT1* 3'-UTR**	20.9 (0.06)	16.48 (0.08)	51.8 (2.25)	77.85 (3.8)
***ACT1* poly(A)-1**	23.85 (0.03)	24.72 (0.05)	5.2 (0.1)	9.3 (0.25)
***ACT1* poly(A)-2/3**	24.2 (0.23)	25.33 (0.13)	5.07 (0.85)	6.1 (0.5)
***ACT1* poly(A)- 4**	27.36 (0.01)	26.65 (0.1)	1.6 (0.01)	5.85 (0.38)
***ACT1* R-T**	24.93 (0.01)	25.58 (0.03)	0.98 (0.01)	2.45 (0.01)
TDH2	18.79 (0.01)	20.29 (0.49)	134.6 1.45)	94.8 (27.85)
ADH1	17.97 (0.04)	15.57 (0.02)	303.75 (10.06)	144.45 (1.85)
CYC1	22.39 (0.06)	22.82 (0.12)	4.1 (0.2)	5.03 (0.4)
YPT1	24.81 (0.05)	24.75 (0.03)	1.5 (0.01)	2.15 (0.05)

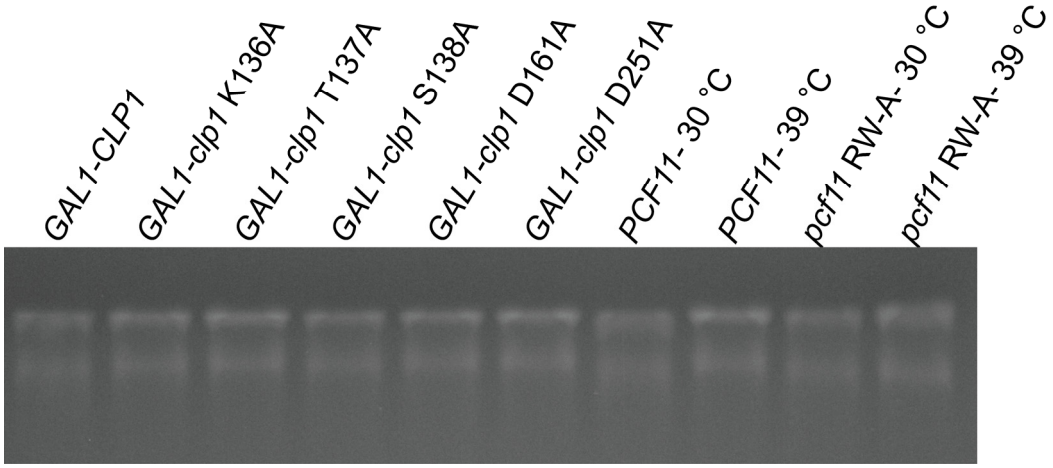
pcf11 RW-A 30°C

Amplicon	C_T Mean (S.D.) Exp1	C_T Mean (S.D.) Exp2	Quantity Mean (S.D.) Exp1	Quantity Mean (S.D.) Exp2
***ACT1* 3'-UTR**	19.87 (0.04)	16.74 (0.09)	126.75 (0.05)	184.1 (11.7)
***ACT1* poly(A)-1**	22.87 (0.19)	23.94 (0.03)	11 (1.33)	13.45 (0.3)
***ACT1* poly(A)-2/3**	23.17 (0.03)	24.59 (0.24)	10.29 (0.19)	9.8 (1.56)
***ACT1* poly(A)- 4**	25.22 (0.09)	25.52 (0.16)	6.7 (0.39)	10.2 (0.95)
***ACT1* R-T**	25.65 (0.03)	26.41 (0.07)	1.25 (0.01)	1.38 (0.06)
TDH2	19.03 (0.07)	20.17 (0.03)	115.1 (5.45)	99.65 (1.8)
ADH1	18.45 (0.01)	16 (0.18)	211.7 (1.6)	109.55 (12.5)
CYC1	21.87 (0.05)	22.21 (0.05)	6.05 (0.01)	7.5 (0.24)
YPT1	23.6 (0.09)	23.52 (0.07)	3.9 (0.28)	5.4 (0.28)

pcf11 RW-A 39°C

Amplicon	C_T Mean (S.D.) Exp1	C_T Mean (S.D.) Exp2	Quantity Mean (S.D.) Exp1	Quantity Mean (S.D.) Exp2
***ACT1* 3'-UTR**	22.95 (0.01)	23.82 (0.09)	15.9 (0.01)	18.35 (1.1)
***ACT1* poly(A)-1**	25.72 (0.08)	26.59 (0.04)	2.45 (0.13)	2.55 (0.05)
***ACT1* poly(A)-2/3**	25 (0.17)	26.66 (0.13)	3.15 (0.07)	2.45 (0.19)
***ACT1* poly(A)- 4**	27.65 (0.071)	28.05 (0.08)	3.11 (0.01)	2.9 (0.14)
***ACT1* R-T**	24.24 (0.06)	25.15 (0.02)	3.1 (0.13)	3.35 (0.01)
TDH2	20.19 (0.03)	21.79 (0.03)	52 (1.2)	34.75 (0.65)
ADH1	19.87 (0.05)	17.48 (0.14)	71.6 (2.5)	41.8 (3.9)
CYC1	23.81 (0.09)	24.12 (0.12)	1.3 (0.1)	2.15 (0.17)
YPT1	26.57 (0.01)	26.45 (0.04)	0.39 (0.004)	0.6 (0.02)

8.7. Analysis of RNA quality by denaturing agarose gel electrophoresis



9. 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, Wittop Koning TH, Schumperli D, Muller B. A 5'-3' exonuclease activity involved in forming the 3' products of histone pre-mRNA processing in vitro. RNA. 1998 Sep;4(9):1034-46.
3. Tarun SZ, Jr., 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 21;10(18):1147-50.
5. Shuman S. Structure, mechanism, and evolution of the mRNA capping apparatus. Prog Nucleic Acid Res Mol Biol. 2001;66:1-40.
6. 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.
7. Itoh N, Yamada H, Kaziro Y, Mizumoto K. Messenger RNA guanylyltransferase from Saccharomyces cerevisiae. Large scale purification, subunit functions, and subcellular localization. J Biol Chem. 1987 Feb 15;262(5):1989-95.
8. Mao X, Schwer B, Shuman S. Yeast mRNA cap methyltransferase is a 50-kilodalton protein encoded by an essential gene. Mol Cell Biol. 1995 Aug;15(8):4167-74.
9. Komarnitsky P, Cho EJ, 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. 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.
11. Kim HJ, Jeong SH, Heo JH, Jeong SJ, Kim ST, Youn HD, et al. mRNA capping enzyme activity is coupled to an early transcription elongation. Mol Cell Biol. 2004 Jul;24(14):6184-93.
12. Schroeder SC, Zorio DA, 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.
13. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001 Feb 15;409(6822):860-921.
14. Finishing the euchromatic sequence of the human genome. Nature. 2004 Oct 21;431(7011):931-45.
15. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, et al. Life with 6000 genes. Science. 1996 Oct 25;274(5287):546, 63-7.
16. Ares M, Jr., Grate L, Pauling MH. A handful of intron-containing genes produces the lion's share of yeast mRNA. RNA. 1999 Sep;5(9):1138-9.
17. Spingola M, Grate L, Haussler D, Ares M, Jr. Genome-wide bioinformatic and molecular analysis of introns in Saccharomyces cerevisiae. RNA. 1999 Feb;5(2):221-34.
18. Moore MJ, Query, C.C., Sharp, P.A. . Splicing of Precursors to mRNA by the Spliceosome. Cold Spring Harbor, NY: Cold Spring Harbor

Lab. Press; 1993.
19. Staley JP, Guthrie C. Mechanical devices of the spliceosome: motors, clocks, springs, and things. Cell. 1998 Feb 6;92(3):315-26.
20. Stark H, Luhrmann R. Cryo-electron microscopy of spliceosomal components. Annu Rev Biophys Biomol Struct. 2006;35:435-57.
21. Jurica MS, Moore MJ. Pre-mRNA splicing: awash in a sea of proteins. Mol Cell. 2003 Jul;12(1):5-14.
22. Michaud S, Reed R. An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. Genes Dev. 1991 Dec;5(12B):2534-46.
23. Du H, Rosbash M. The U1 snRNP protein U1C recognizes the 5' splice site in the absence of base pairing. Nature. 2002 Sep 5;419(6902):86-90.
24. Kandels-Lewis S, Seraphin B. Involvement of U6 snRNA in 5' splice site selection. Science. 1993 Dec 24;262(5142):2035-9.
25. Reyes JL, Kois P, Konforti BB, Konarska MM. The canonical GU dinucleotide at the 5' splice site is recognized by p220 of the U5 snRNP within the spliceosome. RNA. 1996 Mar;2(3):213-25.
26. Query CC, Moore MJ, Sharp PA. Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. Genes Dev. 1994 Mar 1;8(5):587-97.
27. Lacadie SA, Rosbash M. Cotranscriptional spliceosome assembly dynamics and the role of U1 snRNA:5'ss base pairing in yeast. Mol Cell. 2005 Jul 1;19(1):65-75.
28. Hocine S, Singer RH, Grunwald D. RNA processing and export. Cold Spring Harb Perspect Biol. 2010 Dec;2(12):a000752.
29. Furger A, O'Sullivan JM, Binnie A, Lee BA, Proudfoot NJ. Promoter proximal splice sites enhance transcription. Genes Dev. 2002 Nov 1;16(21):2792-9.
30. Lei EP, Silver PA. Intron status and 3'-end formation control cotranscriptional export of mRNA. Genes Dev. 2002 Nov 1;16(21):2761-6.
31. 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.
32. Iglesias N, Stutz F. Regulation of mRNP dynamics along the export pathway. FEBS Lett. 2008 Jun 18;582(14):1987-96.
33. Rondon AG, Jimeno S, Aguilera A. The interface between transcription and mRNP export: from THO to THSC/TREX-2. Biochim Biophys Acta. 2010 Aug;1799(8):533-8.
34. 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.
35. Strasser 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.
36. 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.
37. 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.

38. 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.
39. Strasser 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.
40. 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.
41. Aguilera A. mRNA processing and genomic instability. Nat Struct Mol Biol. 2005 Sep;12(9):737-8.
42. Chekanova JA, Abruzzi KC, Rosbash M, Belostotsky DA. Sus1, Sac3, and Thp1 mediate post-transcriptional tethering of active genes to the nuclear rim as well as to non-nascent mRNP. RNA. 2008 Jan;14(1):66-77.
43. Cabal GG, Genovesio A, Rodriguez-Navarro S, Zimmer C, Gadal O, Lesne A, et al. SAGA interacting factors confine sub-diffusion of transcribed genes to the nuclear envelope. Nature. 2006 Jun 8;441(7094):770-3.
44. Dieppois G, Iglesias N, Stutz F. Cotranscriptional recruitment to the mRNA export receptor Mex67p contributes to nuclear pore anchoring of activated genes. Mol Cell Biol. 2006 Nov;26(21):7858-70.
45. 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.
46. Fasken MB, Stewart M, Corbett AH. Functional significance of the interaction between the mRNA-binding protein, Nab2, and the nuclear pore-associated protein, Mlp1, in mRNA export. J Biol Chem. 2008 Oct 3;283(40):27130-43.
47. Batisse J, Batisse C, Budd A, Bottcher B, Hurt E. Purification of nuclear poly(A)-binding protein Nab2 reveals association with the yeast transcriptome and a messenger ribonucleoprotein core structure. J Biol Chem. 2009 Dec 11;284(50):34911-7.
48. Maniatis T, Reed R. An extensive network of coupling among gene expression machines. Nature. 2002 Apr 4;416(6880):499-506.
49. Bentley D. The mRNA assembly line: transcription and processing machines in the same factory. Curr Opin Cell Biol. 2002 Jun;14(3):336-42.
50. Bentley DL. Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. Curr Opin Cell Biol. 2005 Jun;17(3):251-6.
51. 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.
52. Zorio DA, Bentley DL. The link between mRNA processing and transcription: communication works both ways. Exp Cell Res. 2004 May 15;296(1):91-7.
53. Moore MJ, Proudfoot NJ. Pre-mRNA processing reaches back to transcription and ahead to translation. Cell. 2009 Feb 20;136(4):688-700.
54. 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.
55. Wilusz CJ, Wilusz J. Bringing the role of mRNA decay in the control of gene expression into focus. Trends Genet. 2004 Oct;20(10):491-7.
56. Wells SE, Hillner PE, Vale RD, Sachs AB. Circularization of mRNA by eukaryotic translation initiation factors. Mol Cell. 1998 Jul;2(1):135-40.
57. Mandel CR, Bai Y, Tong L. Protein factors in pre-mRNA 3'-end processing. Cell Mol Life Sci. 2008 Apr;65(7-8):1099-122.
58. Guo Z, Sherman F. 3'-end-forming signals of yeast mRNA. Trends Biochem Sci. 1996 Dec;21(12):477-81.

59. Guo Z, Sherman F. Signals sufficient for 3'-end formation of yeast mRNA. Mol Cell Biol. 1996 Jun;16(6):2772-6.
60. Edmonds M. A history of poly A sequences: from formation to factors to function. Prog Nucleic Acid Res Mol Biol. 2002;71:285-389.
61. Shatkin AJ, Manley JL. The ends of the affair: capping and polyadenylation. Nat Struct Biol. 2000 Oct;7(10):838-42.
62. Wahle E, Keller W. The biochemistry of 3'-end cleavage and polyadenylation of messenger RNA precursors. Annu Rev Biochem. 1992;61:419-40.
63. Butler JS, Platt T. RNA processing generates the mature 3' end of yeast CYC1 messenger RNA in vitro. Science. 1988 Dec 2;242(4883):1270-4.
64. 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.
65. Zaret KS, Sherman F. Mutationally altered 3' ends of yeast CYC1 mRNA affect transcript stability and translational efficiency. J Mol Biol. 1984 Jul 25;177(1):107-35.
66. 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.
67. 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.
68. Guo Z, Sherman F. 3'-end-forming signals of yeast mRNA. Mol Cell Biol. 1995 Nov;15(11):5983-90.
69. Abe A, Hiraoka Y, Fukasawa T. Signal sequence for generation of mRNA 3' end in the Saccharomyces cerevisiae GAL7 gene. EMBO J. 1990 Nov;9(11):3691-7.
70. Irniger S, Egli CM, Braus GH. Different classes of polyadenylation sites in the yeast Saccharomyces cerevisiae. Mol Cell Biol. 1991 Jun;11(6):3060-9.
71. Peterson JA, Myers AM. Functional analysis of mRNA 3' end formation signals in the convergent and overlapping transcription units of the S. cerevisiae genes RHO1 and MRP2. Nucleic Acids Res. 1993 Nov 25;21(23):5500-8.
72. 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.
73. Egli CM, Springer C, Braus GH. A complex unidirectional signal element mediates GCN4 mRNA 3' end formation in Saccharomyces cerevisiae. Mol Cell Biol. 1995 May;15(5):2466-73.
74. Irniger S, Sanfacon H, Egli CM, Braus GH. Different sequence elements are required for function of the cauliflower mosaic virus polyadenylation site in Saccharomyces cerevisiae compared with in plants. Mol Cell Biol. 1992 May;12(5):2322-30.
75. 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.
76. Bennetzen JL, Hall BD. The primary structure of the Saccharomyces cerevisiae gene for alcohol dehydrogenase. J Biol Chem. 1982 Mar 25;257(6):3018-25.
77. 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.
78. 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.

79. Graber JH, Cantor CR, Mohr SC, Smith TF. In silico detection of control signals: mRNA 3'-end-processing sequences in diverse species. Proc Natl Acad Sci U S A. 1999 Nov 23;96(24):14055-60.
80. van Helden J, del Olmo M, Perez-Ortin JE. Statistical analysis of yeast genomic downstream sequences reveals putative polyadenylation signals. Nucleic Acids Res. 2000 Feb 15;28(4):1000-10.
81. 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.
82. Chen J, Moore C. Separation of factors required for cleavage and polyadenylation of yeast pre-mRNA. Mol Cell Biol. 1992 Aug;12(8):3470-81.
83. 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.
84. 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.
85. 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.
86. Ohnacker M, Barabino SM, 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.
87. 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.
88. 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.
89. 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.
90. 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 USA. 1997 Jul 22;94(15):7897-902.
91. 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.
92. Gordon JM, 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. 2011 Nov 29;50(47):10203-14.
93. Bloch JC, Perrin F, Lacroute F. Yeast temperature-sensitive mutants specifically impaired in processing of poly(A)-containing RNAs. Mol Gen Genet. 1978 Oct 4;165(2):123-7.
94. 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;11(6):3075-87.
95. Preker PJ, Keller W. The HAT helix, a repetitive motif implicated in RNA processing. Trends Biochem Sci. 1998 Jan;23(1):15-6.
96. Andrade MA, Bork P. HEAT repeats in the Huntington's disease protein. Nat Genet. 1995 Oct;11(2):115-6.
97. 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.
98. Moreno-Morcillo M, Minvielle-Sebastia 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.
99. 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.
100. Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B. A generic protein purification method for protein complex characterization and proteome exploration. Nat Biotechnol. 1999 Oct;17(10):1030-2.
101. Xu X, Song Y, Li Y, Chang J, Zhang H, An L. The tandem affinity purification method: an efficient system for protein complex purification and protein interaction identification. Protein Expr Purif. 2010 Aug;72(2):149-56.
102. 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.
103. 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.
104. 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.
105. Skaar DA, Greenleaf AL. The RNA polymerase II CTD kinase CTDK-I affects pre-mRNA 3' cleavage/polyadenylation through the processing component Pti1p. Mol Cell. 2002 Dec;10(6):1429-39.
106. Sadowski M, Dichtl B, Hubner 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.
107. 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;107 (Pt 4):913-21.
108. 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. 2000 Jan;262(6):1103-12.
109. 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. 2007 Oct;13(10):1756-64.
110. 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.

111. 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.
112. 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;38(9):3119-32.
113. Deka P, Rajan PK, Perez-Canadillas JM, Varani G. Protein and RNA dynamics play key roles in determining the specific recognition of GU-rich polyadenylation regulatory elements by human Cstf-64 protein. J Mol Biol. 2005 Apr 8;347(4):719-33.
114. Qu X, Perez-Canadillas JM, 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.
115. 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.
116. Noble CG, Beuth B, Taylor IA. Structure of a nucleotide-bound Clp1-Pcf11 polyadenylation factor. Nucleic Acids Res. 2007;35(1):87-99.
117. de Vries H, Ruegsegger U, Hubner W, Friedlein A, Langen H, Keller W. Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors. EMBO J. 2000 Nov 1;19(21):5895-904.
118. Walker JE, Saraste M, Runswick MJ, Gay NJ. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. 1982;1(8):945-51.
119. 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.
120. Jang SB, Jeong MS, Seefeldt LC, Peters JW. Structural and biochemical implications of single amino acid substitutions in the nucleotide-dependent switch regions of the nitrogenase Fe protein from Azotobacter vinelandii. J Biol Inorg Chem. 2004 Dec;9(8):1028-33.
121. 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. 2011 Oct 12.
122. Ghazy MA, Gordon JM, 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. 2011 Oct 12.
123. 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.
124. Valentini SR, Weiss VH, Silver PA. Arginine methylation and binding of Hrp1p to the efficiency element for mRNA 3'-end formation. RNA. 1999 Feb;5(2):272-80.
125. Chen S, Hyman LE. A specific RNA-protein interaction at yeast polyadenylation efficiency elements. Nucleic Acids Res. 1998 Nov 1;26(21):4965-74.
126. Perez-Canadillas JM. Grabbing the message: structural basis of mRNA 3'UTR recognition by Hrp1. EMBO J. 2006 Jul 12;25(13):3167-78.
127. 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.

128. Kim Guisbert K, Duncan K, Li H, Guthrie C. Functional specificity of shuttling hnRNPs revealed by genome-wide analysis of their RNA binding profiles. RNA. 2005 Apr;11(4):383-93.
129. Henry MF, Mandel D, Routson V, Henry PA. The yeast hnRNP-like protein Hrp1/Nab4 sccumulates in the cytoplasm after hyperosmotic stress: a novel Fps1-dependent response. Mol Biol Cell. 2003 Sep;14(9):3929-41.
130. Dichtl B, Blank D, Sadowski M, Hubner 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.
131. 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.
132. Zhao J, Kessler MM, Moore CL. Cleavage factor II of Saccharomyces cerevisiae contains homologues to subunits of the mammalian Cleavage/ polyadenylation specificity factor and exhibits sequence-specific, ATP-dependent interaction with precursor RNA. J Biol Chem. 1997 Apr 18;272(16):10831-8.
133. Ryan K, Calvo O, Manley JL. Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. RNA. 2004 Apr;10(4):565-73.
134. Barabino SM, Hubner 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.
135. Barabino SM, 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.
136. Tacahashi Y, Helmling S, Moore CL. Functional dissection of the zinc finger and flanking domains of the Yth1 cleavage/polyadenylation factor. Nucleic Acids Res. 2003 Mar 15;31(6):1744-52.
137. Ezeokonkwo C, Zhelkovsky A, Lee R, Bohm A, Moore CL. A flexible linker region in Fip1 is needed for efficient mRNA polyadenylation. RNA. 2011 Apr;17(4):652-64.
138. Helmling S, Zhelkovsky A, Moore CL. Fip1 regulates the activity of Poly(A) polymerase through multiple interactions. Mol Cell Biol. 2001 Mar;21(6):2026-37.
139. 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.
140. Rodriguez CR, Cho EJ, Keogh MC, Moore CL, Greenleaf AL, Buratowski S. Kin28, the TFIIH-associated carboxy-terminal domain kinase, facilitates the recruitment of mRNA processing machinery to RNA polymerase II. Mol Cell Biol. 2000 Jan;20(1):104-12.
141. 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.
142. Balbo PB, Bohm A. Mechanism of poly(A) polymerase: structure of the enzyme-MgATP-RNA ternary complex and kinetic analysis. Structure. 2007 Sep;15(9):1117-31.
143. Balbo PB, Meinke G, Bohm A. Kinetic studies of yeast polyA polymerase indicate an induced fit mechanism for nucleotide specificity. Biochemistry. 2005 May 31;44(21):7777-86.

144. 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.
145. Dheur S, Vo le TA, Voisinet-Hakil F, Minet M, Schmitter JM, Lacroute F, et al. Pti1p and Ref2p found in association with the mRNA 3' end formation complex direct snoRNA maturation. EMBO J. 2003 Jun 2;22(11):2831-40.
146. 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.
147. 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. 2004 Jun;10(6):965-77.
148. Zhelkovsky A, Tacahashi Y, Nasser T, He X, Sterzer U, Jensen TH, et al. The role of the Brr5/Ysh1 C-terminal domain and its homolog Syc1 in mRNA 3'-end processing in Saccharomyces cerevisiae. Rna. 2006 Mar;12(3):435-45.
149. 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.
150. 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.
151. Reyes-Reyes M, Hampsey M. Role for the Ssu72 C-terminal domain phosphatase in RNA polymerase II transcription elongation. Mol Cell Biol. 2007 Feb;27(3):926-36.
152. He X, Moore C. Regulation of yeast mRNA 3' end processing by phosphorylation. Mol Cell. 2005 Sep 2;19(5):619-29.
153. Mandart E, Parker R. Effects of mutations in the Saccharomyces cerevisiae RNA14, RNA15, and PAP1 genes on polyadenylation in vivo. Mol Cell Biol. 1995 Dec;15(12):6979-86.
154. 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.
155. 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.
156. 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.
157. Meinke G, Ezeokonkwo C, Balbo P, Stafford W, Moore C, Bohm A. Structure of yeast poly(A) polymerase in complex with a peptide from Fip1, an intrinsically disordered protein. Biochemistry. 2008 Jul 1;47(26):6859-69.
158. 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.
159. Brown CE, Sachs AB. Poly(A) tail length control in Saccharomyces cerevisiae occurs by message-specific deadenylation. Mol Cell Biol. 1998 Nov;18(11):6548-59.
160. 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.
161. 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.
162. Viphakone N, Voisinet-Hakil F, Minvielle-Sebastia L. Molecular dissection of mRNA poly(A) tail length control in yeast. Nucleic Acids Res. 2008 Apr;36(7):2418-33.
163. Brune C, Munchel SE, Fischer N, Podtelejnikov AV, Weis K. Yeast poly(A)-binding protein Pab1 shuttles between the nucleus and the cytoplasm and functions in mRNA export. RNA. 2005 Apr;11(4):517-31.
164. Anderson JT, Paddy MR, Swanson MS. PUB1 is a major nuclear and cytoplasmic polyadenylated RNA-binding protein in Saccharomyces cerevisiae. Mol Cell Biol. 1993 Oct;13(10):6102-13.
165. 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.
166. 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.
167. Sachs AB, Sarnow P, Hentze MW. Starting at the beginning, middle, and end: translation initiation in eukaryotes. Cell. 1997 Jun 13;89(6):831-8.
168. Vitaliano-Prunier A, Babour A, Herissant L, Apponi L, Margaritis T, Holstege FC, et al. H2B Ubiquitylation Controls the Formation of Export-Competent mRNP. Mol Cell. 2012 Jan 13;45(1):132-9.
169. Johnson SA, Cubberley G, Bentley DL. Cotranscriptional recruitment of the mRNA export factor Yra1 by direct interaction with the 3' end processing factor Pcf11. Mol Cell. 2009 Jan 30;33(2):215-26.
170. 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.
171. Perales R, Bentley D. "Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions. Mol Cell. 2009 Oct 23;36(2):178-91.
172. Cramer P, Bushnell DA, Kornberg RD. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. Science. 2001 Jun 8;292(5523):1863-76.
173. Kim H, Erickson B, Luo W, Seward D, Graber JH, Pollock DD, et al. Gene-specific RNA polymerase II phosphorylation and the CTD code. Nat Struct Mol Biol. 2010 Oct;17(10):1279-86.
174. Rosonina E, Kaneko S, Manley JL. Terminating the transcript: breaking up is hard to do. Genes Dev. 2006 May 1;20(9):1050-6.
175. 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.
176. 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.
177. Steinmetz EJ, Brow DA. Ssu72 protein mediates both poly(A)-coupled and poly(A)-independent termination of RNA polymerase II transcription. Mol Cell Biol. 2003 Sep;23(18):6339-49.
178. 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.

179. 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.
180. 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. 2011;6(12):e29139.
181. Hampsey M, Singh BN, Ansari A, Laine JP, Krishnamurthy S. Control of eukaryotic gene expression: gene loops and transcriptional memory. Adv Enzyme Regul. 2011;51(1):118-25.
182. Singh BN, Hampsey M. A transcription-independent role for TFIIB in gene looping. Mol Cell. 2007 Sep 7;27(5):806-16.
183. 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.
184. 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.
185. 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.
186. 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.
187. Danckwardt S, Hentze MW, Kulozik AE. 3' end mRNA processing: molecular mechanisms and implications for health and disease. EMBO J. 2008 Feb 6;27(3):482-98.
188. 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.
189. 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.
190. Higgins DG. CLUSTAL V: multiple alignment of DNA and protein sequences. Methods Mol Biol. 1994;25:307-18.
191. Higgins DG, Bleasby AJ, Fuchs R. CLUSTAL V: improved software for multiple sequence alignment. Comput Appl Biosci. 1992 Apr;8(2):189-91.
192. Higgins DG, Sharp PM. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene. 1988 Dec 15;73(1):237-44.
193. Sambrook J, Russell, D.W., Maniatis, T., Fritsch, E.F. Molecular Cloning: A Laboratory Manual. NY, U.S.A.: Cold Spring Harbor Press; 2001.
194. Wen J, Arakawa T, Philo JS. Size-exclusion chromatography with on-line light-scattering, absorbance, and refractive index detectors for studying proteins and their interactions. Anal Biochem. 1996 Sep 5;240(2):155-66.
195. Van Holde KE, Johnson, W. C., Ho, P.S. Principles of Physical Biochemistry. NJ, U.S.A.: Prentice-Hall; 1998.
196. Reversed Phase Chromatography: Principles and Methods: SNITS & DESIGN AB for Amersham pharmacia biotech; 1997.
197. 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.
198. Amberg DC, Burke, D.J., Strathern, J.N. Methods in Yeast Genetics 2005: A Cold Spring Harbor Laboratory Course Manual. Cold Spring Harbor Laboratory Press,US. 2005.

199. Boeke JD, Trueheart J, Natsoulis G, Fink GR. 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. Methods Enzymol. 1987;154:164-75.
200. Boeke JD, LaCroute F, Fink GR. A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. Mol Gen Genet. 1984;197(2):345-6.
201. Lingner J, Keller W. 3'-end labeling of RNA with recombinant yeast poly(A) polymerase. Nucleic Acids Res. 1993 Jun 25;21(12):2917-20.
202. Holbein S, Freimoser FM, Werner TP, Wengi A, Dichtl B. Cordycepin-hypersensitive growth links elevated polyphosphate levels to inhibition of poly(A) polymerase in Saccharomyces cerevisiae. Nucleic Acids Res. 2008 Feb;36(2):353-63.
203. Kubista M, Andrade JM, Bengtsson M, Forootan A, Jonak J, Lind K, et al. The real-time polymerase chain reaction. Mol Aspects Med. 2006 Apr-Jun;27(2-3):95-125.
204. Gowher H, Brick K, Camerini-Otero RD, Felsenfeld G. Vezf1 protein binding sites genome-wide are associated with pausing of elongating RNA polymerase II. Proc Natl Acad Sci U S A. 2012 Feb 14;109(7):2370-5.
205. Edmunds JW, Mahadevan LC, Clayton AL. Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. EMBO J. 2008 Jan 23;27(2):406-20.
206. Sapojnikova N, Maman J, Myers FA, Thorne AW, Vorobyev VI, Crane-Robinson C. Biochemical observation of the rapid mobility of nuclear HMGB1. Biochim Biophys Acta. 2005 May 25;1729(1):57-63.
207. Ranade K, Chang MS, Ting CT, Pei D, Hsiao CF, Olivier M, et al. High-throughput genotyping with single nucleotide polymorphisms. Genome Res. 2001 Jul;11(7):1262-8.
208. Morrison TB, Weis JJ, Wittwer CT. Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification. Biotechniques. 1998 Jun;24(6):954-8, 60, 62.
209. Holland PM, Abramson RD, Watson R, Gelfand DH. Detection of specific polymerase chain reaction product by utilizing the 5'----3' exonuclease activity of Thermus aquaticus DNA polymerase. Proc Natl Acad Sci U S A. 1991 Aug 15;88(16):7276-80.
210. Applied Biosystems User Bulletin #2- *ABI Prism 7700 Sequence Detection System*. Life Technologies, U.S.A.; 1997; Available from: http://www.appliedbiosystems.com/.
211. IUPAC. Basic definitions of terms relating to polymers. Pure Appl Chem. 1974;40:477-91.
212. IUPAC. Stereochemical definitions and notations relating to polymers (recommendations 1980). Pure Appl Chem. 1981;53:733-52.
213. IUPAC. Note on the terminology for molar masses in polymer science. Makromol Chem. 1984;185.
214. Hohmann S, Magerm W.H. Yeast Stress Responses: Springer; 2003.
215. Johnston M, Davis RW. Sequences that regulate the divergent GAL1-GAL10 promoter in Saccharomyces cerevisiae. Mol Cell Biol. 1984 Aug;4(8):1440-8.
216. Schindelin H, Kisker C, Schlessman JL, Howard JB, Rees DC. Structure of ADP x AIF4(-)-stabilized nitrogenase complex and its implications for signal transduction. Nature. 1997 May 22;387(6631):370-6.
217. Zhou T, Radaev S, Rosen BP, Gatti DL. Structure of the ArsA ATPase: the catalytic subunit of a heavy metal resistance pump. EMBO J. 2000 Sep 1;19(17):4838-45.

218. Zhou T, Radaev S, Rosen BP, Gatti DL. Conformational changes in four regions of the Escherichia coli ArsA ATPase link ATP hydrolysis to ion translocation. J Biol Chem. 2001 Aug 10;276(32):30414-22.
219. Harding HP, Lackey JG, Hsu HC, Zhang Y, Deng J, Xu RM, et al. An intact unfolded protein response in Trpt1 knockout mice reveals phylogenic divergence in pathways for RNA ligation. RNA. 2008 Feb;14(2):225-32.
220. Lappe-Siefke C, Goebbels S, Gravel M, Nicksch E, Lee J, Braun PE, et al. Disruption of Cnp1 uncouples oligodendroglial functions in axonal support and myelination. Nat Genet. 2003 Mar;33(3):366-74.
221. Jain R, Shuman S. Characterization of a thermostable archaeal polynucleotide kinase homologous to human Clp1. Rna. 2009 Mar 19.
222. Fasman GD. Prediction of Protein Structure and the Principles of Protein Conformation. New York: Plenum; 1989.
223. Levin JZ, Yassour M, Adiconis X, Nusbaum C, Thompson DA, Friedman N, et al. Comprehensive comparative analysis of strand-specific RNA sequencing methods. Nat Methods. 2010 Sep;7(9):709-15.
224. 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.
225. Gilchrist DA, Fargo DC, Adelman K. Using ChIP-chip and ChIP-seq to study the regulation of gene expression: genome-wide localization studies reveal widespread regulation of transcription elongation. Methods. 2009 Aug;48(4):398-408.