Regulation of RNA Polymerase II Transcription in *Schizosaccharomyces pombe* by SpELL and Associated Factor SpEAF

Thesis

Regulation of RNA polymerase II transcription in *Schizosaccharomyces pombe* by SpELL and associated factor SpEAF

Charles Antony Scott Banks B.Sc. (Honours)

Submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

Biochemistry

Stowers Institute for Medical Research

The Open University

April 2010

ABSTRACT

RNA polymerase II (pol II) synthesised RNAs encode for proteins (mRNA), aid in splicing and telomere maintenance (snRNPs), and down-regulate gene expression (microRNAs). Therefore, pol II can be a critical target for modulating differential gene expression. Understanding mechanisms that underlie the various stages of transcription will elucidate origins of gene misexpression, which can give rise to a host of human diseases. Eleven-nineteen lysine-rich in leukaemia, (ELL), affects pol II transcription. The *ELL* gene was initially identified as a gene translocation fusion partner of the mixed lineage leukaemia (*MLL*) gene in patients with acute myeloid leukaemia (1). ELL interacts with pol II *in vitro*, increasing the rate at which it elongates nascent transcripts (2). Two ELL-associated factors, EAF1 and EAF2, bind ELL family members and act as strong positive regulators of their transcription activities (3).

Although ELL and EAF orthologs have been identified in metazoan organisms, previous attempts to identify similar elongation factors in lower eukaryotes had been unsuccessful. This thesis describes the identification of two genes in the yeast *Schizosaccharomyces pombe*, *ell1* and *eaf1*, that share some sequence similarity to highly conserved regions in *ELL* and *EAF1/EAF2* (4). Biochemical characterisation of these gene products, SpELL and SpEAF, shows that they have similar activities to their metazoan counterparts *in vitro*. Using a whole genome approach, I identified genes that are likely direct targets for regulation by SpELL/SpEAF in cells. Notably, ELL and EAF functions have not been studied previously at a genomic level. One SpELL/SpEAF target, *sme2*, was used as a model gene to investigate how the SpELL/SpEAF complex is recruited to chromatin.

Together with results of *in vitro* binding assays, these results suggest the SpELL/SpEAF complex is recruited to pol II via the SpELL subunit.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS AND NOMENCLATURE

% (v/v) ..... ml per 100ml (volume/volume)
% (w/v) ..... grams per 100ml (weight/volume)
°C ..... degrees centigrade
aa ..... amino acid(s)
AML ..... acute myelogenous leukaemia
ALL ..... acute lymphoblastic leukaemia
ATP ..... adenosine 5'-triphoshate
BLAST ..... Basic Local Alignment Search Tool
bp ..... base pair
Bq ..... Becquerel
BSA ..... bovine serum albumin
CBP ..... CREB binding protein
Cdk9 ..... cyclin dependent kinase 9
cDNA ..... complimentary deoxyribonucleic acid
ChIP ..... chromatin immunoprecipitation
Ci ..... Curie
clonNAT ..... nourseothricin
CPSF ..... cleavage and polyadenylation specificity factor
CREB ..... cAMP response element binding protein
CSB ..... Cockayne syndrome type B
CstF ..... cleavage stimulation factor
CTD ..... carboxy terminal domain
CTP ..... cytidine 5' triphosphate
Da ..... Dalton
DEAE ..... diethylaminoethyl
dELL ..... *Drosophila melanogaster* ELL
DEPC ..... diethylpyrocarbonate
DNA ..... deoxyribonucleic acid
DRB 5,6-Dichloro-1-β-D-ribofuranosylbenzimidazole
DSIF ..... DRB sensitivity inducing factor
DTT ..... dithiothreitol
Dvl ..... Dishevelled
EAF ..... ELL associated factor

EEC ..... early elongation complex
EDTA ..... diaminoethanetetra-acetic acid
ELL ..... eleven-nineteen lysine rich in leukaemia[1]
EMM..... Edinburgh minimal medium
ESCRT..... endosomal sorting complex required for transport
FACT..... facilitates chromatin transcription
FCP1 ..... TFIIF associating CTD phosphatase 1
FCS ..... fluorescence correlation spectroscopy
GAL4..... positive regulator of galactose inducible genes 4
GTP ..... guanosine triphosphate
g..... gram
h..... hour
HAT..... histone acetyltransferase
HEPES..... N-[hydroxyethylpiperazine –N '-[2-ethanesulphonic acid]
Hox..... homeobox
HPLC..... high performance liquid chromatography
Igh..... immunoglobulin heavy chain
IP ..... immunoprecipitation
IMP..... inosine monophosphate
ITC..... initially transcribing complex
ITP..... inosine triphosphate
kb ..... kilobase
M ..... molar
MDa..... megadalton
ME..... malt extract
mg ..... milligram
min..... minute
miRNA..... micro ribonucleic acid
ml ..... millilitre
MLL..... mixed-lineage leukaemia or myeloid / lymphoid leukaemia
mM..... millimolar
MOPS..... 3-(N-morpholino)propane sulphonic acid
mRNA..... messenger RNA

[1] The term ELL was originally used to label dishes because the name of the 41 year old "anonymous" female patient whose cells were used in the study began "Ell...". The meaning of the acronym was adjusted later for publication (Michael Thirman, personal communication) (1).

NELF..... negative elongation factor
nm..... nanometre
NP-40..... Nonidet P-40
N-terminal..... amino terminal
NTP..... nucleotide triphosphate
NuA4..... nucleosome acetyltransferase of histone H4
OD..... optical density
ORF..... open reading frame
PAGE..... polyacrylamide gel electrophoresis
PBS..... phosphate buffered saline
Pce1..... pombe capping enzyme 1
PCR..... polymerase chain reaction
Pct1..... partner of cdc ten 1
PIC..... pre-initiation complex
pmol..... picomole
pol II..... RNA polymerase II / RNA polymerase B
PP_i..... inorganic pyrophosphate
pre-mRNA..... precursor messenger ribonucleic acid
PSI-BLAST..... position specific iterated BLAST
P-TEFb..... positive elongation factor b
qPCR..... quantitative polymerase chain reaction
RACE..... rapid amplification of cDNA ends
RAP30..... RNA polymerase II-associating protein 30
RNA..... ribonucleic acid
Rpb1..... RNA polymerase B (pol II) subunit 1
Rpd3S..... reduced potassium dependency 3 small
rpm..... revolutions per minute
rRNA..... ribosomal RNA
RSC..... remodels the structure of chromatin
RT-qPCR..... reverse transcription followed by quantitative PCR
SAGA..... Spt-Ada-Gcn5-acetyltransferase
Scr..... sex combs reduced
SDS..... sodium dodecyl sulphate
Set1..... Suppressor of variegation 3-9, Enhancer of zeste, Trithorax 1
shRNA..... short hairpin RNA
SII..... factor stimulating RNA pol II - II (TFIIS)
SIII..... factor stimulating RNA pol II - III (Elongin)
snRNA..... small nuclear RNA
snRNPs..... small nuclear ribonuclear proteins
S. pombe..... Schizosaccharomyces pombe

Spt5..... suppressor of Ty 5
SR..... serine arginine rich protein
SS..... splice site
SSC..... sodium chloride / sodium citrate (buffer)
Su(Tpl)..... Suppressor of Triplo-lethal
SWI/SNF..... switch / sucrose non fermentable
TPG ..... translocation partner gene
TBE ..... Tris-borate-EDTA buffer
TBP ..... TATA binding protein
TE..... Tris-EDTA buffer
TFIIA..... general transcription factor II A[1]
TPG ..... translocation partner gene
Tris..... Tris[hydroxymethyl aminomethane]
tRNA..... transfer RNA
trx ..... trithorax
U..... Units
UTP..... uridine triphosphate
UV..... ultraviolet
V ..... volts
Vps ..... vacuolar protein sorting
Wnt ..... Wingless / int-1 (integration site 1)
XPB..... Xeroderma Pigmentosum B
YES..... yeast extract with supplements

[1] Similarly for transcription factors TFIIB, TFIID, TFIIE, TFIIF, TFIIH and TFIIS

CHAPTER 1. INTRODUCTION

1.1 Regulating gene expression

In lower eukaryotes cells perform different functions at different times during the cell cycle and survive by responding appropriately to different environmental situations. In higher organisms different types of cells perform specialized functions. Different complements of proteins are needed at different times and in different cells from the same organism for survival. Correct orchestration of gene expression is achieved in multiple ways; protein production can be regulated during all stages of gene transcription (initiation, elongation and termination), during RNA processing and transport, during translation of mRNA into protein and through the process of protein degradation. The earliest stages of protein production involve DNA directed synthesis of RNA transcripts by the enzyme RNA polymerase II (pol II) and offer key opportunities for regulating gene expression. This introduction will examine the functions of pol II in transcribing RNA, discuss the transcription elongation factor ELL, and introduce *S.pombe* as a model system for answering questions about the function of ELL in cells.

1.2 RNA polymerase II transcription

In eukaryotes, there are three nuclear enzymes capable of catalysing the polymerisation of ribonucleic acid (RNA) chains from the more stable DNA templates used to store information in cells. Pol I and pol III are largely responsible for synthesis of the structural RNAs used to translate messenger RNAs into proteins (pol I synthesises the large ribosomal RNAs (rRNA), pol III the smaller 5S rRNA and translation RNAs

(tRNAs))[1]. Pol II transcribes precursors of messenger RNA (mRNA), many small nuclear RNAs (snRNAs), and micro RNAs (miRNAs) used to control gene expression (6). As in other eukaryotes, the *S. pombe* pol II is composed of twelve subunits. Structurally, it may more closely resemble human pol II than its counterpart in baker's yeast, in particular in features related to transcription factor interaction and start site selection (7). This section of the introduction will examine the function of pol II as a catalyst and then follow the various roles of pol II through the transcription cycle.

1.2.1 RNA polymerase II as a molecular machine

RNA polymerase II functions to catalyse the addition of ribonucleotides to an RNA transcript in a DNA template dependent manner (Figure 1-1):

$$NTP + RNA_n \rightleftharpoons RNA_{n+1} + PP_i$$

This reaction, though reversible, is usually driven in the direction of nucleotide addition in cells as inorganic pyrophosphate is removed from the reaction by pyrophosphatase. A magnesium ion (Mg^{2+}) at the active site positions the incoming nucleotide by interacting with the α-phosphate of the NTP. A second metal ion more loosely associated with pol II interacts with, and is possibly removed with, the inorganic pyrophosphate of the NTP. The positioning of the NTP by the two metal ions enables

[1] In addition pol III synthesises a number of other types of RNA including U6 snRNA, a component of the spliceosome (5).

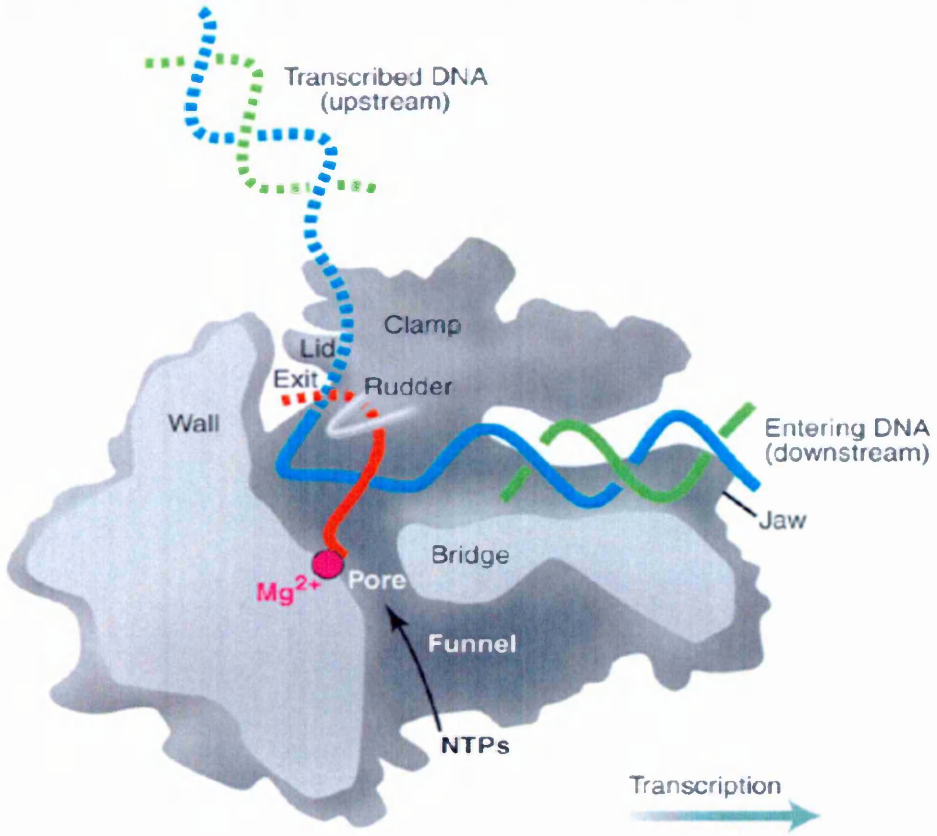


Figure 1-1 A transcribing ternary complex.

Figure is from (10) © Roger Kornberg reprinted with permission. Cross sectional diagram showing some structural features of the elongating RNA polymerase II enzyme important for its function. As pol II transcribes from left to right, the downstream DNA is unwound and the template strand shown in blue is positioned near the 3' end of the growing RNA chain (red) and the active site magnesium ions (pink). Polymerisation of the RNA chain occurs in a template dependent manner as nucleotides access the active site via the "funnel" and "pore" . As each nucleotide is added, a change in conformation of the bridge helix causes a translocation of the enzyme with respect to the DNA / RNA positioning the new 3' end of the RNA chain at the active site ready for further nucleotide addition. The RNA chain is separated from the template by the rudder and guided towards the exit channel. During transcript elongation the enzyme remains stably associated with the DNA and RNA; downstream DNA is gripped by a pair of "jaws" (only one of which is shown here) and nucleic acids are also restrained by a flexible "clamp" which closes over the DNA and RNA hybrid during formation of the transcribing complex (9).

nucleophillic attack by the 3' OH of the RNA transcript, resulting in formation of a phosphodiester bond, addition of the nucleotide, and release of inorganic pyrophosphate (8, 9). *In vitro*, the reverse reaction is favoured when concentrations of PP_i are high and nucleotide concentrations are low.

In addition to this straightforward polymerisation reaction, in which RNA polymerase II elongates a transcript that has already acquired sufficient length to form a stable complex, pol II needs to be able to perform other functions. For example, pol II needs to be able to catalyse the formation of the first phosphodiester bond between two nucleotides in the absence of a polynucleotide transcript (initiation) and then to proceed through early stages of elongation, where the RNA transcript has not grown long enough to occupy the RNA exit channel and stabilize the complex. Pol II also has to be able to remove nucleotides in an attempt to restart the elongation reaction should the 3' end of the transcript become misaligned with the active site (a stalled complex). Finally pol II needs to be able to terminate the elongation reaction and release the transcript at a suitable site.

1.2.2 Initiating transcription

The process of forming a stable complex capable of elongating an RNA transcript starting from an appropriate place on the DNA template is called initiation. The barriers to successful initiation depend on the nature of the DNA template presented to pol II and provide ample opportunities for regulating gene expression. For example, pol II is capable of accurate initiation without the need for accessory factors when presented with a "naked " double stranded piece of DNA to which a single stranded tail has been added to the complimentary strand. In this simple *in vitro* reaction, pol II initiates at the junction between the single and double strand. More problematic, and more physiologically relevant, is the requirement to initiate from a position within a section of double stranded DNA. Here, regulatory DNA sequences such as the TATA box and other core regulatory elements direct pol II to begin transcription at the appropriate site, and a set of general

transcription factors is required to help position pol II, unwind the DNA and form the first phosphodiester bonds.

The first step of this process is the formation of a pre-initiation complex (PIC) at the promoter. For promoters containing the TATA box, the TBP (Tata Binding Protein) subunit of the general transcription factor TFIID binds to, and bends the DNA, most commonly at the sequence 5'-TATAAAA-3'. This serves as a platform for the assembly of other factors, TFIIA and TFIIB. TFIIA, which is not essential for initiation from TATA box containing promoters *in vitro* (11), stabilises the interaction between the DNA and TBP while TFIIB recruits a pol II/TFIIF complex and directs transcription to begin at a defined start site downstream of the TATA box (reviewed in 11). The heterodimer TFIIF was originally identified as a mammalian RNA pol II interacting factor consisting of subunits RAP74 and RAP30 and shown to be essential for accurate initiation (12). The Drosophila factor 5 (TFIIF) was also shown to stimulate the rate of transcription elongation on oligo(dC) tailed templates (13), and this activity will be discussed later. As well as targeting pol II to the PIC, TFIIF also functions to reduce non-specific interactions between pol II and DNA at initiation (14, 15). It has also been proposed that TFIIF dependent bending of DNA within the PIC may facilitate unwinding of promoter DNA (16). TFIIH is recruited to the PIC by TFIIE (17-19).

Once the cast of factors is in place at the promoter and PIC formation has been completed, energy from the hydrolysis of ATP is used to unwind the DNA to form an open but unstable complex in preparation for transcription initiation (20). The XPB subunit of TFIIH, a 3'→5' DNA helicase, unwinds the promoter DNA to form the open complex and prevent premature arrest of early elongation complexes (21). It has been proposed that in *S. cerevisiae* RNA pol II scans the template DNA for an initiator

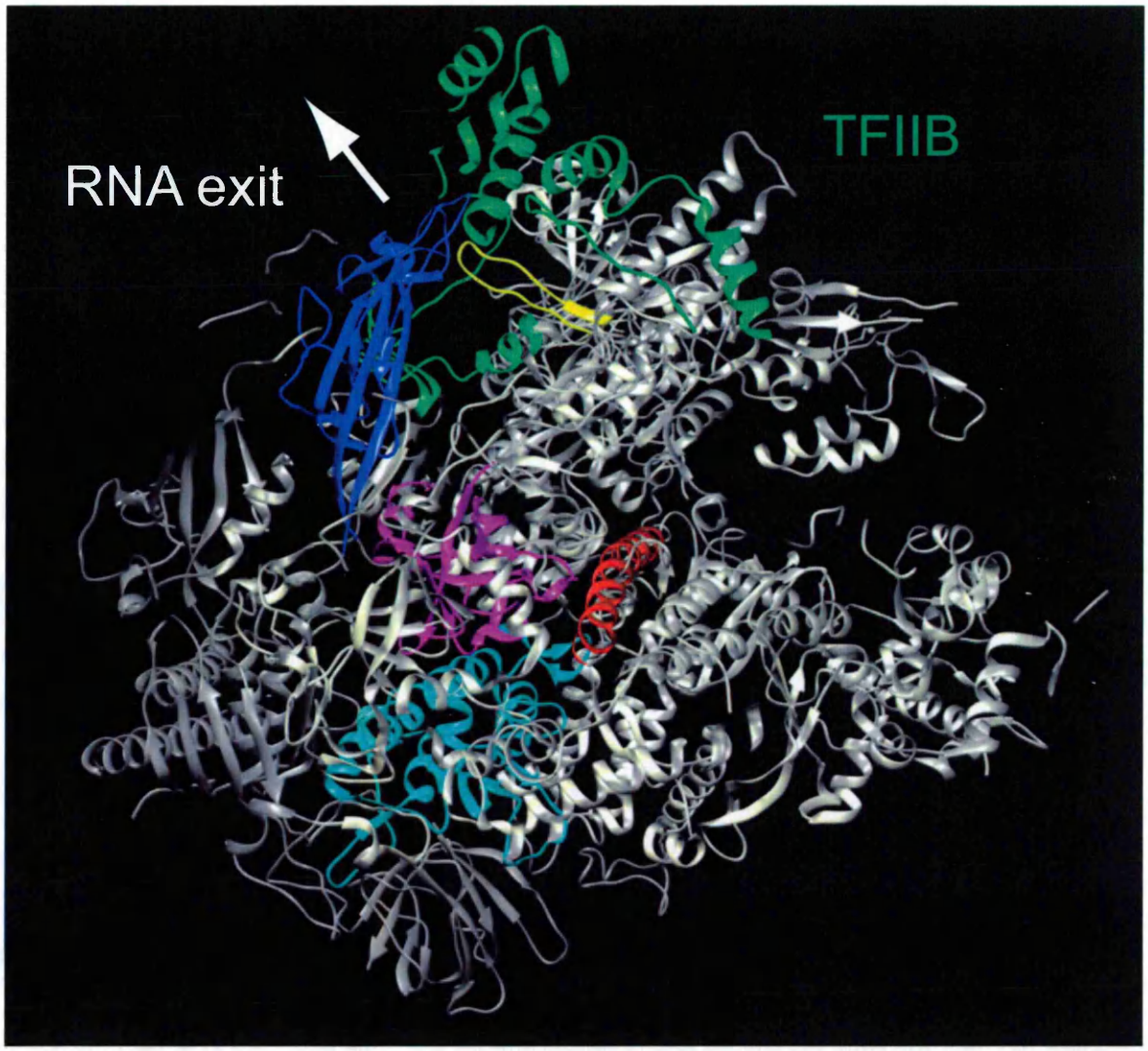
sequence (22); recent structural studies have suggested that start site selection may be determined by sequence specific interactions between the DNA template strand and the B-reader loop and B-reader helix regions (together also designated the B finger) of TFIIB (23, 24). Pol II is now in a position to begin catalysing the polymerisation of nucleotides, but, in the absence of a sufficiently long RNA transcript occupying the enzyme and with pol II still tethered to the initiation factors assembled at the promoter, the complex remains unstable in the early stages of elongation.

Additional barriers serve to limit pol II access to DNA template in cells. Here the DNA is not naked but bound to nucleosomes. The various mechanisms that are known to provide pol II access to DNA packaged in this way will be discussed later.

1.2.3 Early stages of elongation

Figure 1-2 TFIIB binds to RNA pol II and blocks the RNA exit channel.

A cutaway view of RNA pol II (grey) with key features highlighted and residues nearer the viewer removed. The RNA exit channel (white arrow) which lies between the wall (dark blue) and the lid (yellow) is blocked by the B finger domain of TFIIB (bright green). For orientation, other features include the bridge helix (red), the active site (purple) and the pore (light blue). The image has been produced by highlighting structural features in the file 3K1F.pdb (24) using the UCSF Chimera structural package (supported by NIH grant P41 RR-01081) (25) and the final image was processed and rendered in Autodesk Maya.



During the early stages of elongation, there are a number of distinct steps that lead to the formation of a stable ternary complex (reviewed in (26)).

1. *The Initially Transcribing Complex (ITC).* The start of transcription involves the formation of the first phosphodiester bond, after which the elongating pol II is

referred to as the initially transcribing complex (ITC). During polymerisation of the first 4-5 ribonucleotides, the B-finger of transcription factor TFIIB remains in the RNA exit channel and helps to stabilise the binding of short transcripts to pol II (23, 26). The ITC is unstable, and a number of short transcripts may be generated without the initiation complex disengaging from the DNA in a process known as abortive initiation. Abortive initiation may be caused by interference between the growing RNA and the TFIIB B-finger (Figure 1-2) (24). In eukaryotes there is a decrease in abortive initiation after the third phosphodiester bond has been formed (27), and the complex has been described as being committed to escape from the promoter (reviewed in (26)).

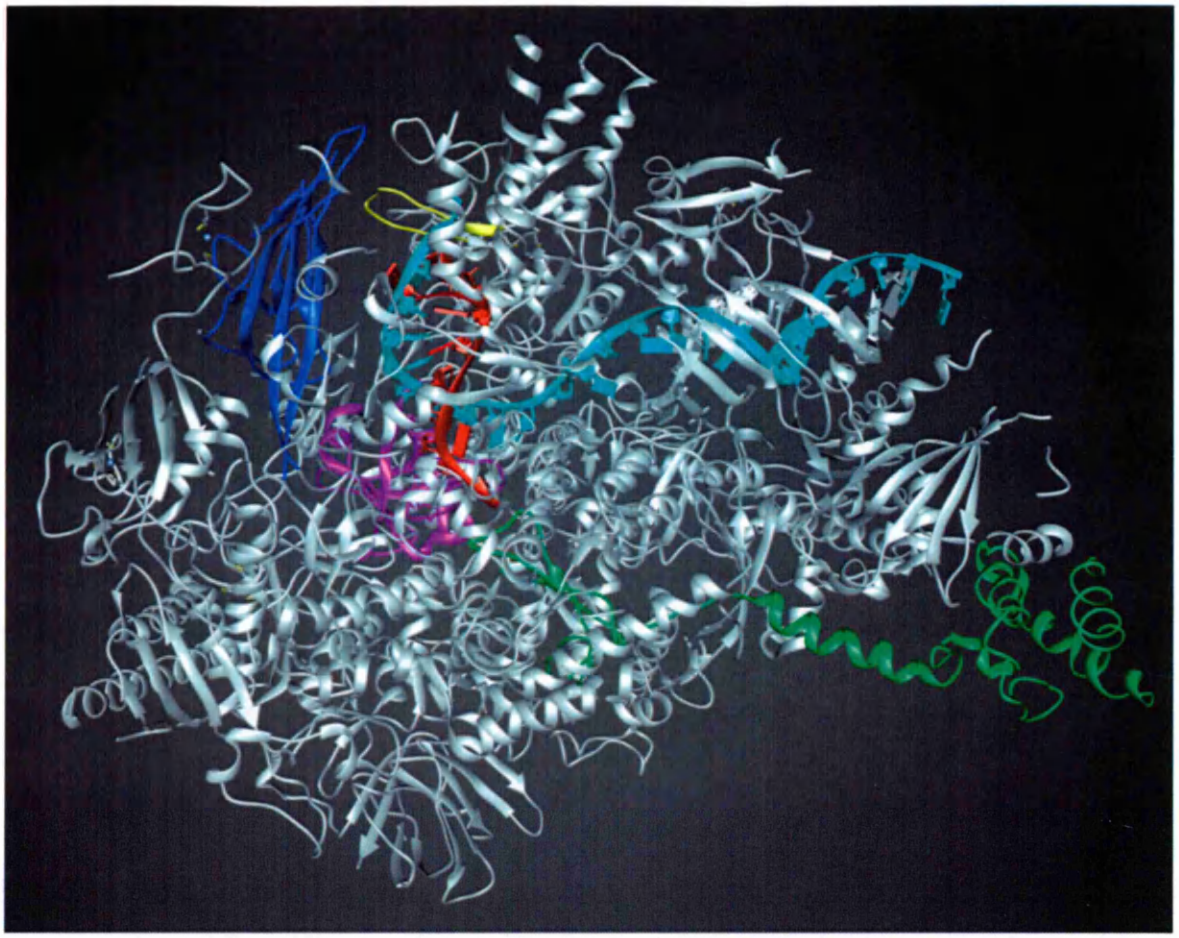


Figure 1-3 A backtracked pol II bound to TFIIS.

A cross sectional representation of pol II is shown in a reverse translocated or backtracked state with the 3' end of the RNA transcript (red) misaligned with the active site (purple). A domain of TFIIS (green) is inserted into the pore/funnel region of pol II and stimulates cleavage of a dinucleotide at the 3' end of the transcript. Other features included for viewer orientation are the wall (dark blue), lid (yellow) and DNA template strand (light blue). The image has been produced by highlighting structural features in the file 3GTM.pdb (28) using the UCSF Chimera structural package (supported by NIH grant P41 RR-01081) (25) and the final image was processed and rendered in Autodesk Maya.

2. *Promoter Clearance.* Kostrewa and co-workers describe a model for promoter clearance (escape) in which the nascent RNA triggers the release of TFIIB and binds stably to pol II, after which the transcribing unit is referred to as the early elongation complex (EEC) (24, 29). Early elongating complexes are no longer susceptible to abortive initiation, and the energy from ATP hydrolysis is no longer required for transcript elongation. The EEC is not as stable as a fully functional

ternary complex and is liable to transcript slippage (until about +23) or backtracking (until about +30), which results in transcriptional arrest as the pol II active site becomes misaligned with the 3' end of the transcript. The transcription factor SII (TFIIS) can induce pol II catalysed transcript cleavage to form a new 3' end properly aligned at the active site allowing elongation to continue (Figure 1-3) (28, 30, 31).

3. *CTD phosphorylation during initiation.* A number of proteins contain largely unstructured protein tails, which can be modified in different ways by reactions involving, for example, phosphorylation or methylation of different amino acid residues. This provides a system for the temporally ordered recruitment of factors that would catalyse a series of events in sequence. The C terminal domain of pol II subunit Rpb1 consists of a series of tandem repeats of the heptapeptide consensus sequence YSPTSPS, the number of repeats ranging from 52 in humans to 29 in *S. pombe* and 27 in *S. cerevisiae* (6, 26). The functional unit may be a heptapeptide pair, since recruited proteins may bind to more than one CTD repeat (32). The primary targets for modification are serine 2 and serine 5, although other modifications include phosphorylation of serine 7, glycosylation of the serines and threonines and isomerisation of the proline residues (reviewed in (33)).

   The ordered phosphorylation and dephosphorylation of Ser2, Ser5, and Ser7 is catalysed by a series of kinases and phosphatases. The phosphorylation of Ser5 marks the presumptive initiating / early elongating form of pol II, whereas phosphorylation of Ser2 is associated with productively elongating pol II (34). During the sequence of events at initiation, Cdk7, a subunit of TFIIH,

phosphorylates Ser5. This phosphorylation reaction begins on formation of the PIC and may continue through the early stages of elongation (35). Cdk8, a subunit of the mediator complex, can also phosphorylate both Ser5 and Ser2. During the progression to elongation, Cdk9, a subunit of P-TEFb, phosphorylates Ser2. The TFIIF associated phosphatase FCP1 binds to a pol II sub module composed of subunits Rpb4 and Rpb7 and can dephosphorylate both Ser2 and Ser5. The modified CTD provides a platform for recruiting capping, splicing, cleavage and polyadenylation machinery in a timely manner. For example the RNA capping enzyme is recruited to pol II during early elongation (+20 to +30) (36). Complexes may be recruited to the CTD either though contacts with the phosphorylated repeats and / or repeats with altered structure (33).

4. *Pausing in the promoter proximal region.* Transcriptionally engaged polymerase is susceptible to controlled pausing in the region +20 to +40 (26). The pausing factors DSIF and NELF cooperate to induce promoter proximal pausing, possibly by binding of NELF to the nascent RNA. The repressive effects of DSIF and NELF can be relieved by the CTD kinase P-TEFb (Positive Transcription Elongation Factor b), allowing the transcribing pol II to enter the phase referred to as productive elongation. Early pausing may provide a mechanism to allow the RNA transcript to be capped (37).

1.2.4 Transient pausing during productive transcript elongation

Experiments with yeast RNA polymerase III suggest that elongation cannot be modelled simply as a sequence of pseudo-first order reactions (38). Instead, observations are consistent with a model in which pol III can switch between a rapidly stepping "celerous" state, $\varepsilon 1$, and a slowly stepping "slothful" state, $\varepsilon 2$, during elongation (38). A

plausible model is that the slothful state, ε2, results from a modest dislocation of the 3' end of the transcript relative to the catalytic site. Realignment of the 3' end of the transcript with the active site converts the complex to state ε1 allowing transcription to resume (38).

One can envision at least two kinds of mechanisms that could lead to transient pausing. First, limitations in the amount of nucleotides available for elongation could lead to pausing, since the rate of nucleotide addition is slowed at positions where a limiting nucleotide must be incorporated. Indeed, *in vitro* pausing patterns can be altered by changing nucleotide concentrations (38, 39). Importantly, however, pausing is not eliminated at saturating nucleotide concentrations, arguing that pausing is not solely due to substrate limitations. Second, pausing can be strongly influenced by template sequence. As the ternary complex engages in productive elongation, the exact nature of the polymerase / nucleic acid complex changes as different sequences[1] of DNA and RNA become threaded through the enzyme (Figure 1-1). Each of these may provide slightly different conditions that could influence conversion of the complex to the more slowly stepping state and thereby could influence the time taken for the catalytic addition of a nucleotide. Although some sequences may act to promote transcription termination, many others cause pausing by the elongating ternary complex but do not favour termination (40). Studies with E. coli polymerase have demonstrated that such pausing occurs at many sites along a DNA template during transcript elongation in vitro, and

[1] Structural studies indicate that in *S. cerevisiae* there are eleven base pairs of unwound DNA within the pol II "jaws", three bases of unwound template downstream of the active site and nine base pairs of DNA - RNA hybrid (9) suggesting a minimum of $4^{23} = 7.0 \times 10^{13}$ possible unique ternary complexes due to the different possible sequences of nucleic acid which may be associated with transcribing polymerase.

pause sites have been mapped for a number of DNA templates (39, 41). Although it is possible to determine pause sites empirically in this way, systematic prediction of pause sites from primary sequence has not been reported.

Pausing at some sites has been found to be dependent on the RNA transcript upstream of the pause site; when polymerase transcribed in the presence of ITP (inosine triphosphate) rather than GTP, pausing was reduced at some sites (39). Because ITP base pairing with CTP is less stable than GTP with CTP, it has been suggested that RNA base pairing may cause pausing at some sites. Consistent with this idea, predicted RNA hairpin structures were found just upstream of a number of pause sites where substitution of ITP for GTP reduced pausing (39). Adding an oligonucleotide complimentary to part of the hairpin also suppresses pausing at sites with RNA hairpin structures (42).

Not all pausing depends on the nature of the RNA transcript upstream of the pause site. Pausing at some sites has been shown to be dependent on DNA sequences downstream of the pause site, and pausing at these sites was insensitive to substitution of GMP with IMP in the transcript. Interestingly, pausing influenced by downstream DNA sequences is not thought to depend on the ease of melting the DNA template as GC rich sequences inserted downstream of a pause site failed to enhance pausing (43). Recent studies indicate that contacts between the "jaw" region of pol II and downstream DNA sequences may be involved in pausing at some sites; mutations in the "jaw" region of *E. coli* RNA polymerase II suppresses pausing at a site for which pausing depends in part on a downstream sequence (44).

In conclusion, although there is a possibility of conversion of the ternary complex from a "celerous to a "slothful" state at any point during transcription, the probability of

pausing is greater at certain sequences. This may be due to base pairing in the RNA transcript, interactions between pol II and the downstream DNA template, or a combination of both. The change in the ability of the ternary complex to catalyse the addition of the next nucleotide may be due at least in part to the 3' end of the transcript becoming misaligned with the pol II active site.

1.2.5 Increasing the overall rate of transcript elongation by reducing transient pausing

Several transcription elongation factors that reduce transient pausing on naked DNA templates *in vitro* have been described.

1. In addition to its role in promoter-specific initiation, TFIIF can interact with transcribing pol II and stimulate the rate of elongation. Structure function studies identified regions of RAP 30 essential for elongation activity including an N-terminal region that binds RAP74 and a central region that binds to pol II (45, 46). The role of TFIIF as an elongation factor *in vivo* remains unclear. Recent chromatin immunoprecipitation studies found TFIIF present at promoters but not in coding regions (47). TFIIF elongation function may be limited to the early stages in transcription. Studies by Yan and co-workers found that a TFIIF mutant selectively lacking elongation activity was unable to suppress abortive initiation (16). TFIIF might increase the processivity of early elongation complexes and prevent premature release of short, unstable transcripts (16).
2. Elongin consists of an active subunit, Elongin A, and two regulatory subunits, Elongin B and Elongin C (48). The Elongin complex can also form an E3 ubiquitin ligase and may be responsible for the polyubiquitination and degradation of Rpb1

in response to DNA damage in yeast and human (49-51). Elongin A is not essential for cell viability but disruption of Elongin A leads to slow growth of mammalian cells (52) and defective development in *Drosophila* (53).

3. ELL is the focus of this thesis and will be discussed in the third section of this introduction.
4. The Cockayne syndrome group B protein, CSB, has been shown to have both chromatin remodelling activity and to stimulate pol II elongation on naked DNA. CSB can interact directly with double stranded DNA and histones and uses the energy from ATP to remodel nucleosomes *in vitro* (54) and *in vivo* (55). CSB can also stimulate transcription elongation on naked DNA templates *in vitro*, probably through direct interaction with elongating pol II (56). In addition to these two activities CSB is also required for transcription-coupled DNA repair and may be recruited to pol II stalled at DNA damage sites.

It is not known how these factors reduce transient pausing. As TFIIF, Elongin, ELL and CSB all stimulate elongation in the presence of only DNA template, pol II and ribonucleotides in highly purified, reconstituted transcription systems, it seems likely that stimulation is the result of interaction with some element of the ternary complex (DNA template, RNA transcript, pol II or some combination of these three) (2, 57).

Three lines of evidence suggest that these elongation factors may reduce the number of pol II molecules in the slothful state, $\varepsilon 2$, by helping to keep the 3' end of the nascent transcript properly aligned with the active site:

1. Maintaining alignment between the 3' end of the transcript with the active site should prevent pausing or arrest caused by backward movement of pol II relative

to the transcript. which could either result in pausing or arrest. TFIIF has been shown to reduce the probability that pol II complexes will arrest at DNA sequences that function as strong arrest sites (58).

2. The elongation factor SII stimulates endonucleolytic cleavage near the 3' end of non-arrested transcripts *in vitro* by the polymerase active site. This reaction would presumably require the active site to have become aligned with the cleavage site rather than the 3' end of the transcript. TFIIF, ELL and Elongin inhibit this SII mediated transcript cleavage (59).
3. RNA pol II can catalyse template directed addition of nucleotides to the 3' ends of duplex DNA *in vitro* in a reaction that seems to mimic transcription elongation (60). In doing so, pol II binds to the 3' OH terminus of one strand of the DNA as if it were the 3' OH end of a nascent transcript. The reaction is strongly stimulated by TFIIF, ELL and Elongin, and these observations are consistent with the idea that these elongation factors promote the alignment of the 3' terminus of the DNA with the pol II active site.

1.2.6 Processing the RNA transcript

Early in the process of transcribing a gene, eukaryotes begin to process the RNA transcript, adding a 7-methylguanosine cap at the 5' end, splicing out introns, and polyadenylating the 3' end of the transcript prior to transcription termination and export of the transcript from the nucleus (reviewed in (61)). Many of the proteins needed for these events have been defined and, as with transcription, the reactions have been investigated using purified proteins *in vitro*. The evidence outlined below suggests that RNA processing factors are recruited to the transcribing complex and that RNA processing and transcription are coordinated.

1. *RNA capping.* Capping the 5' end of the nascent transcript serves both to protect the RNA from degradation by exonucleases and to enhance export of the transcript from the nucleus to the cytoplasm. Capping of many transcripts occurs as the RNA begins to emerge from pol II. Capping enzyme binds to poll II CTD that has been phosphorylated at serine 5 during early elongation. Capping enzyme may also be allosterically activated by binding the CTD (61). Pct1, a triphosphatase that converts the 5' end of a transcript from a triphosphate to a diphosphate, and Pce1, a guanylyltransferase that caps the 5' end of the transcript with a GMP moiety, are recruited to the pol II CTD in fission yeast. Experiments with synthetic CTD peptides have shown that phosphorylation of serine 5 but not serine 2 is required for recruitment of capping enzyme *in vitro*, and *in vivo* interactions between capping enzymes and the CTD in a two hybrid assay require 12 or 14 YSPTS repeats (62). Both Pct1 and Pce1 also bind to the Spt5 subunit of elongation factor DSIF, and Pct1 binds to the Cdk9 subunit of the elongation factor pTEFb, suggesting that coordination between capping and control of early elongation could be a mechanism used to prevent productive elongation of uncapped transcripts (37, 63).
2. *Pre-mRNA Splicing.* Introns are removed from transcripts during transcription by the spliceosome, a complex consisting of snRNPs (small nuclear ribonucleoproteins) and other proteins (non-snRNPs) including members of the SR family. Hyperphosphorylated pol II coimmunoprecipitates with both snRNPs and SR proteins suggesting a physical interaction between pol II and the spliceosome (64). Expression of Rpb1 lacking a CTD in *Xenopus* oocytes results in

inefficient capping, splicing and polyadenylation of pre-mRNA, suggesting a functional role for the CTD in these processing activities (65).

3. *3' end processing.* Before transcription terminates, RNA cleavage and polyadenylation machinery is recruited to the hyperphosphorylated pol II CTD. The precise location of transcript cleavage is determined by DNA sequences including the hexanucleotide sequence AATAAA in higher eukaryotes and less well characterised sequences in *S. pombe* (66). Cleavage is catalysed by CPSF, which binds to the body of pol II and may then scan for the cleavage and polyadenylation hexamer (AAUAAA) on the transcript. Recognition of the polyadenylation hexamer triggers pol II to pause, and CPSF can then catalyse cleavage of the transcript 10-30 nucleotides downstream (67). Pausing does not require the pol II CTD but the CTD is required for cleavage and termination. CPSF may be transferred from the body of pol II to the CTD after the pause by the cleavage stimulation factor CstF (67). Polyadenylation of the cleaved transcript by polyadenylate polymerase, which binds CPSF, serves to protect the 3' end of the transcript from degradation and aids RNA export and translation. The polymerase continues to transcribe after transcript cleavage until directed to terminate but the resulting RNA is quickly degraded.

1.2.7 Transcribing the correct genes in cells

DNA in cells needs to be tightly packaged into chromatin and this has an overall repressive effect. For genes to be transcribed, the DNA needs to be made accessible and the transcriptional machinery recruited to the appropriate promoter. A host of chromatin remodelling complexes and histone modifying enzymes are recruited by DNA binding regulatory proteins (activators and repressors), and modulate access to

chromatin, while some DNA binding regulatory proteins recruit coactivators such as the mediator complex that are responsible for assembly and control of RNA pol II and the general transcription machinery (reviewed in (68, 69)).

Where does transcription occur? The nucleus is not uniform, and areas of condensed, repressive heterochromatin are distinguishable from less condensed euchromatic regions where transcription occurs. There is evidence that transcription occurs in nanostructures called "transcription factories" where active polymerases are localised at discrete sites (70-72). Some models for transcription factories suggest that pol II might remain immobilised as the template moves through it. Recently, 87nm, 10MDa structures that are rich in nitrogen (protein) but low in phosphorous (nucleic acid) have been observed associated with labelled nascent RNA, leading to the suggestion that transcription factories consist of a protein rich core with the DNA templates and RNA transcripts attached to the surface (73).

Unpackaging the DNA template. The DNA template needs to be accessible to the transcription apparatus throughout the transcription cycle. Chromatin remodelling involves the rearrangement of nucleosomes on DNA allowing transcriptional processes to occur. Although nucleosomal DNA is somewhat accessible to DNA binding transcription factors, chromatin remodelling complexes can increase transcription factor binding (reviewed in (68)). Once bound, activators recruit chromatin-remodelling complexes and histone modification enzymes that make nucleosomal DNA more accessible to the general transcription factors. *In vitro*, transcription on templates containing nucleosomes exhibits enhanced sequence specific pausing during elongation, and pol II is unable to achieve physiological rates of gene transcription on chromatin templates, even in the presence of elongation factors (74, 75).

Histone modifications associated with gene activation include Acetyl-Co-A dependent acetylation by histone acetyltransferases (HATs), such as SAGA and NuA4, and methylation by methyltransferases including Set1 and Set2. SAGA and NuA4 can stimulate recruitment of the chromatin remodelling complex SWI/SNF *in vivo* (76). More recently Carey and co-workers investigated transcription elongation on nucleosomal templates *in vitro*. They found that another ATP dependent chromatin remodeler, RSC, stimulates elongation through nucleosomes, and that RSC's effect on elongation was stimulated by either SAGA or NuA4 (77, 78). Methylation of histones at H3K4 by Set1 and at H3K36 by Set2 occurs in the coding regions of actively transcribed genes (68, 79). The H3K36 mark may function to control histone acetylation by helping to recruit Rpd3S, a histone deacetylase. The recruitment of this HDAC is thought to be important because although histone acetylation enables remodelling of chromatin, which in turn enhances transcription elongation, it also allows DNA sequences within transcribed regions to become exposed to transcription initiation machinery resulting in spurious intergenic transcription (80).

In addition to chromatin remodelers moving histones along the DNA, transcription can also be enhanced by removing histones or by replacing them with histone variants using histone chaperones. The histone chaperone complex FACT enhances productive transcription elongation on chromatin templates in *vitro* (81) and is localised to actively transcribed pol II genes on *Drosophila* polytene chromosomes (82). FACT removes histone H2A / H2B dimers from nucleosomes leaving "hexasomes" and may also help redeposit the dimers after the passage of pol II (83, 84).

Activating transcription. In addition to controlling gene expression by modifying chromatin structure and making the DNA accessible, expression is controlled by actively

recruiting the transcription machinery to a gene. Pol II and the general transcription factors are not recruited directly to DNA binding activators but require additional cofactors. Mediator is a multisubunit complex first isolated from yeast. It binds directly to both DNA binding transcriptional regulators and to pol II to control both activation and repression of genes (85, 86). The mediator complex is conserved from yeast to man and is thought to be required for all activated transcription (87). Interestingly, mediator has also been reported in the coding regions of genes, perhaps indicating a role for mediator in transcription elongation (88). Mediator subunit Med13 interacts with transcription elongation factor SII (TFIIS) (89), and the ELL / EAF1 complex binds to Med26 (CASB and Tari J Parmely unpublished data).

1.3 The role of the *ELL* gene product

1.3.1 Discovery of the human *ELL* gene through its link with leukaemia

The *ELL* gene was first identified as a fusion partner with the *MLL* (mixed-lineage leukaemia) gene in the translocation t(11;19)(q23;p13.1) in cells taken from the bone marrow of a patient with acute myeloid leukaemia. The gene resulting from this translocation encodes a protein in which the C terminus of ELL (aa 148-621) is fused to the N terminus of MLL (aa 1-1527) (1, 90). The MLL-ELL chimera causes immortalisation when introduced into haematopoietic progenitor cells. Transplantation into mice of cells immortalised by the MLL-ELL chimera causes acute myeloid leukaemia (91). In humans, ELL (eleven-nineteen lysine-rich leukaemia) is a 621 amino acid protein with a C-terminal domain homologous to the C terminus of occludin, a membrane protein that localises to tight junctions. ELL also includes regions required for binding to proteins called ELL-associated factors (EAFs) and regions that have been found to inhibit transcription

initiation or stimulate transcription elongation (Figure 1-4). Two paralogues of *ELL* have been identified in humans, *ELL2* (92) and *ELL3* (93). Deletion of *ELL* in *Drosophila* (*dELL* or "suppressor of triplo-lethal") is recessive lethal, and *ELL* mutations cause embryonic segmentation defects (94).

Chromosomal rearrangements involving *MLL* translocations occur in a number of acute or therapy induced leukaemias and to date 50 translocation partner genes (TPGs) in addition to *ELL* have been identified (95). The *MLL* gene product, a homolog of *Drosophila* trithorax (trx), is a histone methyltransferase that positively regulates gene expression during development and is known to target *Hox* genes (96, 97). The N terminus, which is retained in the MLL fusion proteins, has domains, including an AT hook motif and a CpG region similar to the DNA binding domain of DNA methyltransferase, that are expected to be able to bind DNA and that may help target *MLL* to genes. The C terminal domain contains regions that may modulate the accessibility of chromatin leading to gene activation. These include an H3K4 methyltransferase SET domain, and a transactivation domain that can recruit the histone acetyltransferase CBP (96-98). In the *MLL-ELL* translocation, the carboxy terminal region of ELL, equivalent to amino acids 497-621 of the human protein, is necessary and sufficient for immortalisation of primary murine haemopoietic cells; MLL-ELL(1-373), which includes the transcription elongation domain but lacks the C terminal occludin domain, does not cause immortalisation (99). There has been speculation that recruitment of the EAF1 protein to MLL-ELL might be necessary for immortalisation, as the C terminus of ELL binds EAF1, and murine primary cells transfected with an MLL-EAF1 chimera become immortalised (although there has been no evidence of MLL-EAF1 induced leukaemias in humans) (100). Overexpression of the ELL protein alone has been reported to cause transformation of rat primary embryo

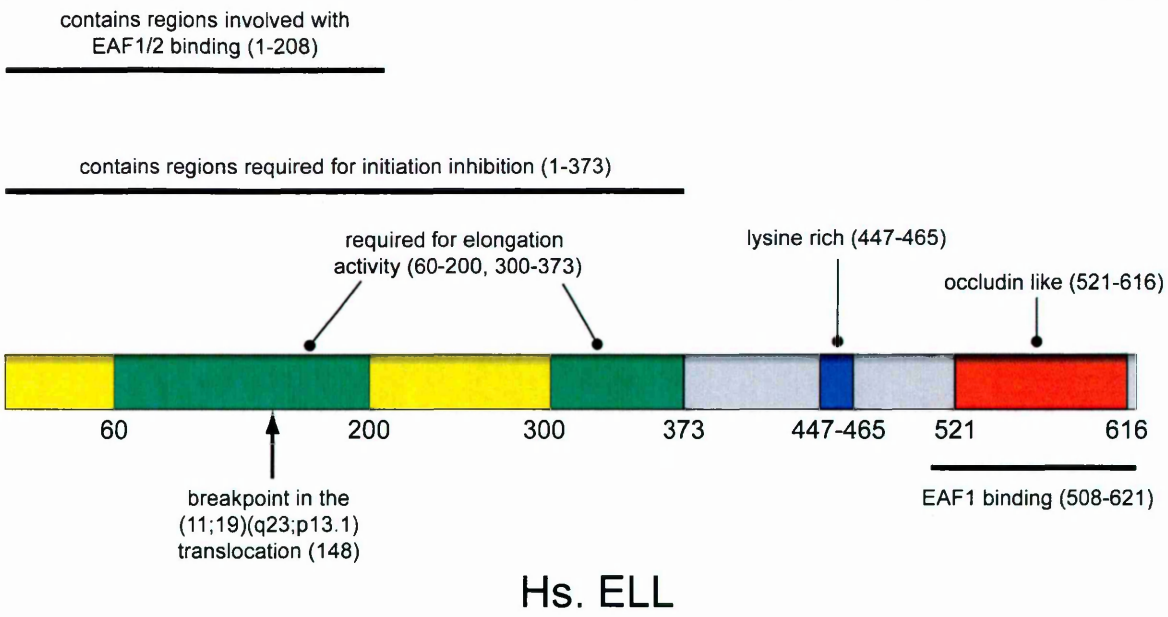


Figure 1-4 The human *ELL* gene product.

The locations of the regions indicated in the figure are described in the following publications: *ELL* full length gene, lysine rich region and translocation breakpoint (1), regions required for inhibition of transcription initiation or for stimulation of transcription elongation (103), occludin like domain (104), regions involved in EAF1 and EAF2 binding (105).

fibroblasts (with transformation dependent on the lysine rich region of ELL); however, in a subsequent study ELL overexpression did not transform primary haemopoietic cells in mice (91, 101). Very recently, a number of common MLL-fusion proteins including MLL-ELL have been reported to form complexes that include AFF4 (AF4/FMR2 family, member 4) and other translocation partner genes (102).

1.3.2 ELL localization

ELL shows broad tissue distribution in humans; transcripts have been detected by Northern blotting in peripheral blood leukocytes, skeletal muscle, placenta and testis and many major organs (1). ELL is conserved across metazoa (mammals, chicken, frog and

fish (1) and fly (106)) but prior to this study had not been identified in lower eukaryotes. Early in murine development, RNA transcripts are diffusely expressed until about day 16, when high levels become evident in the liver and gut. In adult mice, transcript levels are most pronounced in the liver and gastrointestinal tract (107). The ELL associated factor EAF2 is localised to the tissues of the nervous system during mouse embryogenesis (108).

Within cells, both ELL and EAF1 colocalise with Cajal bodies, nuclear compartments that contain components of the RNA processing machinery (109). In contrast two other elongation factors, Elongin and TFIIF show diffuse nuclear localisation. Colocalisation with Cajal bodies is dependent on transcription, with ELL/EAF appearing dispersed throughout the nucleus in cells treated with α-amanitin, a potent inhibitor of pol II transcription. It has been suggested that Cajal bodies may be sites for preassembly of complexes containing both transcriptional and RNA processing machinery, which are subsequently transported to the genes to be transcribed (110).

1.3.3 ELL interacts with pol II and can function as a transcription elongation factor

ELL interacts with RNA polymerase II. Coimmunoprecipitations using purified factors show that ELL binds directly to pol II in vitro both in mammals and in flies (103, 106). Although the presence of the pol II CTD is not required for ELL function or binding in vitro, CTD phosphorylation has been implicated in ELL recruitment *in vivo*. Knockdown of Cdk9 (the kinase subunit of P-TEFb) in flies resulted in reduced CTD phosphorylation and decreased chromosome binding of dELL suggesting that CTD phosphorylation might be required for dELL recruitment (111). Whether recruitment involves a direct interaction between pol II CTD and ELL is not known.

ELL stimulates transcription elongation in vitro. Shilatifard and co-workers originally identified ELL as a transcription elongation factor by purifying an activity from rat liver nuclear extracts over several chromatographic steps by its ability to stimulate transcription elongation by purified pol II on naked DNA templates *in vitro* (2). The activity copurified with a single protein, which was identified as a rat homologue of the 621aa human ELL protein (2). Analysis of recombinant human ELL deletion mutants showed that sequences required for stimulation of elongation were contained in the regions aa 60-200 and aa 300-373 (Figure 1-4) (103). In addition to its elongation activity, ELL also inhibits the activity of the elongation factor SII *in vitro*. The transcription elongation factor SII (TFIIS) acts to reactivate pol II that has become stalled by stimulating a pol II ribonuclease activity that hydrolytically cleaves a misplaced transcript at the active site[1] resulting in a complex competent to continue elongation (113).

$$RNA_n + H_2O \rightleftharpoons RNA_{n-m} + RNA_m$$

SII accesses the pol II active site via the funnel and pore (112) (Figure 1-3). The inhibition of this reaction by ELL could result from ELL preventing SII access to the active site or from ELL maintaining the position of the 3' end of the transcript in alignment with the pol II active site so as to prevent cleavage (59).

[1] The different RNA pol II activities share the same active site which is "tuneable" to catalyse either elongation / pyrophosphorolysis or hydrolytic cleavage (112).

Indirect evidence for a role of ELL as an elongation factor in vivo. Studies with *drosophila* ELL (dELL) have provided evidence that ELL functions during transcription elongation *in vivo*. Like cdk9 (P-TEFb) and Spt5 (DSIF), dELL colocalises with phosphorylated pol II on polytene chromosomes at transcriptionally active puff sites and relocalises to heat shock genes on heat shock (106, 114). The association with elongating pol II in flies only requires an N terminal domain of dELL homologous to Hs. ELL aa 1-373. Eissenberg and co-workers have suggested that dELL may be preferentially targeting large genes. Mutations in *dELL* cause misexpression of *Scr* (23kb), *Notch* (35kb) and *cut* (65kb) but do not affect the smaller *rudimentary* (13kb) and *white* (5.9kb) (94).

1.3.4 ELL associated factors EAF1 and EAF2

Two homologous proteins that interact directly with human ELL have been identified (105, 115). These ELL associated factors EAF1 and EAF2 positively regulate ELL elongation activity in vitro (3). Binding to the EAF proteins is achieved through regions in both the N and C termini of ELL (EAF1) or through the N terminus only (EAF2) (Figure 1-4). Activation of ELL activity by EAF1 requires not only the transcription elongation domains in ELL but also the first 45 N terminal ELL amino acids, which overlap an EAF1/EAF2 interaction domain. Activation can occur in the absence of the C terminal EAF1 binding domain, suggesting that the binding of ELL to EAF1 is not sufficient for stimulation of ELL activity. Upregulation of ELL elongation activity requires only the N terminal 118 aa of EAF1 (CASB unpublished data). As well as stimulating ELL activity, the C terminus of EAF1 contains an acidic transactivation domain that when fused to a GAL4 DNA binding domain, will stimulate transcription from templates containing GAL4 binding sites in whole cell extracts (115).

In addition to the ELL Associated Factors (EAFs) already described, ELL has been found to copurify as a complex with three proteins designated EAP20, EAP30 and EAP45. The association of the EAPs with ELL does not appear to affect the ability of ELL to stimulate elongation (116-118).

1.3.5 ELL interacts with transcription factor p53

The C terminal domain of p53, a transcription factor that can function as either an activator or a repressor, interacts with the ELL transcription elongation domain (119). This interaction negatively regulates ELL stimulated transcription from oligo(dC) tailed templates, and it has been proposed that p53 could repress general transcription at least in part by targeting ELL, while activating promoters containing p53 responsive elements by other, ELL-independent mechanisms (119).

1.3.6 ELL2 affects pre mRNA processing

ELL2 has recently been shown to regulate pre mRNA processing in plasma cells, in which alternative splicing and polyadenylation of the *Igh* gene results in synthesis of mRNAs encoding either the secreted or the membrane-bound form of the immunoglobulin heavy chain (120). ELL2 and the polyadenylation factor CstF-64 normally co-localise with pol II on the *Igh* locus. Reducing levels of ELL2 with ELL2 specific shRNA was found to decrease levels of ELL2 and CstF-64 to the locus. Depletion of ELL2 also results in a reduced proportion of the shorter secretory-specific transcripts (implying a decreased use of a weak promoter proximal polyA site), and an increase in longer transcripts (requiring use of a weak splice site) (Figure 1-5).

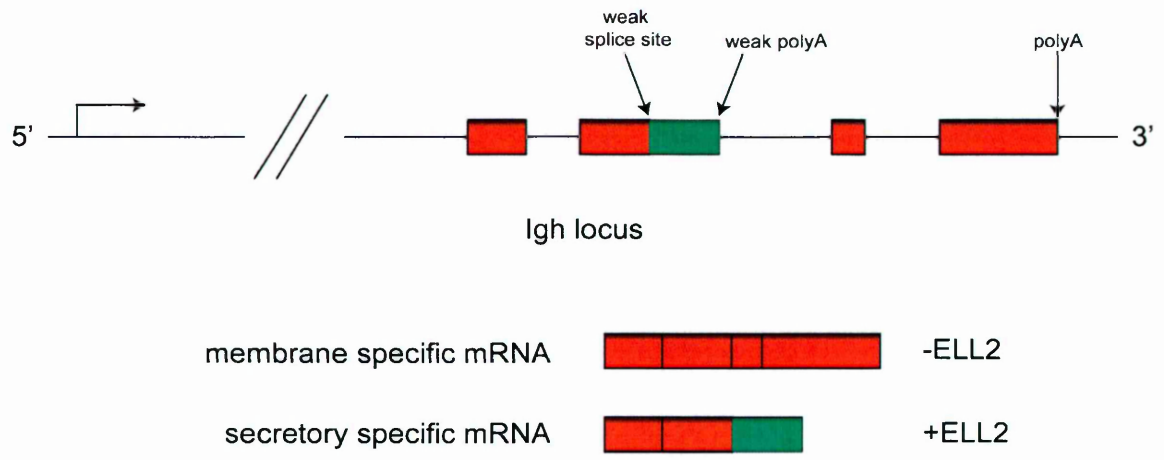


Figure 1-5 ELL2 influences *Igh* pre mRNA processing.

B cells, in which ELL2 expression is low, favour production of the membrane specific form of immunoglobulin heavy chain mRNA, whereas plasma cells, which express higher levels of ELL2, produce a secretory specific form of the mRNA.

It is thought that slowing of elongation near a weak splice site might help its recognition by allowing time for the splicing machinery to assemble. Conversely, rapid elongation promoted by ELL2 could decrease recognition of this weak alternative splice site, which is 5' to the weak polyA site, thereby favouring use of the weak polyA site rather than continued transcription and splicing. Indeed, one model for "exon skipping", in which a weak splice site is ignored, suggests that transcription elongation factors might help favour exon skipping as the increase in the rate of pol II elongation decreases the likelihood that a weak 3' splice site will be used (Figure 1-6) (121). Interestingly expression of ELL2 in non-plasma cells, where ELL2 is normally in low abundance, has been shown to increase exon skipping or to increase use of a weak polyA site in different reporter constructs transfected into cells (120).

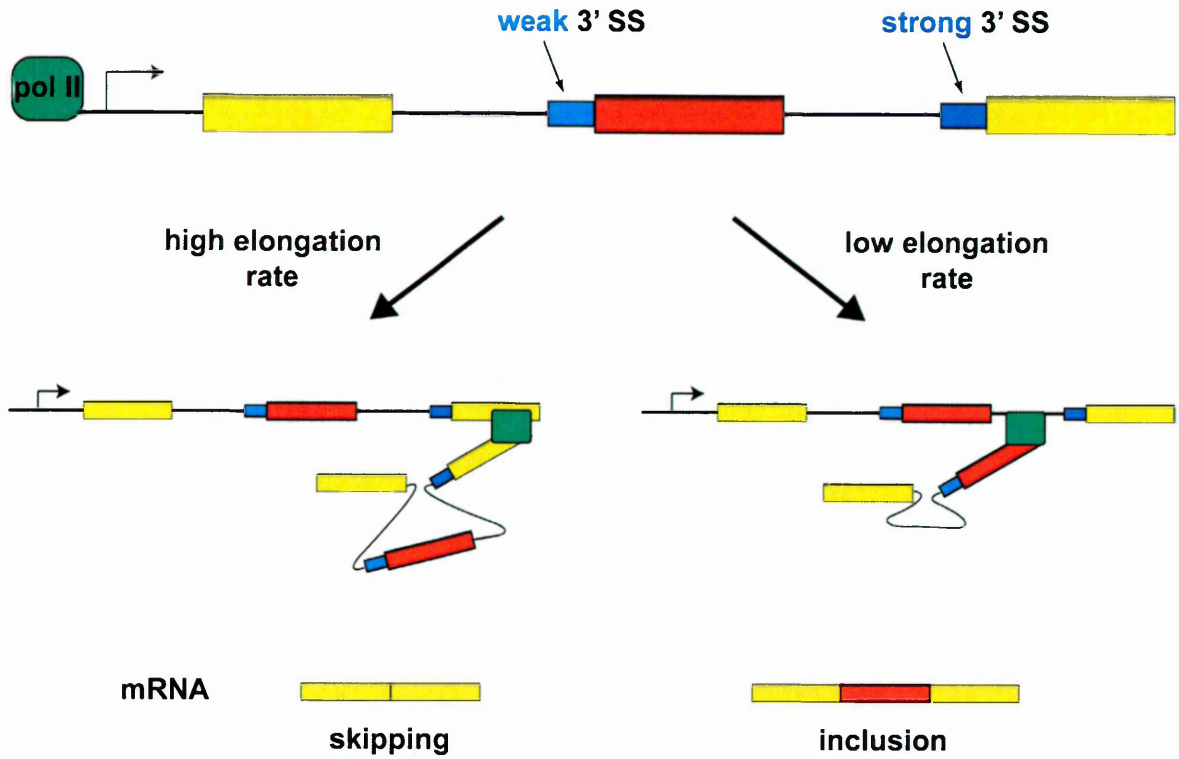


Figure 1-6 Rate of transcription elongation might regulate alternative splicing.

This illustration is based on figure 3 in (121). This model suggests that an increase in the rate of pol II elongation, possibly by factors that might reduce pausing and result in high pol II processivity, leads to an increase in exon skipping. In genes that contain weak alternative 3' splice sites (light blue) a high elongation rate favours transcription through the altenative exon (red) resulting in use of a stronger 3' splice site (dark blue) and removal of the alternative exon. Lower elongation rates favour inclusion of the alternative exon.

1.3.7 ELL role in development

ELL is necessary for the timely expression of some genes during development. Although deletion of *ELL* in mice results in viable heterozygotes (*ELL*+/-), intercrossing heterozygotes fails to generate homozygote mice (*ELL*-/-), with nearly a quarter of embryos exhibiting degenerative changes during early embryogenesis (122). Mutations in *Drosophila ELL* (*dELL*) cause misexpression of several genes during development (including *Scr*, *cut* and *Notch*), causes segmentation defects, and results in late embryonic lethality (maternally loaded *dELL* mRNA is present early in development) (94).

ELL associated factors have also been shown to be necessary for controlling developmental processes. Wnt-4, via the noncanonical Wnt signalling pathway, controls neural differentiation. EAF2 is a target for Wnt-4 signalling, is essential for proper eye development in *Xenopus laevis*, and functions by regulating transcription of the eye specific *Rx* gene (123). In these experiments Maurus and coworkers demonstrated that EAF2 controls levels of *Rx* mRNA but not control (*Pax-6*) RNA suggesting that ELL / EAF2 function might be gene specific.

1.3.8 Summary

ELL stimulates transcription elongation *in vitro* by binding to RNA pol II and suppressing transient pausing of the enzyme at many sites along the DNA. Pausing may be caused by conversion of the ternary complex from an actively elongating (celerous) state to a slowly stepping (slothful) state that results from the misalignment of the 3' end of the transcript and the pol II active site, and ELL might function at least in part by helping to maintain proper alignment. There is evidence that ELL functions together with elongating pol II *in vivo* and that the state of pol II CTD phosphorylation may affect recruitment of ELL either directly or indirectly. Knockdown of an ELL homologue leads to mislocalisation of a polyadenylation factor, an increase in exon skipping, and altered use of polyadenylation sites but it remains unclear whether this is due to a direct function of ELL (such as an altered rate of pol II elongation) or an indirect effect (such as a change in recruitment or function of the splicing and polyadenylation machinery). It is intriguing that ELL has a C terminal domain that appears to have a function independent of transcription elongation. The homology with the membrane-distal cytoplasmic domain of the membrane bound protein occludin, which is a target for modification and binds to several proteins, suggests that the ELL occludin domain may act as a docking site for other

factors including EAF1 (124). Finally, ELL has been shown to be essential for proper development and may be targeted to specific developmental genes.

1.4 Scope of this thesis

1.4.1 Using fission yeast as a model system to explore the role of ELL in cells

Previous attempts to find homologues of ELL had not identified it in lower eukaryotes, and there is no apparent homologue of ELL in *S. cerevisiae*. In this thesis, I will describe the identification and *in vitro* characterisation of homologues of ELL and EAF in the fission yeast *S. pombe*. I will then address questions about the function of these proteins, designated SpELL and SpEAF, in cells. Fission yeast is a valuable model system, having the advantages that it is easy to manipulate genetically and that some cellular processes are remarkably similar to their counterparts in Metazoa (125). *S. pombe* also uses some enzyme systems or components of systems that are not present in *S. cerevisiae*, for example RNA interference and parts of the spliceosome (126).

1.4.2 Questions to be addressed

This thesis begins by demonstrating that ELL and EAF do exist in a single celled eukaryote, identifing the coding regions of the *ell1* and *eaf1* genes, and demonstrating a physical interaction between the SpELL and SpEAF proteins. I then ask whether the *in vitro* activities of the human ELL/EAF complex are shared by SpELL/SpEAF and look at how they interact with pol II *in vitro*. I recently collaborated with Stephanie Kong to demonstrate that human ELL appears to function as a dimer with EAF in stimulating transcription elongation (3) and so, having confirmed that they function this way in *S. pombe*, wanted to ask whether they appeared to function together on chromosomes *in*

vivo. It has not previously been clear whether ELL is involved with regulating a specific subset of genes or whether its function is more general, nor has there been information about what properties SpELL regulated genes might share. Using a whole genome approach, I have attempted to define the most likely candidates for direct regulation by SpELL/SpEAF, and I have used one of these candidate genes to begin to answer mechanistic questions about how the SpELL/SpEAF complex might be recruited to genes in cells.

CHAPTER 2. MATERIALS AND METHODS

2.1 Materials

Unlabeled ultrapure ribonucleoside 5'-triphosphates and [α-32P]CTP (400 mCi/mmol; 1 Ci = 37 GBq) were purchased from Amersham Pharmacia Biosciences. Recombinant RNasin ribonuclease inhibitor was obtained from Promega. Anti- FLAG (M2) monoclonal antibodies, anti-Myc (C-3956) rabbit polyclonal antibodies, anti-FLAG (M2)-agarose, and anti-FLAG peptide were purchased from Sigma. Light chain-specific anti-mouse antibodies were from Bethyl laboratories and labelled with Alexa Fluor 680 (Invitrogen) according to the manufacturer's instructions.

2.2 Cloning of *S. pombe ell1* and *eaf1* genes

The SMART™ RACE cDNA Amplification system (Clontech) was used according to manufacturer's instructions to identify the 5' ends of each gene using the gene specific primers: 5'-GGGTGGAAGGCAAGGATTGCGGAGGAGA-3' (ELL) or

5'-TGCTGGCCGTTTGGGATACTGTAGAGGG-3' (EAF).

RACE products were cloned into pCR®2.1-TOPO® and the inserts were sequenced to identify the 5' end of each gene and to confirm the positions of introns. Sequences corresponding to the coding regions of the mRNA were then subcloned into pBacPAK8 using primers overlapping consecutive exons to remove introns from *S. pombe* genomic DNA sequences.

2.3 Strain construction

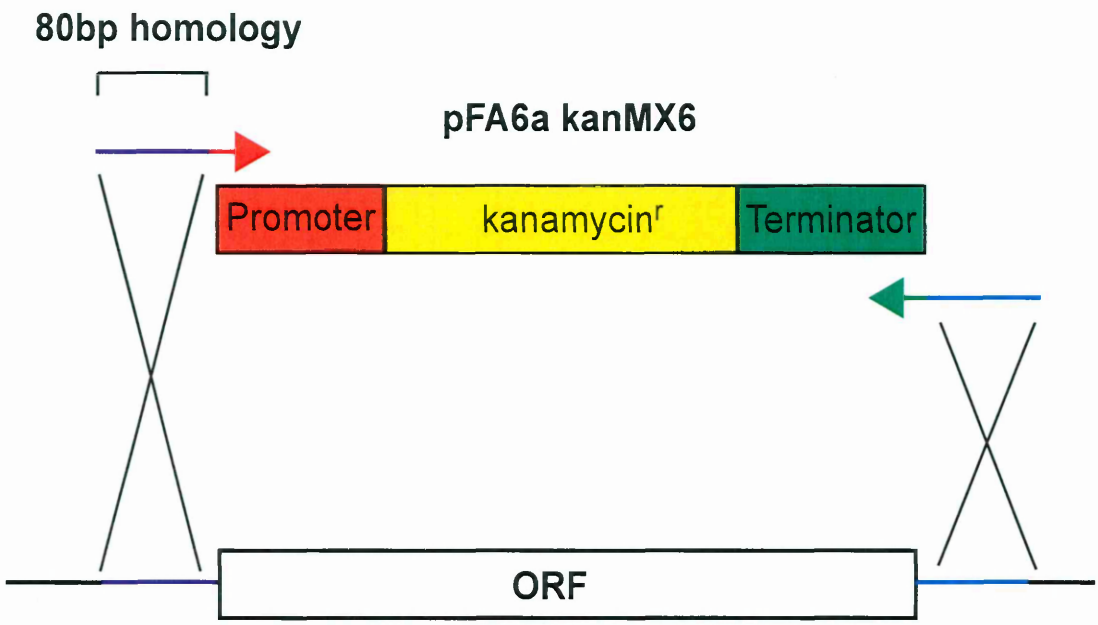


Figure 2-1 Construction of gene deletion strains.

PCR primers (indicated by arrows) were used with the pFA6a kanMX6 plasmid as template to generate a PCR product with 80 bp of homology to regions flanking the sequence to be deleted as indicated. Electroporation of the PCR product into *S.pombe* cells resulted in replacement of the target sequence with a kanamycin resistance cassette under the control of promoter and terminator sequences from *Ashbya gossypii*. Transformed cells were plated on YES containing 100 mg/l G418 / geneticin. G418 resistant transformants were streaked to isolation. Deletion of the target sequence and insertion of the resistance cassete at the correct locus were both confirmed by PCR.

The *S. pombe* strains used are listed in appendix A. *S.pombe* cells were grown at 32°C in rich medium (YES) supplemented with adenine, histidine, leucine and uracil (225 μg/ml) or minimal medium (EMM) with the supplements indicated. The gene deletion strains were generated by replacing the coding region of the *ell1* gene with the kanMX6 marker as described (127) (Figure 2-1). Epitope tagged strains were generated by replacing the stop codon of the target gene with the 13myc natMX6 cassette (127, 128) (Figure 2-2). I constructed strains containing both a gene deletion and an epitope tagged

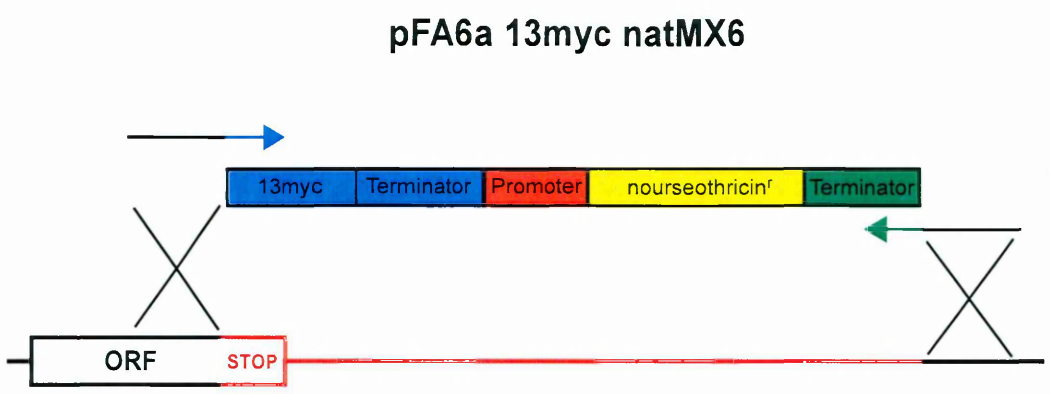


Figure 2-2 Construction of epitope tagged strains.

The target gene is indicated as the ORF (black) including the stop codon (red). PCR primers were designed with 80 bases of homology to the region immediately upstream of the stop codon (5') and a region 20-200 bp downstream of the stop codon (3'). These were used to generate a PCR product using the pFA6a 13myc nat MX6 plasmid as a template. This includes 13 tandem repeats coding for the c-myc epitope with a transcription termination sequence from the *S. cerevisiae ADH1* gene (blue) followed by a nourseothricin antibiotic cassette. The PCR product was used to generate recombinant strains as described in Figure 2-1 .

gene by crossing two strains with a single disruption. A mating mixture of the two stains on malt extract (ME) plates was incubated at 28°C for three days and dilutions of cells were then plated on media containing both geneticin and nourseothricin (clonNAT). An *ell1* rescue strain was constructed by inserting the *ell1* gene into the *leu1* locus of the *ell1Δ* strain using the pDUAL system as described (129).

2.4 Analysis of Proteins Associated With FLAG-tagged SpELL and SpEAF in *S. pombe*

cDNAs encoding SpELL or SpEAF with N-terminal FLAG epitope tags were subcloned into a modified version of pNMT-TOPO (Invitrogen), which carries a thiamine repressible promoter, and transformed into *S. pombe* strain PP138. Cells were grown in

EMM supplemented with 10 μM thiamine and 225 μg/ml each of adenine, histidine, and uracil. Once they had reached a density of ~$5x10^6$ /ml, cells were washed in EMM supplemented with adenine, histidine, and uracil alone to release thiamine repression and grown for an additional 18 hours. Cultures were harvested by centrifugation at 3,000 x g, washed in cold H_2O, and then washed in an extraction buffer containing 0.2 M Tris-HCl (pH 7.5), 0.39 M $(NH_4)_2SO_4$, 10 mM $MgSO_4$, 1 mM EDTA (pH 8.0), 20% (v/v) glycerol, 0.28 μg/ml leupeptin, 1.4 μg/ml pepstatin A, 0.17 mg/ml PMSF, and 0.33 mg/ml benzamidine. Cells were pulverized under liquid nitrogen by mortar and pestle, thawed, and resuspended in extraction buffer. Whole cell extracts were clarified by centrifugation at 150,000 x g before immunoprecipitation. FLAG-tagged SpELL or SpEAF and associated proteins were purified from whole cell extracts as described below under Purification of Recombinant Proteins. FLAG-immunopurified complexes were analyzed by MudPIT mass spectrometry as described previously (130-133).

2.5 Expression of Recombinant Proteins in Insect Cells

cDNAs encoding wild-type SpELL and SpEAF containing N-terminal FLAG or c-Myc epitope tags, as well as an SpEAF deletion mutant lacking the first 59 amino acids, were subcloned into pBacPAK8. Recombinant baculoviruses were generated with the BacPAK expression system (Clontech) using pBacPAK6 viral DNA prepared as described (134), and stored in 50 ml centrifuge tubes spray-painted black to prevent exposure to light during storage (135). The titer of recombinant baculoviruses was determined using the immunological assay developed by Kitts and Green (136). Sf21 insect cells were cultured at 27°C in Sf-900 II SFM (Invitrogen). Flasks containing 1 x 10^8 Sf21 cells were infected with the recombinant baculoviruses. Forty-eight hours after infection, cells were collected and lysed in 15 ml of ice-cold buffer containing 50 mM Hepes-NaOH (pH7.9),

0.5 M NaCl, 5 mM $MgCl_2$, 0.2% Triton-X-100, 20% (v/v) glycerol, 0.28 μg/ml leupeptin, 1.4 μg/ml pepstatin A, 0.17 mg/ml PMSF, and 0.33 mg/ml benzamidine. Lysates were centrifuged 100,000 x g for 30 min at 4 °C.

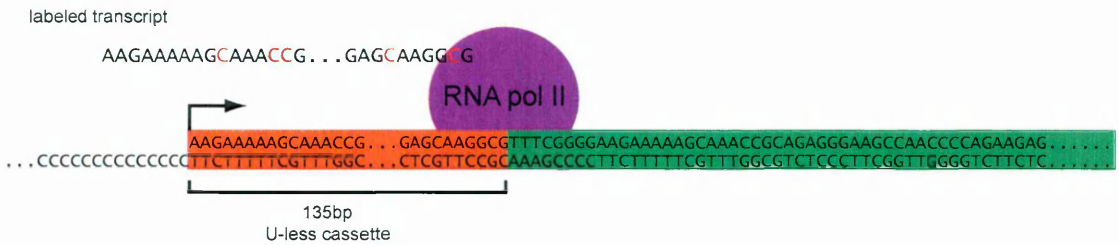
2.6 Purification of Recombinant Proteins

FLAG-tagged proteins were purified from Sf21 cell lysates by anti-FLAG agarose immunoaffinity chromatography. Lysates from 1 x 10^8 cells were incubated with 0.5 ml anti-FLAG (M2) agarose beads overnight at 4°C. The beads were washed three times with Tris-buffered saline (TBS, 25 mM Tris-HCl (pH 7.4), 137 mM NaCl, 2.7 mM KCl), and bound proteins were eluted from the beads with TBS containing 10% (v/v) glycerol and 0.3 mg/ml FLAG peptide. Where indicated, anti-FLAG agarose eluates prepared from Sf21 cells expressing both recombinant FLAG-SpELL and Myc-SpEAF were further purified by anion exchange HPLC. Eluates were adjusted to a conductivity equivalent to that of 0.05 M KCl and applied to a 0.6 ml TSK DEAE-NPR HPLC column (Tosoh-BioSep) equilibrated in Buffer A (40 mM Tris-HCl (pH 7.9), 1 mM EDTA, 1 mM DTT, 10% (v/v) glycerol) containing 0.1 M KCl. The column was eluted with a 6 ml linear gradient from 0.1 to 0.5 M KCl in Buffer A, and 0.2 ml fractions were collected. Concentrations of eluted proteins were estimated by using ImageQuant TL software (G. E. Healthcare) to compare the relative intensity of Coomassie Blue stained bands corresponding to full-length proteins to the intensity of bands corresponding to BSA standards after SDS/PAGE.

2.7 Preparation of RNA Polymerase II

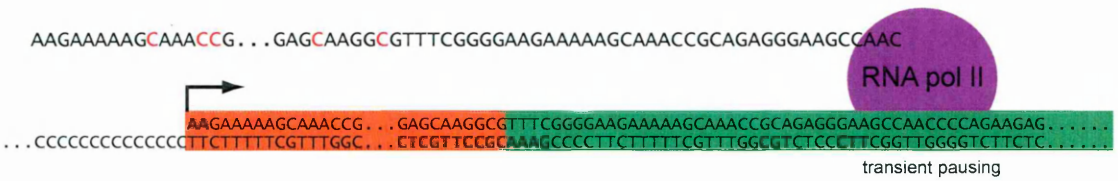
Mammalian pol II was purified from rat liver nuclei as previously described (14), except that a TSK DEAE 5-PW HPLC column was used in place of TSK DEAE-NPR. Pol II from *S. pombe* was a kind gift from Henrik Spåhr, which he purified as described (4).

1. Pulse: add RNA pol II, ATP, GTP, [α32P]CTP



2. Remove nucleotides

3. Chase: continue transcription by adding ATP, GTP, CTP and UTP (and elongation factors)



4. Stop Reaction

Figure 2-3 Pulse chase assay.

The oligo(dC) tailed pAdGR220 template contains a U-less cassette (red) to allow the formation of paused ternary complexes. (1) On the addition of the nucleotides indicated, RNA pol II will start transcription at the junction between the single and double strands of the template without the need for additional factors to initiate transcription and will pause at +135 bp from the start site in the absence of UTP. The 135 nucleotide transcript is labeled with [αP32]CTP. (2) Ternary complexes can then be purified free of nucleotides. This enables lower concentrations of unlabeled nucleotides to be used during the chase phase (3). The RNA pol II continues transcription elongation until the reaction is stopped.

2.7.1 Oligo(dC)-tailed Template Transcription Assays

Oligo(dC) tailed pAd-GR220 was prepared as described (137). 60 μl transcription assays were carried out in the presence of 20 mM Hepes-NaOH (pH 7.9), 20 mM

Tris·HCl (pH7.9), 60 mM KCl, 2 mM DTT, 0.5 mg/ml BSA, 3% (w/v) polyvinyl alcohol (average molecular weight 30,000-70,000 Da), 3% (v/v) glycerol, 8 units of RNasin, 8 mM $MgCl_2$, ~100 ng oligo(dC)-tailed pAd-GR220, ~25 ng pol II, and ribonucleoside triphosphates and transcription factors as indicated in the figure legends. The pulse chase assay used in Figure 3-4 is outlined in Figure 2-3. Reactions were preincubated for 5 min at 28 °C before the addition of ribonucleoside triphosphates. Reactions were stopped by the addition of 60 µl stop buffer (10 mM Tris·HCl, pH 7.2 / 0.5 mM EDTA / 0.3 M NaCl / 0.4% SDS) after incubation at 28 °C for the times indicated in the figures, and transcription products were resolved on 6% polyacrylamide gels containing 7 M urea, 45 mM Tris-borate, and 1 mM EDTA (pH 8.3) and detected with a Molecular Dynamics Typhoon PhosphoImager.

2.8 Preparation of Paused RNA Polymerase II Elongation Complexes

Paused pol II elongation complexes were assembled on oligo(dC) tailed pAd-GR220 by performing transcription reactions (scaled up 10 fold) in the presence of 50 µM ATP, 50 µM GTP, 2 µM CTP and 10 µCi [α-32P]CTP for 30 minutes at 28 °C. Two Sephadex G-50 spin columns were prepared by inserting a 5ml syringe (with the plunger removed) into a 15ml centrifuge tube, pushing a small amount of polyester aquarium filter fibre into the bottom of the syringe, and filling the syringe with Sephadex G-50 gel slurry (prepared as described (138)) equilibrated in buffer containing 20 mM Hepes-NaOH (pH7.9), 20 mM Tris·HCl (pH7.9), 50 mM KCl, 1 mM DTT, 0.5 mg/ml BSA, 2% (w/v) polyvinyl alcohol, 3% (v/v) glycerol, and 5 mM $MgCl_2$. The columns were allowed to drain, spun at 2000 x g to remove excess buffer and inserted into fresh centrifuge tubes. Free ribonucleoside triphosphates were then removed by applying ~600 µl of the reaction

mixture to two sequential Sephadex G-50 spin columns. The columns were spun for 5 min at 2000 x g in a swinging bucket rotor, and 60 µl of eluant/reaction mixture was used in further experiments.

2.9 Chromatin Immunoprecipitations (CHIPs)

Chromatin immunoprecipitations were carried out essentially as described (139) with modifications:

2.9.1 Preparation of Chromatin Extracts

S. pombe strains were grown in 100 ml YES to an OD600 between 0.9 and 1.1. Cells were crosslinked by adding formaldehyde to a final concentration of 1% for 15 minutes at room temperature[1]. The crosslinking reaction was quenched with the addition of 5 ml of 2.5 M glycine. Cells were harvested by centrifugation and the cell pellets were washed twice with ice-cold Tris-buffered saline (TBS) containing protease inhibitors (0.28 µg/ml leupeptin, 1.4 µg/ml pepstatin A, 0.17 mg/ml PMSF, and 0.33 mg/ml benzamidine). Cells were resuspended in 800 µl FA lysis SDS buffer (50 mM HEPES/KOH pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 % Triton X-100, 0.1% sodium deoxycholate, 0.2% SDS, 0.28 µg/ml leupeptin, 1.4 µg/ml pepstatin A, 0.17 mg/ml PMSF, and 0.33 mg/ml benzamidine) and aliquoted into four 1.5 ml microfuge tubes. After the addition of 200 µl acid washed glass beads (Sigma G8772), cells were disrupted by vortexing using a Vortex-Genie® 2 with a Turbomix™ attachment for 15 minutes at 4°C. Lysates were recovered from the beads by piercing the bottom of each tube with a 21G needle, placing each tube over a clean 1.5

[1] No attempt was made to optimise crosslinking conditions.

ml centrifuge tube and centrifuging both tubes for 30 seconds at 3000 x g. Lysates were combined in a 15 ml centrifuge tube and the total volume increased to approximately 2 ml with FA-lysis SDS buffer. Lysates were sonicated at 4 °C using a Bioruptor™ (Diagenode) on setting "high" (7 cycles 30 seconds on, 60 seconds off) to generate chromatin fragments in the range 200-1000 bp. For ChIP-qPCR experiments lysates were sonicated more extensively to generate smaller chromatin fragments of average size ~200bp in order to improve the resolution of ChIP peaks (140). For this more extensive chromatin fragmentation, lysates were sonicated for 20 minutes (30 seconds on 60 seconds off), centrifuged at 4000 x g for 5 minutes at 4 °C, and the supernatant sonicated for a further 20 minutes (30 seconds on 60 seconds off). Cellular debris was removed by centrifuging the sonicated extracts in 1.5 ml microfuge tubes at 4 °C for 20 min at 14000 rpm.

2.9.2 Preparation of Input DNA

An aliquot of chromatin extract was used to prepare DNA to check the degree of chromatin shearing. 60 µl of chromatin extract was combined with 140 µl TE, pH 8.0 (10 mM Tris-HCl, 1 mM EDTA), 2 µl proteinase K (20 mg/ml New England Biolabs) and 200 µl proteinase K buffer (50 mM Tris-HCl, 12.5 mM EDTA, 300 mM NaCl, 1% SDS), and incubated at 55 °C for 1 hour and then at 65 °C overnight to reverse the crosslinks. After addition of 6.2 µl 5 M NaCl, DNA was purified by phenol/chloroform extraction and ethanol precipitation using 2 µl GlycoBlue™ (Ambion) as a carrier. Contaminating RNA was removed by treatment with 2 µl RNaseA (5 mg/ml Epicentre Biotechnologies), followed by phenol/chloroform extraction and ethanol precipitation. DNA was resuspended in TE to a final volume of 100 µl, 10 µl of which was run out on a 1% agarose gel. The majority of DNA fragments were between 200 bp and 1000 bp.

2.9.3 Chromatin Immunoprecipitations

300 µl chromatin solution was mixed with 900 µl FA-lysis buffer (50 mM HEPES/KOH pH 7.5, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate) containing 150 mM NaCl, 9 µl 100x protease inhibitors (28 µg/ml leupeptin, 0.14 mg/ml pepstatin A, 17 mg/ml PMSF, and 33 mg/ml benzamidine), and 60 µl 20% sarkosyl. After centrifugation at 21000 x g for 30 minutes, supernatants were transferred to fresh tubes. Extracts were incubated with 2 µl goat polyclonal anti-Myc antibody (Abcam ab9132) for 2 hours at 4 °C and then for a further 2 hours with 100 µl protein G dynalbeads (Invitrogen) pre-equilibrated in FA-lysis buffer containing 150 mM NaCl. Beads were concentrated with a magnetic particle concentrator, the supernatant removed and the beads washed with:

- FA lysis buffer (150 mM NaCl) for 10 minutes
- FA lysis buffer (1 M NaCl) for 5 minutes

Beads were transferred to clean tubes and further washed:

- FA lysis buffer (500 mM NaCl) for 10 minutes
- TEL buffer (0.25 M LiCl, 1% NP-40, 1% sodium deoxycholate, 1 mM EDTA, 10 mM Tris-HCl pH 8.1) for 10 minutes
- TE pH 8.0 for 5 minutes twice

Bound complexes were eluted from the beads by incubating with 200 µl of elution buffer (10 mM TRIS-HCl pH 8.1, 1 mM EDTA, 250 mM NaCl, 1% SDS) at 65 °C for 30 minutes. The elution was repeated and the combined eluates digested with 2 µl Proteinase K (10 mg/ml) for 1 hour at 55 °C. After reversing the crosslinks by incubating

overnight at 65 °C, DNA was purified by phenol/chloroform extraction followed by ethanol precipitation with 2 µl GlycoBlue™ (Ambion) as a carrier. DNA pellets were resuspended in 50 µl TE and analysed by real time PCR.

2.10 ChIP-chip analysis

Three biological replicate ChIPs were prepared for each tagged strain together with three control ChIPs using the wild type untagged strain. 40 µl was submitted (80%) of immunoprecipitated DNA from each ChIP sample to Karin Zueckert-Gaudenz in the microarray core facility for amplification, labelling and hybridisation to GeneChip® *S. pombe* Tiling 1.0FR Arrays (Affymetrix). 10 µl from each sample was retained to compare original ChIP DNA with amplified ChIP DNA. This was to ensure that less abundant DNA sequences were not significantly amplified relative to more abundant sequences. Samples were processed according to the manufacturer's protocols, and Madeline Gogol analysed the scanned array data using the MAT algorithm to generate a MAT score together with an associated p-value for each probe on the array (141). The p-value indicates the probability that a result would be obtained by chance. Genes with a ChIP signal with a p-value less than 0.05 within the annotated region were classified as being enriched by the bait protein (for a lists of SpELL and SpEAF enriched genes refer to Appendix C).

2.11 Preparation of RNA

RNA was prepared either by hot phenol extraction (larger quantities) or using Epicentre® MasterPure™ yeast RNA purification kits (142). For hot phenol extractions 200 ml cultures were grown to early log phase (OD600 = 0.2-0.4), harvested on ice and

centrifuged at 3000 x g for 5 minutes at 4 °C. Pellets were washed in DEPC treated ddH_2O and resuspended in 6 ml TES (10 mM Tris pH 7.5, 10 mM EDTA pH 8 and 0.5% SDS) in phenol / chloroform resistant Oak Ridge centrifuge tubes (Nalgene #3114 with sealing cap assemblies #DS3131)[1]. After addition of 6 ml acidic (pH 4.3 - 4.7) phenol:chloroform 5:1 (Sigma P1944), samples were vortexed for approximately 5 seconds in a fume hood and incubated at 65 °C in a hybridisation oven for at least 1 hour, vortexing for 10 seconds every 10 minutes. Samples were incubated on ice for 4 minutes, vortexed for 20 seconds and centrifuged at 4000 x g for 20 minutes. Approximately 5 ml from the upper phase was transferred to a 50 ml phase-lock tube (5 Prime) containing 5 ml acidic phenol:chloroform and 5 ml chloroform, mixed by inversion and centrifuged at 4000 x g for 10 minutes. Samples were then washed by adding 5 ml of chloroform:isoamyl alcohol (24:1) to the same tube, centrifuged as before, and RNA precipitated from the aqueous phase ethanol and sodium acetate pH 5.2 using standard procedures. 100 μg of the RNA was purified further using an RNeasy mini spin column (Qiagen) according to the manufacturer's instructions.

2.12 Analysis of RNA from deletion strains using spotted arrays

For each analysis I prepared purified RNA enriched for polyadenylated RNA (polyA+ RNA) from three biological replicates of both the deletion strain and the wild type parental strain. I prepared the polyA+ RNA from 1 mg of total RNA using the Oligotex Direct mRNA Mini Kit (Qiagen). Brian Fleharty in the microarray facility

[1] Obtaining tubes that can withstand vortexing hot phenol is critical for this protocol.

processed the samples according to the protocol described in (143), hybridising labelled cDNA with *70mer S.pombe* spotted arrays designed by Chris Seidel. The scanned data was analysed by Chris Seidel to generate lists of genes with significant changes in RNA levels between wild type and deletion strains essentially as described (144). Lists of genes with greater than 1.40 fold changes in RNA levels in the deletion strains can be found in Appendix C.

2.13 Preparation of cDNA

Except where otherwise noted, I prepared all cDNA from 1 μg of total RNA in 20μl reactions using iScript reverse transcriptase (Bio-Rad) and an oligo(dT) primer according to the manufacturer's instructions. Reaction products were diluted 1:50 for qPCR analysis.

2.14 Real time Polymerase Chain Reactions

Quantitative real time polymerase chain reactions (qPCR) were prepared with iQ SYBR Green Supermix (Bio-Rad) in 96 well plates sealed with Microseal "B" sealing film (Bio-Rad). Typically 5 μl of diluted template was used in each 25 μl reaction to avoid pipetting errors associated with using smaller template volumes. Reactions were cycled on MyiQ thermocyclers (Bio-Rad) and results were analysed with iQ5 analysis software (Bio-Rad) according to the manufacturer's instructions. Quantitation of gene expression was done either with reference to a standards curve or by the ΔΔCt method with the final analysis done with the iQ5 software. ChIP samples were analysed with reference to a standards curve produced from 10 fold serial dilutions of genomic DNA. Values for "relative quantity" were then exported to excel for final analysis.

2.15 Northern Analysis of RNA

Probe preparation. Single stranded radiolabeled RNA probes were prepared by *in vitro* transcription using the MAXIscript® kit (Ambion) according to the manufacturer's instructions. DNA template for the *in vitro* transcription was generated by PCR amplification of the region 26-547 of the *sme2* gene using an antisense primer containing the T7 promoter sequence (5' TAATACGACTCACTATAGGG 3') at the 5' end. Unincorporated nucleotides were removed with NucAway® spin columns (Ambion).

Northern blotting. 30 μg of total RNA per lane was resuspended in NorthernMax® formaldehyde loading dye (Ambion) and incubated at 65 °C for 15 minutes. Samples were fractionated in gels containing 1% agarose, 2.2 M formaldehyde and 1X MOPS running buffer (40 mM MOPS pH 7.0, 10 mM sodium acetate and 1 mM EDTA) run at 5 V/cm in 1x MOPS running buffer. RNA was transferred onto BrightStar-Plus nylon membranes using 20x SSC buffer (3 M NaCL, 0.3 M Na_3citrate·$2H_2O$ pH 7.0) and immobilized by exposure to 120 mJ/cm^2, 254 nm ultraviolet light in a CL-1000 UV crosslinker (UVP). Blots were prehybridised for 30 minutes at 60 °C in ULTRAhyb buffer (Ambion), hybridised overnight at 60 °C in ULTRAhyb buffer containing single stranded radiolabeled RNA probes (~10^6 cpm/ml), washed twice for 5 minutes at room temperature in 2x SSC/0.1% SDS, and washed twice for 15 minutes at 60°C in 0.1x SSC/0.1%SDS. Signals were detected with a Molecular Dynamics Typhoon PhosphoImager.

2.16 Analysis of 3' RNA ends

In order to analyse the 3' ends of both polyadenylated and non-polyadenylated RNA species in samples of total RNA, total RNA samples were incubated in reactions with or without poly(A) polymerase and the RNA products then analysed by 3' RACE.

Polyadenylation. 2.5 μg total RNA was incubated with or without 5 units poly(A) polymerase at 37 °C for 30 minutes in 20 μl reactions in buffer containing 40 units RNase Inhibitor (NEB), 1 mM ATP, 50 mM TRIS-HCl pH 7.9, 250 mM NaCl, and 10 mM $MgCl_2$.

Reverse Transcription. Reaction volumes were increased to 50 μl with the addition of:

2.5 μl dNTP mix (10 mM)
5 μl 3'-CDS primer A (0.5 μM)
3'-AAGCAGTGGTATCAACGCAGAGTAC(T_{30})VN-5'
1 μl (40 units) RNase Inhibitor
1 μl (200 units) SMART™ MMLV reverse transcriptase (Clontech)
2.5 μl DTT (0.1 M)
10 μl TRIS-HCl pH 8.3 (250 mM), KCl (375 mM), $MgCl_2$ (30 mM)
8 μl nuclease free ddH_2O.

Reactions were incubated at 42 °C for 60 minutes and then at 72 °C for 7 minutes. Following the addition of 5 units RNAseH (NEB) reactions were incubated at 37 °C for 30 minutes.

PCR amplification. cDNA products were amplified by polymerase chain reaction over 25 cycles using a gene specific forward primer and a reverse primer complimentary to 3' CDS primer A and with an annealing temperature of 68 °C. Amplified products were resolved by electrophoresis on 1.5% agarose TBE gels. Major bands were excised and the DNA purified using QiaQuick gel extraction kits (Qiagen). DNA fragments were cloned

into the pCR®2.1TOPO® vector (Invitrogen) according to the manufacturer's instructions and plasmid DNA preps from approximately 10 colonies were sequenced to confirm the identity of the cDNA 3' ends.

CHAPTER 3. IDENTIFICATION AND CHARACTERISATION OF THE SpELL/SpEAF COMPLEX[1]

3.1 Introduction

The human *ELL* gene was first identified as a fusion partner of the *MLL* gene in the (11:19)(q23;p13.1) translocation in some patients with acute myeloid leukemia (1, 145). ELL was later purified from rat liver nuclear extracts based on its ability to activate the rate of transcript elongation by RNA polymerase II (pol II) *in vitro*. Mechanistic studies showed that ELL interacts directly with transcribing pol II *in vitro* and functions by suppressing transient pausing by the enzyme at many sites along the DNA (2, 103). Two additional ELL family members, designated ELL2 and ELL3, as well as a single *Drosophila melanogaster* orthologue, have been identified through searches of sequence databases. These proteins all function similarly to activate the rate of elongation by pol II *in vitro* (92, 93, 106). Yeast two-hybrid screens for ELL-interacting proteins led to the identification of two closely related proteins EAF1 and EAF2 (105, 115). EAF1 colocalises with ELL in the nucleus of mammalian cells in Cajal bodies, nuclear structures that are enriched in factors involved in transcription and mRNA processing (109). Subsequently, we showed that EAF1 and EAF2 bind directly to ELL family members and function as strong positive regulators of ELL transcription activity *in vitro* (3).

It is unclear what role the ELL/EAF complex plays in regulating transcription *in vivo*. The *Drosophila* ELL homolog, dELL, is encoded by the Suppressor of Triplo-lethal

[1] This work has been published (J. Biol Chem **282**, 5761-9, 2007) and was featured as "paper of the week" (4).

[Su(Tpl)] locus and is essential for viability in flies. Mutations in Su(Tpl) may preferentially affect synthesis of some long transcripts and it is possible that some Su(Tpl) mutations suppress lethality resulting from overexpression of the *Tpl* gene by impairing synthesis of *Tpl* mRNA (94). dELL also colocalises with hyperphosphorylated (elongating) pol II on polytene chromosomes, consistent with a role for ELL in controlling transcript elongation *in vivo* (106). In *Xenopus laevis*, the EAF2 protein is specifically expressed in the eye and functions to activate transcription of the eye-specific *Rx* gene during development (123).

Although many components of the pol II transcription machinery are highly conserved across species from mammals to yeast, until now attempts to identify orthologs of the ELL and EAF proteins in fungi have been unsuccessful. It has been suggested that elongation factors like ELL might have evolved after the emergence of multicellular organisms, where genes can be many tens of kilobases long, and fine-tuning of the transcriptional program is expected to be particularly important for differentiation and development (94).

Using a bioinformatics approach, Ron Conaway and Arcady Mushegian had previously identified two genes annotated in the *S.pombe* genome (*ell1* and *eaf1*) that appeared to encode proteins similar to ELL and EAF. These genes are found in *S. pombe* and several other fungi with completely sequenced genomes, but not in *Saccharomyces cerevisiae*.

I began by confirming the sequence of *ell1* and *eaf1* mRNA and demonstrating that *eaf1* has a sequence at the N terminus which was missed in the original genome annotation. Like their counterparts from larger eukaryotes, *S. pombe* ELL and EAF (SpELL and SpEAF) interact with one another to form a stable heterodimer that potently

activates transcription elongation by pol II *in vitro*. In common with yeast strains bearing mutations in several other components of the pol II elongation machinery, *S. pombe* lacking the gene encoding ELL exhibit a 6-azauracil-sensitive phenotype, suggesting that ELL could play an important role in transcriptional regulation even in simple, unicellular organisms (4).

3.2 Identification of ELL and EAF Orthologs in *S. pombe*

The predicted *SPBP23A10.14c* ORF was identified by Ron Conaway and Arcady Mushegian as a possible *S. pombe* orthologue of *ELL*. It codes for a 533 amino acid protein specified by a gene with 3 exons on *S. pombe* chromosome II. Sequencing of a cDNA, which I generated by 5'-RACE, confirmed the predicted *SPB23A10.14c* ORF. Multiple sequence alignments indicate that the sequence similarity between *SPBP23A10.14c* product and ELL orthologues from higher eukaryotes extends throughout the entire length of the protein, with the highest similarity concentrated in three predicted globular regions: an N-terminal all-beta domain, a central region that includes a predicted all-alpha region, and the previously defined C-terminal occludin-like domain (Figure 3-1 A). The gene encoding the *SPBP23A10.14c* ORF has been designated *ell1*, and the *ell1* gene product as SpELL. In an attempt to find an *S. pombe* orthologue of EAF1, Ron Conaway compared the full-length 268 amino acid human *EAF1* ORF to sequence databases; although no highly-scoring fungal homologues were found, using the highly conserved N-terminal 110 amino acids of the human EAF1 protein as the query generated a low-scoring match to the predicted *S. pombe SPCC1223.10c* ORF. The predicted *SPCC1223.10c* ORF in the database is 199 amino acids long and lacks 59 N-terminal amino

acids that are highly conserved in metazoan EAF proteins. Using both 5' and 3'-RACE, I was able to isolate a full length cDNA from *S. pombe* total RNA that included a new first exon and encoded 59 N-terminal amino acids missed in the original genome annotation (Figure 3-1 B).

Figure 3-1 Multiple sequence alignment of ELL family proteins.

This alignment was performed by Arcady Mushegian. Distances, in amino acids, between conserved sequence blocks are indicated by numbers. Conserved regions in ELL proteins: N-terminal mostly-beta domain (top), middle alpha-helical domain, and C-terminal occludin-like domain (bottom).Predicted secondary structure is shown below the alignment. Yellow shading indicates conserved bulky hydrophobic residues (aliphatic I, L,V, M or aromatic F, Y, W), red type indicates conserved residues with small side chains or propensity for main-chain turns (A, G, S, P), red-shaded green type indicates conserved residues with carboxy or carboxamide side chains (D, E, N, Q), blue type indicates conserved positively charged residues (K, R), and magenta type indicates a conserved histidine. In the secondary structure lanes, h indicates alpha-helix, l indicates loop and s indicates beta-sheet. Secondary structures are predicted, except for the the occludin domain structure, which that is taken from PDB structure 1XAW (C-terminal domain of human occludin) (124). Species abbreviations are as follows: *Aedae, Aedes aegypti; Arath, Arabidopsis thaliana; Aspni, Aspergillus nidulans; Bos, Bos taurus; Caebr, Caenorhabditis briggsae; Chagl, Chaetomium globosum; Danre, Danio rerio; Drome, Drosophila melanogaster; Gibze, Giberella zeae; Maggr, Magnaporte grisea; Phano, Phaeospheria nodorum; Schja, Schistosoma japonicum; Schpo, Schizosaccharomyces pombe; Strpu, Strongylocentrotus purpuratus; Tetni, Tetraodon nigroviridis; Trica, Triboleum castaneum; Xenla, Xenopus laevis; Yarli, Yarrowia lipolytica.*

A

ELL

```
6723968_Schpo_ELL11 PVALGGNEEE  4 MLVQLPKEFLEGY  1 SGTITDVSLECSDVGT  11 TSVPETAPHEIYRLVD  82
50549573_Yarli       1 YTLLPNQAGG  3 IHVKLTAEILEQI 12 SGTTPCIKIDNTTFAA   0 HMSQEPSFHSLYQRKG  42
111057331_Phano      9 GIALVGAPDS 14 MQVEMTQEIIDEL  5 SGKAPQIFFGKNPQLK   8 HSSPESHRCELYRSSG  76
46108010_Gibze       9 GLQLEGSAGK  8 FALILSDVVIEDM 13 LGANPTFLYGSQSHTI   0 SPPSASNPYDLYLTRP  95
70997213_Maggr       8 GLELETSVDK  8 FAITLSDDMLEDL 13 LGQSPAFLYGGNTAPI   0 TRIPENFSYDLYVTDP  93
38567161_Neucr       8 GLLLEGPTEA  8 FAVPLSAELVAQM 13 LSDDPAIFIEDQEVPL   2 ELLTEPVDYELFYTNP  86
88185964_Chagl       8 GVLLDGSAKV  8 FAISLSGSVIEDM 13 LGSNPKFLFDDHELRI   0 PKTSDPSGYDLFRTDS 124
15991717_Drome      18 NYGMSQSHRY  7 IIVKLTDSAFRAI 13 PGQRAKIQFVGNTGVI  23 GAGGGGGRKFGFTINN 215
41054381_Danre       8 CYGLSCGRVS  7 YHVKLTDSALRAF  8 LTAKPLIGFTGSQGKI   4 SENPNELRTFTFYLSN  86
72679841_Xenla       8 SYGLSCGRVG  5 FHVKLTDSALKAF  9 LNLKPAIRFQGSQGQI   4 PECPTEVRTFTFYLSN  86
53128622_Chick       8 GYGLSCGRLG  7 FHVKLTESALRAF  9 VTSKPVIQFQGSQGHI   4 PDRPTEVRTFTFYLSN  86
6679635_Mouse        8 SYGLSCGRVS  7 FHVKLTDSALKAF  9 VSLRPSIRFEGSQGHI   4 PDCPEEVRAFSFYLSN  86
5729812_Human        8 SYGLSCGRVS  7 FHVKLTDSALRAF  9 VSLRPSIRFQGSQGHI   4 PDCPAEARTFSFYLSN  86
secondary structure lllllllsss    sssssllhhhhhh    lllsssssssllsss     lllllllsssssssll
```

```
6723968_Schpo    82 SSPTNTPSPNLPVSQPSASPH 10 IDLRTRVIQLLAIAPETEDFLRLRTK  7 ALLPEVAWKN--NMNQWELLNPVYKDVRV
50549573_Yarli   42 PSPQPKSHPSSPVVKTKNRTL 13 SSLPMRVLHLLALQPTSPEVVANRTK  7 ALCKEYGVQV--ASGSYKLADKYFKDLRC
111057331_Phano  76 AQHLAPSKLGSPMLGTASSPM 17 NALRVPVIHLLARQPATEASLAETCR  7 ELLSKIAKRSSPDTDKWQLTDKVFRELNP
46108010_Gibze   95 SSAAPRSIPQSPGQVSLGSAT 11 KGDRAMLVHELAVQDRSIDYLSKKWE  7 PTLRKVADYID-DSRKWALKKRAWKDLDV
70997213_Maggr   93 MSTQARSRPTSPAISAVGSPL 16 KELRAPLIHALAVREMTYDELWEKWG 12 NILSKVAEQVK-NSNKYMMKKNHWKELDV
38567161_Neucr   86 TNGTSRSISVSPALSGLGSPS 12 KQQKFPIIHELAAQNLTFEELLERYD  7 PILNKVADFDS-DEQKWVLKKMYWKELDV
88185964_Chagl  124 AGPTPRSLPPSPALSGIGSPS 12 KQQRFPLMHELAVQDLSREELLERWD  8 TVLDKVASFDK-DMQKWVLRKNCWKDLDV
15991717_Drome  215 ASSVISSRNKMPSGGLTSSNS 24 RNIRERLIHLLALKAFKKPELFARLK 11 NILMDISTM---SHNTYNLRRQMWNDVDE
41054381_Danre   86 SDVAPSRRTSRPVIISSSAQK  4 RPLRERLVHLLALKPYRKPELLVRLT 11 SLLQQVANLNS-KDNTFTLKDCLFKEVQK
72679841_Xenla   86 PDAVPSRKHPTPVNLANAIKK  7 RPFKDRVVHLLALKPYKKPELILRLQ 11 SLLQQVANLNA-KDCTYTLKDCLYKEVQK
53128622_Chick   86 SDAVPSRKRPTPVNLASAIKR  7 RPFKDRVVHLLALKPYKKPELILRLQ 11 NLLQQVANLSA-KDSTFTLKDCVYKEVQK
6679635_Mouse    86 ADAVPSRKRATPINLASAIRK 11 RPFRDRVLHLLALRPYRKAELLLRLQ 11 SLLQQVASVNP-KDGTCTLQDCMYKSLQK
5729812_Human    86 TDAVPSRKRATPINLASAIRK 14 RPFRDRVLHLLALRPYRKAELLLRLQ 11 GLLQQVANMSA-KDGTCTLQDCMYKDVQK
secondary structure hhhhhhhlllllllhhhhlll    lhhhhhhhhhhhhhhlllhhhhhhhhh    hhhhhhhhlllllllllhhhhhhhhhhh
```

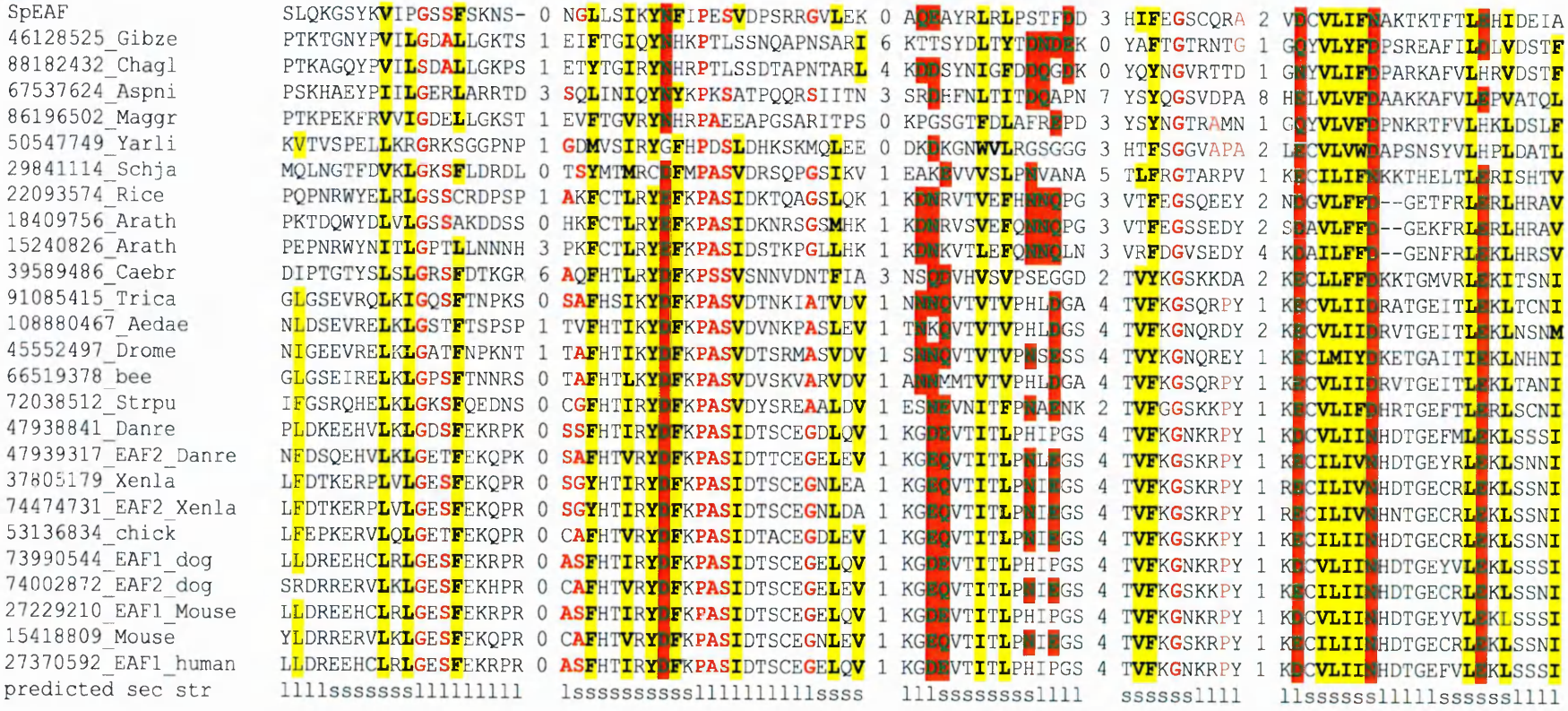
```
6723968_Schpo    212 MRALAKRFRETPRYKNLYLKVSSYYDNNDTNNPNLNKLQDE--LISLHSQLKSWKNTLYDA
50549573_Yarli   228 APHYLSLARQFKESYDEYVKMYKTVAKQPGSDMKK--------LLMMHRQLETWKKELWAY
111057331_Phano  350 FRQTVELSQRFQKYYKKYEELYWQLTESDKSPTEAQRND----LLRMHKKLEEMKREIKAG
46108010_Gibze   327 SDDLVQKAQAFKVCYQKYEALHNEVSALSNPSSRKLDH-----LMEMRNRLKTMKMEIYSE
70997213_Maggr   398 PPDVVAKARRFKEAYSDYERLHYELSGMNNPEESKLNE-----LMNMHRRLEKMKKEIYST
38567161_Neucr   390 DPDILSRAEKFKRAYTRYEALHREVSSAEDPSSEKVES-----LMRMHARLQIEKKEIYSA
88185964_Chagl   402 SQETVELAARFRQFYSRYEQLHQDIASHDNPDPDKLSD-----LLDMHKRLSRMKTEIYAA
15991717_Drome   443 DNGYGDYDHIKRQIVCEYERINNDRTIGEDKER----------FDYLHAKLAHIKQLVMDY
41054381_Danre   312 QQGTDKYKTIHNQILEEYRKIKKTNPNYSQEKNR---------CEYLHNKLAHIKKLIAEY
72679841_Xenla   274 CQGSEDYKTIHDQILQEYRKIKKSNPNYSEEKNR---------CEYLHNKLAHIKKLIAQY
53128622_Chick   289 LQGSEEYKTIHDQILQEYRKIKKTNPNYSQEKNR---------CEYLHNKLAHIKKLIAEY
6679635_Mouse    273 SQGSDEYETTRGQILQEYRKIKKTNTNYSCEKRR---------CEYLHRKLAHIKRLIAEY
5729812_Human    280 SQGSEEYETTRGQILQEYRKIKKTNTNYSQEKHR---------CEYLHSKLAHIKRLIAEY
```

Occludin C-terminus

```
540494_chick     445 LDSITEDSPQYQDVAEEYNQLKDLKRSPDYQSKKQE-------SKVLRNKLFHIKRMVSAY
5833879_Xenla    437 GQSSRKDSEEYRTVADKYNRLKEIKSSADYRNKKKR-------CKGLKTKLNHIKQMVSNY
20987418_1XAW     81 LDDYREESEEYMAAADEYNRLKQVKGSADYKSKKNH-------CKQLKSKLSHIKKMVGDY
secondary structure  lllllllllhhhhhhhhhhhhhhhhhhhhlhhhhhh       hhhhhhhhhhhhhhhhhh
```

B

EAF


SpEAF
46128525_Gibze
88182432_Chagl
67537624_Aspni
86196502_Maggr
50547749_Yarli
29841114_Schja
22093574_Rice
18409756_Arath
15240826_Arath
39589486_Caebr
91085415_Trica
108880467_Aedae
45552497_Drome
66519378_bee
72038512_Strpu
47938841_Danre
47939317_EAF2_Danre
37805179_Xenla
74474731_EAF2_Xenla
53136834_chick
73990544_EAF1_dog
74002872_EAF2_dog
27229210_EAF1_Mouse
15418809_Mouse
27370592_EAF1_human
predicted sec str

The sequence of this full-length *eaf1* cDNA has been submitted to the GenBank with accession number EF192607 (Appendix D).

When this extended protein sequence was compared to the database, the first match, was a protein from *Schistozoma japonicum* (gi 29841114) annotated as similar to Homo sapiens EAF1. When this *S. japonicum* sequence was used as a query, the known and predicted EAF homologues from various eukaryotes were detected, including SpEAF. EAF homologues were found in all metazoa, in higher plants, and in most fungi and protists, but not in *S. cerevisiae*. The gene encoding the extended *SPCC1223.10c* ORF has been designated as *eaf1*, and the *eaf1* gene product as SpEAF.

3.3 Sensitivity of an *ell1Δ* strain to 6-azauracil

Yeast strains carrying mutations in genes encoding a number of proteins implicated in regulation of transcription elongation grow slowly in the presence of the nucleotide-depleting drug 6-azauracil. To determine whether *ell1* or *eaf1* mutants are sensitive to 6-azauracil, I generated strains lacking *ell1* or *eaf1*. The *ell1Δ* and *eaf1Δ* strains were both viable. Although the *eaf1Δ* strain appeared to grow as well as wild type *S. pombe* on plates containing 6-azauracil, the *ell1Δ* strain exhibited a 6-azauracil-sensitive phenotype (Figure 3-2).

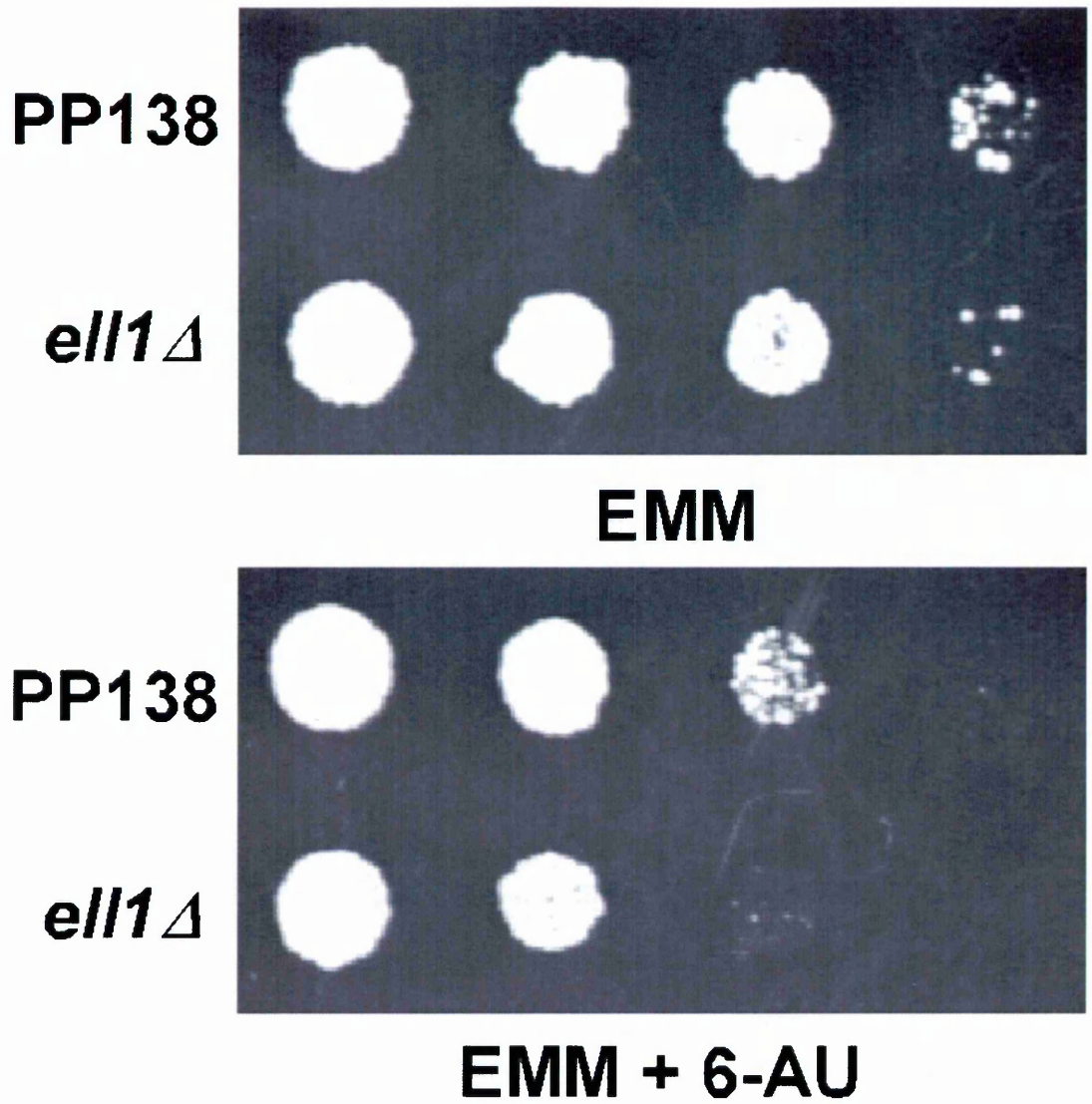


Figure 3-2 6-azauracil sensitivity of *S. pombe* deleted for the gene encoding SpELL.

The parental strain PP138 and the *ell1Δ* strain, both containing the pUR19 plasmid, were grown to mid log phase in rich media, washed in 1.2 M sorbitol, and resuspended in 1.2 M sorbitol at a density of 1 x 10^8 cells/ml. 5 μl of 3-fold serial dilutions of cells were spotted onto EMM plates supplemented with adenine, histidine and leucine (225 μg/ml) with or without 6-azauracil (300 μg/ml).

3.4 Direct interaction of SpELL with SpEAF

To determine whether SpELL and SpEAF interact with one another in cells, I expressed N-terminal FLAG-tagged SpELL or SpEAF individually in *S. pombe* under

control of the inducible *nmt1* promoter. I purified SpELL- and SpEAF-associated proteins from these strains using anti-FLAG agarose immunoaffinity chromatography and these were identified by mass spectrometry using multi-dimensional protein identification technology (MudPIT)[1]. As summarized in Table 3.1, epitope-tagged SpELL and SpEAF specifically copurified with endogenous SpEAF and SpELL, respectively, indicating that the two proteins interact in cells. The NSAF gives an approximate indication of the relative abundance of a protein in each sample and the NSAF values here indicate that the amounts of epitope-tagged subunit and the endogenous partner subunit are of the same order of magnitude.

	FLAG-SpELL			FLAG-SpEAF			FLAG-control		
	peptides	spectra	NSAF	peptides	spectra	NSAF	peptides	spectra	NSAF
SpELL	20	70	0.0094	36	179	0.0123	0	0	-
SpEAF	13	117	0.0335	32	292	0.0425	0	0	-

Table 3.1 SpELL coimmunoprecipitates with SpEAF expressed in *S. pombe*.

Cell lysates from *S. pombe* expressing FLAG-SpELL, FLAG-SpEAF, or FLAG-6His were purified by anti-FLAG immunochromatography and subjected to MudPIT mass spectrometry. The table shows the number of peptides and spectra for SpELL and SpEAF detected in samples purified from cells expressing either FLAG-SpELL or FLAG-SpEAF. *NSAF*, normalized spectral abundance factor. The number of spectra for a given protein detected in a MudPIT run has been shown to be a function of the protein's size and abundance (133); hence, comparison of the spectral counts provides a rough estimation of the relative abundance of different proteins across samples. NSAF is the spectral count for a given protein, normalized to the protein's length and the total number of spectra detected in the MudPIT run. NSAF is calculated using the equation

$$(NSAF)_k = \frac{(SpC / Length)_k}{\sum_{i=1}^{N} (SpC / Length)_i}$$

where *SpC* is the number of spectra detected, *k* denotes a specific protein, and *N* is all proteins detected.

[1] The MudPIT runs and analysis were performed by Skylar Martin-Brown, Laurence Florens and Michael Washburn in the Stowers Institute proteomics facility.

I next subcloned the SpELL and SpEAF ORFs into baculovirus vectors and expressed them in Sf21 insect cells in several epitope-tagged forms. In particular, I coexpressed FLAG-

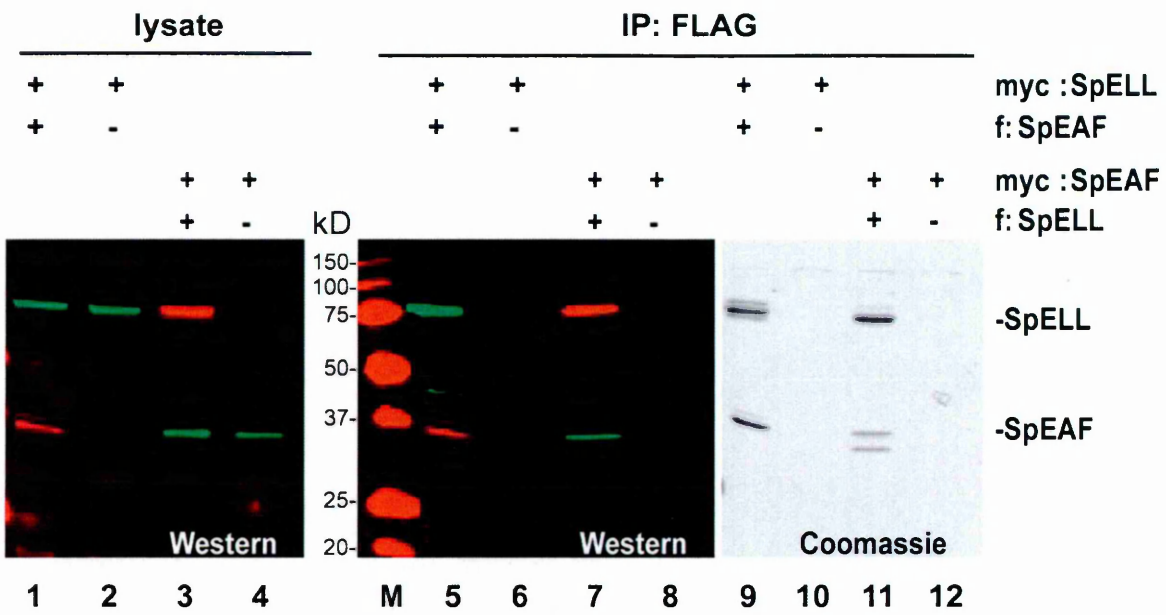


Figure 3-3 Interaction of SpELL and SpEAF.

Sf21 cells coinfected with baculoviruses encoding FLAG-SpELL, Myc-SpELL, FLAG-SpEAF or Myc-SpEAF in the combinations indicated in the figure were prepared and immunoprecipitations were carried out with the antibodies indicated in the figure. Bound proteins were eluted with 150 ng/μl FLAG peptide, analyzed by SDS-PAGE, and detected by either Western blotting or staining with Coomassie Blue R-250.

SpEAF with Myc-SpELL or FLAG-SpELL with Myc-SpEAF (

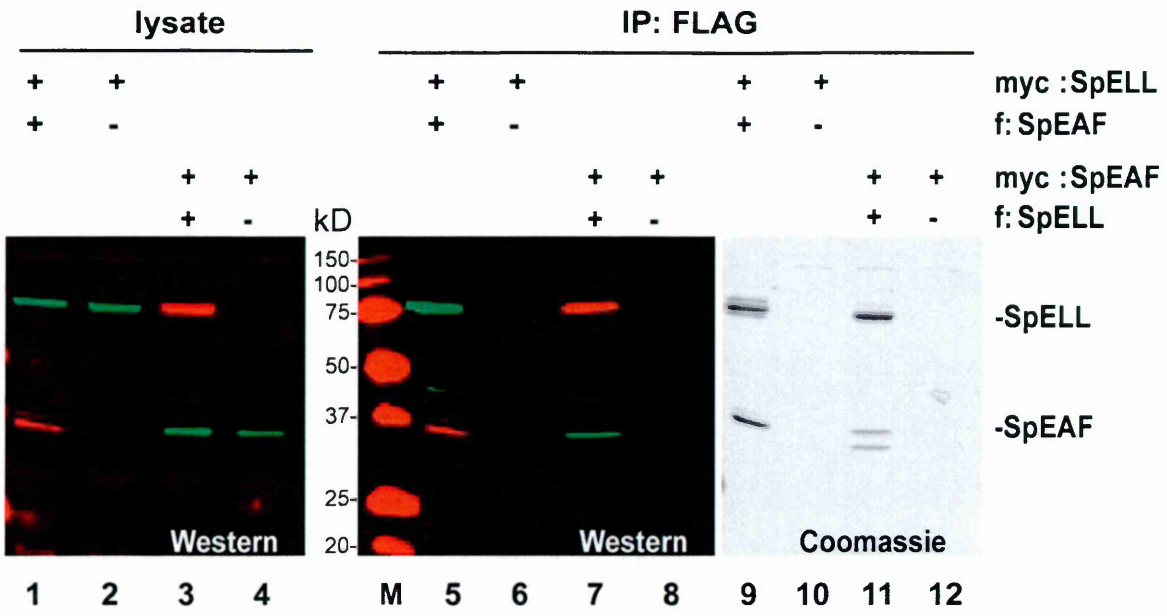


Figure 3-3 lanes 1-4) and purified the resulting protein complexes using anti-FLAG agarose chromatography (lanes 5-12). FLAG-SpEAF copurified with Myc-SpELL (lanes 5 and 9) and FLAG-SpELL copurified with Myc-SpEAF (lanes 7 and 11). The interaction of SpELL and SpEAF appeared to be direct, since only very small amounts of any additional proteins were present in my purified preparations of SpELL/SpEAF (lanes 9 and 11).

3.5 The SpELL/SpEAF Complex Activates Transcription Elongation by *S. pombe* RNA Polymerase II

To test the effects of SpELL and SpEAF on transcription elongation by *S. pombe* pol II, I assayed transcript elongation on a linearised plasmid with a single-stranded 3' oligo(dC)-tail on its template strand. Although purified pol II is unable to initiate from a specific location on double-stranded DNA without assistance from the general

transcription factors TFIID, TFIIB, TFIIE, TFIIF, and TFIIH, it binds to the single-stranded oligo(dC)-tail and initiates transcription at the junction between the single- and double-stranded regions of the template (146). On the template used in my experiments, the first nontemplate strand (dT) residue is 136 nucleotides downstream of the oligo(dC)-tail; thus, transcripts of 135 nucleotides will accumulate when transcription is initiated with just ATP, CTP, and GTP.

In the experiment shown in Figure 3-4, I initiated transcription by the addition of purified *S. pombe* pol II[1] to reaction mixtures containing the oligo(dC)-tailed template, ATP, GTP, and [α-32P]CTP. After a 30-min incubation to allow accumulation of pol II ternary elongation complexes containing radioactively labelled, 135 nucleotide long transcripts, I purified the elongation complexes by gel filtration to remove unincorporated ribonucleoside triphosphates (Figure 3-4 lane 1). I then chased nascent transcripts into longer products by the addition of ATP, GTP, CTP and UTP, in the presence or absence of SpELL, SpEAF, or both (Figure 3-4, lanes 2-13). Like the mammalian ELL- EAF complex (3), the SpELL/SpEAF complex stimulated the rate of transcription elongation by its cognate *S. pombe* pol II, as detected by an increase in the rate that radioactively labelled 135 nucleotide long transcript was chased into longer products when reactions included the SpELL/SpEAF complex. Addition of either SpEAF or SpELL alone had a negligible effect on the rate of transcript elongation. Notably, SpEAF(60-251), which is the truncated version of SpEAF encoded by the predicted *SPCC1223.10c* ORF and which lacks

[1] The purified pol II was a kind gift from Henrik Spahr

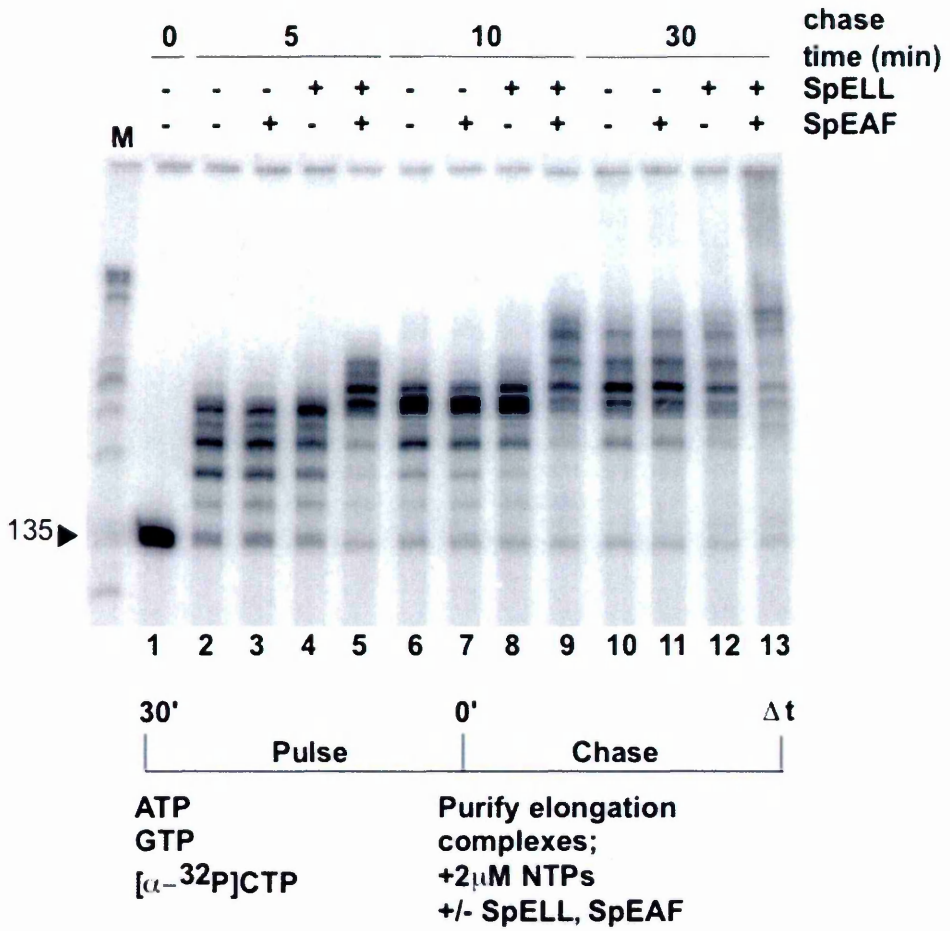
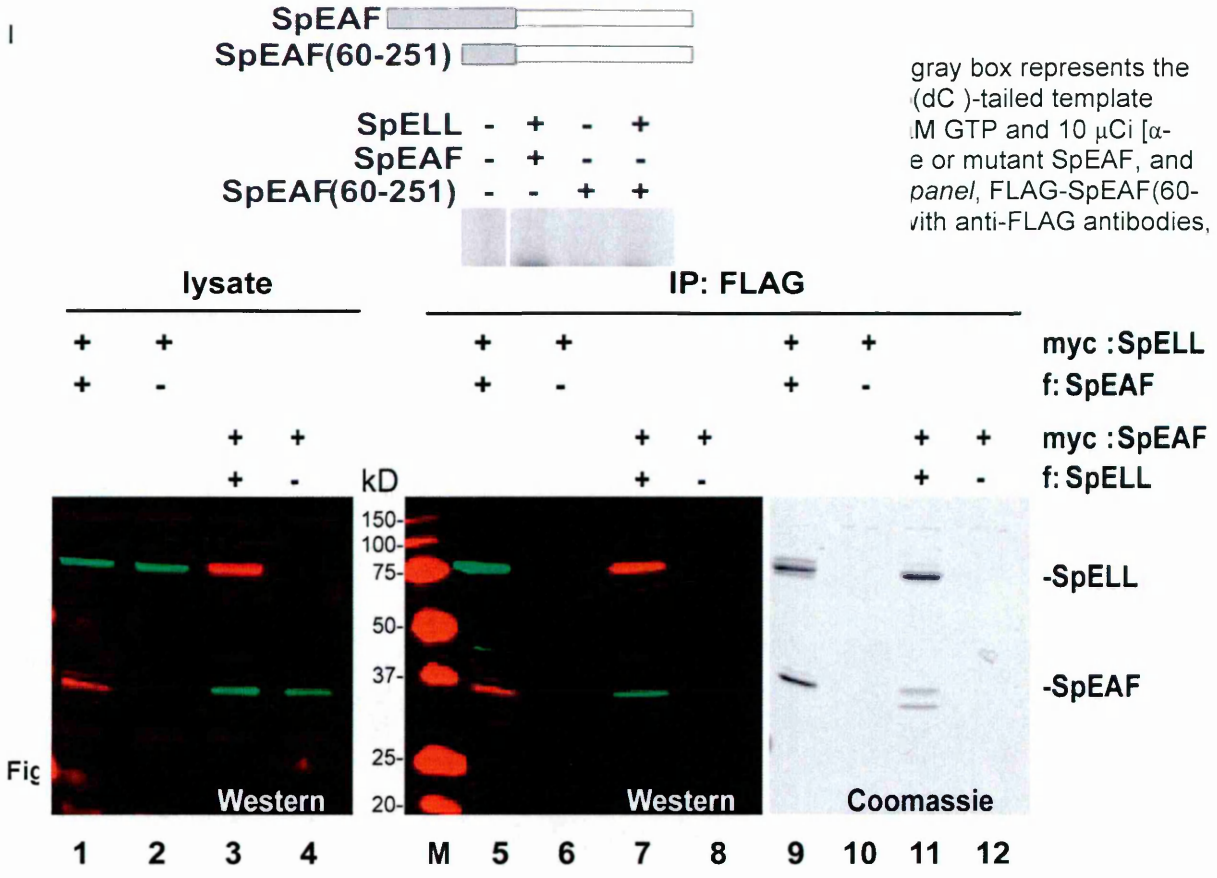


Figure 3-4 SpELL and SpEAF stimulate elongation by *S. pombe* pol II.

Oligo(dC)-tailed template transcription reactions were initiated by the addition of 50 μM ATP, 50 μM GTP, 2 μM CTP and 10 μCi [α-32P]CTP (400 mCi/mmol). After a 30-min incubation, paused pol II elongation complexes were purified, and labeled transcripts were chased into longer products by the addition of 2 μM ATP, 2 μM GTP, 2 μM CTP and 2 μM UTP, with or without 5 pmol SpELL or SpEAF. Reactions were stopped after 5, 10 or 30 minutes of chase. The lane marked M shows ΦX174 Hae III restriction fragments used as size markers.

the most highly evolutionarily conserved N-terminal region of the protein, was inactive even in the presence of SpELL (Figure 3-5 and data not shown).



gray box represents the
(dC)-tailed template
M GTP and 10 μCi [α-
e or mutant SpEAF, and
panel, FLAG-SpEAF(60-
vith anti-FLAG antibodies,

Fig

To obtain additional evidence that the SpELL/SpEAF complex possesses transcription activity, I further fractionated the purified SpELL/SpEAF complex by ion exchange HPLC and monitored copurification of SpELL/SpEAF with pol II stimulatory activity. I coinfected Sf21 insect cells with baculoviruses expressing FLAG-SpELL and Myc-SpEAF, and the resulting cell lysates were subjected to anti-FLAG agarose chromatography followed by chromatography on a TSK DEAE-NPR HPLC column. FLAG-SpELL and Myc-SpEAF proteins co-eluted from the TSK DEAE-NPR column at ~160 mM KCl, with the majority of both proteins being detected in fractions 7, 8 and 9 (Figure 3-6 A, lower panel).

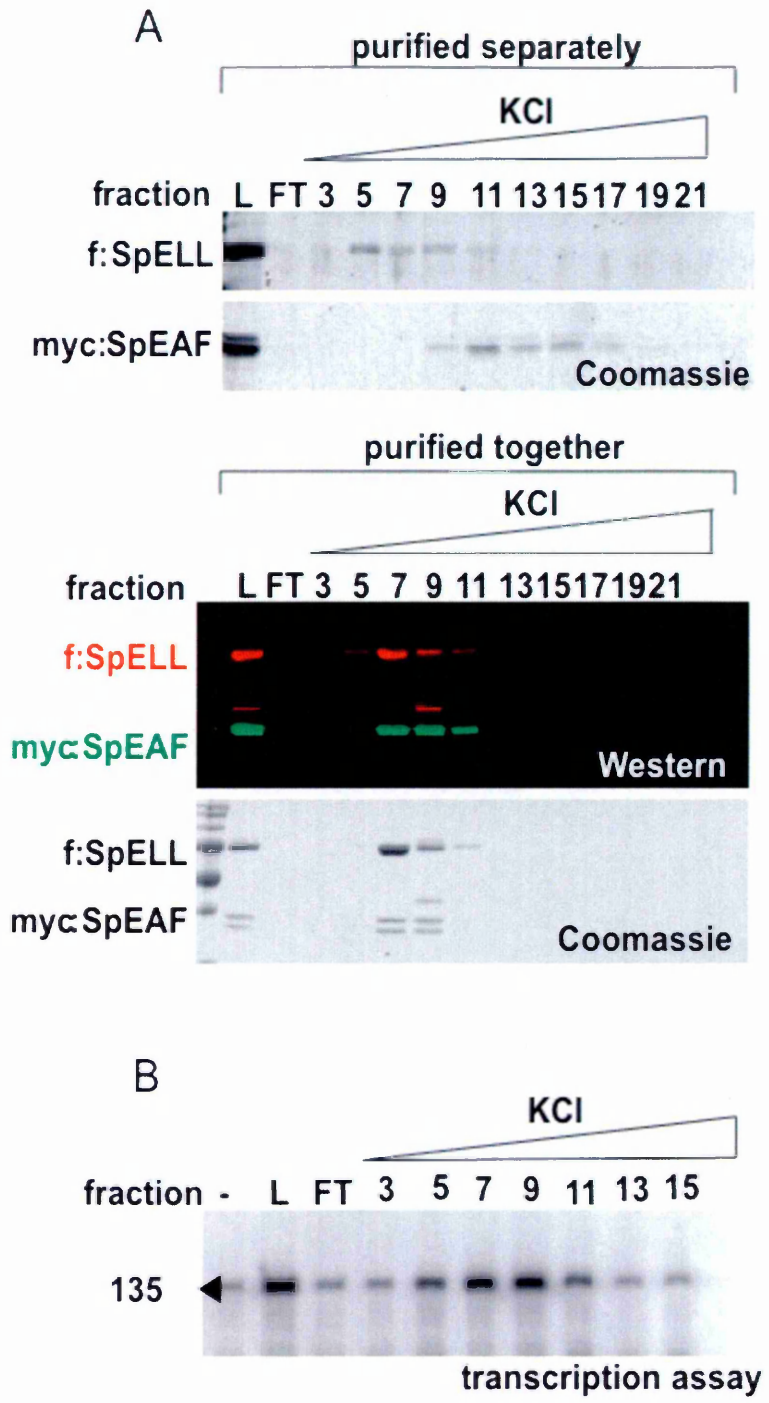


Figure 3-6 Cochromatography of stimulatory activity with the SpELL/SpEAF complex.

FLAG-SpELL and Myc-SpEAF were expressed separately (upper panel) or together (lower panel) in insect cells and purified by anti-FLAG agarose chromatography. The eluate was adjusted to a conductivity equivalent to 0.05 M KCl and applied to a 0.6 ml TSK DEAE-NPR HPLC column pre-equilibrated in buffer A containing 0.1 M KCl. The column was eluted with a 6 ml linear gradient from 0.1 to 0.5 M KCl in buffer A, and 0.2 ml fractions were collected. Aliquots of the indicated fractions were analyzed by SDS-polyacrylamide gel electrophoresis and either stained with Coomassie Blue R-250 or analyzed by Western blotting with anti-FLAG (red) or anti-myc (green) antibodies.

SpELL and SpEAF purified according to this procedure were stably associated with one another, since SpELL and SpEAF eluted from the TSK DEAE-NPR at distinct positions when they were expressed and purified individually (Figure 3-6 A, upper panel). To assay column fractions for their ability to stimulate the rate of transcript elongation, I used a modified version of the tailed template assay, in which I measured accumulation of the 135-nt, U-less transcript initiated from the oligo(dC)-tail on pAd-GR220 in the presence of only ATP, GTP and a limiting amount of CTP. During the short incubation time and with the limiting concentration of nucleotides used in these assays, only a fraction of pol II molecules are able to complete synthesis of 135 nucleotide long transcripts in the absence of elongation factors unless the rate of transcript elongation is stimulated by elongation factors. As shown in Figure 3-7, synthesis of 135 nucleotide long transcripts depends on the presence of pol II and is inhibited by 5 μg/ml α-amanitin, a concentration that is sufficient to inhibit purified S. pombe pol II (data not shown) but that does not inhibit S. pombe pol I or pol III (147, 148).

Thus, this assay provides a reliable and more easily quantifiable alternative to the pulse-chase protocol, since it measures synthesis of a single transcript, rather than a collection of products with variable lengths. As shown in Figure 3-6 C, the SpELL/SpEAF complex co-eluted with pol II stimulatory activity from the TSK DEAE-NPR column, arguing that the SpELL/SpEAF complex possesses intrinsic transcription stimulation activity.

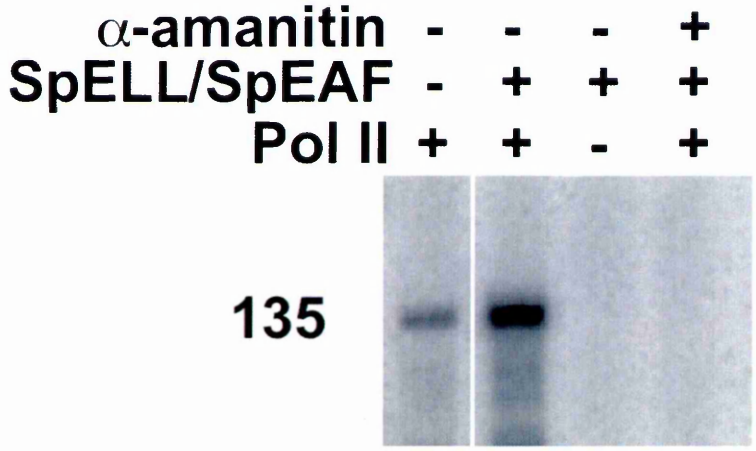


Figure 3-7 Synthesis of 135nt transcripts by pol II.

Oligo(dC)-tailed template transcription reactions were performed in the presence of 50 μM ATP, 50 μM GTP and 10 μCi [α-^{32}P]CTP (400 mCi/mmol) with 3 pmoles SpELL, 3 pmoles SpEAF, 5 μg α-amanitin and 20 ng pol II as indicated . Reactions were stopped after 5 minutes.

3.6 The SpELL/SpEAF Complex Stimulates Pyrophosphorolysis.

During RNA synthesis, RNA polymerases catalyze nucleophillic attack by the 3'-hydroxyl group of the nascent transcript on the α-phosphate of the incoming ribonucleoside triphosphate, resulting in nucleotide addition and pyrophosphate release:

$$NTP + RNA_n \rightleftharpoons RNA_{n+1} + PP_i$$

In the presence of high concentrations of inorganic pyrophosphate, RNA polymerases will catalyze the reverse reaction, pyrophosphorolysis, during which the nascent transcript is shortened with concurrent release of ribonucleoside triphosphates.

These reactions are thought to be carried out by the same active site in pol II. If, as has been proposed for other pol II elongation factors including mammalian TFIIF, Elongin, and ELL (59, 113), SpELL/SpEAF stimulates transcription by stabilizing the active site of pol II in an active configuration, it would be expected to stimulate not only elongation but also pyrophosphorolysis. To determine whether the SpELL/SpEAF complex could stimulate pyrophosphorolysis, I purified pol II ternary transcription complexes free of ribonucleoside triphosphates and incubated them in inorganic pyrophosphate in the presence or absence of the purified SpELL/SpEAF complex.

As shown in Figure 3-8, transcripts grew progressively shorter during incubation with pyrophosphate, and the rate of transcript shortening was greater in reactions that contained the SpELL/SpEAF complex. Transcript shortening was due to pyrophosphorolysis and not to the presence of contaminating ribonuclease, since it was strictly dependent on the presence of pyrophosphate (Figure 3-8 A and B). As observed for the SpELL/SpEAF elongation stimulatory activity, neither SpELL nor SpEAF alone is sufficient to stimulate pyrophosphorolysis (Figure 3-8 B).

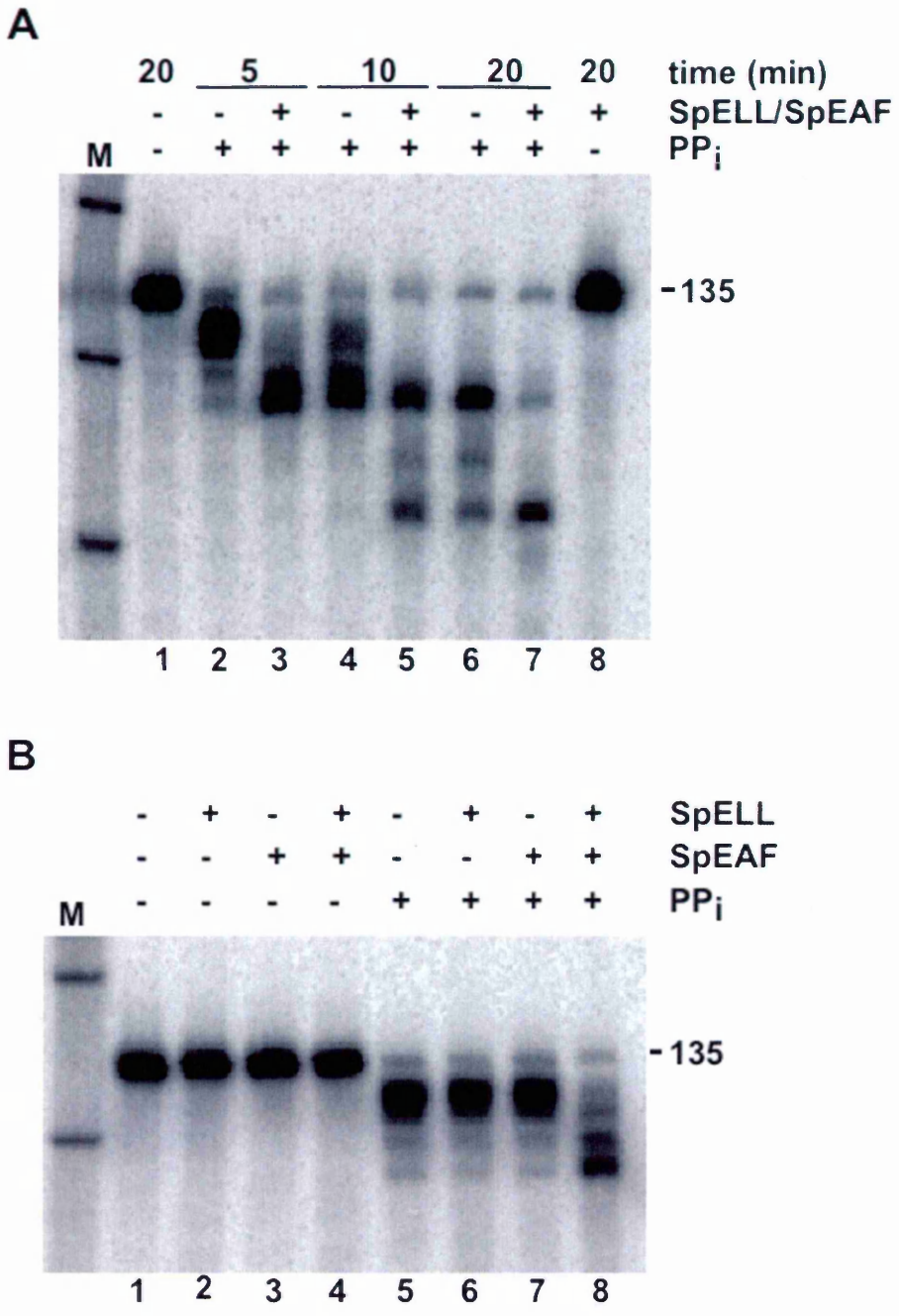


Figure 3-8 The SpELL/SpEAF complex stimulates pyrophosphorolysis by paused pol II elongation complexes.

Paused transcription elongation complexes containing ~135-nucleotide long transcripts were prepared as described in *experimental procedures*. (A) Purified ternary complexes were incubated with 0.3 mM PPi (lanes 2-7) for the times indicated in the presence or absence of an aliquot from the peak fraction of the TSK DEAE-NPR purified SpELL/SpEAF complex. (B) Ternary complexes were incubated for 5 minutes with ~5 pmoles SpELL, SpEAF or both in the presence or absence of 0.3 mM PP_i.

3.7 The SpELL/SpEAF Complex Does Not Activate Elongation By Mammalian RNA Polymerase II.

Previous studies have shown that metazoan ELL can bind directly to pol II in vitro (103, 106). Similarly, SpELL can bind directly to S. pombe pol II, as can SpELL/SpEAF;

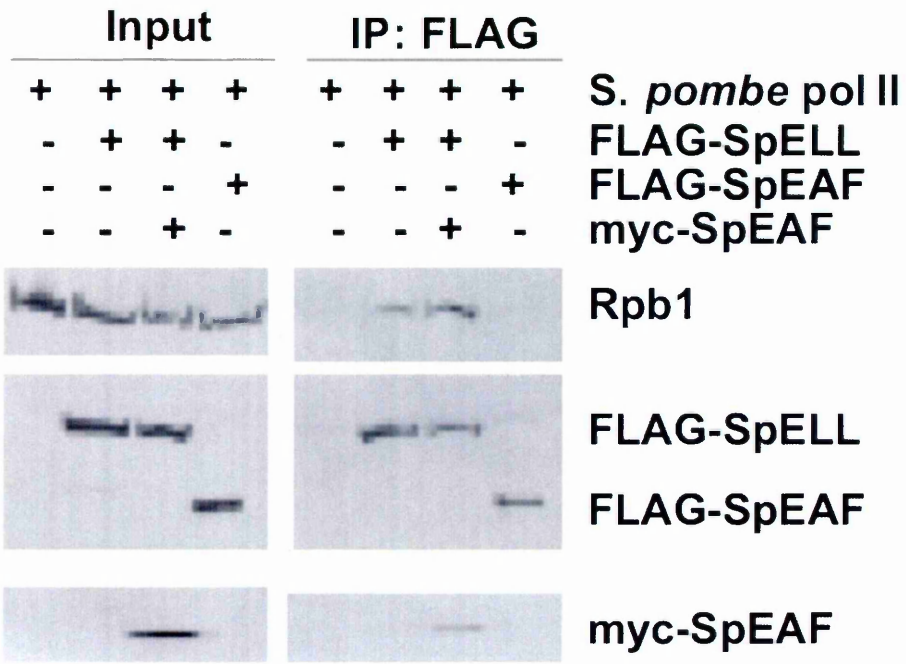


Figure 3-9 Physical interaction of the SpELL/SpEAF complex with pol II.

~0.2 µg *S. pombe* pol II was mixed with ~10 pmoles of the factors indicated, in buffer containing 20 mM Hepes-NaOH, pH7.9, 20mM Tris·HCl, pH7.9, 50 mM KCl, 1 mM DTT, 0.5mg/ml BSA/2% (wt/vol) polyvinyl alcohol (average molecular weight 30,000-70,000 Da), 3% (wt/vol) glycerol, 5 mM $MgCl_2$. Immunoprecipitations were carried out with anti-FLAG agarose, and bound proteins were eluted with 150 µg/ml FLAG peptide, analyzed by SDS-polyacrylamide gel electrophoresis, and detected by Western blotting.

however, free SpEAF binds only poorly to pol II (Figure 3-9). In addition, mutations in human ELL that interfere with pol II binding also interfere with its ability to stimulate elongation (103). Taken together, these observations suggest that ELL increases the elongation activity of transcribing pol II through direct interactions with the enzyme. If this model is correct, one might expect ELL/EAF to stimulate pol II in a species-specific

fashion. To determine whether the activity of SpELL and SpEAF is specific for *S. pombe* pol II or whether it can also stimulate mammalian pol II, I assayed the ability of purified pol II from rat liver to generate 135-nt transcripts initiated from the oligo (dC)-tail on pAd-GR220, in the presence of recombinant SpELL, SpEAF or both. As shown in Figure 3-10 A, transcription by mammalian pol II was unaffected by SpELL/SpEAF, although the same amount of SpELL/SpEAF stimulated elongation by *S. pombe* pol II. Similarly, transcription by *S. pombe* pol II was unaffected by recombinant human ELL/EAF1, although the same amount of human ELL/EAF1 stimulated elongation by rat pol II (Figure 3-10 B). Therefore, stimulation of elongation by ELL/EAF depends on specific interactions between ELL/EAF and pol II.

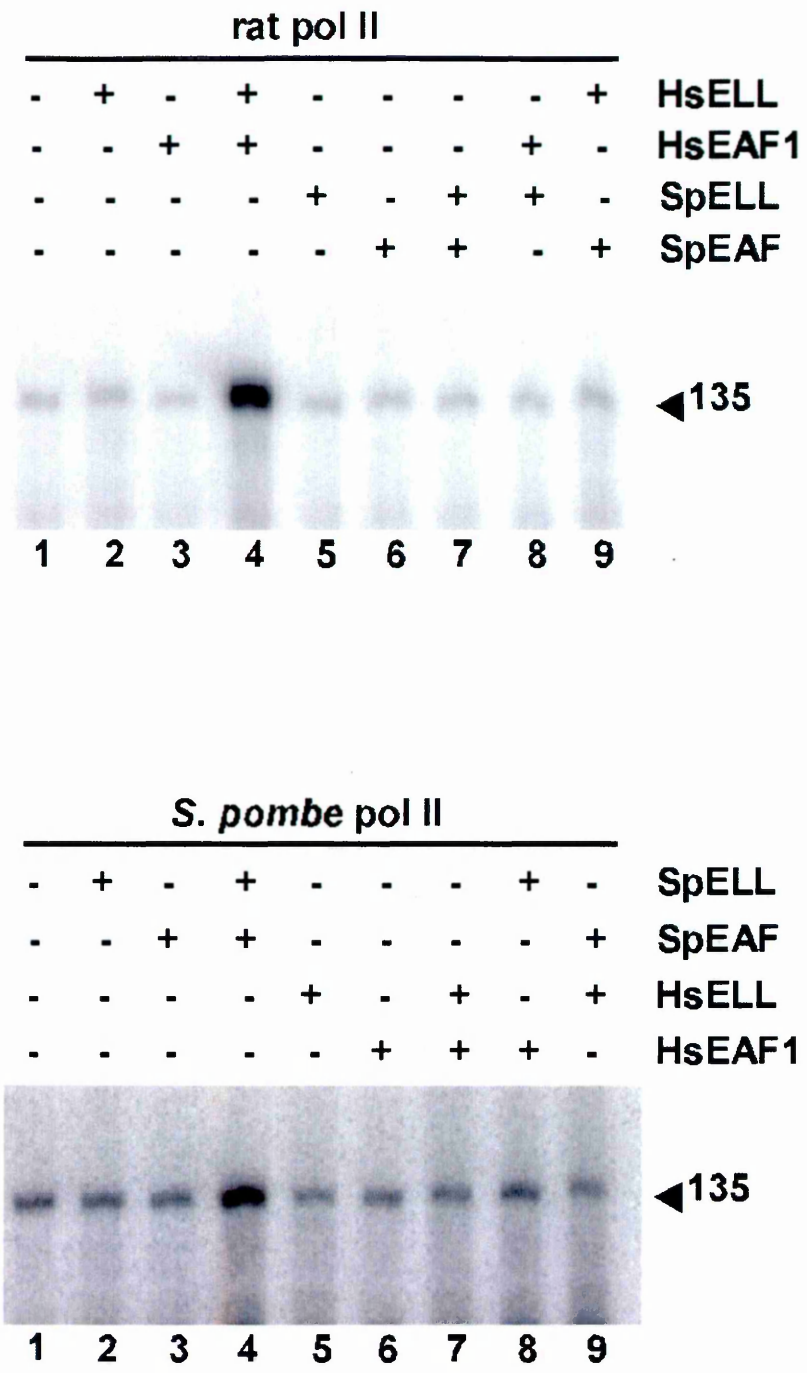


Figure 3-10 Functional interaction of the SpELL/SpEAF complex with pol II.

(A) Insect cells were infected with recombinant baculoviruses encoding FLAG-HsELL, FLAG-HsEAF1, FLAG-SpELL or FLAG-SpEAF. Proteins were then purified by anti-FLAG agarose chromatography, and 3 pmoles of each factor was assayed for its ability to stimulate transcription by pol II purified from rat liver nuclear extracts as indicated. Transcription reactions were initiated from the T-less cassette of oligo(dC)-tailed template pAd-GR220 by the addition of 50 μM ATP, 50 μM GTP, 2 μM CTP and 10 μCi [α-^{32}P]CTP (400 mCi/mmol). Reactions were stopped after 5 minutes. (B) Functional interaction of the SpELL-SpEAF complex with S. pombe pol II. 3 pmoles of each factor was assayed for its ability to stimulate transcription by pol II purified from S. pombe. Reactions were initiated by the addition of 50 μM ATP, 50 μM GTP, and 10 μCi [α-^{32}P]CTP (400 mCi/mmol) and stopped after 5 minutes.

3.8 Discussion

The SpELL/SpEAF complex is a new transcription elongation factor that shares structural and functional properties with the ELL/EAF complex found in higher eukaryotes. As with the mammalian ELL/EAF, the *S.pombe* SpELL/SpEAF complex can stimulate the rate of transcription elongation of its associated RNA polymerase (Figure 3-4) and can positively regulate the reverse reaction, the removal of nucleotides through pyrophosphorolysis (Figure 3-8). Previously it had been thought that transcription elongation factors of this type were limited to higher eukaryotes and may have been used particularly for transcribing genes many kilobases long. Eissenberg and coworkers suggested "ELL family proteins appear to have co-evolved with the dramatic growth in transcription unit lengths that accompanied the evolution of metazoan eukaryotes." (94). The discovery of the SpELL/SpEAF complex in *S.pombe*, a single celled eukaryote with an average gene size of 1430bp[1] (150) shows that this is not the case. It is curious that ELL and EAF homologues exist in higher, multicellular eukaryotes and in some fungi including *S. pombe* but not in *S. cerevisiae* and related yeasts. *S. cerevisiae* is distinct in that it lacks the apparatus needed for many splicing processes and for RNAi. This raises the possibility that in vivo SpELL/SpEAF function is related to RNA processing or to gene silencing.

Deletion of the gene encoding SpELL renders the cells sensitive to growth on media containing the drug 6-azauracil (6AU) (Figure 3-2). Although this phenotype is seen with

[1] Compare with the average human gene length of 27000 bp (149).

mutations of genes involved with other cellular processes (151), 6AU sensitivity is often associated with mutations in genes encoding proteins involved in transcription elongation. These include pol II subunits Rpb1 (152), Rpb2 (153), Rpb9 (154), the transcription elongation factors SII (155), Spt4 and Spt5 (similar to human DSIF), and Spt6 (similar to a subunit of human FACT) (156). In cells 6AU is converted to 6-azaUMP, which inhibits GTP biosynthesis and 6AU might affect elongation rate by reducing the cellular pools of GTP (157).

These results provide new insight into the biochemical mechanism by which the ELL/EAF complex stimulates the rate of transcript elongation. Previous coimmunoprecipitation experiments have suggested that ELL interacts directly with pol II (103, 106). I have shown that the *S. pombe* and human ELL/EAF complexes specifically activate *S. pombe* and mammalian pol II, providing functional evidence that ELL stimulates elongation through a direct interaction with pol II (Figure 3-10). In addition, while confirming that ELL coimmunoprecipitates pol II without the need for additional factors, I have shown that SpEAF binds pol II only very weakly or not at all (Figure 3-9). This suggests that EAF might not stimulate elongation by direct interaction with pol II but instead might be modulating ELL activity. EAF might induce a conformational change in ELL. This could allow ELL to stimulate pol II elongation activity more effectively, and/or could change the ELL/pol II binding affinity allowing ELL a greater probability of interacting with and stimulating pol II. The effect of EAF on ELL/pol II binding affinity could be investigated with a more quantitative assay, perhaps by using fluorescence correlation spectroscopy (FCS). It has not previously been clear which regions of EAF are required for binding to ELL and for elongation activity. I have demonstrated that without the highly conserved N terminus SpEAF can neither bind to SpELL nor can upregulate SpELL stimulation of pol II elongation

activity (

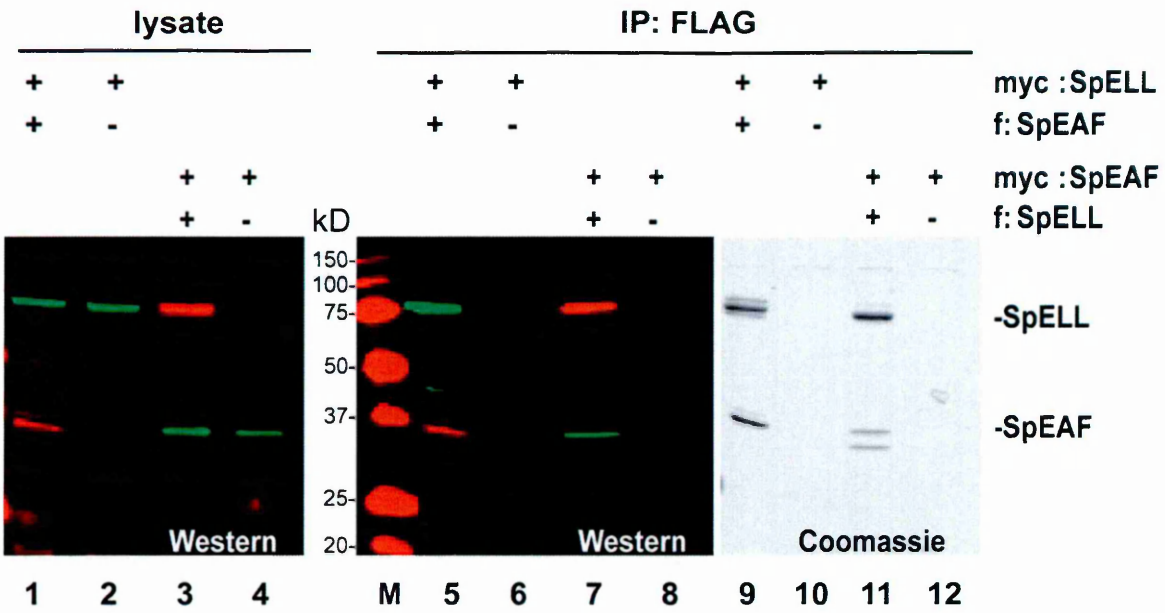


Figure 3-3). Further investigation could reveal whether other regions of SpEAF, for example the acidic C terminus, are dispensable[1].

In vitro the SpELL/SpEAF complex interacts directly with RNA pol II via the SpELL subunit to stimulate the rate of transcription elongation by reducing pol II pausing. The next chapter will begin to investigate the role of the SpELL/SpEAF complex in cells.

[1] Interaction studies with truncation mutants of human EAF1 suggest that the conserved N terminal region is sufficient for binding ELL (CASB unpublished data).

CHAPTER 4. IDENTIFICATION OF GENES REGULATED BY THE SpELL/SpEAF COMPLEX IN *S.POMBE*

4.1 Introduction

I have demonstrated that SpELL and SpEAF function as a complex to stimulate the rate of pol II catalysed transcript elongation *in vitro*. Their role in cells could be to function as a general activator of elongation or the SpELL/SpEAF complex may have evolved to fulfil some alternative function during transcript elongation that necessitates binding to pol II (and perhaps stabilising it) in its elongating state. Previous studies have supported some role for ELL/EAF in transcription in cells. The complex is localised in the nucleus in Cajal bodies, perhaps together with other machinery involved with transcription and RNA processing (109). Studies in drosophila have shown that dELL and dEAF colocalise with pol II engaged in transcription on chromatin and may be relocalised to genes when they need to be transcribed in a timely fashion (106, 158). Furthermore, evidence from a very recent study suggests that human ELL/EAF, functioning as part of a large complex that also includes P-TEFb and several other *MLL* translocation partners, regulates heat shock and several *hox* genes in cells (102).

Many questions remain about what subset of a genome ELL and EAF might be targeting and why. Although there has been no systematic study of ELL/EAF's contribution to gene regulation, a previous study of ELL function in *Drosophila* provided some evidence ELL/EAF may contribute more to regulation of longer, intron containing genes such as *Notch*, *cut* and *Scr* (size range 23-65 kb) rather than shorter genes such as *white* and *rudimentary* (5.9-13 kb) (94). The existence of the complex in *S. pombe*, which has a mean gene length of 1.4 kb and no genes longer than 15 kb, suggests that the ELL/EAF complex

might have a role in the transcription of much shorter genes. Defining which genes in a genome are regulated by ELL/EAF will answer questions including whether the complex regulates all or just some genes, whether regulation depends on some property of the gene (for example the presence of introns or the gene length), or whether specific pathways or genes fulfilling a specific function are affected. In order to define SpELL/SpEAF regulated genes for such an analysis, I have used spotted microarrays to investigate global changes in gene expression after deletion of either SpELL or SpEAF, and I sought to define genes that colocalise with the complex by hybridising chromatin immunoprecipitations to tiling arrays.

In addition to defining which genes are regulated, the results address questions about how the complex functions at genes, for example to what extent SpELL and SpEAF colocalise at a gene and function in concert, and whether they appear to operate in a distinct region such as at the promoter or the 3' end of a gene. Importantly, identification of a set of genes that are thought to be targeted by the SpELL/SpEAF complex provides candidate genes for further analysis.

4.2 Misregulation in *ell1Δ* or *eaf1Δ* cells of a common set of genes, including a number of genes involved in cell separation.

In order to determine which genes or pathways might be regulated either directly or indirectly by the SpELL/SpEAF complex, I compared RNA from either *ell1Δ* or *eaf1Δ* cells with RNA from wild type cells using microarrays spotted with 70mer DNA probes (described in section 2.12). Genes with a greater than 1.40 fold difference between mutant and wild type signals and an adjusted p-value less than 0.025 were further analysed

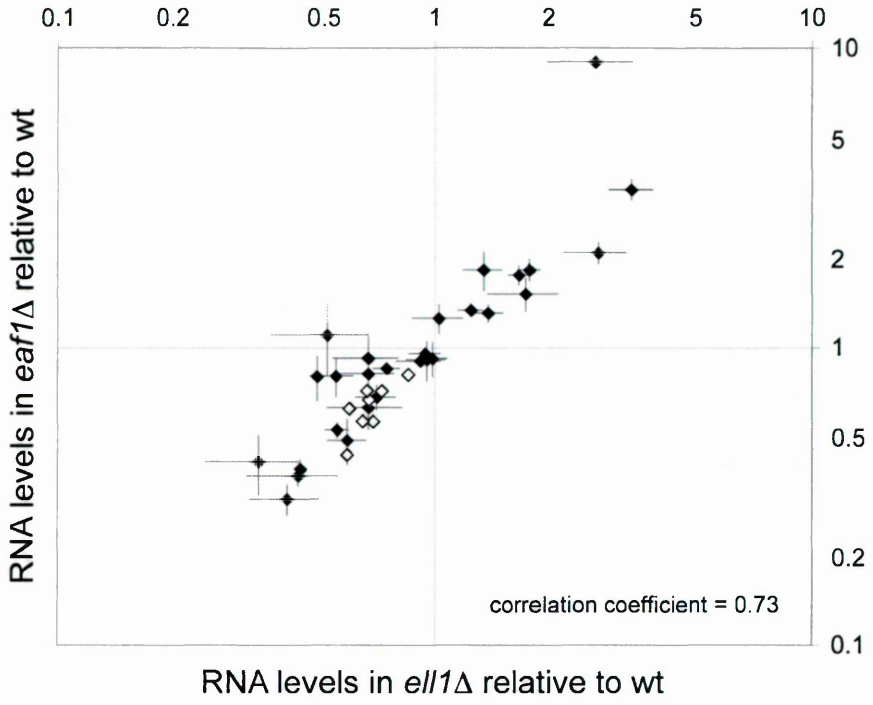


Figure 4-1 Correlation between changes in mRNA levels of 36 genes in *ell1Δ* and *eaf1Δ* cells relative to wt cells.

Genes identified by spotted arrays whose expression changed in either the *ell1Δ* or *eaf1Δ* strains relative to wild type cells were further analysed by RT / qPCR. Reverse transcription was primed with an oligo(dT) primer. Levels of cDNA for each gene analysed were calculated relative to the level of actin cDNA for each experiment. Data points shown as open dots represent genes in the *ace2* pathway. Error bars represent the standard deviation of three independent experiments.

by RT / qPCR and are listed in Appendix C. Of the 36 genes identified by the expression array analysis, 27 (75%) had a greater than 1.40 fold difference between mutant (either *ell1Δ* or *eaf1Δ)* and wild type RNA levels as following analysis by qPCR. When I compared the changes in RNA levels in *ell1Δ* cells relative to wild type cells with changes in RNA levels in *eaf1Δ* cells relative to wild type cells, I found a strong positive correlation (correlation coefficient 0.73 Figure 4-1).

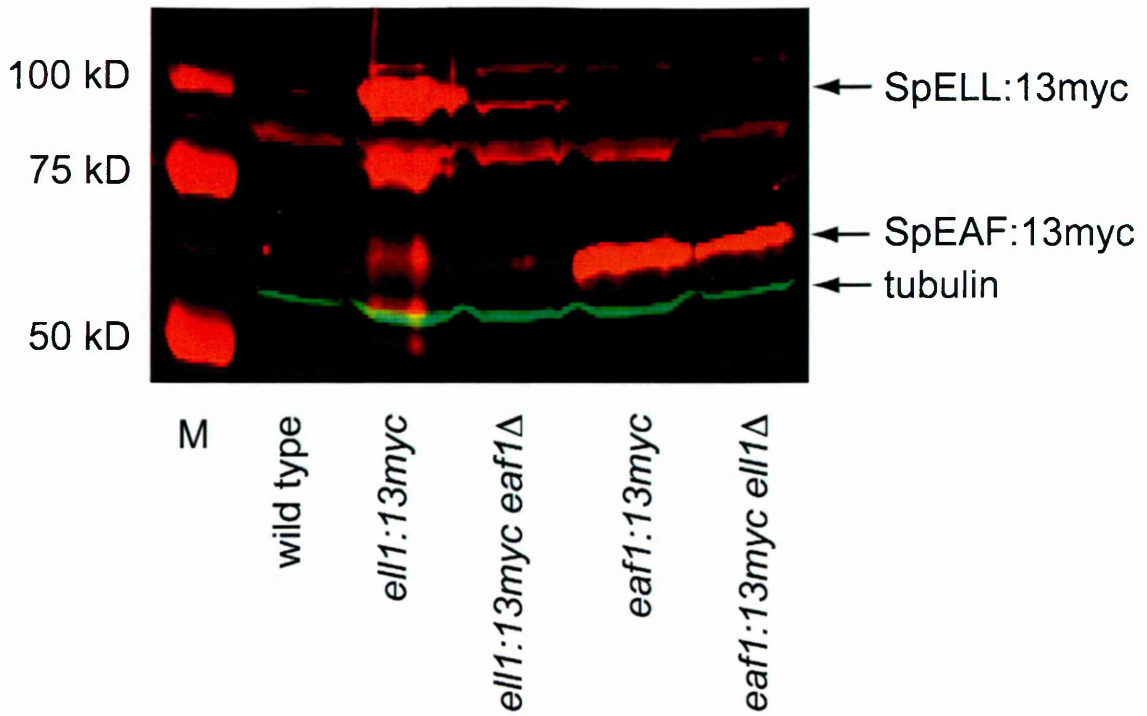


Figure 4-2 Deletion of the *eaf1* gene results in a decrease in SpELL protein.

Whole cell extracts from approximately equal numbers of cells of the strains indicated were prepared as described in Chapter 2. and analysed by SDS-PAGE followed by Western blotting. Membranes were probed with anti-tubulin (mouse monoclonal, Sigma T9026), anti-myc (goat polyclonal, AbCam Ab9132) primary antibodies and IRDye®800 anti-mouse (green) and Alexa Fluor® 680 anti-goat (red) secondary antibodies prior to scanning with a Li-COR® Odyssey® infra red imaging system.

4.3 Deletion of *eaf1* results in a change in SpELL protein levels

The correlation between global effects on RNA levels caused by the deletion of the *ell1* gene or the *eaf1* gene could be due to the requirement for both the SpELL and SpEAF proteins for the correct expression of the genes affected by deletion of either *ell1* or *eaf1*. Alternatively this correlation could be the result of a direct requirement for only one subunit of the SpELL/SpEAF complex for proper gene regulation, but that deletion of the other subunit has an indirect effect on global patterns of gene expression by causing a change in levels of its partner subunit. In order to test whether deletion of either *ell1* affects SpEAF protein levels or deletion of *eaf1* affects SpELL protein levels, I analysed

whole cell extracts from various strains by SDS PAGE followed by Western blotting as detailed in Figure 4-2. I found a significant decrease in SpELL:13myc in the *ell1:myc eaf1Δ* strain compared with the *ell1:myc* strain. There was a more modest effect on levels of SpEAF:13myc in the absence of *ell1*. While these results do not exclude the possibility that SpELL and SpEAF affect global patterns of expression of a common set of genes independently, they suggest that the common pattern of changes in global gene expression may, at least in part, be due to reduced levels of one subunit of the SpELL/SpEAF complex in both the *ell1Δ* and *eaf1Δ* strains. Having shown that deletion of one subunit decreases levels of the partner subunit, I wondered whether this co-regulation was at the transcriptional level. To test whether, for example, deletion of *eaf1* leads to a decrease in *ell1* mRNA levels, I performed oligo(dT) primed reverse transcription on total RNA isolated from the strains indicated (Figure 4-3) followed by qPCR analysis of both *ell1* and *eaf1* cDNA levels. I did not detect a decrease in *ell1* mRNA in *eaf1Δ* cells (Figure 4-3 A). Similarly, deletion of *ell1* does result in decreased levels of *eaf1* mRNA (Figure 4-3 B). Taken together, these results suggest that the downregulation of SpELL protein in the *eaf1Δ* strain (and to a lesser extent of SpEAF protein in the *ell1Δ* strain) is a result of events post transcription.

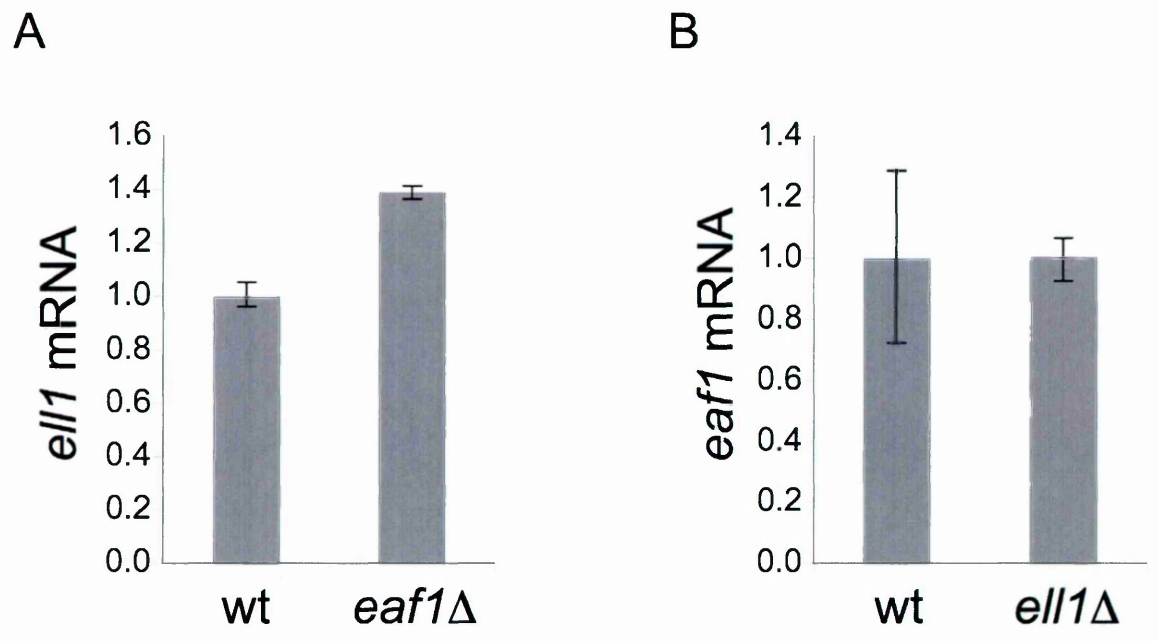


Figure 4-3 Levels of *ell1* mRNA are not reduced in *eaf1Δ* cells.

Amounts of *ell1* or *eaf1* mRNA were normalised to *act1* (actin). Error bars represent the range of three independent experiments.

4.4 Reintroducing *ell1* at the *leu1* locus partially rescues altered patterns of gene expression

To confirm that the global changes in RNA levels we had observed were due to disruption of the SpELL/SpEAF complex and not, for example, caused simply by the insertion of the antibiotic marker, I inserted the *ell1* gene at the *leu1* locus into cells in which the endogenous *ell1* gene had been deleted. This insertion included regions upstream and downstream of the *ell1* coding region, which may be necessary for proper transcriptional regulation of *ell1*. I then looked at the RNA levels of a number of genes whose expression was altered in the *ell1Δ* cells to see whether insertion of *ell1* at the *leu1* locus could rescue the altered gene expression. The level of *ell1* mRNA in the *ell1* rescue

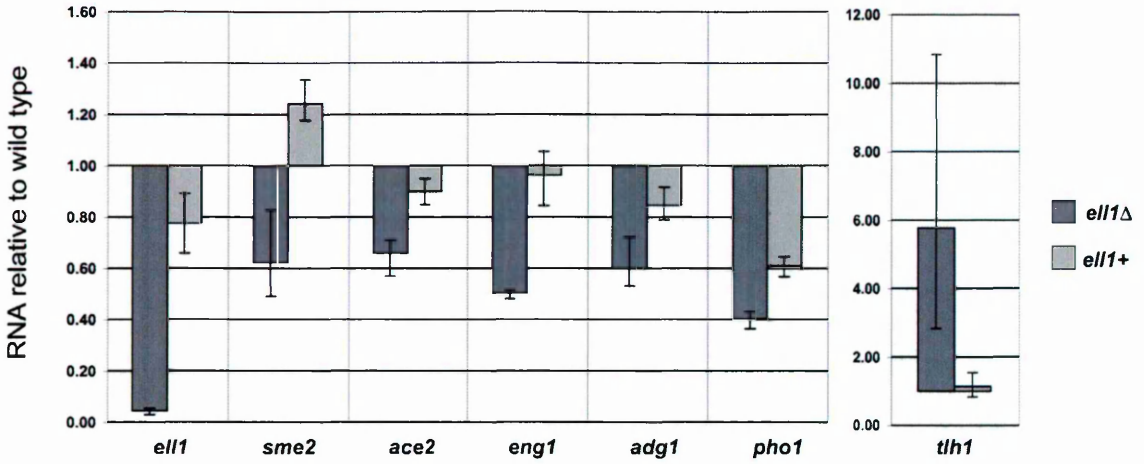


Figure 4-4 Rescue of altered gene expression in the *ell1Δ* strain

The *ell1* gene was cloned into the pDUAL vector (129) and inserted at the *leu1* locus in *ell1Δ* cells to generate the *ell1+* strain. Total RNA from three independent replicates of wild type (PP138), *ell1Δ*, and *ell1+* cells was prepared by hot phenol extraction as described on page 44. 1 μg of total RNA was reverse transcribed with an oligo(dT) primer and the cDNA was analysed by qPCR. RNA levels were calculated relative to the levels of actin (*act1*) using the ΔΔCt function in the iQ5 analysis software (Bio-RAD). Error bars represent the range of the three biological replicates.

cells (indicated as *ell1+*) was between 70% and 90% of the level of ell1 mRNA in wild type cells (Figure 4-4). The lower levels of *ell1* mRNA could be due to a failure to reintroduce all of the control regions required for proper *ell1* transcription in the *ell1+* cells. Alternatively there may be a difference in the chromatin environment between *leu1* and *ell1* loci, which could result in altered transcription of genes in these regions.

I next tested a selection of genes that were either downregulated (*sme2, ace2, ace2* regulated genes *eng1 and adg1*, and *pho1*) or upregulated (*tlh1*) in the *ell1Δ* cells. The changes in RNA levels seen in *ell1Δ* cells for these genes were reversed in *ell1+* cells (Figure 4-4), although in some cases, most notably for *pho1*, the rescue was partial. These results suggest that the altered patterns of gene expression we had observed were primarily due to disruption of the SpELL/SpEAF complex in our deletion strains and not,

for example, to the presence of the antibiotic cassette. In the future, it will be interesting to investigate the possibility *pho1* is particularly sensitive to SpELL/SpEAF levels in cells.

In addition to these rescue experiments, I have tested an additional *S. pombe ell1Δ* strain and an *S. japonicus ell1Δ* strain.[1] In these additional strains, I also observe the same changes in RNA levels, confirming that these effects are not specific to the PP138 parental strain (data not shown). The same effects are also observed when *ell1Δ* cells are grown in minimal rather than rich media.

4.5 Chromosomal localisation of SpELL and SpEAF suggest a direct role in regulating of some genes downregulated in the *ell1Δ* and *eaf1Δ* strains.

I hypothesise that genes that colocalise with the SpELL and SpEAF proteins are more likely to be directly regulated by the SpELL/SpEAF complex. In order to determine the chromosomal location of SpELL and SpEAF proteins, I first constructed the epitope tagged strains SpELL:myc and SpEAF:myc (as described in section 2.3). As previously described, *ell1Δ* cells have a 6-azauracil sensitive growth phenotype so I confirmed that the presence of the epitope tag in the ell1:myc cells did not have an effect on cell growth on EMM media supplemented with the drug 6-azauracil (Figure 4-5). In addition to the *ell1:myc* and *eaf1:myc* strains, I generated a tagged pol II strain, *rpb3:myc*, and a tagged TBP strain, *tbp1:myc*, to use as controls in the ChIP-chip experiments. C terminal myc epitope

[1] Additional strains are *ell1Δ* (972h-) and Sj*ell1Δ* (CBS354/OY1) see appendix A. The *Schizosaccharomyces japonicus ell1* gene was identified by sequence comparison with *S. pombe ell1* and deleted using the methods described for *S. pombe* gene deletion. The *S. japonicus eng1* gene was identified by comparison with *S. pombe* sequences; RNA levels of *S. japonicus eng1* were reduced in *S. japonicus ell1Δ* cells.

tagged *S.cerevisiae* strains for TBP-myc and Rpb3-myc have been described previously (47, 159).

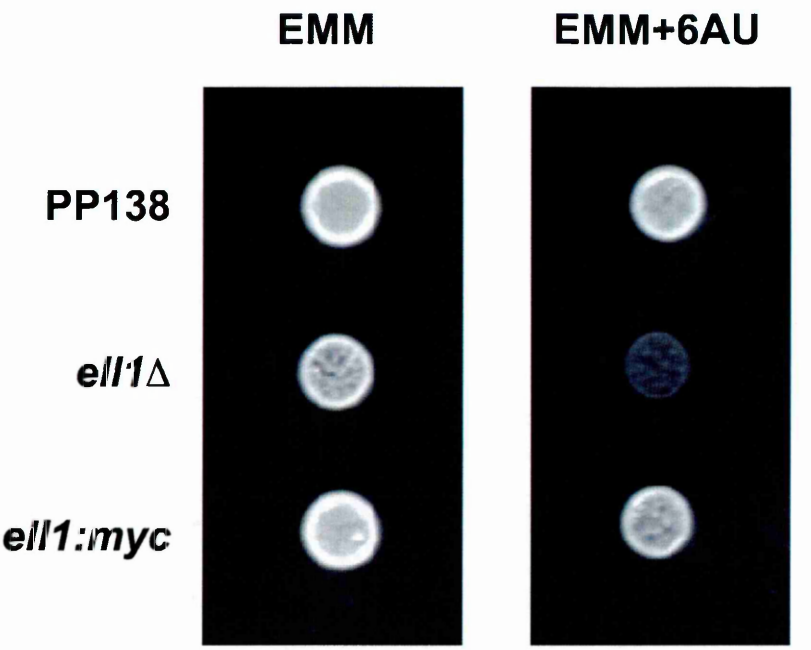


Figure 4-5 *ell1:myc* epitope tagged cells are not sensitive to growth in the presence of 6-azauracil.

The strains indicated, each containing the pUR19 plasmid, were grown to mid log phase in rich media, washed in 1.2 M sorbitol, and resuspended in 1.2 M sorbitol at a density of 1 x 10^8 cells/ml. 5 μl of cells were spotted onto EMM plates supplemented with adenine, histidine and leucine (225 μg/ml) with or without 6-azauracil (300 μg/ml).

4.5.1 ChIP-chip analysis

I performed chromatin immunoprecipitations using chromatin from strains ell1:myc, eaf1:myc, rpb1:myc and tbp1:myc (as detailed in sections 2.9). Immunoprecipitated DNA was then analysed using Afymetrix GeneChip® *S. pombe* Tiling 1.0FR Arrays (section 2.10). Previous studies have suggested that some regions of chromatin, particularly transcribed regions, are coimmunoprecipitated more efficiently than others and that different approaches to analysing the same data can produce different results ((160) Figure2). For these ChIP-chip analyses I have controlled for over

representing regions that might be more likely to immunoprecipitated nonspecifically. For each experiment, DNA enriched with a myc antibody from cells expressing a myc-tagged protein was compared with DNA enriched nonspecifically from wild type (untagged) cells with the same antibody. Data analysis was performed by Madelaine Gogol using the MAT algorithm (as described (141)). The MAT algorithm generates a score based on the intensity value of each 25mer probe on the array. The intensity value is modified to take into account probe sequence characteristics and copy-number variation (as some probes map to multiple locations in the genome). The MAT score at any position is calculated from the modified intensity values of all probes within a 500 nucleotide window. Lists of genes enriched above background with an associated p-value[1] below 0.05 are designated "enriched genes" and are presented in Appendix C.

4.5.2 Do SpEAF:myc ChIPs enrich genes that are also enriched by SpELL:myc ChIPs?

There are 824 SpELL enriched genes, 490 SpEAF enriched genes with 383 genes enriched in both the SpELL:myc ChIP and SpEAF:myc ChIP samples (Figure 4-9). The probability that at least 383 genes would be enriched in these two gene sets if the genes were selected at random can be calculated using hypergeometric distribution calculator http://stattrek.com/Tables/Hypergeometric.aspx[2]):

[1] The p-value indicates the probability that a result would be obtained by chance. A p value of 0.05 is equivalent to a window with a MAT score above 95% of the MAT scores in the null distribution (estimated from a sample of 600 windows).

[2] Using a population size = 5542, a sample size = 824 (all SpELL enriched genes), successes in the population = 490 (all SpEAF enriched genes) and successes in sample = 383 (genes that are both SpELL and SpEAF enriched genes).

$$P(\text{SpELL ChIP} \cap \text{SpEAF ChIP} \geq 383) = 3.24 \times 10^{-252}$$

That is the probability of selecting 824 genes at random from a population of 5542 genes, then selecting 490 genes at random from the same population of 5542 genes, and getting an overlap of at least 383 genes is $\sim 10^{-252} \approx 0$. This suggests that the overlap between the set of SpELL and SpEAF enriched genes would be unlikely to have happened by chance. In addition, visual inspection of the ChIP profiles of the ell1-myc and eaf1-myc strains using a genome browser suggested that many genes classified only as a "SpELL-myc enriched gene" had a ChIP signal with the eaf1-myc strain just above the 0.05 p value cutoff and vice versa (Figure 4-6). Thus, although there are a number of genes that have been classified as being occupied by only SpELL or SpEAF, there might be a number of these that are in fact occupied by both SpELL and SpEAF. Reduced ChIP signals for one of the two subunits could perhaps be caused by partial masking of the myc epitope tag at some locations.

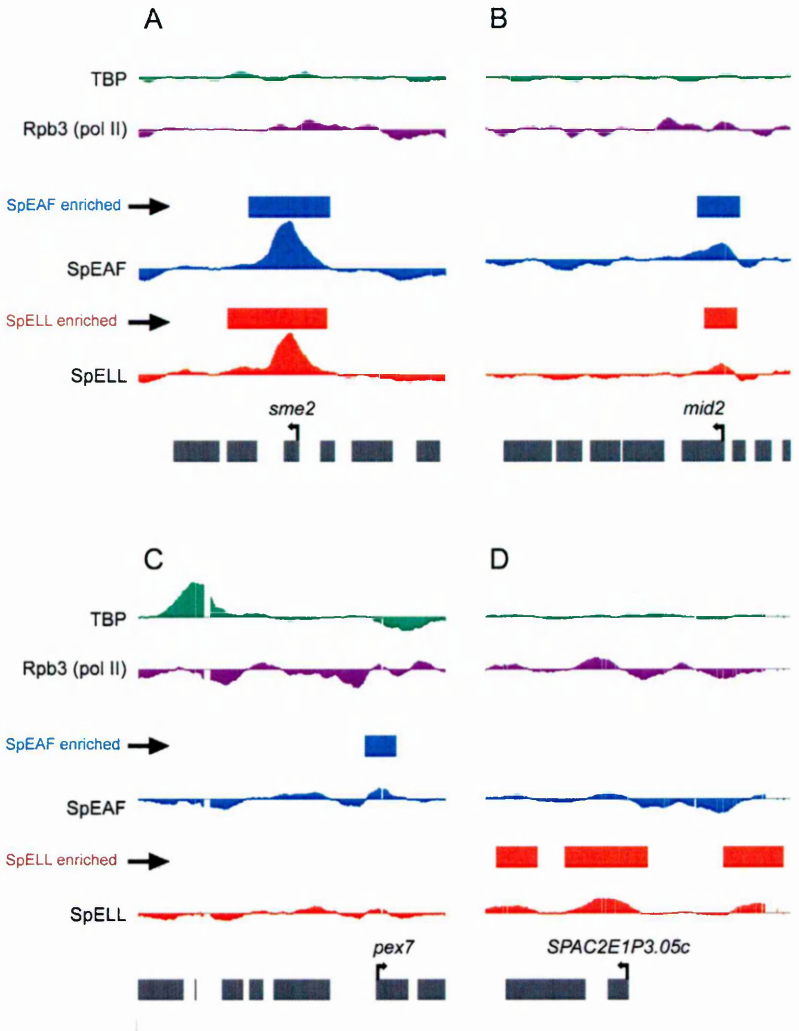


Figure 4-6 SpELL and SpEAF enriched genes have similar ChIP-chip profiles.

ChIP profiles (myc tagged strain / untagged strain) for four genes are shown with MAT scores on the y axis. Annotated genes are depicted by gray boxes. Control ChIPs use strains *tbp1:myc* (green) and *rpb3:myc* (purple). Experimental ChIPs use strains *eaf1:myc* (blue) and *ell1:myc* (red). SpEAF enriched regions (calculated with a p value cutoff of 0.05) are shown with blue boxes, and SpELL enriched regions with red boxes. **A** and **B**. Genes *sme2* and *mid2* are both SpEAF and SpELL enriched. **C**. *pex7* is SpEAF enriched and not SpELL enriched. **D**. *SPAC2E1P3.05c* is SpELL enriched and not SpEAF enriched. All four genes have reduced mRNA expression in the *ell1Δ* and/or *eaf1Δ* strains.

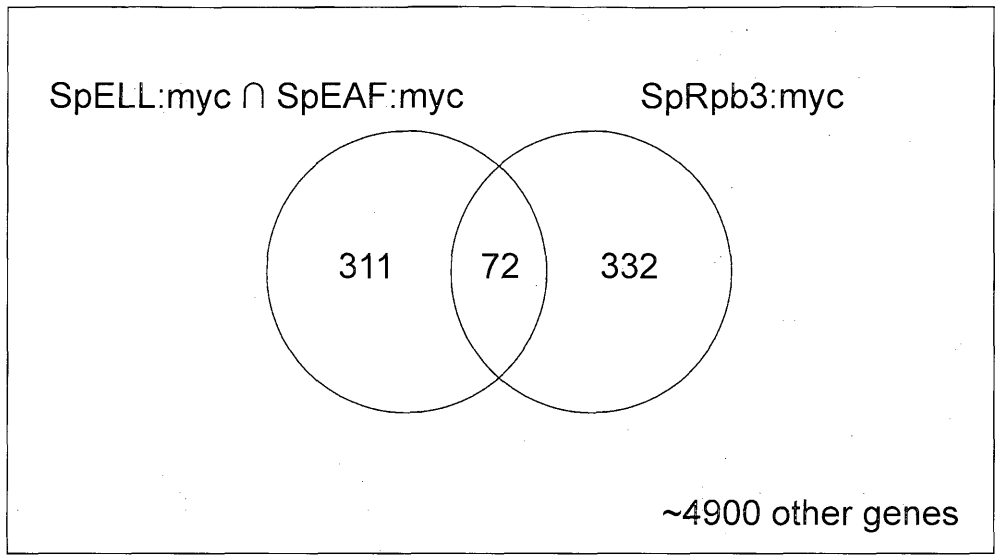


Figure 4-7 Genes enriched by both RNA pol II and the SpELL/SpEAF complex. SpELL:myc ∩ SpEAF:myc describes the set of genes enriched by both SpELL:myc and SpEAF:myc chromatin immunoprecipitations.

4.5.3 The SpELL/SpEAF complex is enriched at genes with epitope tagged pol II

To ask whether the set of SpELL/SpEAF genes correlate with genes enriched by immunoprecipitated RNA pol II, I used a strain in which the third largest subunit of pol II was tagged with 13 tandem repeats of the myc epitope at the C terminus to perform an Rpb3:myc ChIP-chip study. I compared the SpELL/SpEAF enriched genes with the SpRpb3 enriched genes (Figure 4-7) and found 72 genes common to both groups. The probability of 72 or more genes being common to these groups if the genes had been selected at random from the genome is 6.00×10^{-15}, suggesting that the SpELL/SpEAF complex colocalises with epitope tagged pol II at a set of genes in cells. I use a more quantitative approach to examine the relationship between SpELL/SpEAF localisation and pol II occupancy in sections 4.6.2 and 4.6.3.

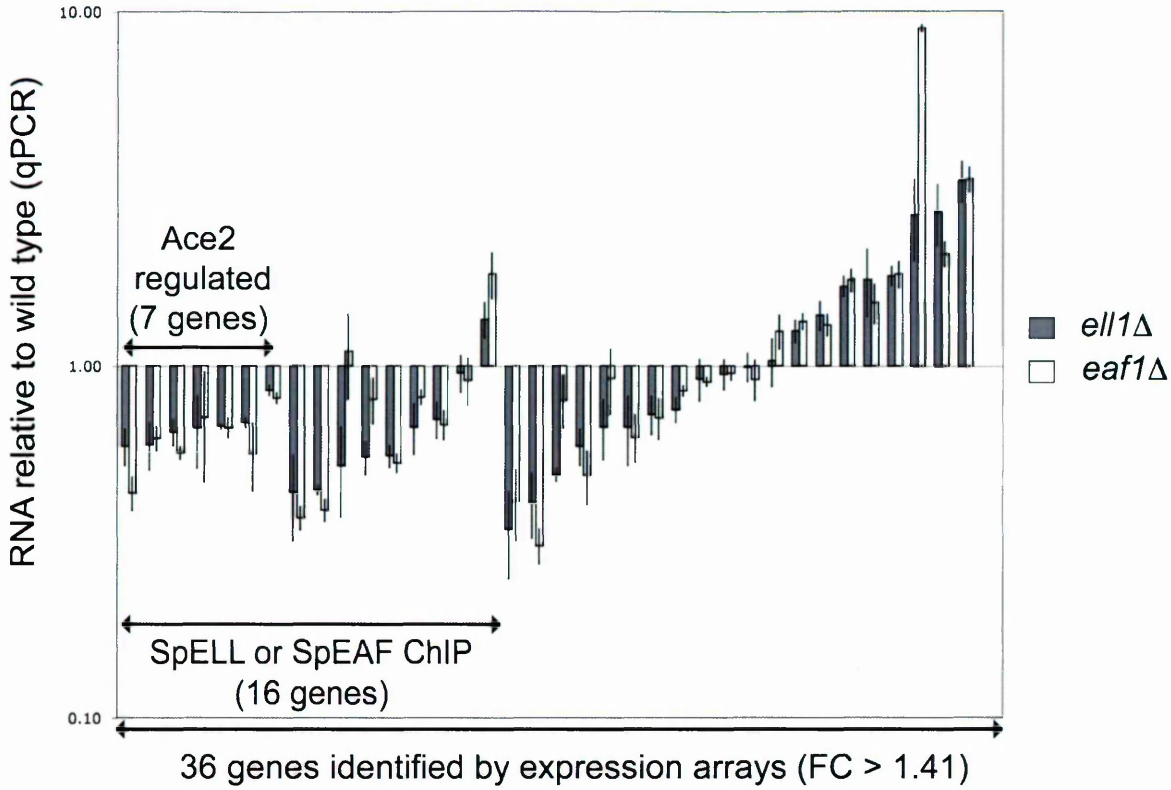


Figure 4-8 SpELL or SpEAF localises to a number of the genes that are downregulated in the *ell1* and *eaf1* deletion strains.

All genes (n = 36) which were identified by spotted microarrays with a change in RNA levels in the *ell1Δ* or *eaf1Δ* strains have been compared with the set genes classified as either SpELL enriched or SpEAF enriched genes (n=383). The change in RNA levels in deletion cells relative to wild type cells measured by qPCR is shown. Results are the mean of three biological replicates and error bars represent range.

4.5.4 Are the genes that are downregulated in the deletion strains also SpELL/SpEAF enriched genes?

Of the 37 genes exhibiting altered RNA levels in the *ell1* and/or *eaf1* deletion strains, 16 were either SpELL and/or SpEAF enriched genes. Among these are genes encoding the transcription factor Ace2 and 6 Ace2 regulated genes (Figure 4-8). The identities of these genes are listed in appendix C. 27 genes have reduced RNA levels in the *ell1Δ* or *eaf1Δ* strains. Of these, 10 are also both SpELL and SpEAF enriched genes

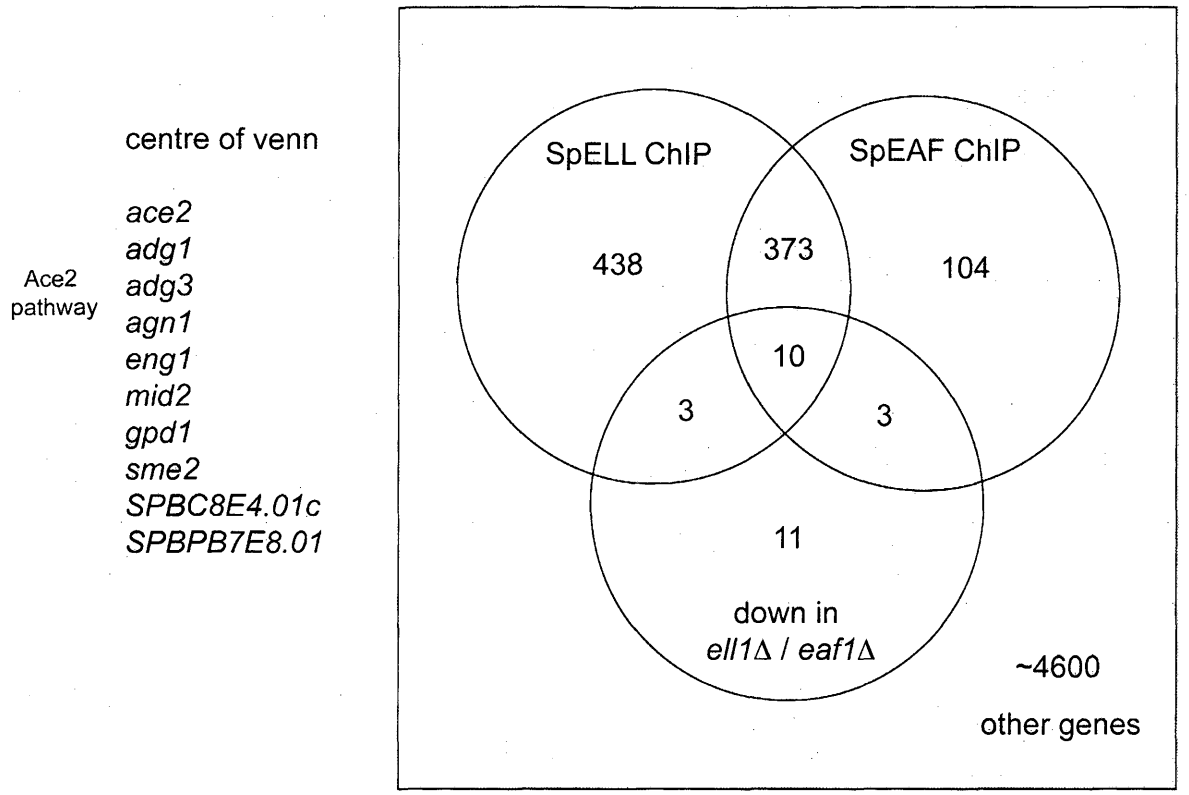


Figure 4-9 SpELL and SpEAF enriched genes identified by tiling arrays.

The SpELL and SpEAF enriched genes have been compared with the 26 genes identified as being downregulated in the *ell1* or *eaf1* deletion arrays as well as with the *sme2* gene which was not spotted onto the array but was later confirmed to be downregulated in the deletion strains by qPCR. N.B. there are 16 genes which are both downregulated in both the *ell1Δ* and *eaf1Δ* strains, and are also either SpELL or SpEAF enriched genes. This includes *sme2* in addition to the the 15 genes fulfilling this condition in Figure 4-8.

(Figure 4-9). The probability that 27 genes selected at random from the genome would yield at least 10 of the 383 (SpELL ∩ SpEAF) enriched genes can be calculated as before:

$$P(\text{center of venn} \geq 10) = 6.43 \times 10^{-6}$$

This suggests that the overlap between the two gene sets is not due to chance and that SpELL/SpEAF complex coimmunoprecipitates a set of genes that is downregulated in the absence of the complex.

4.6 A set of candidate genes for SpELL/SpEAF regulation

The 10 genes that are downregulated in *ell1* and/or *eaf1* deletion strains and have both an SpEAF and an SpELL ChIP signal (listed in Figure 4-8) define a set of candidate genes that I believe might be regulated directly by the SpELL/SpEAF complex. I will refer to these as "SpELL/SpEAF candidate genes". These include *ace2* and the Ace2 regulated genes *adg1, adg3, agn1, eng1* and *mid2*, a glycerol phosphate dehydrogenase, *gpd1*, which is induced by osmotic stress, *sme2* a non coding RNA involved in meiosis, a putative phosphate transporter *SPBC8E4.09c*, and an uncharacterised gene *SPBPB7E8.01*.

4.6.1 Does the SpELL/SpEAF complex preferentially regulate longer genes?

It has been proposed that dELL preferentially affects the expression of large genes in the fruit fly (94). To investigate the possibility that the SpELL/SpEAF complex preferentially regulates longer genes, we used both a statistical one sample t-test and a Mann-Whitney U test to compare the ORF lengths of our candidate genes to the mean ORF length of all other *S. pombe* genes[1,2](Table 4.1). The probability that the mean ORF length of a random sample (candidate genes), drawn from a population of median ORF length 1155, would be greater than 1829 is calculated to be 0.0583 (using one sample t-

[1] I performed these tests at the suggestion of and in parallel with Hua Li, a biostatistician.

[2] The two tests make different assumptions; the t-test assumes that the data are normally distributed while the Mann-Whitney test is non-parametric and does not make assumptions about the underlying distribution of the data. The Mann-Whitney U test ranks all data and then compares the rankings of one data set to the rankings of the other data set. It is less likely that the Mann-Whitney test will give a false positive result due to outlying data points.

SpELL / SpEAF enriched genes

	ORF length (bp)	mRNA level	Relative pol II occupancy
mean	1452	5488	2.55
standard deviation	1089	3976	2.01

All other genes

	ORF length (bp)	mRNA level	Relative pol II occupancy
mean	1416	2578	1.25
standard deviation	1115	2635	1.01

Compare SpELL / SpEAF enriched genes to all other S.pombe genes

	ORF length	mRNA level	Relative pol II occupancy
two sample t-test p value	0.5421	< 0.0001	< 0.0001
Mann-Whitney U test p value	0.386	< 0.0001	< 0.0001
is the SpELL/SpEAF enriched gene mean value significantly greater than the mean value of all other genes at the 95% significance level?	NO	YES	YES

Table 4.1 Hypothesis test to determine whether SpELL/SpEAF candidate genes have significantly higher than average values for ORF length, mRNA level and pol II occupancy.

Values for ORF length, mRNA level and relative pol II occupancy in *S. pombe* have previously been determined in a study by Lackner and co-workers (161). I used these values to perform a one sample t-test (using the t-test calculator at http://www.graphpad.com/quickcalcs/OneSampleT1.cfm?Format=SD), and a Mann Whitney U test (using the calculator at http://www.fon.hum.uva.nl/Service/Statistics/Wilcoxon_Test.html). The *sme2* gene was not included in the analysis as this gene had not been included in the study by Lackner et al. Values for mRNA level and pol II occupancy are in arbitrary units.

test), or 0.0759 (using a Mann-Whitney U test). This is larger than the significance level of 0.05 for both tests and the null hypothesis that SpELL/SpEAF regulated genes do not have larger ORF length is accepted suggesting that the evidence is not sufficient to support the hypothesis that the SpELL/SpEAF complex is preferentially recruited to longer genes (Table 4.1).

As the results of a statistical analysis such as this become more valid as the number of genes in the data set increases, I decided to repeat the analysis using the set of 383 SpELL/SpEAF enriched genes (the genes with a significant ChIP signal for both the *ell1:myc* and *eaf1:myc* strains). Many of these genes may be regulated directly by the

SpELL / SpEAF enriched genes

	ORF length (bp)	mRNA level	Relative pol II occopancy
mean	1452	5488	2.55
standard deviation	1089	3976	2.01

All other genes

	ORF length (bp)	mRNA level	Relative pol II occopancy
mean	1416	2578	1.25
standard deviation	1115	2635	1.01

Compare SpELL / SpEAF enriched genes to all other S.pombe genes

	ORF length	mRNA level	Relative pol II occopancy
two sample t-test p value	0.5421	< 0.0001	< 0.0001
Mann-Whitney U test p value	0.386	< 0.0001	< 0.0001
is the SpELL/SpEAF enriched gene mean value significantly greater than the mean value of all other genes at the 95% significance level?	NO	YES	YES

Table 4.2 Hypothesis test to determine whether SpELL/SpEAF enriched genes have significantly higher than average values for ORF length, mRNA level and pol II occupancy.

Values for ORF length, mRNA level and relative pol II occupancy in S. pombe have previously been determined (161). I used these values to perform a two sample t-test (using the t-test calculator at http://www.graphpad.com/quickcalcs/ttest1.cfm?Format=SD), and a Mann Whitney U test (using the calculator at http://www.fon.hum.uva.nl/Service/Statistics/Wilcoxon_Test.html). Genes not included in the study by Lackner et al were excluded from this analysis. Values for mRNA level and pol II occupancy are in arbitrary units.

SpELL/SpEAF complex even though there is no significant change observed in the RNA levels of these genes in the *ell1Δ* or *eaf1Δ* strains under the conditions used to perform the expression array analysis. As the sample size is now significant compared with the population we have replaced the one-sample t-test with the two-sample t-test. The results of the hypothesis tests comparing the SpELL/SpEAF enriched genes to the set of all other *S. pombe* genes are shown in Table 4.2. The mean ORF length of the SpELL/SpEAF enriched genes (1452) is not significantly greater than the mean ORF length of all other *S. pombe* genes (1416) at the 95% significance level ($p<0.05$).

4.6.2 Does the SpELL/SpEAF complex preferentially regulate genes with high mRNA levels or genes with a high RNA pol II occupancy?

Gerber et al. have established that dELL co-localises with hyperphosphorylated RNA pol II on polytene chromosomes and that RNA pol II from fly extracts copurifies with dELL (106). Also, Shilatifard and co-workers showed that purified ELL binds pol II isolated from rat liver (103). In addition, I have shown that SpELL interacts physically and functionally with purified *S. pombe* pol II in vitro (Figure 3-9 and Figure 3-10). Taken together, these observations suggest that ELL/EAF and pol II might be part of a common complex during elongation in vivo. For this reason, I decided to ask whether SpELL/SpEAF occupancy correlates with RNA pol II occupancy and/or with levels of mRNA expression genome-wide. When I compared the mean values of both of these quantities between the SpELL/SpEAF candidate gene set and all other *S.pombe* genes I found that the candidate genes had significantly higher levels of mRNA compared with the average value for the genome. However, it was not clear whether pol II occupancy was greater for candidate genes as the two statistical tests used had different outcomes (Table 4.1). If I compare the larger set of SpELL/SpEAF enriched genes with all genes in the *S.pombe* genome, the mean values of mRNA level (5488) and pol II occupancy (2.55) are clearly significantly greater than for all other *S.pombe* genes (2578 and 1.25 respectively)(Table 4.2).

Figure 4-10 illustrates the differences in distributions of ORF length, mRNA level and pol II occupancy between SpELL/SpEAF enriched genes and all *S.pombe* genes. Both gene sets have a similar distribution for ORF length but there is a greater proportion of the SpELL/SpEAF enriched gene set with high mRNA and pol II occupancy levels compared with the set of all genes. Taken together the analyses of Table 4.1, Table 4.2 and Figure 4-10 suggest that the SpELL/SpEAF complex is not preferentially recruited to

longer genes only, but is preferentially recruited to genes with higher pol II occupancy and mRNA levels.

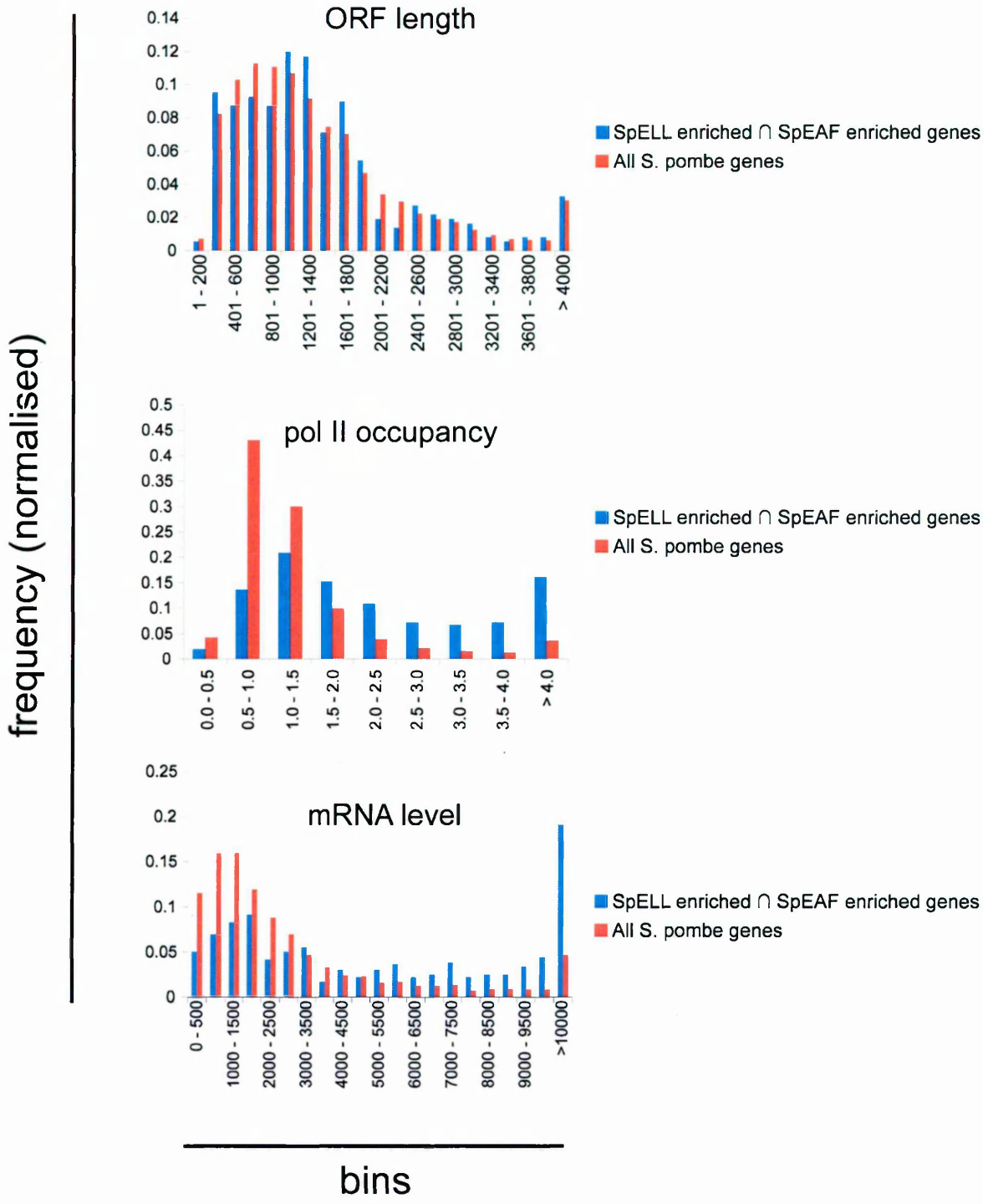


Figure 4-10 Histograms illustrating the frequency distribution of values for ORF length, pol II occupancy and mRNA level for both SpELL/SpEAF enriched genes and all genes.

Values for ORF length, mRNA level and relative pol II occupancy in S. pombe have previously been determined in a study by Lackner and co-workers (161). ORF length is in base pairs. mRNA level and relative pol II occupancy are both in arbitrary units.

A

SpELL / SpEAF enriched genes with high pol II occupancy

	ORF length (bp)	mRNA level
mean	1433	9906
standard deviation	1074	2645

Other genes with high pol II occupancy

	ORF length (bp)	mRNA level
mean	742	10002
standard deviation	1312	3691

B

Compare SpELL / SpEAF enriched genes to other *S.pombe* genes with high pol II occupancy

	ORF length	mRNA level
two sample t-test p value	0.0006	0.860
Mann-Whitney U test p value	<0.0001	0.088
is the SpELL/SpEAF enriched gene mean value significantly greater than the mean value of other genes at the 95% significance level?	YES	NO

C

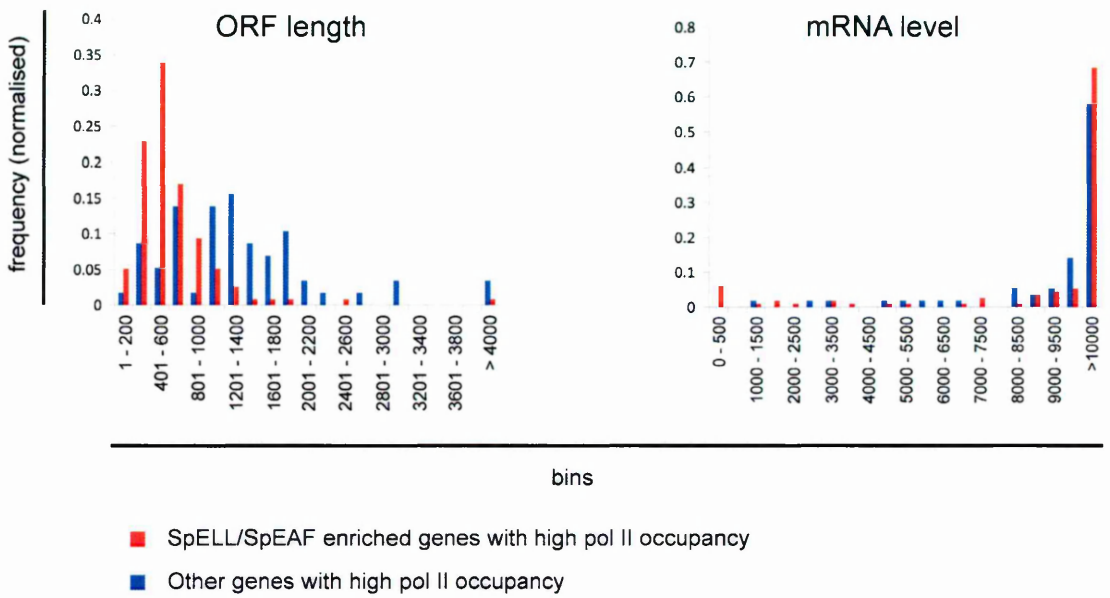


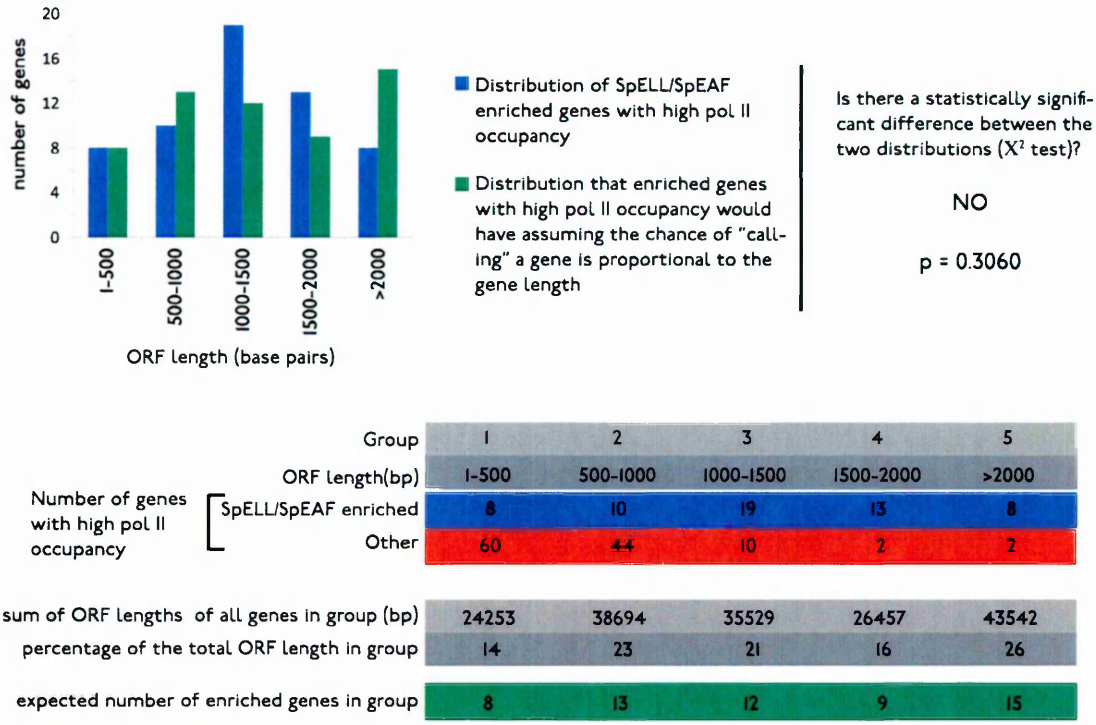
Figure 4-11 SpELL/SpEAF enriched genes with a high pol II occupancy have a larger ORF length than other genes with a high pol II occupancy.

I repeated the analysis of Table 4.2 and Figure 4-10 using only genes with a pol II occupancy greater than 4.00 (arbitrary units). ORF length is in base pairs and mRNA level in arbitrary units. There are 58 SpELL/SpEAF enriched and 118 other genes with a pol II occupancy greater than 4.00 units.

4.6.3 Is the SpELL/SpEAF complex recruited to a class of longer genes with a high pol II occupancy?

It is possible that the SpELL/SpEAF complex is recruited to more than one distinct category of genes. In particular actively transcribed genes with pause sites or regions where pol II moves slowly might be expected to have a high pol II occupancy resulting

from pol II piling up. The SpELL/SpEAF complex might be required specifically to help transcribe longer genes in this class, since longer genes could contain more pause sites and take a longer time to transcribe. With this in mind I analysed all genes with a pol II occupancy greater than 4.00 (arbitrary units) to ask whether SpELL/SpEAF might be recruited to the longer genes with a high pol II occupancy. There were 58 SpELL/SpEAF enriched genes and 119 other genes with a pol II occupancy greater than 4.00. The SpELL/SpEAF enriched genes which had high pol II occupancy had a mean ORF length of 1433 which is significantly longer than the other genes with high pol II occupancy, which had a mean ORF length of 742 (Figure 4-11). If the SpELL/SpEAF complex is promoting more efficient transcription of these longer genes we might expect similar levels of mRNA to be generated for the longer (SpELL/SpEAF regulated) and shorter genes. Table 3.1 suggests that there is no difference in the amounts of mRNA synthesised between longer and shorter genes in this class.



Group	1	2	3	4	5
ORF length(bp)	1-500	500-1000	1000-1500	1500-2000	>2000
Number of genes with high pol II occupancy: SpELL/SpEAF enriched	8	10	19	13	8
Number of genes with high pol II occupancy: Other	60	44	10	2	2
sum of ORF lengths of all genes in group (bp)	24253	38694	35529	26457	43542
percentage of the total ORF length in group	14	23	21	16	26
expected number of enriched genes in group	8	13	12	9	15

Figure 4-12 Measured vs modelled distribution of gene lengths for SpELL/SpEAF enriched genes with a high pol II occupancy.

SpELL/SpEAF enriched (blue) or other (red) genes with a high pol II occupancy were grouped into 5 groups according to ORF length (in base pairs) as shown in the table. The total length occupied by all genes within a group was calculated (in base pairs) for each group and expressed as a percentage of the sum of all genes in all groups. The expected distribution of the 58 SpELL/SpEAF enriched genes (green) assumes that the genes are distributed among the groups in proportion to the sum of the lengths of all the genes within a group (i.e. the percentage of the total ORF length in each group).
The measured (blue) and modelled (green) distributions were compared for independence using the chi squared test (162), and found not to be significantly different (p = 0.30).

One possibility is that SpELL/SpEAF might be more likely to be detected at longer genes simply because longer genes provide a larger target for colocalisation with a peak, and perhaps thus have a higher probability of containing pause sites. In collaboration with Hua Li, we asked whether the distribution of the lengths of genes with SpELL/SpEAF peaks and a high pol II occupancy was statistically significantly different to the distribution of the lengths of these genes that would be expected assuming a random distribution of SpELL/SpEAF peaks along the sum total length of all of these genes. The analysis of Figure 4-12 suggests that there is no statistical difference between the actual and modelled

distributions of gene lengths for the SpELL/SpEAF enriched genes with high pol II occupancy.

4.6.4 Is the SpELL/SpEAF complex is recruited to a class of longer genes with a high mRNA level?

I decided to do a similar analysis to ask whether SpELL/SpEAF might be recruited to a class of genes with high associated mRNA levels. I analysed all genes with mRNA levels greater than10000 (arbitrary units); there were 69 SpELL/SpEAF enriched and 154 other genes in this category (Figure 4-13 A-C). The SpELL/SpEAF enriched genes which had high mRNA levels had a mean ORF length of 1156 which is significantly longer than the other genes with high mRNA level, which had a mean ORF length of 803. This suggests that of the SpELL/SpEAF enriched genes with high cellular mRNA levels are significantly longer than other genes with high mRNA levels.

As with the class of SpELL/SpEAF enriched genes with high pol II occupancy, it is possible that SpELL/SpEAF enriched genes with high associated mRNA levels are more likely to be longer genes simply because longer genes offer a larger target for colocalisation with a peak. We modelled the distribution of SpELL/SpEAF peaks among the genes with high associated mRNA levels as we had done for the genes with high pol II occupancy (Figure 4-13 D). Again there was no significant statistical difference between the measured and modelled distribution of genes according to length. This suggests that SpELL/SpEAF might be more likely to enrich longer than shorter genes in this class simply because they provide a larger target for colocalisation. The analysis of does not eliminate the possibility that SpELL/SpEAF is specifically targeted to these longer genes via a mechanism independent of ORF length (target size).

A

SpELL / SpEAF enriched genes with high mRNA level

	ORF length	Relative pol II occupancy
mean	1156	4.97
standard deviation	671	2.62

Other genes with high mRNA level

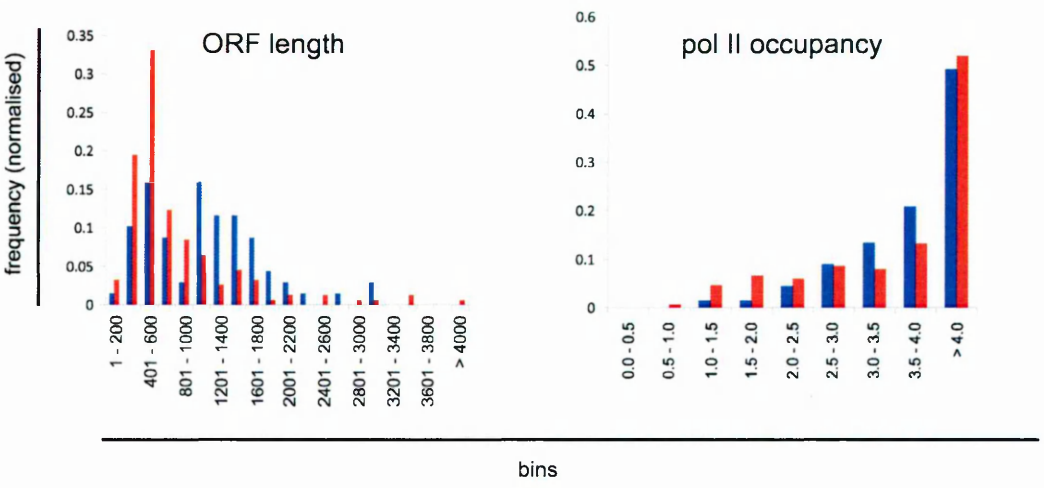
	ORF length	Relative pol II occupancy
mean	803	4.06
standard deviation	687	1.66

B

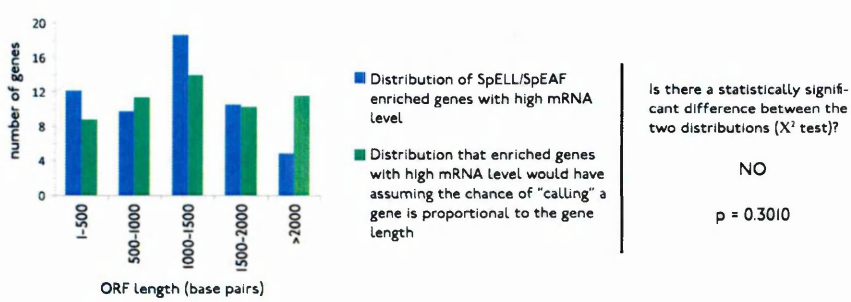
Compare SpELL / SpEAF enriched genes to all other S.pombe genes

	ORF length	Relative pol II occupancy
one sample t-test p value	0.0004	0.0023
Mann-Whitney U test p value	<0.0001	0.1043
is enriched gene mean value significantly greater than the median value other genes at the 95% significance level?	YES	YES/NO

C



D



Group	1	2	3	4	5
ORF length (bp)	1-500	500-1000	1000-1500	1500-2000	>2000
Number of genes with high mRNA level — SpELL/SpEAF enriched	15	12	23	13	6
Number of genes with high mRNA level — Other	71	47	18	9	9
sum of ORF lengths of all genes in group (bp)	31875	41433	50799	37308	42042
percentage of the total ORF length in group	16	20	25	18	21
expected number of enriched genes in group	11	14	17	13	14

Figure 4-13 SpELL/SpEAF enriched genes with a high mRNA level have a larger ORF length than other genes with a high mRNA level.

I repeated the analysis of Table 4.2, Figure 4-10, and Figure 4-12 using only genes with an mRNA level greater than 10000 (arbitrary units). ORF length is in base pairs and pol II occupancy in arbitrary units. There are 69 SpELL/SpEAF enriched and 154 other genes with an

4.6.5 SpELL/SpEAF distribution across coding regions

Although previous studies have examined chromosomal localisation of the ELL at a resolution that allows large chromosomal regions associated with ELL to be identified (94, 163) it is not clear how the SpELL/SpEAF complex might be distributed across genes. One possibility could be that the distribution of the complex might vary according to the nature of individual genes, perhaps depending on the distribution of intrinsic pause sites within the transcribed region. Another possibility could be that the complex might be consistently located at the promoter or at the 3' end of the gene. For example one study demonstrated that ELL shows a higher enrichment at the 3' end of the human c-Fos gene (164). To address these questions we[1] analysed the distribution of the SpELL and SpEAF proteins across all genes, SpELL/SpEAF enriched genes or SpELL/SpEAF candidate genes as shown in Figure 4-14. The complex appears to be distributed within the coding regions of genes and is not located exclusively in either the promoter or 3' end regions. There is a slight increase in localisation of both proteins towards the 5' end of the coding region, in particular in the analysis of the set of candidate genes.

[1] The analysis was done in collaboration with Madelaine Gogol who designed a script using the programming language R to process the data and generate the graphs shown.

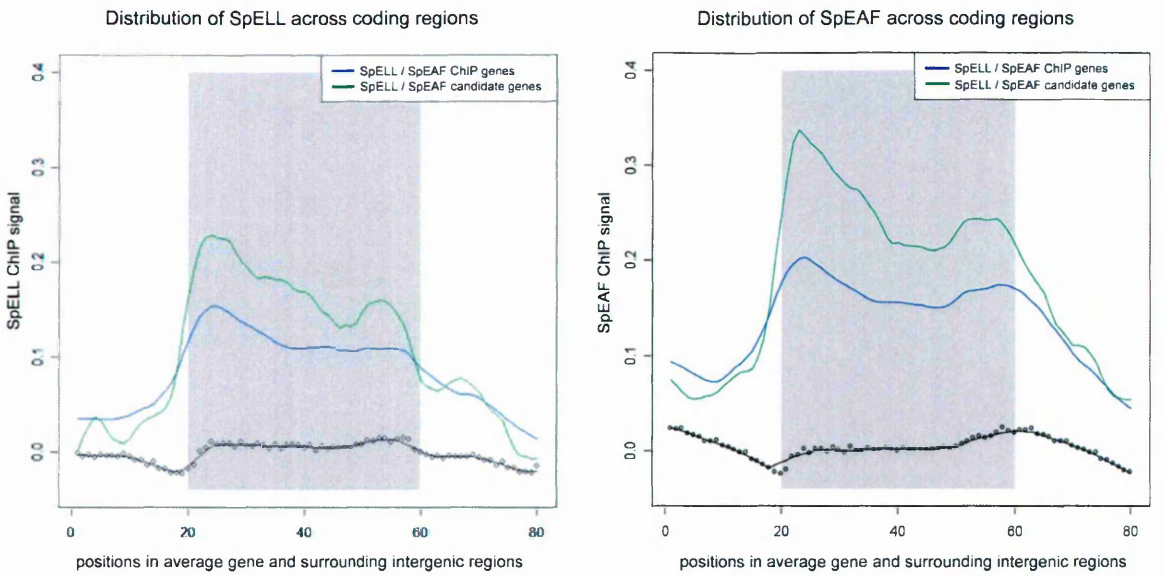


Figure 4-14 Distribution of SpELL and SpEAF across coding regions.

Each ORF was divided into 40 equally sized bins and upstream and downstream intergenic regions into 20 equally sized bins. Genes are usually annotated from the translation start codon (ATG) to the stop codon. ChIP signals for each gene were assigned to the closest bin and empty bins for each gene filled by linear interpolation between the closest surrounding bins with non zero values. The average bin value for all genes in the gene set being analysed is plotted as the ordinate. The black line represents the set of all *S. pombe* genes.

4.6.6 The SpELL/SpEAF complex is not exclusively located at intron containing genes

In light of recent evidence that human ELL2 may influence splice site selection (120), I asked whether the SpELL/SpEAF complex was localised exclusively at intron containing genes. Of the set of 10 candidate genes, none are annotated as containing introns suggesting that the SpELL/SpEAF complex has roles other than (and possibly in addition to) regulating splicing.

4.6.7 The SpELL/SpEAF candidate genes are enriched for genes involved in cell separation.

The Gene Ontology Consortium has developed a vocabulary for describing the roles of genes and their products in cells (165). In order to determine whether the SpELL/SpEAF complex was responsible for regulating genes with a common function in cells I analysed the list of candidate genes for enrichments in gene ontology terms as described in Table 4.3. I found a significant enrichment for genes involved in cell separation during cytokinesis. These included the gene encoding the transcription factor Ace2 and a subset of Ace2 regulated genes, *agn1*, *eng1*, *mid2*, *adg1* and *adg3* (166). In light of the physical and functional interactions of the SpELL/SpEAF complex with RNA polymerase II, it is interesting that this set of Ace2 regulated genes is also downregulated in cells with reduced levels of the pol II subunit Rpb4 (167), The deletion of the *rpb4* gene in *S. cerevisiae* renders the cells sensitive to growth on 6-AU, and Rpb4 may have a distinct role in transcription elongation (168). This set of Ace2 pathway genes have also been shown to be particularly sensitive to mutations in various Mediator subunits (169, 170), including components of the head module (Med8, Med17, Med20 and Med27) and Med31, which appears to be at or near the interface between the head and middle modules (Chieri Sato, Shigeo Sato unpublished data).

Biological Process

GO Term	P-value	Sample frequency	Background frequency	Genes
GO:0000920 cell separation during cytokinesis	2.99E-09	5/10 (50.0%)	36/5178 (0.7%)	eng1 ace2 adg3 agn1 mid2
GO:0032506 cytokinetic process	7.97E-07	5/10 (50.0%)	107/5178 (2.1%)	eng1 ace2 adg3 agn1 mid2
GO:0000910 cytokinesis	3.74E-06	5/10 (50.0%)	146/5178 (2.8%)	eng1 ace2 adg3 agn1 mid2
GO:0051301 cell division	6.65E-06	5/10 (50.0%)	164/5178 (3.2%)	eng1 ace2 adg3 agn1 mid2
GO:0016052 carbohydrate catabolic process	3.39E-04	3/10 (30.0%)	76/5178 (1.5%)	eng1 agn1 gpd1

Cellular Component

GO Term	P-value	Sample frequency	Background frequency	Genes
GO:0009986 cell surface	7.60E-07	5/10 (50.0%)	106/5178 (2.0%)	eng1 adg3 agn1 adg1 SPBPB7E8.01

Molecular Function

GO Term	P-value	Sample frequency	Background frequency	Genes
GO:0004553 hydrolase activity, hydrolyzing O-glycosyl compounds	7.54E-05	3/10 (30.0%)	46/5178 (0.9%)	eng1 adg3 agn1
GO:0016798 hydrolase activity, acting on glycosyl bonds	1.36E-04	3/10 (30.0%)	56/5178 (1.1%)	eng1 adg3 agn1

Table 4.3 Gene Ontology terms enriched in the set of SpELL / SpEAF candidate genes.

The list of SpELL/SpEAF enriched genes was used as an input for the GO Term Enrichment tool (http://www.geneontology.org 165) with the maximum p-value setting of 0.001 and a minimum of three gene products.

4.7 Discussion

The SpELL/SpEAF complex appears to function at a number of genes regulating different cellular processes and disruption of the complex leads to reduced levels of RNA associated with some of these genes.

I have defined sets of genes likely to be regulated by the SpELL/SpEAF complex, and analysis of the properties of these genes provides insight into how the complex might function in cells. Whole genome analyses necessarily identify a range of signals varying across the genome requiring an arbitrary cut off to be selected to classify genes as belonging to a particular gene set. Thus, the gene sets do not define those genes and only those genes that are SpELL/SpEAF regulated rather sets of genes that are most likely to be SpELL/SpEAF regulated.

I have previously shown that SpELL and SpEAF function as a stable complex *in vitro* (Figure 3-6). In cells, a common set of genes is misregulated by the deletion of either subunit (Figure 4-1), and deletion of one subunit results in a decrease in the protein levels, but not mRNA levels, of the other subunit (Figure 4-2 and Figure 4-3). Together,

these findings provide further support for the idea that SpELL and SpEAF function together as a complex *in vivo*. It is possible that the reduced levels of SpELL protein in *eaf1Δ* cells might be the result of a decrease in stability of the SpELL in the absence of its partner subunit. Alternatively, SpEAF might be needed for proper processing or translation of *ell1* mRNA.

Studies with purified proteins *in vitro* have demonstrated that the SpELL/SpEAF complex functions through a direct interaction with pol II (Figure 3-9,Figure 3-10), and *drosophila* ELL colocalises with active pol II on polytene chromosomes (106). Consistent with these observations, I have shown that the SpELL/SpEAF complex is located at a significant number of genes with both high levels of pol II occupancy and high mRNA levels (Table 4.1, Table 4.2, and Figure 4-10). Consistent with this, SpELL/SpEAF complex is distributed throughout the coding regions of genes (Figure 4-14).

As ELL/EAF functions to stimulate the rate of transcription elongation *in vitro*, it has been proposed that the complex might be used for the timely transcription of longer genes (94). The function(s) of the SpELL/SpEAF complex are not limited to enabling transcription of longer genes, as such genes are not overrepresented in the SpELL/SpEAF gene sets (Table 4.1, Table 4.2). The histogram showing the distribution of the SpELL/SpEAF enriched genes ranked by ORF length, mRNA level and pol II occupancy illustrates that genes with high pol II occupancy and high mRNA levels are overrepresented in the SpELL/SpEAF enriched gene set but that longer genes are not (Figure 4-10). It has also been suggested that ELL/EAF might be specifically required for transcribing intron containing genes (4, 94). However, the set of SpELL/SpEAF candidate genes does not contain any genes with introns, suggesting that the SpELL/SpEAF complex does not exclusively control splicing.

There is a possibility that SpELL/SpEAF complex may be involved with regulating more than one class of genes, including longer genes with high pol II occupancy. Zeitlinger and co-workers recently considered this possibility in analysing CHIP-chip data sets investigating pol II binding to *drosophila* chromosomes and found three distinct classes of genes including a class of transcribed genes with pol II binding across the gene and a class of genes with pol II stalled near the promoter poised for activation (171). Analysing subsets of the SpELL/SpEAF enriched gene dataset, limited to only the genes with a high pol II occupancy or high associated mRNA level, shows a correlation between such genes and ORF length (Figure 4-11 - Figure 4-13). I cannot exclude the possibility that this correlation may result, at least in part, from a greater probability of a peak overlapping a longer gene; longer genes provide a larger target and would be more likely than shorter genes to coincide with randomly distributed peaks. Hence the SpELL/SpEAF complex may be important for regulation of longer, highly transcribed genes in addition to other classes of genes (perhaps those that are of mixed length but need to be transcribed quickly at certain times), but is not clear why this might be true. This could be either due to specific recruitment of SpELL/SpEAF to this set of longer genes or could simply due to an increased probability of SpELL/SpEAF occupancy as the length of any contiguous DNA sequence increases.

It is interesting that most of the genes in the set of candidate genes do seem to be used only at a distinct period during the cell cycle. There are six genes that are part of the Ace2 pathway and are downregulated in *ell1Δ* and *eaf1Δ* strains, are coimmunoprecipitated by both SpELL and SpEAF, and whose products are required for cell separation during cytokinesis (Figure 4-9 and Table 4.3). These include *ace2, adg1, adg3, agn1, eng1* and *mid2*. Interestingly, these genes have been identified as being

misregulated in strains expressing low levels of the pol II subunit Rpb4 (172), which has also been suggested to be important for proper transcription elongation (168). These Ace2 pathway genes are also misregulated in cells with mutations in a variety of mediator subunits. Taken together these findings suggest a joint role for pol II, mediator and the SpELL/SpEAF complex in the control of a pathway of genes required for the proper completion of cell division. In the future, it will be of interest to explore a potential role for Mediator in recruiting and/or regulating the function of SpELL/SpEAF in *S. pombe*.

The complex may not function at every gene. 942 of the ~5000 genes in *S. pombe* were coimmunoprecipitated by either SpELL or SpEAF (Figure 4-9), and, of the genes associated with pol II, most were not also associated with both SpELL and SpEAF (Figure 4-7). Alternatively the complex could function at many more genes under different growth conditions. Notably, only a minority of genes were associated with pol II, suggesting that a proportion of the genome was not actively transcribed in the majority of cells at the time of the experiment; thus, many of these genes may use the SpELL/SpEAF complex when they are transcribed.

The results of this whole genome study are consistent with a role of SpELL/SpEAF acting in concert with pol II to stimulate transcription of a set of genes *in vivo*, although additional roles for the complex cannot be excluded. In addition, the study has identified a number of genes that are likely to be regulated by SpELL/SpEAF and provides models that can be used in more targeted analyses to investigating the mechanism of SpELL/SpEAF function *in vivo*.

CHAPTER 5. RECRUITMENT OF SpELL AND SpEAF SUBUNITS TO THE *sme2* LOCUS

5.1 Introduction

Using whole genome approaches, I have identified genes that are enriched and likely to be regulated by the SpELL/SpEAF complex. By using a targeted approach to study a gene found to be highly occupied by SpELL/SpEAF, I have been able to characterise in more detail the recruitment of SpELL and SpEAF to regions within a candidate transcribed gene.

The biochemical studies described in Chapter 3 have shown that free SpELL binds to pol II and that free SpEAF does not (Figure 3-9, Figure 3-10). These observations predict that SpELL might localise to transcribed genes in the absence of SpEAF by binding directly to pol II and that SpEAF localisation might require SpELL. To test this hypothesis, I used SpELL- and SpEAF-dependent chromatin immunoprecipitations analysed by qPCR to study the recruitment of SpELL and SpEAF to the *sme2* gene.

The *sme2* gene in particular provides a suitable target for investigating recruitment for three reasons. First, the peak MAT scores within the SpELL and SpEAF ChIP profiles associated with *sme2* were the largest of all genes tiled on the arrays. Second, the profile of each peak appears well resolved, with no other significant peaks in the surrounding genomic regions thus reasonably accurate information about the presence and location of each peak is possible (Figure 5-1). This is especially important in a model system such as *S. pombe*, where the average gene size is considerably smaller than for higher eukaryotes and is close to the same order of magnitude as the sizes of sonicated chromatin fragments produced for chromatin immunoprecipitations. Third, the *sme2* gene can be induced in

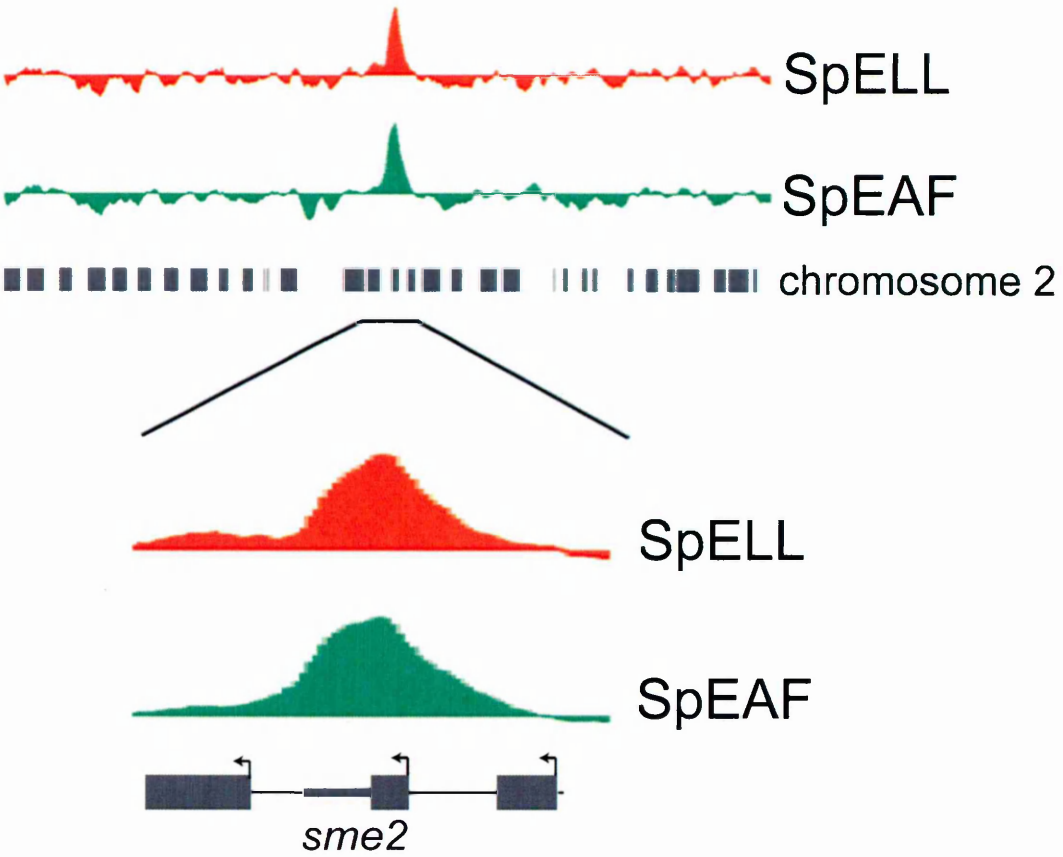


Figure 5-1 SpELL and SpEAF colocalise at the *sme2* locus.

The SpELL and SpEAF proteins coimmunoprecipitate chromatin that is enriched for sequences within the *sme2* gene on chromosome 2. Ordinate values for each ChIP-chip profile indicate MAT scores (141) based on three biological replicate experiments. Experimental ChIPs were with strains containing myc epitope tagged SpELL or SpEAF. Control ChIPs using the same myc antibody and the wild type strain were used as a reference to normalise signals. The transcribed region of *sme2* is indicated; the decrease in the height of this region corresponds to the low amounts of longer species of *sme2* transcript (discussed in 5.2)

response to a readily manipulated transcription signal such as oxidative and heat stress and nitrogen starvation.

5.2 Characterisation of *sme2* transcripts

Sme2 codes for a non-coding RNA, meiRNA, that is transcribed by pol II and is subsequently polyadenylated. It functions by binding to the *mei2* gene product during

meiosis and promotes its translocation to the nucleus. It can be induced in response to heat shock or oxidative stress as well as in response to conditions, such as nitrogen starvation, that promote meiosis in *S. pombe*. Before investigating the localisation of pol II, SpELL and SpEAF on the *sme2* gene, I wanted to confirm which regions of the gene were being transcribed either under conditions where *sme2* is expressed at basal levels or in cells exposed to oxidative stress, which results in the accumulation of larger amounts of *sme2* transcript. Data from Lackner *et al.* suggests that this accumulation is not due to a decrease in RNA transcript degradation but rather due to an increase in transcription, since whole genome experiments showed no significant reduction in *sme2* transcript levels after cells were treated with 1, 10-phenantholine[1] (161).

The *sme2* gene, as characterised previously by Watanabe and co-workers (173), has a putative TATA box ~30 bp upstream of the two major transcription start sites at T1 and A2. In addition, it has and several poly(A) sites. Northern analysis of the *sme2* transcript after induction by nitrogen starvation showed two doublets, with the majority of transcripts in the lower doublet corresponding to products that are 439/440 and 507/508 nucleotides long. Although the lengths of the transcripts in the second, upper doublet were not determined in this paper, they appeared to be ~1200 nt in length. The authors suggest that the less abundant, longer transcripts probably reflect read-through products that overlap with the shorter products (173).

[1] *sme2* transcript levels were reduced to 93% (standard deviation = 10%) 28 minutes after transcription was blocked by 300 μg/ml 1,10-phenanthroline. 1,10-phenantholine at 100 μg/ml reduces transcription to 10% of normal levels.

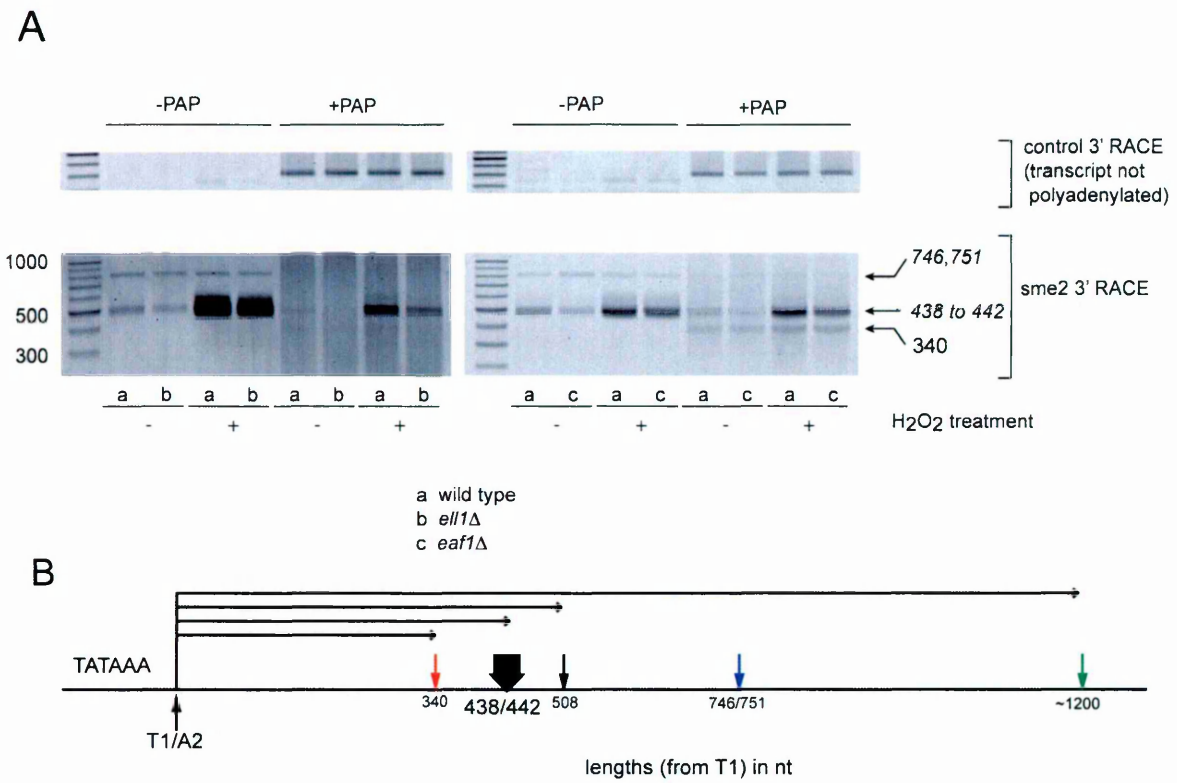


Figure 5-2 Identification of the 3' ends of *sme2* transcripts by RACE

A. The 3' ends of *sme2* transcripts were identified by sequencing 3' RACE products generated from cDNA prepared from wild type, *ell1Δ* or *eaf1Δ* cells incubated for 30 minutes with or without 6 mM H_2O_2 (described in section 2.16). In order to identify both polyadenylated and non-polyadenylated species of transcripts, RNA was initially incubated in reactions with or without poly(A) polymerase. Reverse transcription was with the "lock docking" oligo(dT) primer described in materials and methods. cDNA was amplified using an sme2 specific forward primer (sme2 +26) and a reverse primer complimentary to the 5' end of the primer used for reverse transcription.
B. Summary of data from this study and from the study by Watanabe and co-workers (173) characterising the *sme2* transcripts. Note the transcription start site (T1/A2), non-polyadenylated transcript at +340 (red arrow), majority of transcripts polyadenylated at +438 to +442, polyadenylated transcript at +508, and longer transcripts which may or may not be polyadenylated at +748/+751 (blue arrow) and transcripts detected by Northern blot which are ~1200 nt long (green arrow)

5.2.1 Identification of the 3' ends of *sme2* by 3' RACE

To confirm the poly(A) sites for the shorter transcripts described by Watanabe *et al.* and to define the positions of the poly(A) sites of the longer doublet, I amplified 3' RACE products generated with an internal *sme2* forward primer at +26 nt and a modified

"lock docking" oligo(dT) reverse primer (174) that anneals to the 5' end of the poly(A) sequence during reverse transcription. 3' polyadenylation sites were mapped by fractionating RACE products generated from wild type and *ell1Δ* cells by agarose gel electrophoresis, cloning the DNA from individual gel bands, and sequencing the cloned DNA. In addition to the reported poly(A) sites at +440 and +508 (relative to the T1 at the transcription start site), I identified additional polyadenylation sites at +438, +442, +746 and +751. Treatment of cells with 6 mM H_2O_2 resulted in an increase in the intensity of the band that resolved just below the 500 bp marker. The majority of *sme2* sequences represented transcripts polyadenylated at between +438 and +442 suggesting that these poly(A) sites are used when the gene is induced (Figure 5-2). There was no apparent difference between wild type and *ell1Δ* in the exact choice of polyadenylation site. Only 1 of the 28 *sme2* sequences represented a transcript polyadenylated at +508.

Pre-treatment of RNA with polyA polymerase resulted in an additional band indicative of a prevalent non-polyadenylated species of transcript whose 3' end was at +340. The band representing these transcripts can be seen in all samples pre-treated with polyA polymerase in Figure 5-2 A, although the bands are fainter relative to background for the experiment comparing wild type and *ell1Δ* cells. This is probably due to slight variation in pcr amplification between the wt/*ell1Δ* samples and the wt/*eaf1Δ* samples. Subsequent experiments using nested primers for pcr amplification confirmed the presence of the 340 nt transcripts in all samples. It is not clear whether this transcript represents a prematurely terminated transcript, a transcript associated with paused pol II, or is the result of some alternative process.

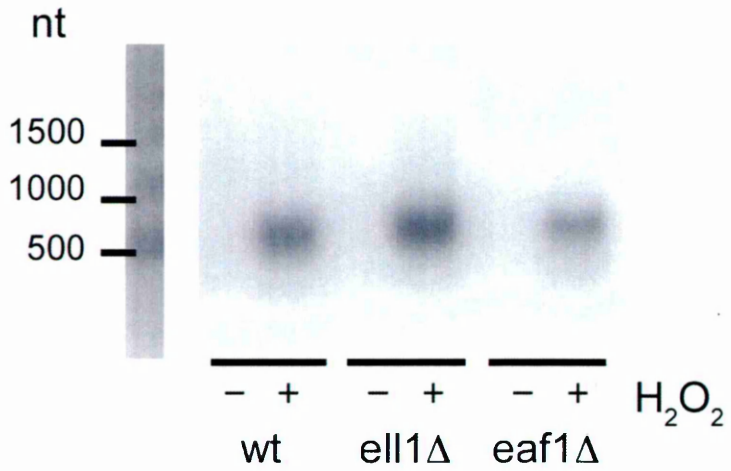


Figure 5-3 Northern analysis of *sme2* transcripts

30 μg of total RNA prepared from either wild type, *ell1Δ* or *eaf1Δ* cells incubated at 32 °C for 30 minutes with or without 6 mM H_2O_2 was fractionated on a 1% agarose gel, and *sme2* transcripts were detected by Northern blotting as described in section 2.15. Single stranded radiolabeled RNA probes contained a sequence complimentary to the region +26 to +545 of the *sme2* gene.

Analysis of RACE products by PCR is semi-quantitative. To confirm that ~440 nt *sme2* transcripts were the main species generated in response to treatment of cells with 6 mM H_2O_2, I analysed samples of total RNA by Northern blotting (Figure 5-3). I was unable to detect *sme2* transcripts transcribed at basal levels in the absence of H_2O_2. In contrast, stressed cells contained detectable levels of *sme2* transcript, which appeared as a single diffuse band. Taken together with the results of Figure 5-2, these results suggest that the majority of RNA products generated on exposure of cells to stress in Figure 5-3 are greater than 500 nt and less than 800 nt long and probably represent polyadenylated species of the dominant 438/442 nt transcripts characterised by 3' RACE.

5.3 Co-localisation of SpELL, SpEAF and pol II at the *sme2* gene

Having characterised the transcripts synthesised at the *sme2* gene, I wanted to see if there was a correlation between the localisation of SpELL, SpEAF and pol II under conditions where the gene is expressed at basal levels or induced.

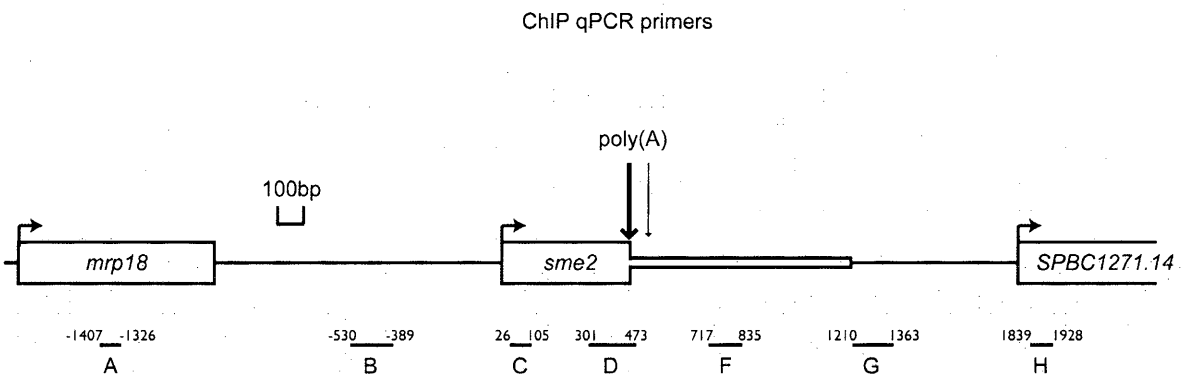


Figure 5-4 Location of primers for ChIP qPCR across the *sme2* locus.

Sequences for appropriate primer pairs were generated by using approximately 400 bp genomic DNA sequences as an input for PrimerQuest software (IDTDNA). Primer Quest was used with the default settings for generating primer sequences for real-time PCR and generated lists of suggested primer pairs with a PCR product range 80-200 bp from which the most appropriate primer pairs were selected. Primer sequences are listed in Appendix B. Arrows indicate the poly(A) sites at +438 to +442 nt (thick arrow) and +508nt (thin arrow).

In order to better characterise the distribution of the SpELL/SpEAF complex in and around the *sme2* locus, I designed a series of qPCR primer sets covering a region from 1407 bp upstream to 1928 bp downstream of the transcription start site. This region overlaps the two neighbouring ORFs (Figure 5-4). In addition, I used an amplicon within the K region as a control. The K region is a transcriptionally silent locus that has been used previously as a control for ChIP assays in *S. pombe* (79). Appropriate primer pairs were selected from a list generated by PrimerQuest software (IDTDNA www.idtdna.com). In initial experiments I used sonication conditions that generated chromatin in the range 200-1000 bp as suggested in the Affymetrix® ChIP assay protocol

and by Lee and Young (175). These conditions produced unacceptable variation and very broad peaks. More recent studies carried out by Fan and Struhl (140) found that extensive chromatin fragmentation increases enrichment of binding sites relative to control regions. Likewise, I was able to get more reproducible data with greater resolution by increasing sonication time to generate chromatin fragments with an average size of 200 bp (data not shown).

5.3.1 SpELL enriches chromatin within the transcribed region of *sme2*

In agreement with the microarray analysis (Figure 5-1), qPCR analysis of chromatin fragments coimmunoprecipitated with SpELL revealed an enrichment of sequences near the polyA sites of the shorter *sme2* transcripts relative to the surrounding regions and a distant control region (the K region) as shown in Figure 5-5 A. Control chromatin generated from the parental strain without epitope tagged SpELL was less efficiently immunoprecipitated with the myc antibody and did not produce a significant enrichment of sequences within the *sme2* gene relative to the surrounding regions. To control for experimental variation, I have normalised all experimental ChIP signals to the K region control ChIP signal (Figure 5-5 B and following figures).

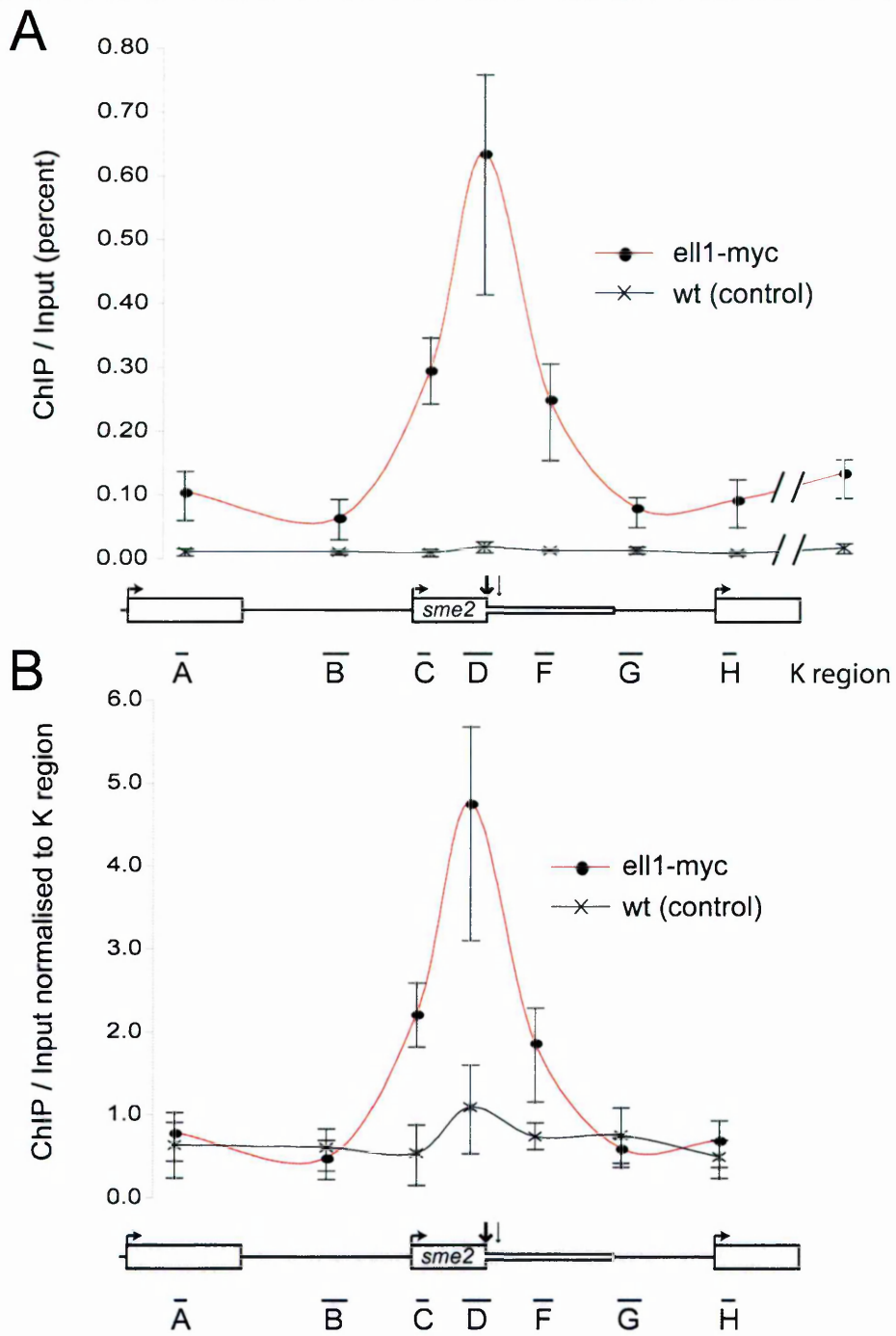


Figure 5-5 SpELL enriches sequences within the *sme2* region.

A Chromatin immunoprecipitations (ChIPs) were performed with antibodies specific for the myc epitope using a strain with the *ell1* gene modified to encode 13 tandem repeats of the myc epitope at the C terminus. As a negative control ChIPs were performed with myc antibodies and chromatin extracts prepared from wild type cells. ChIPs were performed as described in section 2.9 with an increased sonication time which resulted in sheared chromatin with a reduced average size of ~200 bp. ChIP signals show enrichment of DNA in the immunoprecipitated sample relative to the input DNA expressed as a percentage. All data points are the mean of three biological replicates. Error bars represent range.

B The data represented in Figure 5-5 A with experimental ordinate values divided by the K region ordinate value. All data points are the mean of three biological replicates. Error bars represent range.

5.3.2 Is there a correlation between the localisation of SpELL, SpEAF and RNA pol II at the *sme2* locus?

The results of Figure 3-6 suggest that the SpELL and SpEAF proteins function as a complex *in vitro,* and the ChIP-chip analysis (Figure 5-1) suggests co-localisation of SpEAF with SpELL at the *sme2* gene in cells. Consistent with this, a comparison of Figure 5-6 A (red line) with Figure 5-6 B (red line) shows that myc-antibody-mediated chromatin immunoprecipitations from *S. pombe* strains encoding either SpEAF-myc or SpELL-myc give similar profiles when analysed by qPCR.

If the SpELL subunit is recruited to chromatin as a result of an interaction between SpELL and pol II bound to the chromatin, one might expect pol II to be present on the chromatin in the same location as SpELL. To test this, I immunoprecipitated pol II with the antibody 8WG16, which recognises the pol II CTD sequence YSPTSPS. Like SpELL and SpEAF, pol II coimmunoprecipitated genomic sequences that were enriched for region D (Figure 5-6 A to C red line).

5.3.3 The effect of a transcription signal on SpELL and SpEAF and pol II recruitment to the *sme2* locus

As oxidative stress induces *sme2*, it may also cause changes in the landscape of pol II, SpELL and SpEAF occupancy across the *sme2* region. To characterise these variations, I performed chromatin immunoprecipitations using antibodies against the hypophosphorylated pol II C-terminal domain, the CTD phosphorylated on serine5, the CTD phosphorylated on serine 2 and epitope tagged SpELL and SpEAF, and compared occupancy levels during normal logarithmic growth and after 30 minutes exposure to hydrogen peroxide. Initiating pol II or pol II engaged in early elongation is characterised by

phosphorylation of serine 5 in the YSPTSPS repeats within the CTD of the Rpb1 subunit of pol II, while productively elongating pol II is characterised by phosphorylation of serine 2. Consistent with *sme2* gene induction, exposure of the cells to H_2O_2 resulted in a significant increase in *sme2* sequences enriched by antibodies to the hyperphosphorylated pol II CTD. Notably, oxidative stress caused a marked increase in initiating RNA polymerase II phosphorylated at serine 5 in the promoter region (Figure 5-6 D amplicon C), as well as productively elongating pol II, which is phosphorylated at serine 2, in the body of the *sme2* gene (Figure 5-6 E amplicons C and D).

SpELL and SpEAF occupancy also increased significantly under *sme2*-inducing conditions (Figure 5-6 A and B). Although SpELL and SpEAF function with elongating pol II *in vitro*, the ChIP-qPCR profiles of these proteins do not parallel those of CTD-phospho-serine 5 nor CTD-phospho-serine 2. This suggests that there is not a direct relationship between pol II phosphorylation state and the presence of SpELL or SpEAF. [This is consistent with previous data that suggest that ELL can stimulate transcription elongation by pol II lacking a CTD *in vitro* (JWC personal communication)].

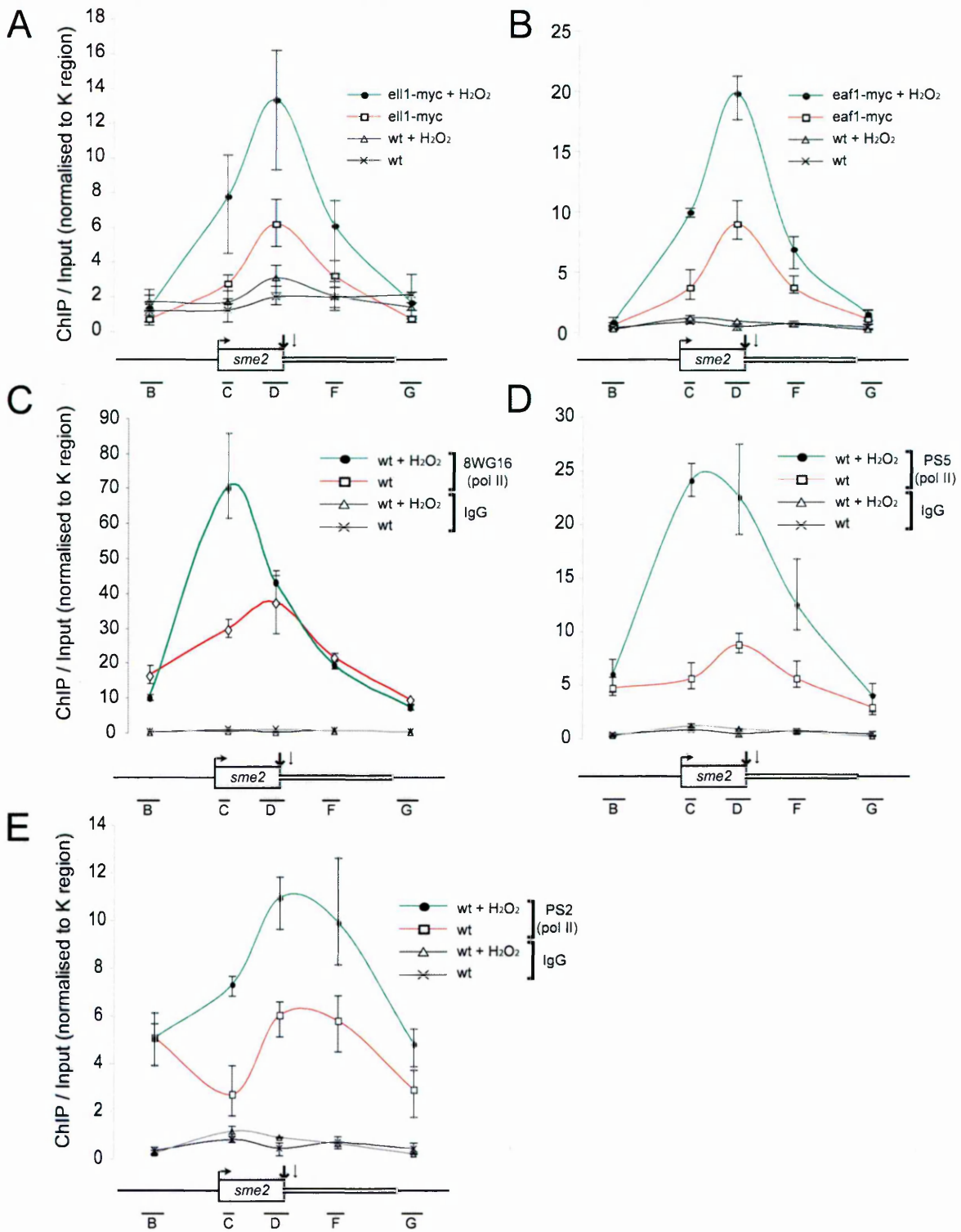


Figure 5-6 Co-localisation of SpELL, SpEAF and RNA pol II at the *sme2* gene.

Chromatin extracted from cells of the strains indicated treated with (green) or without (red) 6 mM H_2O_2 was immunoprecipitated using the following antibodies: anti myc (abcam ab9132) (A, B, C to E controls), pol II 8WG16 (C), pol II CTD phospho S5 (abcam ab5131) (D), pol II CTD phospho S2 (abcam ab5095) (E). All ChIP signals have been normalised to the K region as before. All data points are the mean of three biological replicates. Error bars represent range.

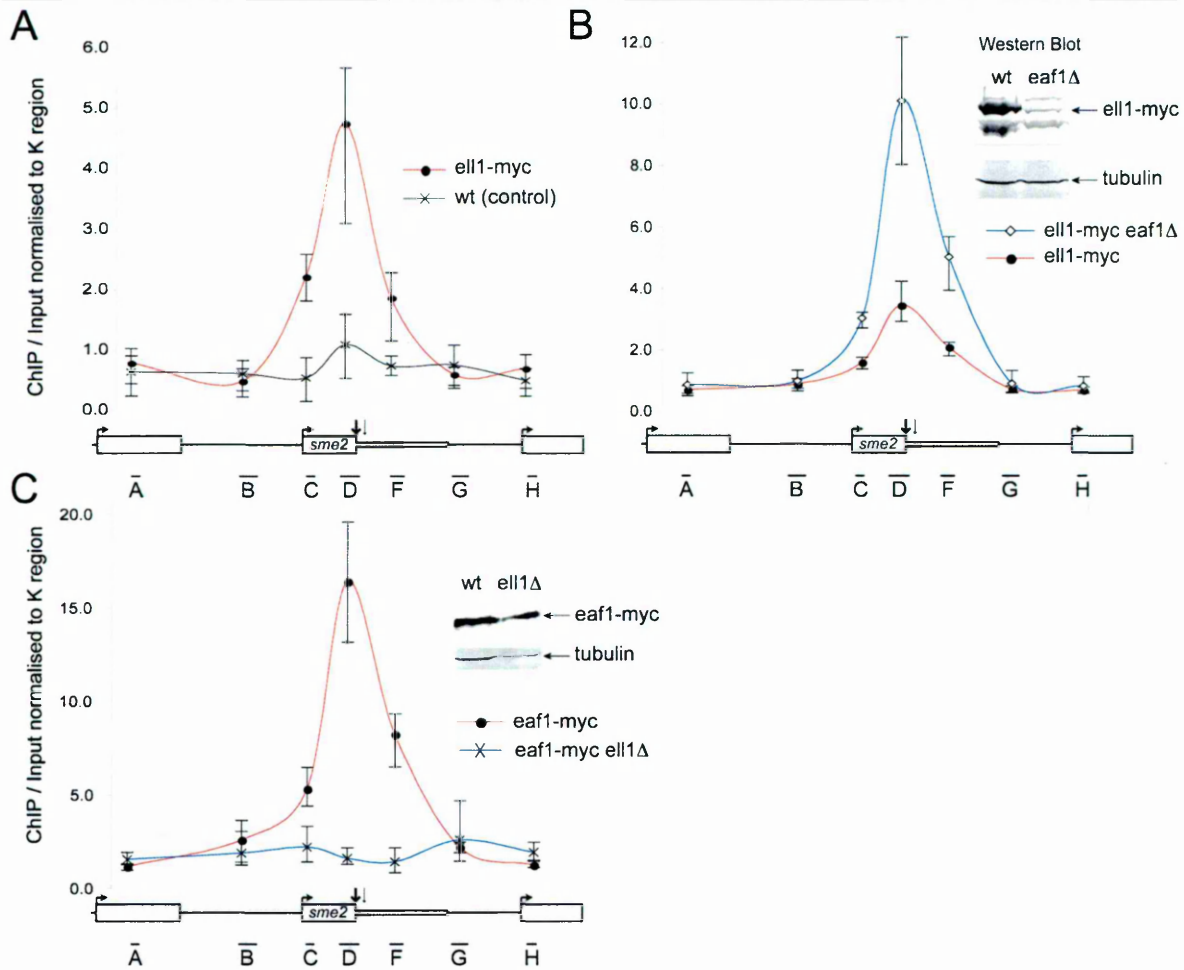


Figure 5-7 SpEAF enrichment at the *sme2* region depends on SpELL

Chromatin immunoprecipitations were performed with antibodies specific for the myc epitope using chromatin extracts from the strains indicated. Deletion of the *eaf1* gene results in reduced levels of the SpELL-myc protein (as shown by Western blot (inset B)). To account for these differences in levels of SpELL-myc protein in different strains, ChIP signals have been normalised. Values on the y axis are enrichments in immunoprecipitated DNA (ChIP / Input) normalised to enrichment at an amplicon within the transcriptionally silent K region. All data points are the mean of three biological replicates. Error bars represent range.

5.4 SpEAF enrichment at the *sme2* locus depends on SpELL

To begin to explore how the SpELL/SpEAF complex is recruited to the *sme2* gene, I decided to investigate the contribution of each subunit to complex localisation. As my previous studies have suggested that SpELL interacts physically with pol II *in vitro* in the absence of SpEAF (Figure 3-9), I wanted to ask whether the SpELL subunit was recruited to the *sme2* gene in the absence of SpEAF in *S.pombe* cells. To test this possibility, I used a

strain with the *eaf1* gene deleted and with a sequence coding for thirteen tandem copies of the myc epitope inserted before the *ell1* stop codon. Even though the strain lacking the *eaf1* gene has reduced levels of SpELL-myc protein, ChIPs with *ell1-myc eaf1Δ* cells enrich for sequences within the *sme2* locus. This suggests that SpEAF is not required for recruitment of SpELL to *sme2 in vivo* (Figure 5-7 B). It is not clear why the enrichment relative to the K region is larger in the *eaf1Δ* strain. This may reflect lower background at the K region due to the very low expression of SpELL in *eaf1Δ* cells. Alternatively the myc epitope tag on SpELL may be partially masked by SpEAF. Importantly the profiles of the two peaks shown in Figure 5-7 B are qualitatively similar.

To test the model that SpEAF is recruited to chromatin by SpELL *in vivo*, I used a strain with the *ell1* gene deleted and with a sequence coding for thirteen tandem copies of the myc epitope inserted before the *eaf1* stop codon. Consistent with the ChIP-chip assay (Figure 5-1), I found that SpEAF-myc coimmunoprecipitated sequences enriched for a region within the *sme2* transcribed region, with the peak enrichment detected by primer set D (sme2 301-473) (Figure 5-7 C red line), a region coincident with the location of peak enrichment by SpELL-myc cells (Figure 5-7 A and B). In the absence of the *ell1* gene there was no chromatin enrichment by the SpEAF-myc protein above background levels with respect to the surrounding regions (Figure 5-7 C blue line). As there is only a relatively small decrease in SpEAF-myc proteins when *ell1* is deleted (Figure 4-2), and I have normalised to the K region to account for changes in protein levels in lysates, these observations suggest that SpELL is required for localisation of SpEAF within the *sme2* gene.

5.5 Quantification of different species of *sme2* transcript in *ell1Δ* or *eaf1Δ* strains

Taken together, evidence from the 3'RACE and Northern analysis indicate that induction of the *sme2* gene in response to oxidative stress causes an increase in pol II transcription of *sme2*, which results in the production of transcripts polyadenylated at between +438 and +442 nt relative to the transcription start site. To test the possibility that there is a change in the level of these transcripts in cells in which the SpELL/SpEAF complex has been disrupted, I used two complementary assays, in which RNA from wild type, *ell1Δ* or *eaf1Δ* cells was reverse transcribed and the resulting cDNA analysed by qPCR. In the first assay, I used an oligo(dT) primer for reverse transcription to enrich for cDNA generated from polyadenylated transcripts, and, in the second assay, I used a series of gene specific primers for reverse transcription to detect both polyadenylated and non-polyadenylated transcripts.

5.5.1 Quantification of *sme2* transcript in *ell1Δ* or *eaf1Δ* strains using oligo(dT) primed RT-qPCR

To determine whether levels of the shorter *sme2* transcripts were altered in cells with either the *ell1* or *eaf1* genes deleted, I used oligo(dT) primed reverse transcription to generate cDNA complimentary to polyadenylated RNA from wild type, *ell1Δ* or *eaf1Δ* cells treated with or without 6mM H_2O_2. The cDNA samples were then analysed by quantitative PCR (qPCR) using primer pairs to amplify sequences containing either region C or region L as indicated in Figure 5-8 A. All correctly initiated transcripts longer than 105 nt should be detected by primers that amplify region C, whereas only transcripts

longer than +535 nt should be detected by the primers that amplify region L. I detected reduced levels of cDNA containing the region C sequence (sme2 26-105) in *eaf1Δ* cells treated with H_2O_2 compared to wild type cells and only slightly reduced levels in *ell1Δ* cells treated with H_2O_2 (Figure 5-8 B).

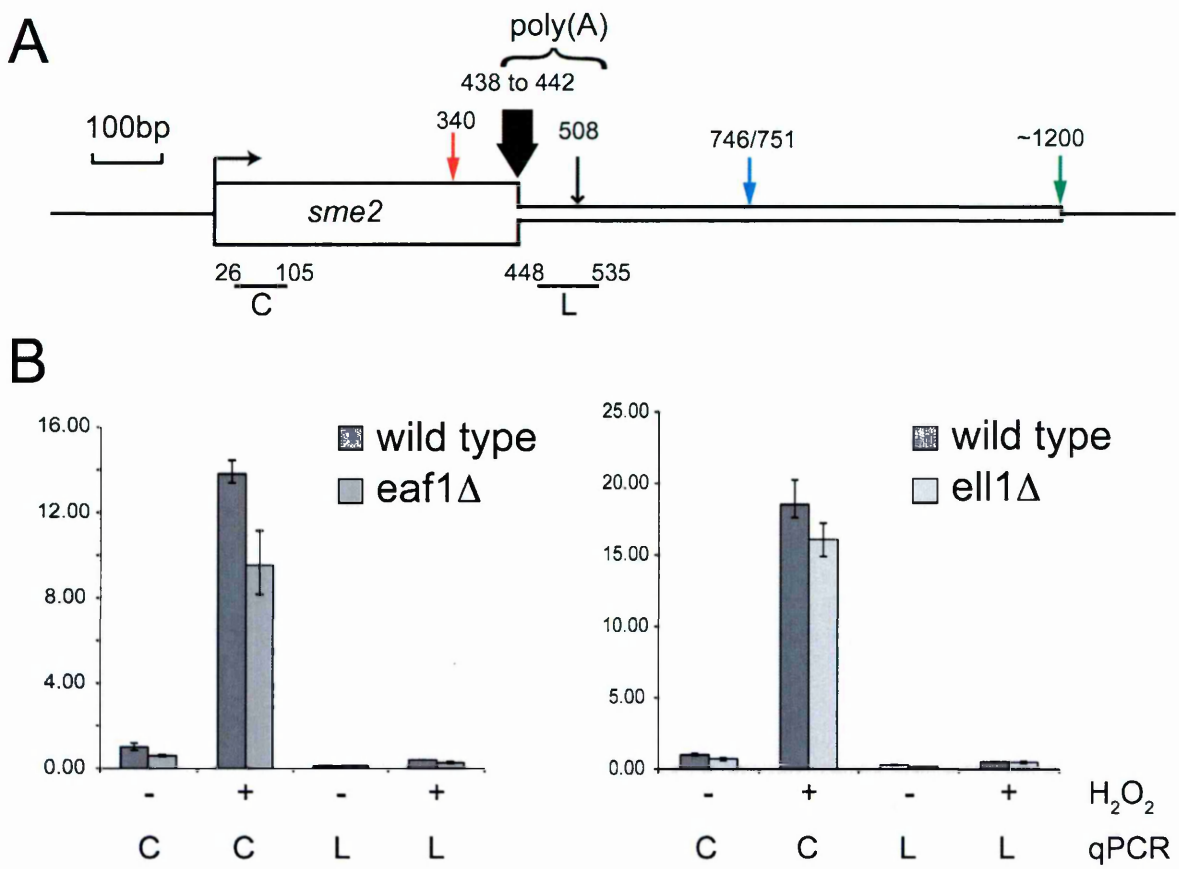


Figure 5-8 Reverse Transcription qPCR at *sme2*.

Cells were incubated at 32 °C with or without 6 mM H_2O_2 for 30 minutes and total RNA extracted with acidic phenol as described in section 2.11. 1 μg of total RNA was reverse transcribed with an oligo(dT) primer and the resulting cDNA was analysed by qPCR with the indicated primer pairs and a primer pair specific to the coding region of the *act1* gene (actin). Signals relative to genomic DNA standards are normalised first to actin then to wild type untreated cells. Error bars represent the range of three biological replicates.

The majority of transcripts identified by the primer set 26-105 in cells subjected to oxidative stress are likely shorter transcripts (polyadenylated at 439, 443 or 508) and not the longer ~1.2 kb transcripts detected by Watanabe and co-workers. The primers

amplifying region L (*sme2* 448-535) did not indicate a significant amount of the longer transcripts in cells whether or not they were treated with H_2O_2.

5.5.2 Quantifying transcription across *sme2*

As not all *sme2* transcripts identified by 3' RACE appeared to be polyadenylated, I wanted to use an assay that did not depend on the use of an oligo(dT) primer for reverse transcription to quantify the relative amounts both polyadenylated and non-polyadenylated transcripts. Gullerova and co-workers recently described an assay to assess transcription patterns across genes using only gene specific primers (176). In a similar assay, I performed separate reverse transcription reactions for each RNA sample using a series of primers located across *sme2* as indicated in Figure 5-9 A. Following reverse transcription, cDNA samples were analysed by qPCR with the reverse primer used for reverse transcription and a corresponding forward primer as indicated (Figure 5-9 A). I confirmed that levels of *sme2* transcripts that include the region *sme2* +26 to +105, were lower in *ell1Δ* or *eaf1Δ* cells than in wild type cells under conditions where the *sme2* gene is induced by oxidative stress Figure 5-9 B. All correctly initiated transcripts longer than 105 nt include the region *sme2* +26 to +105 and are detected by amplicon C. Interestingly, under conditions where *sme2* is transcribed at basal levels we did not detect a decrease in levels in transcripts containing the +25 to +105 region as we had done when specifically analysing polyadenylated transcripts using the oligo(dT) primer for reverse transcription (compare Figure 5-8 B with Figure 5-9 B). This may be due do the presence of non-polyadenylated transcripts containing the *sme2* +26 to +105 region which are not significantly reduced in the mutant strains.

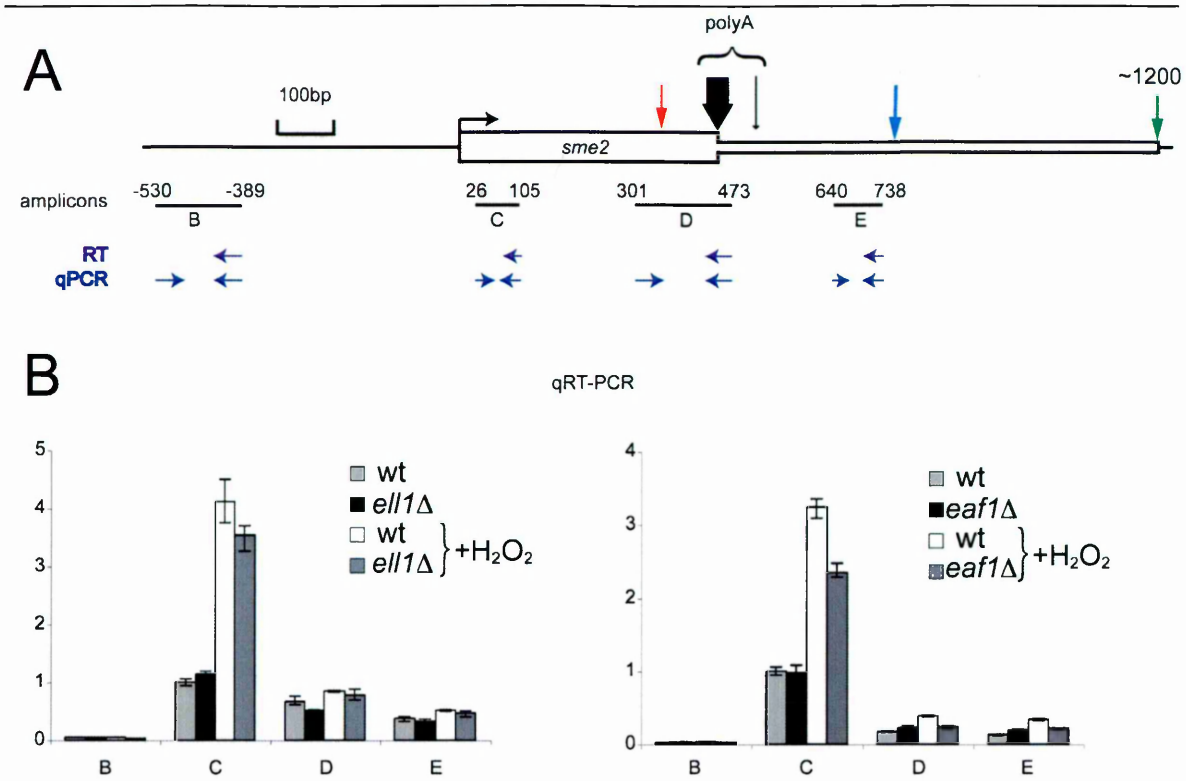


Figure 5-9 Transcription across the *sme2* locus

This assay is essentially as described by Gullerova and co-workers (176). 1 μg of total RNA from wild type, *ell1Δ* or *eaf1Δ* cells, incubated for 30 minutes with or without 6 mM H_2O_2, was reverse transcribed in four separate reactions (B to E) with the reverse primer corresponding to the regions B-E within the *sme2* gene as indicated, [together with a reverse primer corresponding to an amplicon within the actin gene (*act1*).] The resulting cDNA was analysed by qPCR using the reverse primer used for reverse transcription together with its corresponding forward primer. Signals relative to genomic DNA standards have been normalised to actin. Results from three separate experiments are shown with error bars representing the range.

5.6 Discussion

The major goal of the experiments described in this chapter was to define some of the basic requirements for recruiting SpELL/SpEAF to specific chromosomal locations. Although deletion of *ell1* or *eaf1* had only a very modest effect on the accumulation of *sme2* RNA, the *sme2* locus stood out as the gene with the highest apparent SpELL/SpEAF occupancy in initial screens for candidate genes using tiling arrays. Accordingly, I chose to

use *sme2* as a model gene to address questions about how SpELL and SpEAF are recruited to genes.

sme2: A complex locus with multiple polyadenylated species.

In order to use *sme2* as a model gene it was important first to define more completely the nature of the transcripts initiated at the *sme2* promoter under basal and inducing conditions. As discussed in the results section, a prior study (173) concluded that there are two major classes of transcripts: short ones, with major poly A sites at ~440 nt, and long ones, which hadn't been characterised and which were proposed to result from read-through of the upstream polyA sites. Consistent with previous studies, I found that there are major polyA sites at +438, +440 and +442 nt corresponding to the short products previously reported by Watanabe *et al.* with an additional, less frequently used polyA site immediately downstream at +508 nt. Although I did not detect polyadenylation sites at +1.2 kb, I identified cDNAs that appeared to correspond to transcripts polyadenylated at +746 and +751 nt. The 3' ends of these are immediately upstream of runs of T-residues in the template strand, thus, templated runs of As in the transcript could allow oligo(dT) priming from sites that do not correspond to *bona fide* polyadenylation sites. All basally transcribed *sme2* products were also detected under inducing conditions by qPCR with a significant increase in amounts of the shorter, polyadenylated transcripts. Although Northern blot analysis could detect only RNAs from cells treated with H_2O_2, these RNAs migrated as a diffuse band of intermediate size, and this band would be consistent with the presence of the +438 to +442 nt and the +508 nt polyadenylated transcripts and the absence of the longer or shorter transcripts.

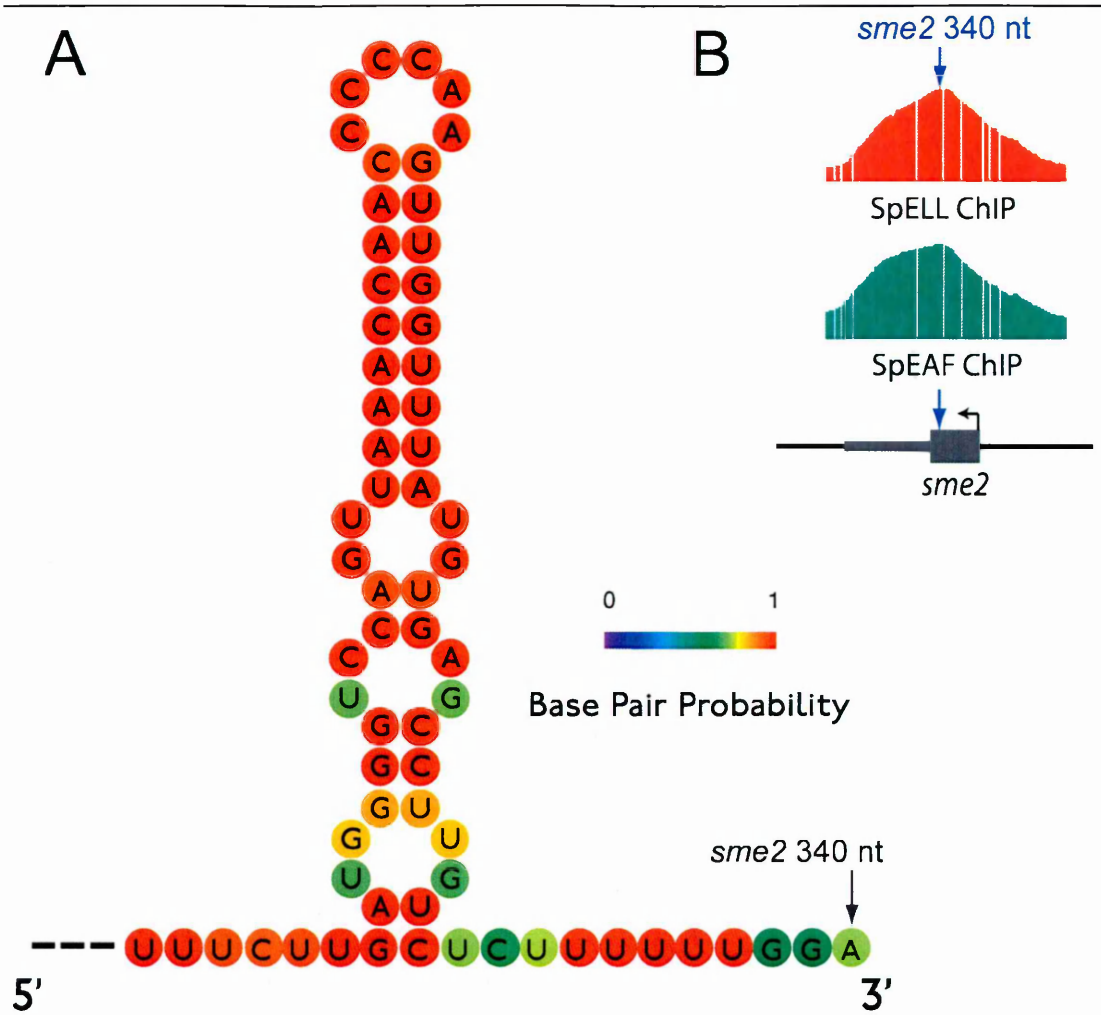


Figure 5-10 Predicted structure of the 3' end of the 340 nt *sme2* transcript.

A. Structure prediction was with the RNA Vienna package 1.8.2 at the University of Vienna RNA fold server : http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi. B. The 3' end of the 340 nt transcript is coincident with the maximum SpELL and SpEAF ChIP-chip MAT scores.

Evidence for stalling and/or premature termination upstream of the major polyadenylation site.

In addition to polyadenylated transcripts, I detected a non-polyadenylated species with the 3' end at +340 nt, and future studies will clarify whether this transcript could be associated with stalled polymerase or result from prematurely terminated transcription. Curiously, there is a region that is predicted to form a hairpin structure just upstream of

the 3' end of this transcript (Figure 5-10 A). This structure could act in a similar fashion to the prokaryotic intrinsic termination sequence, which destabilises the transcription elongation complex by forming a hairpin structure just before a poly(U) region in the transcript prior to termination (177). The location of the 3' end of this species is also coincident with the location of the maximum MAT score associated with chromatin enriched by SpELL and SpEAF (Figure 5-10 B). In addition, ChIP-qPCR suggested that this site, together with the major polyA sites, falls within the region of highest SpELL/SpEAF occupancy. In the future it would be interesting to investigate pol II transcription through this DNA sequence *in vitro* to see whether the +340 nt site might represent an intrinsic pause site or a premature termination site, and then to assess the effect of SpELL/SpEAF on transcription through this sequence.

How do SpELL and SpEAF affect each other's recruitment to sme2?

SpELL and SpEAF interact with RNA pol II via the SpELL subunit to stimulate transcription *in vitro*, suggesting a model in which pol II recruits an SpELL/SpEAF complex to DNA by binding to the SpELL subunit. I used ChIP-qPCR to assess the contribution of each subunit to recruitment of the SpELL/SpEAF complex in vivo. The results of the ChIP-qPCR experiments confirm results from the ChIP-chip experiments suggesting that SpELL is localised near to a region corresponding to primer set D (+301 to +473 nt) (Figure 5-5). This is consistent with the ChIP-chip data, which suggests a bias for SpELL/SpEAF towards the 3' end of *sme2*. SpELL enriches for this region even in cells with the *eaf1* gene deleted, suggesting that SpELL can be recruited to this region of the gene in the absence of SpEAF (Figure 5-7 B). In contrast, SpEAF enrichment at the *sme2* locus depends on SpELL (Figure 5-7 C). These results support a model whereby pol II recruits an SpELL/SpEAF complex to *sme2 via* an interaction with the SpELL subunit.

This model would also predict pol II localisation at the same locus, and so I asked where SpELL/SpEAF was on the gene relative to pol II. The results of Figure 5-6 indicate that like SpELL and SpEAF, pol II enriches for the +301 to +473 region of *sme2*. The characteristics of *sme2* transcripts detected by 3' RACE and Northern blotting do not indicate the 5' ends of transcripts in this region suggesting that the pol II here would be transcriptionally engaged and not, for example, part of a preinitiation or early elongation complex. Induction of sme2 by oxidative stress leads to an increase in SpELL/SpEAF localisation together with increased pol II phosphorylated at serine 2 and serine 5 on the CTD in the transcribed region of *sme2*. Future studies will help to determine whether SpELL/SpEAF localisation is dependent upon or is affected by the state of pol II CTD phosphorylation.

Analysis of SpELL/SpEAF contribution to sme2 regulation.

Although I had originally chosen to use *sme2* as a model to investigate recruitment of the SpELL/SpEAF complex based on the large enrichment of *sme2* sequences in ChIP-chip assays, I considered the possibility that deletion of either the *ell1* or *eaf1* genes could affect either transcription or processing of *sme2*. To assess any changes in polyadenylated and non polyadenylated *sme2* transcripts in the *ell1Δ* and *eaf1Δ* strains, I performed RT-qPCR analysis with reverse transcription primed with either an oligo(dT) primer or with gene specific primers (176). The results suggest a modest decrease in levels of the shorter, polyadenylated transcripts both under conditions where *sme2* is expressed at basal levels or is induced (Figure 5-8). Using an assay which detects both polyadenylated and non-polyadenylated transcripts, I detected no significant difference between levels of *sme2* transcripts in the different strains used under non-inducing conditions, suggesting that the presence of the non-polyadenylated transcripts might reduce any apparent

difference between transcript levels detected in the oligo(dT) primed RT-qPCR assay (Figure 5-9). This would be consistent with the results of Figure 5-2, which indicate similar amounts of +340 (non-polyadenylated) and +438 to +442 (polyadenylated) transcripts for unstressed cells in both mutant and wild type strains. Taken together, the results of Figure 5-2, Figure 5-3, Figure 5-8, and Figure 5-9 indicate:

1. That under conditions where *sme2* is transcribed at basal levels, cells contain relatively low levels of two species of transcripts, a ~340 nt non-polyadenylated transcript and +438 to +442 nt polyadenylated transcripts.
2. When cells are exposed to oxidative stress, there is a significant increase in the levels of the +438 to +442 nt transcripts but not the +340 non polyadenylated transcripts.
3. There are slightly lower levels of the +438 to +442 transcripts in *ell1Δ* and *eaf1Δ* strains compared with levels in the wild type strain under both conditions where *sme2* is expressed at basal levels or is induced.

The changes in *sme2* transcript levels are small but reproducible. There are other genes identified by the global genome studies presented in chapter 4 that have greater changes in transcript levels in the mutant strains, and these may be more suitable to use as a model to investigate the causes of these changes in transcript levels in the mutant strains in the future.

Possible separate functions for SpELL and SpEAF

As *ell1Δ* cells grow less well in media containing the drug 6AU, whereas *eaf1Δ* cells grow normally, it is likely that an SpELL function that does not require SpEAF causes the 6AU growth phenotype. If SpELL recruitment to genes is necessary to fulfil this function, then

consistent with the results of Figure 5-7, SpELL would need to be recruited in the absence of SpEAF. It is not clear whether SpELL/SpEAF recruitment to this locus affects the level of *sme2* transcripts.

A hint that SpEAF might have an independent function comes from the observations of Figure 5-8, which suggest that there might be a greater reduction in *sme2* transcript levels in *eaf1Δ* cells than in *ell1Δ* cells. One possibility is that there is an SpEAF function that does not require the presence of SpELL and would not depend on SpEAF co-localisation with SpELL. The absence of this function might cause the reduced levels of *sme2* transcripts in the *eaf1Δ* strain. Consistent with the idea that SpEAF may have a function that does not involve its interaction with SpELL, it has been reported previously that EAF proteins contain a transactivation domain that is separate from the ELL interaction domain (115).

In conclusion, I have used a model gene, *sme2*, to answer questions about the recruitment of the SpELL/SpEAF complex to genes in cells and propose a model, consistent with *in vitro* data, that the SpELL/SpEAF is recruited *via* an interaction between pol II and the SpELL subunit.

CHAPTER 6. GENERAL DISCUSSION

Prior to this study, the status of ELL family members in lower, single celled eukaryotes had been unclear, with orthologues of ELL not apparent in either of the best characterised fungal model organisms *S. cerevisiae* or *S. pombe*. This had led to speculation that transcription elongation factors of this class may have appeared at a later evolutionary stage and may have evolved for functions specific to higher, multi-cellular organisms, such as the transcription of particularly long genes, or complex developmental programs that require very precise timing of transcriptional activation or repression. In addition, no previous studies had attempted to define the role of ELL globally – whether ELL is used for the regulation of all genes in an organism or is limited to regulating transcription of only a subset of genes. The function of ELL in vivo had also been unclear. Although ELL clearly stimulates transcription elongation in vitro by reducing transient pausing, it had not been known whether ELL functions in this way on genes in cells, whether ELL performs alternative functions, and how ELL function might affect the quantity and nature of RNA transcripts.

The results presented in this thesis have shed light on a number of these questions. First, I have confirmed that a single celled eukaryote, *S. pombe*, uses ELL, confirming that this class of transcription elongation factors did not co-evolve with higher eukaryotes. Furthermore, ELL does not contribute exclusively to the transcription of genes tens of kilobases long (as the average gene in *S. pombe* is 1416 bp with only six being annotated as being greater than 10 kb and none longer than 15 kb). Like their cousins in higher organisms, the SpELL and SpEAF proteins function together as a

complex to stimulate transcription elongation *in vitro*. SpELL and SpEAF are co-enriched on similar regions of chromatin from *S. pombe* cells, which suggests that they colocalise on chromatin. They appear not to localise to all genes, and deletion of the *ell1* and *eaf1* genes is not lethal, arguing that the SpELL/SpEAF complex is not universally required. The exact function of the complex in cells and how the nature of RNA transcripts might be regulated remains unclear although as I will discuss below there are some indications that the SpELL/SpEAF complex might function at a pause site at the *sme2* gene in cells. Future experiments will elucidate whether the sequence of the transcribed region of this gene influences recruitment of the complex to the 3' end of the *sme2* gene.

6.1 Future Directions

6.1.1 The function of SpELL and SpEAF at *sme2*

I have demonstrated that SpELL/SpEAF is recruited to *sme2* and colocalises with pol II near the 3' end of the gene. This colocalisation appears to coincide with a non-polyadenylated *sme2* transcript, which could represent a nascent transcript associated with paused pol II. Intriguingly, near the 3'-end of the non-polyadenylated transcript is a sequence predicted to adopt a hairpin structure, which could influence pausing.

Although my results suggest that the complex is recruited to the gene through the SpELL subunit, it has not been demonstrated whether there is a direct interaction with pol II at the 3' end of *sme2*, or indeed whether pol II is required for recruitment of the SpELL/SpEAF complex. To address this, I propose conducting experiments that would change pol II occupancy of *sme2* so that I could ask whether there would be corresponding changes in SpELL/SpEAF occupancy. Changes in pol II occupancy might be achieved by mutating promoter sequence elements needed for pol II recruitment, such as

the TATA box, which has been mapped previously. Alternatively it might be feasible to block transcription of pol II specifically with a drug such as α-amanitin. If it is established that pol II is required for SpELL/SpEAF complex recruitment, one could then ask whether pol II CTD modifications correlate with SpELL/SpEAF occupancy. The complex appears to colocalise with pol II with serine-2 and/or serine-5 phosphorylated. It may be possible to manipulate phosphorylation of the pol II CTD by mutations in genes involved in CTD kinase pathways such as *csk1*[1], *cdk7* or *cdk8*. In addition to asking whether CTD phosphorylation is important for SpELL/SpEAF localisation in cells, one could ask whether the complex is able to bind CTD repeats *in vitro* and whether phosphorylation of serine residues within CTD repeats would affect binding. This could be tested by first mixing the purified complex with biotinylated CTD peptides (with or without phosphorylated serine 2 and/or serine 5 residues), and then purifying CTD associated complexes with streptavidin coated magnetic beads.

The enrichment of a region near the 3' end of *sme2* by SpELL, SpEAF, and pol II together with the existence of a non-polyadenylated *sme2* transcript ending near this region, raises the intriguing possibility that there is paused pol II here. To investigate this possibility and to determine whether the 340 nt non-polyadenylated transcript results from paused pol II, it may be possible to use a techniques such as permanganate footprinting to test the presence of single stranded DNA representing a transcription bubble and nuclear run-on assays to determine whether 340nt transcripts can continue to

[1] Csk1 is a CDK activating kinase, which can activate Cdk9 (a subunit of P-TEFb), which in turn phosphorylates the pol II CTD at serine 2. Deletion of *csk1* has been reported to result in a 10 fold reduction of pol II CTD serine 2 phosphorylation (179).

be elongated. Previous studies have suggested that pausing at some sites depends, at least in part, in sequences within the RNA transcript upstream of the active site and/or sequences in the DNA template just downstream of the active site. If there is paused polymerase at +340 nt then sequences in the vicinity may be responsible for pausing. If so, mutations in the coding region here might reduce pausing and reveal the importance of any nucleic acid sequences in promoting the pause. This could either involve replacing the whole of the transcribed region of *sme2* with an alternative sequence or mutating smaller regions. There is a particularly promising region just upstream of 340 nt that appears capable of forming a hairpin structure that might promote pausing and so I would propose investigating this region of the transcript. Mutations that would disrupt the hairpin structure might reduce the accumulation of +340 nt transcripts and might affect the accumulation of SpELL, SpEAF and pol II in this region. As well as mutating this region in *S. pombe* cells, complementary experiments could be done using the *sme2* sequence as a DNA template for reconstituted elongation reactions *in vitro*. This would not only determine whether the hairpin serves as a strong pause site, but in addition the transcripts associated with pausing could be characterized and any structures that might cause the pause could be investigated. In studies using E. coli RNA polymerase, Chamberlin and coworkers used nuclease mapping to provide evidence of secondary structure in RNA transcripts from paused complexes (39). Fisher and Yanofsky describe using DNA oligomers complementary to an RNA hairpin associated with a strong pause site to reduce pausing *in vitro* (42), and similar experiments could be done using DNA oligomers complementary to the predicted hairpin in the *sme2* transcript. Experiments using *S.pombe* cells might also elucidate whether sequences within the transcribed region are important for accumulation of pol II, SpELL and SpEAF at a particular location. For

example, one could ask whether transferring the *sme2* transcribed region to a different genomic location with a different promoter would result in recruitment of SpELL, SpEAF and pol II to an alternative location.

Many RNA processing reactions are regulated cotranscriptionally. The location of SpELL and SpEAF near the 3' end of the *sme2* gene raises the possibility that the complex might be involved in processing the 3' end transcript and so it would be interesting to investigate the possibility that *sme2* transcripts are not processed correctly in strains lacking the *ell1* gene. For example correct polyadenylation of *sme2* could be assessed using the assay described by Kusov and co-workers (180) to determine poly(A) tail lengths of specific genes. In this assay, samples of total mRNA are elongated with guanosine residues, cDNA is generated using an oligo(dC9T6) primer, which anneals to the junction between the 3' end of the poly(A) tail and the added guanosines, and the poly(A) tail for a particular gene is amplified using a gene specific sense primer. The range of tail lengths can then be estimated by gel electrophoresis or determined more precisely by cloning and sequencing the pcr products.

Finally, indications that the SpELL/SpEAF complex may be important for transcription of *sme2*, a gene required for meiosis I, raise the possibility that disruption of the SpELL/SpEAF complex may compromise meiosis. It would be interesting to see whether a diploid homozygote *ell1* deletion strain was capable of progressing through meiosis and sporulating. I could also compare levels of *sme2* transcripts in wild type and *ell1Δ* or *eaf1Δ* cells under conditions, such as nitrogen starvation, which are known to stimulate progression into meiosis.

6.1.2 Analysis of genes other than *sme2*

To see whether the SpELL recruitment depends on SpEAF at genes other than *sme2*, I propose extending my ChIP-qPCR analysis to other genes. The candidate genes that are involved in cell separation, for example *ace2* or *eng1*, would be suitable for further investigation by ChIP-qPCR.

Changes in gene expression in the *ell1Δ* and *eaf1Δ* strains detected by expression array analysis were modest. As only a few probes corresponding to each predicted protein coding gene were spotted on the arrays, a more comprehensive picture of changes in levels of all transcribed RNA might be achieved if the experiments were repeated using Affymetrix tiling arrays. Tiling arrays have a large number of probes covering most of the genome with no bias towards predicted ORFs of protein coding genes and, for example, might detect non-coding RNA transcripts or changes in transcript levels corresponding to the 3' end of a gene relative to the 5' end of a gene.

6.1.3 Further investigations into the biochemical mechanisms of an SpELL/SpEAF / pol II complex

Investigating the biochemical mechanisms controlling pol II pausing.

The results of this study suggest that SpELL/SpEAF functions via a direct interaction with pol II. More detailed structural information regarding the interactions between the SpELL/SpEAF complex and the ternary complex might give insight into how

the complex functions in modulating pol II pausing. Indeed, the nature of the interactions between pol II and other factors involved in regulating pol II pausing has helped in developing models to explain how these factors work. In particular, models have been proposed to explain the biochemical mechanisms involved in the control of pol II pausing near promoters by DSIF/NELF, and to explain the control of transcription through arrest sites by SII (TFIIS). During promoter proximal pausing, a subunit of NELF might bind to the pol II clamp domain, and DSIF is thought to bind Rpb7 (181). This would be consistent with the idea that DSIF/NELF complex is positioned near the pol II RNA exit channel. Binding of NELF to the RNA transcript as it emerges from this channel might explain why elongation becomes inhibited by DSIF/NELF once the transcript reaches 20 – 30 nt long (reviewed in (182)). After transcriptional arrest, a region of SII can enter the funnel region of pol II to help position a water molecule at the active site and thus to help to catalyse hydrolysis of the 3' end of an arrested transcript.

Further investigation into ELL-pol II interactions.

The experiment of Figure 3-10 suggests a species specific functional interaction between the ELL/EAF complex and its cognate pol II, when equimolar amounts of *S. pombe* and/or human ELL and/or EAF were used to stimulate transcription elongation *in vitro*; however, it is not clear whether this species specificity is absolute. Despite conservation of structural domains between *S. pombe* and human ELL/EAF proteins, the amino acid sequences are dissimilar (Figure 3-1) and may result in different characteristics between the surfaces of the human and yeast proteins. Nevertheless, it would be interesting to see whether higher concentrations of human ELL or EAF could stimulate transcription by *S. pombe* pol II. The results of Figure 3-10 B lane 8 are consistent with a possible very modest positive regulation of SpELL by human EAF1 and this experiment

could be repeated with increasing concentrations of EAF1. In addition, experiments could be conducted in *S. pombe* cells to test whether overexpression of the human ELL or EAF proteins in *ell1*Δ or *eaf1*Δ cells might rescue the altered patterns of gene expression in the deletion strains.

To assess which pol II subunits SpELL might bind, I propose coexpressing epitope tagged SpELL with individual *S. pombe* pol II subunits (or smaller regions of pol II subunits) using recombinant baculoviruses. I have conducted similar experiments with human ELL, and preliminary results suggest that human ELL interacts with human pol II subunits Rpb1, Rpb2 and Rpb3. Another approach might be to use Far-western blotting, using labeled SpELL or SpEAF to probe for subunits of RNA pol II that have been hybridised to a membrane. More revealing information about SpELL/SpEAF pol II function might come from a more precise knowledge of the complex structure, perhaps using X-ray crystallography or electron microscopy. Such studies would involve collaboration with groups with expert knowledge in these procedures.

In light of the physical interaction between pol II and SpELL/SpEAF *in vitro*, it is interesting that cells expressing reduced levels of the pol II subunit Rpb4, and cells with either *ell1* or *eaf1* deleted, have reduced expression of a number of genes in the Ace2 pathway. This raises the possibility that overexpression of Rpb4 might suppress downregulation of these Ace2 pathway genes in the *ell1*Δ and *eaf1*Δ strains, I propose expressing the *rpb4* gene on a plasmid under the control of the inducible *nmt1* promoter in *ell1*Δ or *eaf1*Δ cells and comparing RNA levels of the Ace2 regulated genes before and after induction of *rpb4*.

ELL/EAF could enhance elongation in part via interactions with RNA.

It remains unclear whether the complex also interacts with nucleic acids. There is some evidence that another transcription elongation factor that reduces transient pausing, Elongin, does not need to interact with RNA to stimulate elongation (183). Binding of a nascent RNA transcript to SpELL and/or SpEAF could be assessed by performing *in vitro* transcription reactions with both radiolabeled CTP and the UV crosslinkable analogue 5-bromo-UTP. This method was used recently to show that the Spt5 subunit of DSIF binds the nascent transcript which was +22 nt long (but not +18 nt long) as it emerged from the pol II exit channel (181). Binding of the complex to RNA in the absence of pol II and DNA could also be assayed, perhaps using a selection of biotinylated RNA oligos, some with sequences that might be predicted to form hairpin structures. Complementary studies *in vivo* using RNA immunoprecipitations might establish whether transcripts from a set of genes are bound to the SpELL/SpEAF complex and if so, whether these genes are those enriched in SpELL/SpEAF ChIPs.

6.1.4 Why are the *ell1* and *eaf1* genes non-essential in *S. pombe?*

Like many yeast genes, *ell1* and *eaf1* are non-essential under the conditions in which cells are grown in the laboratory. They have a function that is not required for haploid cells growing in rich media at optimal temperatures. Perhaps the SpELL/SpEAF complex is essential for meiosis (*sme2* which is highly enriched by SpELL and SpEAF in chromatin immunoprecipitations is essential for meiosis I) and it would be straightforward to test the ability of homozygous mutant diploids to sporulate. Homozygote diploids can be generated by mixing cells from haploid cells of opposite mating type and incubating on

malt agar. Diploids from wild type cells generated in this way usually sporulate with complete asci after incubation for 3 days at 25°C, and sporulation can be visualized by exposing plate cultures to iodine vapour (184). A reduction in the percentage of viable spores in mutant cells might indicate defects in meiosis (185), and, if mutant cells are able to soprulate, the viability of the spores can be analysed. To do this, spores need to be isolated from the mating mixture, which contains both asci (containing spores) and vegetative cells. Digestion with glusulase removes the walls of the asci, which releases spores. Glusulase also digests the walls of vegetative cells, killing them. Dilutions of spores can then be plated and incubated under conditions where viable spores will generate colonies of cells. The percentage of viable spores can be calculated from estimates of the number of spores plated and the number of colonies formed (185).

Additional experiments might investigate expression of different classes of genes involved in meiosis in mutant cells grown in conditions propitious for the expression of such genes. Wild type and *ell1Δ* or *eaf1Δ* cells could be grown in minimal media lacking a nitrogen source. RNA prepared from these cells could be analysed either by RT-qPCR using primers corresponding to a selection of meiotic genes or by reverse transcription followed by hybridisation to expression arrays.

SpELL and SpEAF also appear to be enriched at, and modulate the transcript levels of, a number of genes involved in cell separation after cytokinesis. It has been proposed that regulation of these Ace2 pathway genes might have evolved to control whether filamentous fungi grow as single cells or filamentous hyphae in response to environmental stimuli (186). Switching from single cellular growth to hyphal growth may be a mechanism that fission yeast use to forage for nutrients, and downregulating cell separation might be a mechanism used to promote the formation of filamentous structures. Expression of the

Ace2 pathway genes changes during the cell cycle (186). So far I have analysed RNA levels of Ace2 pathway genes in unsynchronised cells growing in rich media. I might expect to see greater variation in levels of Ace2 pathway genes between wild type and mutant cells for populations of cells that are synchronised at a stage in the cell cycle where these genes are normally expressed. I propose synchronising cells by centrifugal elutriation (187) and allowing the synchronised population of cells to progress through the cell cycle. Expression of Ace2 pathway genes in synchronised cells could then be measured by analysing RNA from samples of cells taken at different times during progression through the cell cycle by RT-qPCR.

Alternatively, SpELL and SpEAF might fulfil redundant roles. It is possible that other proteins or complexes take over these roles in the absence of SpELL/SpEAF. To investigate this possibility I suggest disrupting the *ell1* or *eaf1* genes in the *S. pombe* deletion strain library to look for genetic interactions between *ell1*/*eaf1* and other genes. Roguev and Krogan have developed a high-throughput platform to screen for genetic interactions between genes of interest (188). This technique could be used to identify both negative interactions (for example pairs of deleted genes that result in a sick or lethal phenotypes), and positive interactions (for example where deletion of *ell1*/*eaf1* suppresses a sick phenotype caused by deletion of another gene, or where the 6-AU sensitivity of the *ell1*Δ cells is suppressed by deletion of a second gene). If I find a meiotic or sporulation phenotype in the *ell1*Δ cells, I could also look for genes that when disrupted suppress or enhance that phenotype.

Appendix A: *S. pombe* Strains used in this study

Strain (parental)	Genotype
972	h-
ell1Δ (972)	ell1::kanMX6 h-
PP138	ade6-M216 leu1-32 ura4-D18 his3-D1 h-
PP137	ade6-M216 leu1-32 ura4-D18 his3-D1 h+
ell1Δ (PP138)	ell1::kanMX6 ade6-M216 leu1-32 ura4-D18 his3-D1
ell1Δ (PP137)	ell1::kanMX6 ade6-M216 leu1-32 ura4-D18 his3-D1
eaf1Δ (PP138)	eaf1::kanMX6 ade6-M216 leu1-32 ura4-D18 his3-D1 h-
ell1-myc (PP138)	ell1::ell1 13myc/natMX6 ade6-M216 leu1-32 ura4-D18 his3-D1 h-
ell1-myc (PP137)	ell1::ell1 13myc/natMX6 ade6-M216 leu1-32 ura4-
eaf1-myc (PP138)	eaf1::eaf1 13myc/natMX6 ade6-M216 leu1-32
ell1-myc eaf1Δ (PP137/8)	ell1::ell1 13myc/natMX6 eaf1::kanMX6 ade6-M216 leu1-32 ura4-D18 his3-D1
eaf1-myc ell1Δ (PP137/8)	eaf1::eaf1 13myc/natMX6 ell1::kanMX6 ade6-M216 leu1-32 ura4-D18 his3-D1
rpb3-myc (PP138)	rpb3::rpb3 13myc/natMX6 ade6-M216 leu1-32
tbp1-myc (PP138)	tbp1::tbp1 13myc/natMX6 ade6-M216 leu1-32

pof4-myc (PP137)	pof4::pof4 13myc/natMX6 ade6-M216 leu1-32
ell1 rescue (PP138)	ell1::kanMX6 ade6-M216 leu1-32::ell1/leu1ura4-
CBS354/OY1*	
Sj *ell1Δ* (CBS354/OY1) *	Sj ell1::kanMX6

*S. japonicus

Appendix B: Primers Used in this Study

sme2 ChIP and RT / qPCR analysis (Chapter 5.)

Primer Name	Sequence
sme2 -1407 fwd (A)	TAC CAC GGT ACA ATG GTC GGA GAT
sme2 -1326 rev (A)	CCC ATG GGA GCA TTA TCG GTT GAT
sme2 -530 fwd (B)	CGG AGA ACG CAA GCT TGA AAT GGA
sme2 -389 rev (B)	TGG TAC CGT ATT GTG AGC GA
sme2 +26 fwd (C)	TGT TGG TCA ATC TTC TGC CGT CTT G
sme2 +105 rev (C)	GCG ATC TTG CAT GCA TAT TCC GTC
sme2 +301 fwd (D)	CCC AAG TTG GTT TAT GTG AGC CTT GTC
sme2 +473 rev (D)	GTG CTT TCA AGG ATA ACA ATG CAG CC
sme2 +640 fwd (E)	ACC CGC AGT TTG TTT AAA CGC T
sme2 +738 rev (E)	GGG TTT AAC GAA TGG TTT CAG CAC G
sme2 fwd +717 (F)	GCT GAA ACC ATT CGT TAA ACC CTG C
sme2 rev +835 (F)	CAG CAC AAC CGA AGA CCA ATG CAA
sme2 fwd +1210 (G)	TGG TTG GAC TTT GCC GAT TTC ACG
sme2 rev +1363 (G)	ATG CTG GGA TCT GTC TGT TCT GCT

sme2 fwd 1839 (H)	TTG TTG GTA CGA TGC CCG ACA AGA
sme2 rev 1928 (H)	CGC CGG AAG CAA TAC CAT TTA AGC
K region fwd	GTT GCG CAA GCG AAG TTA TGG AGT
K region rev	ATA CGG TGC TTG GGC TTA GTC CTT

S. pombe strain construction (Chapter 3. , Chapter 4. and Chapter 5.)

Primer Name	Sequence
ell1Δ MX6 fwd	CTT AAA CAA GCT TCT CCG TTG TGC CAT CTA GCT AAT ATA ATC ATT TTG AGA GGC TTT TAC TAT CGA TCT ATT TGG GTT GAC GGA TCC CCG GGT TAA TTA A
ell1:tag MX6 fwd	AAG ATG AAC TTA TAT CTC TAC ATT CTC AGC TCA AAA GCT GGA AAA ATA CAC TTT ACG ATG CTT CCT CGG AGC TAG CCC TCC GGA TCC CCG GGT TAA TTA A
ell1 MX6 rev	TTA AAC GAA ACA CTA GCC TTA ACG AAA GCA GGA AAA AAC AGC CTG CAC TCG TTG TAG GCA CCG TAC AGT ATT GAA ACA ATG AAT TCG AGC TCG TTT AAA C
eaf1Δ MX6 fwd	TCT TAC CTT ACG TTA TTT ATT TGA TTT ATA TCG AAA TTT CCA ATT TCG TAC AGG CCT GAC TTT TAC CAT TAT AAA CAA TCC GGA TCC CCG GGT TAA TTA A

eaf1:tag MX6 fwd	GGG GTC TAT CTT CGC AAG AGA GGG ATT ATG CTT CTT CTG CTC AGG CAG AGG GTA TCA GCA GCG CTT CCG AGG ATG AGG ATC GGA TCC CCG GGT TAA TTA A
eaf1 MX6 rev	AAA ATG ACA CAG AGT GAA ACG AGT TTA TTA CCG CCG ATT CGC TGT CGA GTA GTG ATG GAA AAG ACT TGT CTT GCT TTA CTG AAT TCG AGC TCG TTT AAA C
rpb3:tag MX6 fwd	TTT CTG CTA ACG AAC TGA ATA TGG AGG AGA ATG CAG AAA TGA ACT GGT CTC CCT ATC AAA ATG GTG AAG AAA ACA CGT GGC GGA TCC CCG GGT TAA TTA A
rpb3 MX6 rev	ATC TAA AGC TAT AAT AAA TCC AAG TGA AGA AGA ACA AAA TTA TGA TAT GAG GGA GGA ATA TAT ATC TCT ATA TAT GCA TAG AAT TCG AGC TCG TTT AAA C
tbp1:tag MX6 fwd	TAA CTG GTG CGA AAG TCC GTG AGG AAA TTT ACC AAG CTT TTG AAG CCA TTT ATC CAG TAT TGT CTG AAT TTC GAA AAC ATC GGA TCC CCG GGT TAA TTA A
tbp1 MX6 rev	AGG ATG TCT CTA TCA TCT GCA TCA TAT AAA AAT ACT CAT AAA TGA TGA CTT ACA GTG AAG GGG ATT CTC CGG AAA GCT TTG AAT TCG AGC TCG TTT AAA C
pof4Δ MX6 fwd	GCA CTC TTC CAA TTG AAG ACT ATA GAG CTC TAA TTC GAT CGT TTT CTT TGA AAT TTT TAA AAA TTA CTA TCA TAT GAG CGC GGA TCC CCG GGT TAA TTA A
pof4 MX6 rev	CCT CGA AGG ATT TGG AAA AAA GGA CAT GGG AAA AGA CGT AGG GCG AGA GAT TGA TGG AGA GGA GGA AGA TGA AGG TTG AGG AAT TCG AGC TCG TTT AAA C

Sj ell1Δ MX6 fwd	TCT GCA TCC TTC AAA CTT CTC GTT GCG TTT GCT GTT TGA CCT GAC GTT GAA TTT GTC TGT GCT CAA GTG CTG CTG CTG GAC GGA TCC CCG GGT TAA TTA A
Sj ell1 MX6 rev	AAG AGT AAG AAA AAG GGG ACT AGG TGG TGG ACG AAT AGC AGA ACG AAT GGA CAA ACC GTC CAA GGG GCG GAG AAG GAA AGG AAT TCG AGC TCG TTT AAA C

PCR confirmation of strains (Chapter 3. , Chapter 4. and Chapter 5.)

Primer Name	Sequence
pFA6A PTEF rev	CTC CAT GTC GCT GGC CGG GTG AC
ell1 coding fwd	CCC AAA CCT TCC CGT CTC GCA ACC GTC TGC ATC ACC
ell1 coding rev	GAG GGC TAG CTC CGA GGA AGC ATC GTA AAG TG
ell1 upstream fwd	CTG GTC CCG TAA CCC TAT TCC GCC GC
eaf1 coding fwd	GAT CTC GAG ATG AAT TCA TTA CAG AAG GGA TCA TAC AAA GTT ATT CCC GGG TCC AGC
eaf1 coding rev	CAT GGT ACC TTA ATC CTC ATC CTC GGA AGC GCT GCT GAT ACC CTC
eaf1 upstream fwd	CAG GCT AGC GTT GTC AAT AGC AGC AAT AAA TCT AAC AGT CAT CAT ACT C
rpb3 fwd	ACA CCC TAT TCT CGC CGA TCC AAA
tbp1 fwd	ATG CGT ATC CGT GAA CCC AAG TCT

pof4 upstream fwd	ACT CGC ACT TAG ATT CTC CGC GTT
pof4 coding fwd	CAG CTC GAG ATG TAT TCT TTA AAA GAT TTG TGC ATC CAA GTC GCG
pof4 coding rev	CAC GGT ACC TTA AAT TCG TTT AGG AAC TTG ACT AGA AAA CCT CGA AGG
Sj ell1 coding fwd	AGT GGG AGC TTC GTC CTG AAA TGT
Sj ell1 coding rev	TGC ATC AGG CGG AAG AGA AAG AGA
Sj ell1 upstream fwd	TTA TGC ACC ACC GAC AGT CTC CTT

Primers used to confirm changes in RNA levels in *ell1Δ* and *eaf1Δ* strains by RT / qPCR (Chapter 4.)

Primer Name	Sequence
abp2 fwd	AGA GCG TCC CAT CAC ATG TCT GAA
abp2 rev	ATG TGA ATT GTT GGC TGC GGA AGG
ace2 fwd	ATG GCT AGC TTA TAT TGC GGC CCT
ace2 rev	CCA GCT TTG CAT AAG TCA CAG CGA
act1 fwd	ACT GCT GCT CAA TCT TCC TCC CTT
act1 rev	ACT GCT GCT CAA TCT TCC TCC CTT
adg1 fwd	TCA AGC TCT GAA GTC ACC TCC CAA

adg1 rev	ACC GGC AGT CTC AGT AAC AGT GAA
adg3 fwd	GAT GGT GTG GCA GTC AAG CCA TTT
adg3 rev	ATT ACC CGG GAG AAC CGT TTG ACA
agn1 fwd	GCA CTG GAA GAA AGC GTT GCA GAT
agn1 rev	AGG CGA GAC TGG TGC AAC ATA CAA
ats1 fwd	GTT CAG TTC GCA TTC GCT CTG TCA
ats1 rev	ATG CCA CAG GAG TAA CCT CGT CAA
cdc45 fwd	CGC CTT GGA CGA TGT CGA TTC ATT
cdc45 rev	GGT GAT TCC AGT ACG AAC AAT GGC
chr1 fwd	TTG GCC AAT GAC AAG GGT AAT GCG
chr1 rev	TTC CCG AGG ACA TAA CAA GCA GGA
eng1 fwd	TTC GCA ATA TGG CTG CGA CAA TGG
eng1 rev	ATG CCA GTT GAC TGT GCG GTA GTA
fio1 fwd	AGC AGT GCA TGT TGG ATG GTG TTC
fio1 rev	ACA GGC TGA AAT CAC ACA TGC TGC
gpd1 fwd	AAG TTG CCC GTG AGC AAT TCT GTG
gpd1 rev	AAC GTC GTC AAC GGA GAC AAC TGA
mak10 fwd	GAC GGT GCT TTG TAT GCG GTT GAA

mak10 rev	AGC TTC CGA ACG AAG TAG TGG CTT
mei2 fwd	ATT CAG TTA GCA CCT CTT CGG CCT
mei2 rev	TAG CAG CAG TGG GAG TTT CTT GGT
mfm2 fwd	GCC TAC AAC AAC AAT CCT ACC GAT G
mfm2 rev	ACA CAC ATA TAA GGA ACC TTT GGA GT
mid2 fwd	ATC CAC CTA AGG TTC GCA CTC GTT
mid2 rev	AGT GCG GTT CGG CTT ACA GAA TCA
pex7 fwd	ATT AGG CGG AAG CGG ACG TCT TTA
pex7 rev	AAA GAA CCA TCA CCA CAG CAA GCG
pho1 fwd	GCT TTC GAT GCT GTT GGA ATC GCT
pho1 rev	AAT AAC TCG ACA CGG CCA CTG CTA
SPAPB1A10.14 fwd	ACC TTG TCT CGT TCT TCC TGC GAT
SPAPB1A10.14 rev	TCC AGG AGC AGC ATC AAC TCT TGT
SPAC1002.19 fwd	TGG TAT TGA GGA AGC TGC CAA GGA
SPAC1002.19 rev	AGT CAC CAG TAC GCT TAC GCT TGT
SPAC1039.02 fwd	ATG GTT GAC ACC GGT GAT TTG CAC
SPAC1039.02 rev	AAT GGC GGC TTG GTA CAA TTC GTG
SPAC15E1.02c fwd	AGT CGT ACT CGA TCG CTG CAT TGT

SPAC15E1.02c rev	CGG GCA ATC CAG ATG CAC CAA ATA
SPAC27D7.09c fwd	CAC TGT TGA AGG CAG CAA TGG TGT
SPAC27D7.09c rev	ATC GCA ACC AGA GCT AGA GCA GTT
SPAC27D7.10c fwd	AAT GGC ACT GGC CGA TTC ATT CTC
SPAC27D7.10c rev	ATC GCA ACC AGA GCT AGA GCA GTT
SPAC2E1P3.05c fwd	TCT AGT GCT GCT TCT TCC ACC ACA
SPAC2E1P3.05c rev	ACT GCT TGT CGA CAC ACT GGA AGA
SPAC2H10.01 fwd	TTC ACC TTC TCT TGA CGC AGA CGA
SPAC2H10.01 rev	GCA ATT GGC TTC CAC TTC CTT GGT
SPBC1289.14 fwd	CCA TAT GCT GCT GGC GTT TCA TGT
SPBC1289.14 rev	TGT GAC AGG ACC AGA TTG GTG TGA
SPBC1861.01c (cnp3) fwd	ACG CGA AGA AGA TGT CCA ACG AGA
SPBC1861.01c (cnp3) rev	TAC GTT TCG ACC TTC GGA CAC CAA
SPBC1E8.04 fwd	CGC TGG AAA TGG ACA CCA ACA CAA
SPBC1E8.04 rev	ATC AAA GTG TCG TAG CAC CGG AGA
SPBC215.06c fwd	AGC TGT ATG ACA GAA GCT CAA CGC
SPBC215.06c rev	TCG TCA AGC GAG TGT TTA GTG GGA
SPBC8E4.01c fwd	ACT GGT ACG AAC GAA TAC CGC ACA

SPBC8E4.01c rev	TAT CCA GGA ACA TAG CCG GCA ACA
SPBPB10D8.01 fwd	AAA CCA AAG TGC CTT TCA CGG AGC
SPBPB10D8.01 rev	TGT GTA TCC AGC ACA AGA GGC TGT
SPBPB7E8.01 fwd	TAC TGG CGA GTC TGT TTG TTC CCA
SPBPB7E8.01 rev	ACT GCA AGC CAT GGG AAT TTC ACG
SPBPB7E8.02 fwd	CAA GCA ACG CAT TGA CTT CCG TGA
SPBPB7E8.02 rev	TGT ACC ACT GAC AGC ACA CAT CCA
str1 fwd	TGT CAC AGC CCT TGT TAG TAC CGT
str1 rev	TTT CTA AAC GCT TGG GCA TCA CGC
str3 fwd	GAC CAA ATT TCG CGG CTA CAG GAA
str3 rev	ACC ACA CTG AAG GAA CCA CCC ATA
tlh1/2 fwd	GGC GCA CTA TGG ACA ACT TGC ATT
tlh1/2 rev	TGG GTG CTT TGT TTC CAC CAA CAG

Appendix C: Gene Lists

Genes identified by microarray analysis with a greater than 1.40 fold change in RNA levels in *ell1Δ* or *eaf1Δ* strains (Figure 4-1)

Gene	ell1Δ / wt (RT / qPCR*)	eaf1Δ / wt (RT / qPCR*)	Product[178]
abp2	0.59	0.44	ARS binding protein Abp2; phenotype aberrant septa
ace2	0.60	0.62	transcription factor Ace2; transcriptionally regulates cell separation during cytokinesis
adg1	0.67	0.67	sequence orphan; cell surface glycoprotein transcriptionally regulated by Ace2
adg3	0.67	0.71	beta-glucosidase Adg3 (predicted) ; cell surface glycoprotein transcriptionally regulated by Ace2
agn1	0.69	0.56	glucan endo-1,3-alpha-glucosidase Agn1; transcriptionally regulated by Ace2
ats1	1.25	1.33	N-acetyltransferase Ats1 (predicted)
cdc45	0.95	0.95	DNA replication pre-initiation complex subunit Cdc45 (Cell division control protein 45)
chr1	0.73	0.71	chitin synthase regulatory factor Chr1 (predicted); transcriptionally regulated by Ace2
eng1	0.65	0.56	endo-1,3-beta-glucanase Eng1; transcriptionally regulated by Ace2
fio1	1.68	1.74	iron transport multicopper oxidase Fio1
gpd1	0.70	0.68	glycerol-3-phosphate dehydrogenase Gpd1
mak10	0.99	0.92	NatC N-acetyltransferase complex subunit Mak10 (predicted)
mei2	0.41	0.31	RNA-binding protein involved in meiosis Mei2
mfm2	0.67	0.63	M-factor precursor Mfm2 (M-factor is a mating pheromone produced by M-type mating cells.)
mid2	0.85	0.81	anillin homologue Mid2; septin ring organizing protein transcriptionally regulated by Ace2
pex7	0.52	1.10	peroxin-7 (predicted); peroxisomal targeting signal 2 receptor
pho1	0.44	0.39	secreted acid phosphatase Pho1
SPAPB1A10.14	0.67	0.92	F-box protein
SPAC1002.19	0.75	0.85	GTP cyclohydrolase II (predicted)
SPAC1039.02	0.67	0.81	phosphoprotein phosphatase (predicted)
SPAC15E1.02c	1.75	1.50	DUF1761 family protein
SPAC27D7.09c	1.79	1.81	But2 family protein (Ubiquitin 3 binding protein But2)
SPAC27D7.10c	1.03	1.25	But2 family protein
SPAC2E1P3.05c	0.44	0.37	fungal cellulose binding domain protein

* mean value of three independent experiments

SPAC2H10.01	0.34	0.41	transcription factor, zf-fungal binuclear cluster type (predicted). Adjacent to adg1
SPBC1289.14	2.72	2.07	adducin
SPBC1861.01c (cnp3)	0.59	0.49	centromere protein CENP-C
SPBC1E8.04	1.39	1.30	pseudogene
SPBC215.06c	0.92	0.90	human LYHRT homolog (zinc finger protein)
SPBC8E4.01c	0.55	0.53	inorganic phosphate transporter (predicted); adjacent to pho1
SPBPB10D8.01	0.49	0.80	cysteine transporter (predicted)
SPBPB7E8.01	0.55	0.80	sequence orphan
SPBPB7E8.02	0.95	0.91	PSP1 family protein
str1	1.35	1.81	siderophore-iron transporter Str1
str3	3.34	3.37	siderophore-iron transporter Str3
tlh1	2.67	8.96	RecQ type DNA helicase (sub-telomeric helicase)

SpELL enriched genes (identified as described in section 4.5.1)

aar2	cox18	hht2	mmm1	ptn1	rps601	SPAC13G6.06c	SPAC24B11.09
ace2	cox5	hht3	msa1	pub1	rps602	SPAC13G6.10c	SPAC24B11.12c
adg1	cpd1	his2	msh6	pvg1	rps801	SPAC13G6.13	SPAC24B11.13
adg3	cpy1	his3	mug142	pvg3	rrn11	SPAC144.12	SPAC24H6.13
adh1	csn2	hmg1	mug5	pvg5	rrn5	SPAC144.18	SPAC25B8.03
adh4	csn4	hmt2	mug80	pyk1	rrn7	SPAC1565.07c	SPAC25B8.11
adk1	csx1	hob1	myh1	pyp1	rsc58	SPAC1687.09	SPAC25B8.12c
agn1	ctr4	hsf1	nda3	pyr1	rsd1	SPAC16E8.01	SPAC25B8.19c
ags1	cut1	hsp10	ned1	pzh1	rst2	SPAC1705.03c	SPAC25G10.09c
alg2	cwf22	hsp90	nep2	qcr1	rsv2	SPAC1782.05	SPAC26F1.05
alg3	cyr1	hst4	nif1	qcr2	sap1	SPAC1782.06c	SPAC26F1.07
amt1	cys2	hta1	nnt1	qcr7	sat1	SPAC1786.02	SPAC26F1.08c
amt2	dak1	hta2	nop12	qcr8	sce3	SPAC17A2.05	SPAC26H5.03
anc1	dbp2	htb1	nrd1	rad26	scr1	SPAC17A2.12	SPAC26H5.04
ape1	dbp3	hxk2	ofd1	rad8	scw1	SPAC17A5.05c	SPAC26H5.09c
app1	dfp1	idh1	ogm1	ran1	sdh2	SPAC17A5.09c	SPAC27D7.11c
aps1	dic1	idp1	orc4	rav1	sdh4	SPAC17A5.10	SPAC27E2.04c
apt1	dld1	ipk1	p23fy	rcd1	seb1	SPAC17G6.11c	SPAC27E2.11c
arc1	drs1	isp4	pac2	rec10	sec10	SPAC17G6.18	SPAC27E2.13
arc15	eca39	isp6	pap1	rep2	sec26	SPAC17G8.07	SPAC27F1.07
arg12	efl1a-c	isp7	pas1	res1	sec31	SPAC17G8.08c	SPAC29A4.03c
arg5	efc25	isu1	pch1	rga3	sec62	SPAC17H9.11	SPAC29A4.09
aro1	egd2	its3	pex13	rga7	sec74	SPAC17H9.12c	SPAC29B12.05c
arp9	elf1	iws1	pgk1	rgf2	sen1	SPAC1805.16c	SPAC2C4.17c
atf1	ell1	kap123	php2	rhp23	set1	SPAC18G6.09c	SPAC2E1P3.04
atl1	eng1	klp2	phx1	rpa2	set8	SPAC19B12.02c	SPAC2E1P3.05c
atp2	eno101	ksp1	pka1	rpb1	sfi1	SPAC19G12.05	SPAC2F3.09
aur1	erg1	leu2	pla1	rpb9	sfp1	SPAC19G12.09	SPAC2F3.10
bem46	erg12	leu3	plb1	rpc1	sgo1	SPAC19G12.16c	SPAC2F7.02c
bfr1	erg7	lkh1	plr1	rpc40	sgt1	SPAC1A6.10	SPAC2F7.05c
bgl2	esc1	lsk1	pma1	rpc53	shk2	SPAC1B3.02c	SPAC30.10c
bgs1	esf2	lys1	pma2	rpl1001	sid2	SPAC1B3.04c	SPAC30.11
bgs2	exg1	lys3	pmp31	rpl101	sir2	SPAC1F12.03c	SPAC30.13
bgs3	fba1	lys4	pmt3	rpl1101	sks2	SPAC1F12.04c	SPAC30.14c
bip1	fep1	mac1	pof14	rpl15	slp1	SPAC1F5.02	SPAC30D11.14c
bir1	fkh1	mad3	pof4	rpl1603	slu7	SPAC1F7.09c	SPAC31A2.04c
bpb1	fma2	mae2	pof8	rpl2101	smb1	SPAC1F8.07c	SPAC31G5.04
brf1	fmn1	maf1	pol1	rpl24	snf30	SPAC1F8.08	SPAC31G5.05c
btf3	for3	mag1	pop3	rpl26	SPAC1002.18	SPAC20G8.07c	SPAC31G5.15
but2	fur4	mal3	ppi1	rpl302	SPAC1002.20	SPAC21E11.07	SPAC31G5.21
cbh1	gcv1	mam3	ppk1	rpl3703	SPAC1006.07	SPAC222.13c	SPAC328.04
ccs1	ggt1	mbx1	ppk11	rpl401	SPAC1039.02	SPAC227.17c	SPAC3A11.07
cct1	ght2	mcp60	ppk32	rpl402	SPAC1093.02	SPAC22A12.06c	SPAC3A12.06c
cda1	ght8	mcs4	ppk4	rpl4301	SPAC1093.03	SPAC22A12.16	SPAC3A12.19
cdc17	gln1	mde3	prl12	rpl4302	SPAC10F6.03c	SPAC22F8.08	SPAC3F10.03
cdc18	gpd1	mde7	prl13	rpl502	SPAC10F6.04	SPAC23A1.02c	SPAC3H1.05
cdc22	gpd2	met26	prl32	rpp203	SPAC10F6.07c	SPAC23A1.14c	SPAC3H1.06c
cdc25	gpm1	meu14	prl39	rps002	SPAC1142.06	SPAC23A1.17	SPAC3H1.07
cdc48	gpt2	meu15	prl42	rps101	SPAC11D3.13	SPAC23C11.01	SPAC3H1.08c
cdc5	gsa1	mfm1	prl52	rps1401	SPAC11D3.14c	SPAC23C11.03	SPAC4C5.05c
cdr2	gst2	mfm3	prl53	rps1502	SPAC11E3.13c	SPAC23C11.05	SPAC4F10.10c
cek1	hal4	mgr2	prp1	rps2201	SPAC11G7.01	SPAC23D3.12	SPAC521.03
cfh4	hhf1	mid2	prp22	rps23	SPAC1250.07	SPAC23H3.09c	SPAC521.04c
clc1	hhf2	mis4	prr1	rps2601	SPAC12B10.18	SPAC23H3.11c	SPAC56E4.03
cnx1	hhf3	mlo3	psi1	rps3002	SPAC13F5.05	SPAC23H3.12c	SPAC56F8.03
cox17	hht1	mmf1	psu1	rps403	SPAC13G6.05c	SPAC24B11.07c	SPAC56F8.12

SPAC57A7.02c
SPAC57A7.04c
SPAC57A7.05
SPAC5H10.03
SPAC5H10.04
SPAC5H10.12c
SPAC630.08c
SPAC630.15
SPAC637.06
SPAC683.02c
SPAC694.02
SPAC6C3.02c
SPAC6C3.08
SPAC6C3.09
SPAC6F12.04
SPAC6G10.03c
SPAC6G10.07
SPAC6G10.09
SPAC6G10.10c
SPAC6G9.03c
SPAC6G9.14
SPAC7D4.08
SPAC7D4.12c
SPAC806.02c
SPAC824.04
SPAC8C9.04
SPAC8E11.09c
SPAC8E11.10
SPAC9.07c
SPAC9.08c
SPAC959.04c
SPAC959.06c
SPAC977.13c
SPAC977.14c
SPAC9E9.01
SPAC9E9.05
SPAC9E9.06c
SPAC9E9.09c
SPAC9G1.10c
SPACUNK4.10
SPACUNK4.15
SPACUNK4.16c
SPAP11E10.01
SPAPB15E9.02c
SPAPB17E12.06
SPAPB18E9.06c
SPAPB1A10.05
SPAPB1A10.06c
SPAPB1A10.08
SPAPB1A10.14
SPAPB1A10.15
SPAPB1E7.04c
SPAPB1E7.07
SPAPB24D3.03
SPAPB2B4.04c
SPBC119.05c
SPBC119.10
SPBC1198.06c
SPBC11B10.03
SPBC11B10.04c
SPBC11C11.05
SPBC11C11.06c
SPBC11C11.10
SPBC11G11.01
SPBC1271.13
SPBC1271.14
SPBC1289.06c
SPBC12C2.03c
SPBC1347.11
SPBC1347.13c
SPBC1347.14c
SPBC13A2.02
SPBC13A2.04c
SPBC13E7.03c
SPBC13G1.09
SPBC14C8.02
SPBC14C8.04
SPBC14F5.10c
SPBC14F5.11c
SPBC1539.07c
SPBC15D4.02
SPBC15D4.05
SPBC15D4.08c
SPBC15D4.09c
SPBC1685.13
SPBC16A3.02c
SPBC16A3.19
SPBC16C6.04
SPBC16C6.05
SPBC16D10.08c
SPBC16E9.16c
SPBC1703.13c
SPBC1711.04
SPBC1711.05
SPBC1711.08
SPBC1711.09c
SPBC1711.12
SPBC1734.04
SPBC1734.11
SPBC18E5.07
SPBC19C2.08
SPBC19C7.04c
SPBC19F5.04
SPBC19G7.04
SPBC19G7.07c
SPBC19G7.17
SPBC1A4.04
SPBC1A4.05
SPBC1A4.09
SPBC1E8.05
SPBC215.13
SPBC21C3.07c
SPBC21C3.08c
SPBC21H7.06c
SPBC23G7.07c
SPBC24C6.04
SPBC25B2.03
SPBC25H2.06c
SPBC27.05
SPBC27B12.12c
SPBC29A10.08
SPBC29A3.09c
SPBC29A3.19
SPBC29A3.21
SPBC29B5.04c
SPBC2A9.04c
SPBC2D10.04
SPBC2F12.02c
SPBC2F12.03c
SPBC2G5.05
SPBC30B4.02c
SPBC30B4.03c
SPBC30B4.10
SPBC30D10.17c
SPBC31A8.01c
SPBC32F12.07c
SPBC32F12.10
SPBC32F12.12c
SPBC32F12.13c
SPBC32F12.15
SPBC336.11
SPBC337.02c
SPBC337.07c
SPBC36.02c
SPBC365.04c
SPBC365.16
SPBC3B8.03
SPBC3B8.09
SPBC3B9.01
SPBC3D6.01
SPBC3D6.16
SPBC3E7.11c
SPBC3H7.02
SPBC3H7.03c
SPBC4.03c
SPBC428.11
SPBC428.12c
SPBC428.14
SPBC428.15
SPBC4C3.01
SPBC4C3.03
SPBC4F6.17c
SPBC530.01
SPBC530.02
SPBC530.08
SPBC56F2.08c
SPBC56F2.12
SPBC577.11
SPBC609.03
SPBC646.15c
SPBC646.16
SPBC660.06
SPBC660.10
SPBC6B1.12c
SPBC713.06
SPBC713.07c
SPBC713.13c
SPBC725.10
SPBC83.17
SPBC839.16
SPBC887.15c
SPBC8D2.15
SPBC8D2.16c
SPBC8D2.18c
SPBC8E4.01c
SPBC8E4.02c
SPBC947.01
SPBP16F5.08c
SPBP18G5.02
SPBP18G5.03
SPBP23A10.09
SPBP23A10.11c
SPBP23A10.12
SPBP35G2.02
SPBP35G2.14
SPBP4H10.14c
SPBP4H10.15
SPBP8B7.15c
SPBP8B7.17c
SPBPB2B2.05
SPBPB2B2.06c
SPBPB7E8.01
SPBPB7E8.02
SPBPB7E8.03
SPBPJ4664.02
SPBTRNAHIS.02
SPBTRNAPRO.06
SPCC1020.08
SPCC1020.09
SPCC1183.07
SPCC1259.02c
SPCC1259.08
SPCC126.07c
SPCC126.08c
SPCC1281.03c
SPCC1281.06c
SPCC1322.02
SPCC1322.03
SPCC1322.04
SPCC1322.05c
SPCC1322.10
SPCC1322.16
SPCC1393.02c
SPCC1393.08
SPCC1442.12
SPCC1442.13c
SPCC1442.16c
SPCC1450.09c
SPCC162.01c
SPCC162.02c
SPCC1672.01
SPCC16C4.02c
SPCC1739.10
SPCC1795.10c
SPCC18.02
SPCC18.03
SPCC18.17c
SPCC1840.07c
SPCC188.10c
SPCC1906.05
SPCC191.05c
SPCC297.05
SPCC297.06c
SPCC306.08c
SPCC306.11
SPCC320.06
SPCC330.06c
SPCC417.15
SPCC4B3.02c
SPCC4B3.07
SPCC4B3.08
SPCC4F11.03c
SPCC4G3.01
SPCC4G3.13c
SPCC4G3.14
SPCC4G3.16
SPCC4G3.17
SPCC550.07
SPCC550.08
SPCC550.09
SPCC550.11
SPCC550.14
SPCC553.10
SPCC553.12c
SPCC576.04
SPCC576.05
SPCC576.12c
SPCC576.13
SPCC584.01c
SPCC594.01
SPCC594.02c
SPCC5E4.05c
SPCC5E4.05c
SPCC5E4.10c
SPCC622.12c
SPCC63.10c
SPCC63.14
SPCC645.09
SPCC645.10
SPCC70.10
SPCC736.15
SPCC757.12
SPCC794.15
SPCC830.10
SPCC962.05
SPCC965.13
SPCC970.06
SPCC970.08
SPCP31B10.06
SPCPB1C11.03
SPCPB1C11.04c
SPCTRNASER.07
spf30
SPNCRNA.103
SPNCRNA.108
SPNCRNA.114
SPNCRNA.120
SPNCRNA.122
SPNCRNA.133
SPNCRNA.129
SPNCRNA.80
SPNCRNA.90
SPNCRNA.92
SPNCRNA.93
SPNCRNA.99
spo15
spp42
sre2
srp1
ssa1
ssa2
ssc1
ssn6
ssp1
ssr2
ssr4
ste11
ste20
sti1
sts5
sua1
suc22
sum3
sup45
tal1
tas3
tbp1
tcg1
tdh1
tef3
tif224
tif512
tim40
tnr3
tom40
top2
tpi1
tpp1
tpr1
tps1
tpx1
trm1
trx1
trx2
tub1
ubc14
ubc8
ubi1
ubi3
ubr1
ubx2
ucp6
ura1
ura3
ura4
urg1
usp101
vht1
vip1
vph1
vps24
vps35
vps5
vps902
vrp1
wee1
wis1
wis2
wos2
wsc1
wsp1
yak3
yam8
ybt1
ypt2
zfs1
zip1

SpEAF enriched genes (identified as described in section 4.5.1)

aar2	elf1	mis16	rpc11	SPAC11E3.13c	SPAC2F7.05c	SPBC1347.13c
abp2	eng1	mlo3	rpc40	SPAC12B10.18	SPAC30.10c	SPBC1347.14c
ace2	eno101	mok12	rpc53	SPAC13G6.10c	SPAC30.11	SPBC13A2.04c
act1	erg12	mrpl16	rpl101	SPAC13G6.13	SPAC31G5.15	SPBC15D4.05
adg1	esc1	mug5	rpl1101	SPAC144.01	SPAC31G5.21	SPBC1604.06c
adg3	exg1	nep2	rpl15	SPAC144.12	SPAC3H1.07	SPBC1685.13
adh1	fas1	nif1	rpl24	SPAC144.18	SPAC3H1.08c	SPBC16D10.08c
agn1	fba1	nnt1	rpl26	SPAC1556.08c	SPAC4C5.03	SPBC1711.04
ags1	fep1	npp106	rpl3202	SPAC1565.07c	SPAC4F10.10c	SPBC1711.05
alg12	fft3	nrf1	rpl401	SPAC1635.01	SPAC4F10.21	SPBC1711.08
alg2	fma2	nse1	rpl402	SPAC1705.02	SPAC4F8.06	SPBC1711.09c
alg3	fta5	nup107	rpp202	SPAC1705.03c	SPAC4H3.09	SPBC1734.11
alp16	gar2	oma1	rpp203	SPAC1782.05	SPAC589.09	SPBC19C2.08
anc1	gcv1	p23fy	rps002	SPAC1782.06c	SPAC589.10c	SPBC19C7.04c
ape1	ght2	pac2	rps101	SPAC1786.02	SPAC5H10.03	SPBC19G7.04
app1	ght8	pap1	rps102	SPAC17A2.15	SPAC5H10.04	SPBC19G7.17
aps1	gln1	pch1	rps1201	SPAC18G6.13	SPAC664.12c	SPBC21C3.07c
aro1	gpd1	pex13	rps1401	SPAC19B12.02c	SPAC6C3.02c	SPBC21C3.08c
arp9	gpm1	pex7	rps1402	SPAC19G12.05	SPAC6F12.04	SPBC23G7.11
atp2	gpt2	pfk1	rps1502	SPAC19G12.16c	SPAC6G10.10c	SPBC23G7.14
bfr1	hhf1	pgi1	rps2201	SPAC1B3.02c	SPAC6G9.08	SPBC25B2.08
bgl2	hhf2	pgk1	rps23	SPAC1B3.04c	SPAC7D4.08	SPBC25H2.04c
bgs3	hhf3	pho1	rps2601	SPAC1F12.03c	SPAC806.02c	SPBC27.05
bgs4	hht1	pka1	rps3002	SPAC1F12.04c	SPAC806.11	SPBC27B12.12c
bip1	hht2	plb1	rps402	SPAC1F5.02	SPAC821.05	SPBC29A10.08
bos1	hht3	plg7	rps403	SPAC1F8.07c	SPAC824.04	SPBC29B5.04c
but2	his3	plr1	rps601	SPAC1F8.08	SPAC8E11.10	SPBC2G5.05
cbh1	hmt2	pma1	rps7	SPAC222.08c	SPAC9.07c	SPBC30B4.02c
cdc22	hsp10	pof14	rps801	SPAC222.13c	SPAC9.08c	SPBC30B4.03c
cdc48	hsp90	ppi1	rpt6	SPAC227.17c	SPAC922.04	SPBC32F12.07c
cfh4	hsr1	ppk1	rrp16	SPAC22A12.16	SPAC926.05c	SPBC32F12.10
chs1	hst4	prl01	rsd1	SPAC22F8.08	SPAC959.04c	SPBC32F12.12c
cnx1	hta1	prl03	sam1	SPAC22F8.08	SPAC959.05c	SPBC32F12.15
cox17	hta2	prl12	sap1	SPAC22H12.05c	SPAC959.06c	SPBC32H8.08c
csn2	htb1	prl32	scw1	SPAC23A1.02c	SPAC977.14c	SPBC359.03c
csn4	hxk2	prl39	sdh2	SPAC23A1.09	SPAC9E9.01	SPBC36.03c
csx1	ipk1	prl42	seb1	SPAC23C11.01	SPAC9E9.09c	SPBC365.16
ctr4	isu1	prl53	sec26	SPAC23C11.03	SPACUNK4.15	SPBC3B8.03
ctr5	iws1	prl63	sec31	SPAC23D3.12	SPACUNK4.16c	SPBC3B9.01
cut11	kap109	prp12	sec62	SPAC23H3.09c	SPAPB15E9.02c	SPBC3D6.01
cut2	kap123	psi1	sen1	SPAC24B11.07c	SPAPB18E9.06c	SPBC3E7.11c
cwf28	leu2	pss1	set8	SPAC24B11.13	SPAPB1A10.05	SPBC3E7.13c
cwf7	lys1	psu1	sfi1	SPAC24H6.13	SPAPB1A10.06c	SPBC3H7.02
cyr1	lys3	psy1	sfp1	SPAC25B8.11	SPAPB1A10.14	SPBC409.08
cys2	lys4	ptc2	sgt1	SPAC25B8.12c	SPAPB1A10.15	SPBC428.11
dbp2	mac1	pvg1	sir2	SPAC26A3.17c	SPAPB1E7.04c	SPBC4F6.17c
dfp1	mae1	pvg3	sks2	SPAC26F1.05	SPAPB1E7.07	SPBC530.08
dic1	mae2	pyk1	slp1	SPAC26F1.07	SPBC119.10	SPBC56F2.08c
dld1	maf1	pyp1	smb1	SPAC26F1.08c	SPBC1198.06c	SPBC56F2.12
dpm2	mal3	rad26	snz1	SPAC27E2.04c	SPBC11C11.05	SPBC646.15c
eaf1	mcp60	rad8	sod1	SPAC27E2.11c	SPBC11C11.06c	SPBC646.16
eca39	mde3	ravl	SPAC1002.20	SPAC27E2.13	SPBC1271.13	SPBC713.09
efl a-b	mep33	rcd1	SPAC1006.07	SPAC27F1.07	SPBC1271.14	SPBC83.17
efl a-c	met26	rlc1	SPAC10F6.07c	SPAC29B12.05c	SPBC1289.06c	SPBC839.16
efc25	mgr2	rpb9	SPAC1142.06	SPAC2C4.17c	SPBC12C2.03c	SPBC887.15c
egd2	mid2	rpc1	SPAC11D3.13	SPAC2F3.09	SPBC1347.11	SPBC8D2.18c

SPBC8E4.01c	*SPCC1322.10*	*SPCC550.14*	*SPNCRNA.122*	*str1*	*trx1*
SPBC8E4.02c	*SPCC1322.16*	*SPCC553.10*	*SPNCRNA.129*	*sua1*	*tub1*
SPBP23A10.11c	*SPCC1393.02c*	*SPCC553.11c*	*SPNCRNA.133*	*suc22*	*ubi1*
SPBP4H10.15	*SPCC1393.08*	*SPCC576.04*	*SPNCRNA.134*	*sum3*	*ubi3*
SPBPBB7.17c	*SPCC1672.01*	*SPCC576.12c*	*SPNCRNA.136*	*sup45*	*ubx2*
SPBPB2B2.04	*SPCC1682.15*	*SPCC594.02c*	*SPNCRNA.80*	*sxa1*	*ura1*
SPBPB2B2.05	*SPCC1795.10c*	*SPCC622.10c*	*SPNCRNA.90*	*tbp1*	*ura4*
SPBPB7E8.01	*SPCC18.02*	*SPCC645.08c*	*SPNCRNA.92*	*tdh1*	*vht1*
SPBPB7E8.03	*SPCC1906.05*	*SPCC645.13*	*SPNCRNA.99*	*tef3*	*vip1*
SPBTRNAHIS.02	*SPCC24B10.22*	*SPCC70.10*	*spp42*	*tif51*	*vps24*
SPCC1223.13	*SPCC297.05*	*SPCC736.15*	*sre2*	*tnr3*	*vps901*
SPCC1259.08	*SPCC297.06c*	*SPCC757.12*	*ssa1*	*tom40*	*wis2*
SPCC1281.06c	*SPCC306.11*	*SPCPB1C11.03*	*ssa2*	*tpi1*	*wos2*
SPCC1322.02	*SPCC330.06c*	*SPCTRNAASN.06*	*ssc1*	*tpp1*	*wsc1*
SPCC1322.03	*SPCC417.15*	*SPCTRNAILE.09*	*ssn6*	*tps1*	*ybt1*
SPCC1322.04	*SPCC4G3.16*	*SPNCRNA.103*	*ste20*	*tpx1*	*ypt2*
SPCC1322.05c	*SPCC4G3.17*	*SPNCRNA.117*	*sti1*	*trm1*	

(SpELL OR SpEAF enriched genes) ∩ (genes with altered RNA levels in *ell1Δ* OR *eaf1Δ* arrays)

These are the first 16 genes depicted in Figure 4-8 (in order from left to right on the figure)

abp2
ace2
eng1
adg3
adg1
agn1
mid2
SPAC2E1P3.05c
pho1
pex7
SPBPB7E8.01
SPBC8E4.01c
SPAC1039.02
gpd1
SPBPB7E8.02
str1

Appendix D: Sequence of the *eaf1* mRNA

(characterised by RACE and submitted to genbank)

S. pombe EAF (*eaf1*) mRNA, gi|122934794| gb|EF192607.1|

ACTTTTACCATTATAAACAATCATGAATTCATTACAGAAGGGATCATACAA
AGTTATTCCCGGGTCCAGCTTCTCAAAGAATTCAAATGGACTTTTATCTATTAAA
TACAATTTTATACCAGAAAGTGTGGACCCTTCTCGTCGAGGTGTTTTAGAAAAG
GCTCAGGAAGCCTATCGTCTACGTCTTCCTTCAACTTTTGATGATGACAGGCCT
CATATATTTGAAGGCTCATGCCAACGAGCCAGGAATGTGGACTGTGTACTAATT
TTTAACGCTAAAACAAAGACATTCACACTGGAGCATATTGATGAGATTGCTCGA
TTAAATGCTTTACGCAATCCAAAAGTCTCTAAAACTGTACCTTCTAATGCCATCA
CTCAATCTGATAATTCCCAAATCTCTGAATCCAAATCGACCTCTCAAAGTGCGG
TCACGACTAATTCCACGAGACGTAAAGAAAGGAATTGGAAGCTTCAAAGGATG
GAAAAATAAAACCTTCATCGAGCAATACTCGATATCCTGCCATATCTAGCAAGG
GACCAATCACCACCGATACCAACGATGAGCCTGACATGGAGGTCATGGAGTTG
GATGACTTTGCCAAAGAGTTAGAGCTAGGATTTGATCAGGAATTTAATTCCATA
GACGACCCCTCTACAGTATCCCAAACGGCCAGCAAACCAATCTCTTTGAGGGG
TCTATCTTCGCAAGAGAGGGATTATGCTTCTTCTGCTCAGGCAGAGGGTATCAG
CAGCGCTTCCGAGGATGAGGATTAAGTATGCTTTATGCTGTGTTTCGTATACAC
TTTCCAAACACGATAATGAGAAAATCTGTTAATGAACGACTTGCAACGAAATTC
CAGGGAGTAAAGTAAAGCAAGACAAGTCTTTTCCATCACTACTCGACAGCGAA
TCGGCGGTAATAAACTCGTTTCACTCTGTGTCATTTTTTTAATTATATACGAACG
TGTCTACGCCAAGGGAAAAACATATATCGTTTTAAAATGTTTTCATCTTTTTATA
TTTCTTTTTATTATTATTAATTCATAATGACTATAAAAAACATGGCTTGTCTTAAA
AAAAAAAAAAAAAAAAAAAA

ACKNOWLEDGEMENTS

An impossible task to give adequate thanks to all those who have given so much of their time and provided insight and motivation, but I am most grateful to the many who have made this thesis a reality.

Firstly, to my thesis advisors, Dr. Joan Conaway and Dr. Ron Conaway for encouraging me to start, for allowing me room to make mistakes and grow, and for providing me with guidance along my scientific research path. To all those with whom I have collaborated: Arcady Mushegian, Skylar Martin-Brown, Laurence Florens, Michael Washburn, Henrik Spahr, Chris Seidel, Madelaine Gogol, Hua Li, Karen Zueckert-Gaudenz, Brian Fleharty, and to my wife and collaborator, Stephanie Kong, whose encouragement has enabled me to keep going and whose knowledge and insight has kept me inspired.

To the other members of my advisory committee who have given their time and shared with me the benefit of their experience - Jerry Workman, Leanne Wiedemann and Peter Baumann.

Lastly to my dear father for encouraging in me a sense of wonder at the world around and encouraging me to think about how it all works - I wish you were still here to share this with.

Thank you most sincerely.

REFERENCES

1. Thirman M. J., Levitan D. A., Kobayashi H., Simon M. C. & Rowley J. D. (1994). Cloning of ELL, a gene that fuses to MLL in a t(11;19)(q23;p13.1) in acute myeloid leukemia. Proceedings of the National Academy of Sciences of the United States of America 91, 12110-12114.

2. Shilatifard A., Lane W. S., Jackson K. W., Conaway R. C. & Conaway J. W. (1996). An RNA polymerase II elongation factor encoded by the human ELL gene. Science 271, 1873-1876.

3. Kong S. E., Banks C. A., Shilatifard A., Conaway J. W. & Conaway R. C. (2005). ELL-associated factors 1 and 2 are positive regulators of RNA polymerase II elongation factor ELL. Proc Natl Acad Sci U S A 102, 10094-10098.

4. Banks C. A., Kong S. E., Spahr H., Florens L., Martin-Brown S., Washburn M. P., Conaway J. W., Mushegian A. & Conaway R. C. (2007). Identification and Characterization of a Schizosaccharomyces pombe RNA Polymerase II Elongation Factor with Similarity to the Metazoan Transcription Factor ELL. J Biol Chem 282, 5761-5769.

5. Dieci G., Fiorino G., Castelnuovo M., Teichmann M. & Pagano A. (2007). The expanding RNA polymerase III transcriptome. Trends Genet 23, 614-622.

6. Kimura, Makoto, Mitsuzawa, Hiroshi & Ishihama, Akira (2003). RNA Polymerases and Accessory Factors. In The Molecular Biology of Schizosaccharomyces Pombe: Genetics, Genomics and Beyond (Egel, Richard, eds), pp. 329-342, Springer-Verlag Berlin and Heidelberg GmbH & Co. K, Germany.

7. Spåhr H., Calero G., Bushnell D. A. & Kornberg R. D. (2009). Schizosacharomyces pombe RNA polymerase II at 3.6-A resolution. Proceedings of the National Academy of Sciences of the United States of America 106, 9185-9190.

8. Cramer P., Bushnell D. A. & Kornberg R. D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. Science 292, 1863-1876.

9. Gnatt A. L., Cramer P., Fu J., Bushnell D. A. & Kornberg R. D. (2001). Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution. Science 292, 1876-1882.

10. Klug A. (2001). STRUCTURAL BIOLOGY: A Marvellous Machine for Making Messages. Science 292, 1844-1846.

11. Roeder R. G. (1996). The role of general initiation factors in transcription by RNA polymerase II. Trends Biochem Sci 21, 327-335.

12. Burton Z. F., Killeen M., Sopta M., Ortolan L. G. & Greenblatt J. (1988). RAP30/74: a general initiation factor that binds to RNA polymerase II. Molecular and Cellular Biology 8, 1602-1613.

13. Price D. H., Sluder A. E. & Greenleaf A. L. (1989). Dynamic interaction between a Drosophila transcription factor and RNA polymerase II. Molecular and Cellular Biology 9, 1465-1475.

14. Conaway J. W. & Conaway R. C. (1990). An RNA polymerase II transcription factor shares functional properties with Escherichia coli sigma 70. Science 248, 1550-1553.

15. Killeen M. T. & Greenblatt J. F. (1992). The general transcription factor RAP30 binds to RNA polymerase II and prevents it from binding nonspecifically to DNA. Molecular and Cellular Biology 12, 30-37.

16. Yan Q., Moreland R. J., Conaway J. W. & Conaway R. C. (1999). Dual roles for transcription factor IIF in promoter escape by RNA polymerase II. J Biol Chem 274, 35668-35675.

17. Serizawa H., Conaway R. C. & Conaway J. W. (1992). A carboxyl-terminal-domain kinase associated with RNA polymerase II transcription factor delta from rat liver. Proceedings of the National Academy of Sciences of the United States of America 89, 7476-7480.

18. Lu H., Zawel L., Fisher L., Egly J. M. & Reinberg D. (1992). Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II. Nature 358, 641-645.

19. Maxon M. E., Goodrich J. A. & Tjian R. (1994). Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: a model for promoter clearance. Genes & Development 8, 515-524.

20. Wang W., Carey M. & Gralla J. D. (1992). Polymerase II promoter activation: closed complex formation and ATP-driven start site opening. Science 255, 450-453.

21. Moreland R. J., Tirode F., Yan Q., Conaway J. W., Egly J. M. & Conaway R. C. (1999). A role for the TFIIH XPB DNA helicase in promoter escape by RNA polymerase II. J Biol Chem 274, 22127-22130.

22. Giardina C. & Lis J. T. (1993). DNA melting on yeast RNA polymerase II promoters. Science 261, 759-762.

23. Bushnell D. A., Westover K. D., Davis R. E. & Kornberg R. D. (2004). Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Angstroms. Science 303, 983-988.

24. Kostrewa D., Zeller M. E., Armache K. J., Seizl M., Leike K., Thomm M. & Cramer P. (2009). RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature* 462, 323-330.

25. Pettersen E. F., Goddard T. D., Huang C. C., Couch G. S., Greenblatt D. M., Meng E. C. & Ferrin T. E. (2004). UCSF Chimera - A visualization system for exploratory research and analysis. Journal of Computational Chemistry 25, 1605-1612.

26. Saunders A., Core L. & Lis J. (2006). Breaking barriers to transcription elongation. Nat Rev Mol Cell Biol 7, 557-567.

27. Kugel J. F. & Goodrich J. A. (2002). Translocation after synthesis of a four-nucleotide RNA commits RNA polymerase II to promoter escape. Molecular and Cellular Biology 22, 762-773.

28. Wang D., Bushnell D. A., Huang X., Westover K. D., Levitt M. & Kornberg R. D. (2009). Structural basis of transcription: backtracked RNA polymerase II at 3.4 angstrom resolution. Science 324, 1203-1206.

29. Dvir A., Conaway J. W. & Conaway R. C. (2001). Mechanism of transcription initiation and promoter escape by RNA polymerase II. Current Opinion in Genetics & Development 11, 209-214.

30. Izban M. G. & Luse D. S. (1992). The RNA polymerase II ternary complex cleaves the nascent transcript in a 3'----5' direction in the presence of elongation factor SII. Genes & Development 6, 1342-1356.

31. Pal M. & Luse D. S. (2003). The initiation-elongation transition: lateral mobility of RNA in RNA polymerase II complexes is greatly reduced at +8/+9 and absent by +23. Proceedings of the National Academy of Sciences of the United States of America 100, 5700-5705.

32. Stiller J. W. & Cook M. S. (2004). Functional unit of the RNA polymerase II C-terminal domain lies within heptapeptide pairs. Eukaryotic Cell 3, 735-740.

33. Meinhart A., Kamenski T., Hoeppner S., Baumli S. & Cramer P. (2005). A structural perspective of CTD function. Genes & Development 19, 1401-1415.

34. Egloff S. & Murphy S. (2008). Cracking the RNA polymerase II CTD code. Trends Genet 24, 280-288.

35. Prelich G. (2002). RNA polymerase II carboxy-terminal domain kinases: emerging clues to their function. Eukaryotic Cell 1, 153-162.

36. Rasmussen E. B. & Lis J. T. (1993). In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. Proceedings of the National Academy of Sciences of the United States of America 90, 7923-7927.

37. Pei Y., Schwer B. & Shuman S. (2003). Interactions between fission yeast Cdk9, its cyclin partner Pch1, and mRNA capping enzyme Pct1 suggest an elongation checkpoint for mRNA quality control. J Biol Chem 278, 7180-7188.

38. Matsuzaki H., Kassavetis G. A. & Geiduschek E. P. (1994). Analysis of RNA chain elongation and termination by Saccharomyces cerevisiae RNA polymerase III. Journal of Molecular Biology 235, 1173-1192.

39. Levin J. R. & Chamberlin M. J. (1987). Mapping and characterization of transcriptional pause sites in the early genetic region of bacteriophage T7. Journal of Molecular Biology 196, 61-84.

40. von Hippel P. H. & Yager T. D. (1991). Transcript elongation and termination are competitive kinetic processes. Proceedings of the National Academy of Sciences of the United States of America 88, 2307-2311.

41. Arndt K. M. & Chamberlin M. J. (1990). RNA chain elongation by Escherichia coli RNA polymerase. Factors affecting the stability of elongating ternary complexes. Journal of Molecular Biology 213, 79-108.

42. Fisher R. P. & Yanofsky C. (1983). A complementary DNA oligomer releases a transcription pause complex. J Biol Chem 258, 9208-9212.

43. Lee D. N., Phung L., Stewart J. & Landick R. (1990). Transcription pausing by Escherichia coli RNA polymerase is modulated by downstream DNA sequences. J Biol Chem 265, 15145-15153.

44. Ederth J., Artsimovitch I., Isaksson L. A. & Landick R. (2002). The downstream DNA jaw of bacterial RNA polymerase facilitates both transcriptional initiation and pausing. J Biol Chem 277, 37456-37463.

45. Tan S., Aso T., Conaway R. C. & Conaway J. W. (1994). Roles for both the RAP30 and RAP74 subunits of transcription factor IIF in transcription initiation and elongation by RNA polymerase II. J Biol Chem 269, 25684-25691.

46. Tan S., Conaway R. C. & Conaway J. W. (1995). Dissection of transcription factor TFIIF functional domains required for initiation and elongation. Proceedings of the National Academy of Sciences of the United States of America 92, 6042-6046.

47. Pokholok D. K., Hannett N. M. & Young R. A. (2002). Exchange of RNA polymerase II initiation and elongation factors during gene expression in vivo. Mol Cell 9, 799-809.

48. Aso T., Lane W. S., Conaway J. W. & Conaway R. C. (1995). Elongin (SIII): a multisubunit regulator of elongation by RNA polymerase II. Science 269, 1439-1443.

49. Ribar B., Prakash L. & Prakash S. (2007). ELA1 and CUL3 Are Required Along with ELC1 for RNA Polymerase II Polyubiquitylation and Degradation in DNA-Damaged Yeast Cells. Molecular and Cellular Biology 27, 3211-3216.

50. Yasukawa T., Kamura T., Kitajima S., Conaway R. C., Conaway J. W. & Aso T. (2008). Mammalian Elongin A complex mediates DNA-damage-induced ubiquitylation and degradation of Rpb1. EMBO J 27, 3256-3266.

51. Harreman M., Taschner M., Sigurdsson S., Anindya R., Reid J., Somesh B., Kong S. E., Banks C. A., Conaway R. C., Conaway J. W. & Svejstrup J. Q. (2009). Distinct

ubiquitin ligases act sequentially for RNA polymerase II polyubiquitylation. Proceedings of the National Academy of Sciences of the United States of America .

52. Yamazaki K., Aso T., Ohnishi Y., Ohno M., Tamura K., Shuin T., Kitajima S. & Nakabeppu Y. (2003). Mammalian elongin A is not essential for cell viability but is required for proper cell cycle progression with limited alteration of gene expression. J Biol Chem 278, 13585-13589.

53. Gerber M., Eissenberg J., Kong S., Tenney K., Conaway J. W., Conaway R. C. & Shilatifard A. (2004). In vivo requirement of the RNA polymerase II elongation factor elongin A for proper gene expression and development. Molecular and Cellular Biology 24, 9911-9919.

54. Citterio E., Van Den Boom V., Schnitzler G., Kanaar R., Bonte E., Kingston R. E., Hoeijmakers J. H. & Vermeulen W. (2000). ATP-dependent chromatin remodeling by the Cockayne syndrome B DNA repair-transcription-coupling factor. Molecular and Cellular Biology 20, 7643-7653.

55. Newman J. C., Bailey A. D. & Weiner A. M. (2006). Cockayne syndrome group B protein (CSB) plays a general role in chromatin maintenance and remodeling. Proceedings of the National Academy of Sciences of the United States of America 103, 9613-9618.

56. Selby C. P. & Sancar A. (1997). Cockayne syndrome group B protein enhances elongation by RNA polymerase II. Proceedings of the National Academy of Sciences of the United States of America 94, 11205-11209.

57. Bradsher J. N., Jackson K. W., Conaway R. C. & Conaway J. W. (1993). RNA polymerase II transcription factor SIII. I. Identification, purification, and properties. J Biol Chem 268, 25587-25593.

58. Gu W. & Reines D. (1995). Identification of a decay in transcription potential that results in elongation factor dependence of RNA polymerase II. J Biol Chem 270, 11238-11244.

59. Elmendorf B. J., Shilatifard A., Yan Q., Conaway J. W. & Conaway R. C. (2001). Transcription factors TFIIF, ELL, and Elongin negatively regulate SII-induced nascent transcript cleavage by non-arrested RNA polymerase II elongation intermediates. J Biol Chem 276, 23109-23114.

60. Takagi Y., Conaway J. W. & Conaway R. C. (1995). A novel activity associated with RNA polymerase II elongation factor SIII. SIII directs promoter-independent transcription initiation by RNA polymerase II in the absence of initiation factors. J Biol Chem 270, 24300-24305.

61. Ho C. K. & Shuman S. (1999). Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. *Mol Cell* 3, 405-411.

62. Pei Y., Hausmann S., Ho C. K., Schwer B. & Shuman S. (2001). The length, phosphorylation state, and primary structure of the RNA polymerase II carboxyl-terminal domain dictate interactions with mRNA capping enzymes. J Biol Chem 276, 28075-28082.

63. Pei Y. & Shuman S. (2002). Interactions between fission yeast mRNA capping enzymes and elongation factor Spt5. J Biol Chem 277, 19639-19648.

64. Kim E., Du L., Bregman D. B. & Warren S. L. (1997). Splicing factors associate with hyperphosphorylated RNA polymerase II in the absence of pre-mRNA. J Cell Biol 136, 19-28.

65. Bird G., Zorio D. A. & Bentley D. L. (2004). RNA polymerase II carboxy-terminal domain phosphorylation is required for cotranscriptional pre-mRNA splicing and 3'-end formation. Molecular and Cellular Biology 24, 8963-8969.

66. Hadcroft & Proudfoot (2003). Transcription Termination. In The Molecular Biology of Schizosaccharomyces Pombe: Genetics, Genomics and Beyond (Egel, Richard, eds), pp. 369-380, Springer-Verlag Berlin and Heidelberg GmbH & Co. K, Germany.

67. Nag A., Narsinh K. & Martinson H. G. (2007). The poly(A)-dependent transcriptional pause is mediated by CPSF acting on the body of the polymerase. Nat Struct Mol Biol 14, 662-669.

68. Li B., Carey M. & Workman J. L. (2007). The role of chromatin during transcription. Cell 128, 707-719.

69. Gustafsson C. M. & Samuelsson T. (2001). Mediator--a universal complex in transcriptional regulation. Molecular Microbiology 41, 1-8.

70. Martin S. & Pombo A. (2003). Transcription factories: quantitative studies of nanostructures in the mammalian nucleus. Chromosome Res 11, 461-470.

71. Cook P. R. (1999). The organization of replication and transcription. Science 284, 1790-1795.

72. Cook P. R. (2009). A Model for all Genomes: The Role of Transcription Factories. *Journal of Molecular Biology* 395, 1-10.

73. Eskiw C. H., Rapp A., Carter D. R. & Cook P. R. (2008). RNA polymerase II activity is located on the surface of protein-rich transcription factories. Journal of Cell Science 121, 1999-2007.

74. Izban M. G. & Luse D. S. (1991). Transcription on nucleosomal templates by RNA polymerase II in vitro: inhibition of elongation with enhancement of sequence-specific pausing. Genes & Development 5, 683-696.

75. Izban M. G. & Luse D. S. (1992). Factor-stimulated RNA polymerase II transcribes at physiological elongation rates on naked DNA but very poorly on chromatin templates. J Biol Chem 267, 13647-13655.

76. Hassan A. H., Neely K. E. & Workman J. L. (2001). Histone acetyltransferase complexes stabilize swi/snf binding to promoter nucleosomes. Cell 104, 817-827.

77. Cairns B. R., Lorch Y., Li Y., Zhang M., Lacomis L., Erdjument-Bromage H., Tempst P., Du J., Laurent B. & Kornberg R. D. (1996). RSC, an essential, abundant chromatin-remodeling complex. Cell 87, 1249-1260.

78. Carey M., Li B. & Workman J. L. (2006). RSC exploits histone acetylation to abrogate the nucleosomal block to RNA polymerase II elongation. Mol Cell 24, 481-487.

79. Morris S. A., Shibata Y., Noma K., Tsukamoto Y., Warren E., Temple B., Grewal S. I. & Strahl B. D. (2005). Histone H3 K36 methylation is associated with transcription elongation in Schizosaccharomyces pombe. Eukaryotic Cell 4, 1446-1454.

80. Carrozza M. J., Li B., Florens L., Suganuma T., Swanson S. K., Lee K. K., Shia W. J., Anderson S., Yates J., Washburn M. P. & Workman J. L. (2005). Histone H3

methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell 123, 581-592.

81. Orphanides G., LeRoy G., Chang C. H., Luse D. S. & Reinberg D. (1998). FACT, a factor that facilitates transcript elongation through nucleosomes. Cell 92, 105-116.

82. Saunders A., Werner J., Andrulis E. D., Nakayama T., Hirose S., Reinberg D. & Lis J. (2003). Tracking FACT and the RNA polymerase II elongation complex through chromatin in vivo. Science 301, 1094-1096.

83. Svejstrup J. Q. (2003). Transcription. Histones face the FACT. Science 301, 1053-1055.

84. Belotserkovskaya R., Oh S., Bondarenko V. A., Orphanides G., Studitsky V. M. & Reinberg D. (2003). FACT facilitates transcription-dependent nucleosome alteration. Science 301, 1090-1093.

85. Björklund S. & Kim Y. J. (1996). Mediator of transcriptional regulation. Trends Biochem Sci 21, 335-337.

86. Asturias F. J., Jiang Y. W., Myers L. C., Gustafsson C. M. & Kornberg R. D. (1999). Conserved structures of mediator and RNA polymerase II holoenzyme. Science 283, 985-987.

87. Sato S., Tomomori-Sato C., Parmely T. J., Florens L., Zybailov B., Swanson S. K., Banks C. A., Jin J., Cai Y., Washburn M. P., Conaway J. W. & Conaway R. C. (2004). A set of consensus mammalian mediator subunits identified by multidimensional protein identification technology. Mol Cell 14, 685-691.

88. Zhu X., Wirén M., Sinha I., Rasmussen N. N., Linder T., Holmberg S., Ekwall K. & Gustafsson C. M. (2006). Genome-wide occupancy profile of mediator and the Srb8-11 module reveals interactions with coding regions. Molecular Cell 22, 169-178.

89. Wery M., Shematorova E., Van Driessche B., Vandenhaute J., Thuriaux P. & Van Mullem V. (2004). Members of the SAGA and Mediator complexes are partners of the transcription elongation factor TFIIS. EMBO J 23, 4232-4242.

90. Ziemin-van der Poel S., McCabe N. R., Gill H. J., Espinosa R., Patel Y., Harden A., Rubinelli P., Smith S. D., LeBeau M. M. & Rowley J. D. (1991). Identification of a gene, MLL, that spans the breakpoint in 11q23 translocations associated with human leukemias. Proceedings of the National Academy of Sciences of the United States of America 88, 10735-10739.

91. Lavau C., Luo R. T., Du C. & Thirman M. J. (2000). Retrovirus-mediated gene transfer of MLL-ELL transforms primary myeloid progenitors and causes acute myeloid leukemias in mice. Proceedings of the National Academy of Sciences of the United States of America 97, 10984-10989.

92. Shilatifard A., Duan D. R., Haque D., Florence C., Schubach W. H., Conaway J. W. & Conaway R. C. (1997). ELL2, a new member of an ELL family of RNA polymerase II elongation factors. Proceedings of the National Academy of Sciences of the United States of America 94, 3639-3643.

93. Miller T., Williams K., Johnstone R. W. & Shilatifard A. (2000). Identification, cloning, expression, and biochemical characterization of the testis-specific RNA polymerase II elongation factor ELL3. J Biol Chem 275, 32052-32056.

94. Eissenberg J., Ma J., Gerber M. A., Christensen A., Kennison J. A. & Shilatifard A. (2002). dELL is an essential RNA polymerase II elongation factor with a general role in development. Proceedings of the National Academy of Sciences of the United States of America 99, 9894-9899.

95. Meyer C., Schneider B., Jakob S., Strehl S., Attarbaschi A., Schnittger S., Schoch C., Jansen M., Van Dongen J., Den Boer M., Pieters R., Ennas M., Angelucci E., Koehl U., Greil J., Griesinger F., Zur Stadt U., Eckert C., Szczepański T., Niggli F., Schäfer B., Kempski H., Brady H., Zuna J., Trka J., Nigro L., Biondi A., Delabesse E., Macintyre E., Stanulla M., Schrappe M., Haas O., Burmeister T., Dingermann T., Klingebiel T. & Marschalek R. (2006). The MLL recombinome of acute leukemias. Leukemia 20, 777-784.

96. Yokoyama A., Wang Z., Wysocka J., Sanyal M., Aufiero D. J., Kitabayashi I., Herr W. & Cleary M. L. (2004). Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with menin to regulate Hox gene expression. Molecular and Cellular Biology 24, 5639-5649.

97. Hess J. L. (2004). MLL: a histone methyltransferase disrupted in leukemia. Trends in molecular medicine 10, 500-507.

98. Ernst P., Wang J., Huang M., Goodman R. H. & Korsmeyer S. J. (2001). MLL and CREB bind cooperatively to the nuclear coactivator CREB-binding protein. Molecular and Cellular Biology 21, 2249-2258.

99. DiMartino J. F., Miller T., Ayton P. M., Landewe T., Hess J. L., Cleary M. L. & Shilatifard A. (2000). A carboxy-terminal domain of ELL is required and sufficient for immortalization of myeloid progenitors by MLL-ELL. Blood 96, 3887-3893.

100. Luo R. T., Lavau C., Du C., Simone F., Polak P. E., Kawamata S. & Thirman M. J. (2001). The elongation domain of ELL is dispensable but its ELL-associated factor 1 interaction domain is essential for MLL-ELL-induced leukemogenesis. Molecular and Cellular Biology 21, 5678-5687.

101. Kanda Y., Mitani K., Kurokawa M., Yamagata T., Yazaki Y. & Hirai H. (1998). Overexpression of the MEN/ELL protein, an RNA polymerase II elongation factor, results in transformation of Rat1 cells with dependence on the lysine-rich region. J Biol Chem 273, 5248-5252.

102. Lin C., Smith E. R., Takahashi H., Lai K. C., Martin-Brown S., Florens L., Washburn M., Conaway J. W., Conaway R. C. & Shilatifard A. (2010). AFF4, a Component of the ELL/P-TEFb Elongation Complex and a Shared Subunit of MLL Chimeras, Can Link Transcription Elongation to Leukemia. Mol Cell 37, 429-437.

103. Shilatifard A., Haque D., Conaway R. C. & Conaway J. W. (1997). Structure and function of RNA polymerase II elongation factor ELL. Identification of two overlapping ELL functional domains that govern its interaction with polymerase and the ternary elongation complex. J Biol Chem 272, 22355-22363.

104. Shilatifard A. (1998). Factors regulating the transcriptional elongation activity of RNA polymerase II. FASEB J 12, 1437-1446.

105. Simone F., Luo R. T., Polak P. E., Kaberlein J. J. & Thirman M. (2003). ELL-associated factor 2 (EAF2), a functional homolog of EAF1 with alternative ELL binding properties. Blood 101, 2355-2362.

106. Gerber M. A., Ma J., Dean K., Eissenberg J. C. & Shilatifard A. (2001). Drosophila ELL is associated with actively elongating RNA polymerase II on transcriptionally active sites in vivo. EMBO J 20, 6104-6114.

107. Thirman M. J., Diskin E. B., Bin S. S., Ip H. S., Miller J. M. & Simon M. C. (1997). Developmental analysis and subcellular localization of the murine homologue of ELL. Proceedings of the National Academy of Sciences of the United States of America 94, 1408-1413.

108. Li M., Wu X., Zhuang F., Jiang S., Jiang M. & Liu Y. (2003). Expression of murine ELL-associated factor 2 (Eaf2) is developmentally regulated. Dev Dyn 228, 273-280.

109. Polak P. E., Simone F., Kaberlein J. J., Luo R. T. & Thirman M. (2003). ELL and EAF1 are Cajal body components that are disrupted in MLL-ELL leukemia. Molecular Biology of the Cell 14, 1517-1528.

110. Gall J. G., Bellini M., Wu Z. & Murphy C. (1999). Assembly of the nuclear transcription and processing machinery: Cajal bodies (coiled bodies) and transcriptosomes. Molecular Biology of the Cell 10, 4385-4402.

111. Eissenberg J., Shilatifard A., Dorokhov N. & Michener D. (2007). Cdk9 is an essential kinase in Drosophila that is required for heat shock gene expression, histone methylation and elongation factor recruitment. Mol Genet Genomics 277, 101-114.

112. Kettenberger H., Armache K. J. & Cramer P. (2003). Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. Cell 114, 347-357.

113. Wang D. & Hawley D. K. (1993). Identification of a 3'-->5' exonuclease activity associated with human RNA polymerase II. Proceedings of the National Academy of Sciences of the United States of America 90, 843-847.

114. Shilatifard A., Conaway R. C. & Conaway J. W. (2003). The RNA polymerase II elongation complex. Annual Review of Biochemistry 72, 693-715.

115. Simone F., Polak P. E., Kaberlein J. J., Luo R. T., Levitan D. A. & Thirman M. J. (2001). EAF1, a novel ELL-associated factor that is delocalized by expression of the MLL-ELL fusion protein. Blood 98, 201-209.

116. Shilatifard A. (1998). Identification and purification of the Holo-ELL complex. Evidence for the presence of ELL-associated proteins that suppress the transcriptional inhibitory activity of ELL. J Biol Chem 273, 11212-11217.

117. Schmidt A. E., Miller T., Schmidt S. L., Shiekhattar R. & Shilatifard A. (1999). Cloning and characterization of the EAP30 subunit of the ELL complex that confers derepression of transcription by RNA polymerase II. J Biol Chem 274, 21981-21985.

118. Kamura T., Burian D., Khalili H., Schmidt S. L., Sato S., Liu W. J., Conrad M. N., Conaway R. C., Conaway J. W. & Shilatifard A. (2001). Cloning and characterization of ELL-associated proteins EAP45 and EAP20. a role for yeast EAP-like proteins in regulation of gene expression by glucose. J Biol Chem 276, 16528-16533.

119. Shinobu N., Maeda T., Aso T., Ito T., Kondo T., Koike K. & Hatakeyama M. (1999). Physical interaction and functional antagonism between the RNA polymerase II elongation factor ELL and p53. J Biol Chem 274, 17003-17010.

120. Martincic K., Alkan S. A., Cheatle A., Borghesi L. & Milcarek C. (2010). Transcription elongation factor ELL2 directs immunoglobulin secretion in plasma cells by stimulating altered RNA processing. Nature Immunology 10, 1102-1109.

121. Kornblihtt A. R., de la Mata M., Fededa J. P., Munoz M. J. & Nogues G. (2004). Multiple links between transcription and splicing. RNA 10, 1489-1498.

122. Mitani K., Yamagata T., Iida C., Oda H., Maki K., Ichikawa M., Asai T., Honda H., Kurokawa M. & Hirai H. (2000). Nonredundant roles of the elongation factor MEN in postimplantation development. Biochem Biophys Res Commun 279, 563-567.

123. Maurus D., Héligon C., Bürger-Schwärzler A., Brändli A. & Kühl M. (2005). Noncanonical Wnt-4 signaling and EAF2 are required for eye development in Xenopus laevis. EMBO J 24, 1181-1191.

124. Li Y., Fanning A. S., Anderson J. M. & Lavie A. (2005). Structure of the conserved cytoplasmic C-terminal domain of occludin: identification of the ZO-1 binding surface. Journal of Molecular Biology 352, 151-164.

125. Sipiczki (2003). Fission Yeast Phylogenesis and Evolution. In The Molecular Biology of Schizosaccharomyces Pombe: Genetics, Genomics and Beyond (Egel, Richard, eds), pp. 431-443, Springer-Verlag Berlin and Heidelberg GmbH & Co. K, Germany.

126. Aravind L., Watanabe H., Lipman D. J. & Koonin E. V. (2000). Lineage-specific loss and divergence of functionally linked genes in eukaryotes. Proceedings of the National Academy of Sciences of the United States of America 97, 11319-11324.

127. Bähler J., Wu J. Q., Longtine M. S., Shah N. G., McKenzie A., Steever A. B., Wach A., Philippsen P. & Pringle J. R. (1998). Heterologous modules for efficient and versatile PCR-based gene targeting in Schizosaccharomyces pombe. Yeast 14, 943-951.

128. Hentges P., Van Driessche B., Tafforeau L., Vandenhaute J. & Carr A. M. (2005). Three novel antibiotic marker cassettes for gene disruption and marker switching in Schizosaccharomyces pombe. Yeast 22, 1013-1019.

129. Matsuyama A., Shirai A., Yashiroda Y., Kamata A., Horinouchi S. & Yoshida M. (2004). pDUAL, a multipurpose, multicopy vector capable of chromosomal integration in fission yeast. Yeast 21, 1289-1305.

130. Washburn M. P., Wolters D. & Yates J. R. (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nature Biotechnology 19, 242-247.

131. Wolters D. A., Washburn M. P. & Yates J. R. (2001). An automated multidimensional protein identification technology for shotgun proteomics. Anal. Chem. 73, 5683-5690.

132. Jin J., Cai Y., Yao T., Gottschalk A. J., Florens L., Swanson S. K., Gutiérrez J. L., Coleman M. K., Workman J. L., Mushegian A., Washburn M., Conaway R. C. & Conaway J. W. (2005). A mammalian chromatin remodeling complex with similarities to the yeast INO80 complex. J Biol Chem 280, 41207-41212.

133. Zybailov B., Mosley A. L., Sardiu M. E., Coleman M. K., Florens L. & Washburn M. P. (2006). Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae. J Proteome Res 5, 2339-2347.

134. Kitts P. A. & Possee R. D. (1993). A method for producing recombinant baculovirus expression vectors at high frequency. BioTechniques 14, 810-817.

135. Jarvis D. L. & Garcia A. (1994). Long-term stability of baculoviruses stored under various conditions. BioTechniques 16, 508-513.

136. Kitts P. A. & Green G. (1999). An immunological assay for determination of baculovirus titers in 48 hours. Analytical Biochemistry 268, 173-178.

137. Rice G. A., Kane C. M. & Chamberlin M. J. (1991). Footprinting analysis of mammalian RNA polymerase II along its transcript: an alternative view of transcription elongation. Proceedings of the National Academy of Sciences of the United States of America 88, 4245-4249.

138. Boyer R. F. (1993). Separation and Purification of Biomolecules by Chromatograpy. In Modern experimental biochemistry , pp. 59-114, Benjamin/Cummings Pub. Co, Redwood City, Calif. .

139. Strahl-Bolsinger S., Hecht A., Luo K. & Grunstein M. (1997). SIR2 and SIR4 interactions differ in core and extended telomeric heterochromatin in yeast. Genes & Development 11, 83-93.

140. Fan X., Lamarre-Vincent N., Wang Q. & Struhl K. (2008). Extensive chromatin fragmentation improves enrichment of protein binding sites in chromatin immunoprecipitation experiments. Nucleic Acids Research 36, e125.

141. Johnson W. E., Li W., Meyer C. A., Gottardo R., Carroll J. S., Brown M. & Liu X. S. (2006). Model-based analysis of tiling-arrays for ChIP-chip. Proceedings of the National Academy of Sciences of the United States of America 103, 12457-12462.

142. Hoffman L. & Jarvis B. W. A Novel, Rapid Method to Release Intact Yeast RNA Using the MasterPure Yeast RNA Purification Kit. Epicentre Forum 11.

143. Cook M., Marchin M. & Seidel C. (2006). S. pombe 70-mer Genome Set. Poster presentation available online at: http://research.stowers-institute.org/microarray/S_pombe/poster.pdf

144. Sansó M., Gogol M., Ayté J., Seidel C. & Hidalgo E. (2008). Transcription factors Pcr1 and Atf1 have distinct roles in stress- and Sty1-dependent gene regulation. Eukaryotic Cell 7, 826-835.

145. Mitani K., Kanda Y., Ogawa S., Tanaka T., Inazawa J., Yazaki Y. & Hirai H. (1995). Cloning of several species of MLL/MEN chimeric cDNAs in myeloid leukemia with t(11;19)(q23;p13.1) translocation. Blood 85, 2017-2024.

146. Kadesch T. R. & Chamberlin M. J. (1982). Studies of in vitro transcription by calf thymus RNA polymerase II using a novel duplex DNA template. J Biol Chem 257, 5286-5295.

147. Hamada M., Sakulich A. L., Koduru S. B. & Maraia R. J. (2000). Transcription termination by RNA polymerase III in fission yeast. A genetic and biochemically tractable model system. J Biol Chem 275, 29076-29081.

148. Kedinger C., Gniazdowski M., Mandel J. L., Gissinger F. & Chambon P. (1970). Alpha-amanitin: a specific inhibitor of one of two DNA-pendent RNA polymerase activities from calf thymus. Biochem Biophys Res Commun 38, 165-171.

149. Venter J. C., Adams M. D., Myers E. W. & Li P. W. (2001). The sequence of the human genome. *Science* 291, 1304-1351.

150. Bahler J. & Wood V. (2003). The Genome and Beyond. In The Molecular Biology of Schizosaccharomyces Pombe: Genetics, Genomics and Beyond (Egel, Richard, eds), pp. 13-23, Springer-Verlag Berlin and Heidelberg GmbH & Co. K, Germany.

151. Riles L., Shaw R. J., Johnston M. & Reines D. (2004). Large-scale screening of yeast mutants for sensitivity to the IMP dehydrogenase inhibitor 6-azauracil. Yeast 21, 241-248.

152. Malagon F., Kireeva M. L., Shafer B. K., Lubkowska L., Kashlev M. & Strathern J. N. (2006). Mutations in the Saccharomyces cerevisiae RPB1 gene conferring hypersensitivity to 6-azauracil. Genetics 172, 2201-2209.

153. Powell W. & Reines D. (1996). Mutations in the second largest subunit of RNA polymerase II cause 6-azauracil sensitivity in yeast and increased transcriptional arrest in vitro. J Biol Chem 271, 6866-6873.

154. Hemming S. A., Jansma D. B., Macgregor P. F., Goryachev A., Friesen J. D. & Edwards A. M. (2000). RNA polymerase II subunit Rpb9 regulates transcription elongation in vivo. J Biol Chem 275, 35506-35511.

155. Williams L. A. & Kane C. M. (1996). Isolation and characterization of the Schizosaccharomyces pombe gene encoding transcript elongation factor TFIIS. Yeast 12, 227-236.

156. Hartzog G. A., Wada T., Handa H. & Winston F. (1998). Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in Saccharomyces cerevisiae. *Genes & Development* 12, 357-369.

157. Exinger F. & Lacroute F. (1992). 6-Azauracil inhibition of GTP biosynthesis in Saccharomyces cerevisiae. Curr Genet 22, 9-11.

158. Smith E. R., Winter B., Eissenberg J. & Shilatifard A. (2008). Regulation of the transcriptional activity of poised RNA polymerase II by the elongation factor ELL. Proceedings of the National Academy of Sciences of the United States of America 105, 8575-8579.

159. Qiu H., Hu C., Wong C. M. & Hinnebusch A. G. (2006). The Spt4p subunit of yeast DSIF stimulates association of the Paf1 complex with elongating RNA polymerase II. Molecular and Cellular Biology 26, 3135-3148.

160. Pokholok D. K., Harbison C. T., Levine S., Cole M., Hannett N. M., Lee T. I., Bell G. W., Walker K., Rolfe P. A., Herbolsheimer E., Zeitlinger J., Lewitter F., Gifford D. K. & Young R. A. (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. Cell 122, 517-527.

161. Lackner D. H., Beilharz T. H., Marguerat S., Mata J., Watt S., Schubert F., Preiss T. & Bahler J. (2007). A network of multiple regulatory layers shapes gene expression in fission yeast. Mol Cell 26, 145-155.

162. Preacher K. J. Calculation for the chi-square test: an interactive calculation tool for chi-square tests of goodness of fit and independence. Available from http://www.quantumpsy.org .

163. Gerber M. A., Shilatifard A. & Eissenberg J. (2005). Mutational analysis of an RNA polymerase II elongation factor in Drosophila melanogaster. Molecular and Cellular Biology 25, 7803-7811.

164. Byun J. S., Wong M. M., Cui W., Idelman G., Li Q., De Siervi A., Bilke S., Haggerty C. M., Player A., Wang Y. H., Thirman M. J., Kaberlein J. J., Petrovas C., Koup R. A., Longo D., Ozato K. & Gardner K. (2009). Dynamic bookmarking of primary response genes by p300 and RNA polymerase II complexes. Proceedings of the National Academy of Sciences of the United States of America 106, 19286-19291.

165. Ashburner M., Ball C. A., Blake J. A., Botstein D., Butler H., Cherry J. M., Davis A. P., Dolinski K., Dwight S. S., Eppig J. T., Harris M. A., Hill D. P., Issel-Tarver L., Kasarskis A., Lewis S., Matese J. C., Richardson J. E., Ringwald M., Rubin G. M. & Sherlock G. (2000). Gene Ontology: tool for the unification of biology. 25, 25-29.

166. Alonso-Nuñez M. L., An H., Martín-Cuadrado A. B., Mehta S., Petit C., Sipiczki M., del Rey F., Gould K. L. & de Aldana C. R. (2005). Ace2p controls the expression of genes required for cell separation in Schizosaccharomyces pombe. Molecular Biology of the Cell 16, 2003-2017.

167. Sharma N., Marguerat S., Mehta S., Watt S. & Bähler J. (2006). The fission yeast Rpb4 subunit of RNA polymerase II plays a specialized role in cell separation. *Mol Genet Genomics* 276, 545-554.

168. Verma-Gaur J., Rao S. N., Taya T. & Sadhale P. (2008). Genomewide recruitment analysis of Rpb4, a subunit of polymerase II in Saccharomyces cerevisiae, reveals its involvement in transcription elongation. *Eukaryotic Cell* 7, 1009-1018.

169. Miklos I., Szilagyi Z., Watt S., Zilahi E., Batta G., Antunovics Z., Enczi K., Bahler J. & Sipiczki M. (2008). Genomic expression patterns in cell separation mutants of Schizosaccharomyces pombe defective in the genes sep10 (+) and sep15 (+) coding for the Mediator subunits Med31 and Med8. Mol Genet Genomics 279, 225-238.

170. Linder T., Rasmussen N. N., Samuelsen C. O., Chatzidaki E., Baraznenok V., Beve J., Henriksen P., Gustafsson C. M. & Holmberg S. (2008). Two conserved modules of Schizosaccharomyces pombe Mediator regulate distinct cellular pathways. Nucleic Acids Research 36, 2489-2504.

171. Zeitlinger J., Stark A., Kellis M., Hong J. W., Nechaev S., Adelman K., Levine M. & Young R. A. (2007). RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. *Nat Genet* 39, 1512-1516.

172. Sharma N., Marguerat S., Mehta S., Watt S. & Bähler J. (2006). The fission yeast Rpb4 subunit of RNA polymerase II plays a specialized role in cell separation. Mol Genet Genomics 276, 545-554.

173. Watanabe Y. & Yamamoto M. (1994). S. pombe mei2+ encodes an RNA-binding protein essential for premeiotic DNA synthesis and meiosis I, which cooperates with a novel RNA species meiRNA. Cell 78, 487-498.

174. Borson N. D., Salo W. L. & Drewes L. R. (1992). A lock-docking oligo (dT) primer for 5'and 3'RACE PCR. PCR Methods Appl 2, 144-148.

175. Lee T. I., Johnstone S. E. & Young R. A. (2006). Chromatin immunoprecipitation and microarray-based analysis of protein location. Nat. Protoc. 1, 729-748.

176. Gullerova M. & Proudfoot N. J. (2008). Cohesin complex promotes transcriptional termination between convergent genes in S. pombe. Cell 132, 983-995.

177. (2009). RNA polymerases as molecular motors (Buc, H. & Strick, T., eds), RSC, Cambridge.

178. Hertz-Fowler C., Peacock C. S., Wood V., Aslett M., Kerhornou A., Mooney P., Tivey A., Berriman M., Hall N., Rutherford K., Parkhill J., Ivens A. C., Rajandream M. - A. & Barrell B. (2004). GeneDB: a resource for prokaryotic and eukaryotic organisms. Nucl Acids Res 32, D339-D343.

179. Pei Y., Du H., Singer J., Stamour C., Granitto S., Shuman S. & Fisher R. (2006). Cyclin-dependent kinase 9 (Cdk9) of fission yeast is activated by the CDK-activating kinase Csk1, overlaps functionally with the TFIIH-associated kinase Mcs6, and associates with the mRNA cap methyltransferase Pcm1 in vivo. Molecular and Cellular Biology 26, 777-788.

180. Kusov Y. Y., Shatirishvili G., Dzagurov G. & Gauss-Müller V. (2001). A new G-tailing method for the determination of the poly(A) tail length applied to hepatitis A virus RNA. Nucleic Acids Research 29, e57.

181. Missra A. & Gilmour D. S. (2010). Interactions between DSIF (DRB sensitivity inducing factor), NELF (negative elongation factor), and the Drosophila RNA polymerase II transcription elongation complex. Proceedings of the National Academy of Sciences of the United States of America 107, 11301-11306.

182. Gilmour D. S. (2009). Promoter proximal pausing on genes in metazoans. Chromosoma 118, 1-10.

183. Moreland R. J., Hanas J. S., Conaway J. W. & Conaway R. C. (1998). Mechanism of action of RNA polymerase II elongation factor Elongin. Maximal stimulation of elongation requires conversion of the early elongation complex to an Elongin-activable form. J Biol Chem 273, 26610-26617.

184. Leupold U. (1970). Genetical Methods for Schizosaccharomyces Pombe. In Methods in Cell Biology (Prescott, D. M., eds), pp. 169-177, Elsevier Inc.

185. Smith G. R. (2009). Genetic analysis of meiotic recombination in Schizosaccharomyces pombe. Methods Mol Biol 557, 65-76.

186. Bähler J. (2005). A transcriptional pathway for cell separation in fission yeast. Cell Cycle 4, 39-41.

187. Walker G. M. (1999). Synchronization of yeast cell populations. Methods Cell Sci 21, 87-93.

188. Roguev A., Wiren M., Weissman J. S. & Krogan N. J. (2007). High-throughput genetic interaction mapping in the fission yeast Schizosaccharomyces pombe. Nat Meth 4, 861-866.