Characterisation of nuclear Cap-Binding Complex in *Drosophila melanogaster*

Lubna Arif

PhD Thesis

University of Edinburgh

November, 2001

For my parents,

Rafique and Bilquis

Abstract

All RNAs transcribed by RNA polymerase II (RNAP II) are characterised by a monomethyl guanosine cap. RNA cap formation occurs co-transcriptionally and is facilitated by the recruitment of the capping enzyme to the phosphorylated carboxyl-terminal domain (CTD) of the largest subunit of RNAP II. The cap is functionally important in various aspects of RNA metabolism including RNA stability, pre-mRNA splicing, 3'-end formation, RNA transport and cap-dependent translation. The effects of the cap in pre-mRNA splicing, 3'-end formation and U snRNA export are mediated by the nuclear cap-binding complex (CBC). CBC is a heterodimer comprised of cap-binding proteins CBP20 and CBP80. Homologues have been identified in a variety of eukaryotes and CBC function appears to be evolutionarily conserved in metazoans. Although CBC mediates the cap's effects in three diverse nuclear processes, little is known about the proteins that CBC interacts with or the mechanisms involved. This thesis aims to characterise CBC in *Drosophila melanogaster* and uses genetic approaches to further understand the mechanisms by which the cap facilitates multiple aspects of RNA metabolism through CBC.

Firstly, P-element induced mutagenesis was used to create mutants in the CBC heterodimer. A P-element residing in *DmCBP20* was mobilised to generate *DmCBP20* excision alleles. This showed *DmCBP20* is an essential gene that functions in both somatogenesis and gametogenesis. Intriguingly, the P-hop generated a surprisingly high number of *DmCBP20* male steriles compared to *DmCBP20* female steriles. Analysis of male steriles showed the early stages of spermatogenesis that correlate to active transcription and RNA processing are normal. Instead, defects might arise in late spermatid differentiation. These findings suggest novel function(s) for *DmCBP20* in spermatogenesis outwith RNA processing.

A second divergent CBP80 homologue named DmCBP80 Related (DmCBP80R) is identified in fruit-flies. In contrast to *DmCBP80*, *DmCBP80R* appears to be predominantly expressed in adult testes. These results suggest DmCBP80R, together with DmCBP20, might have acquired novel functions in spermatogenesis as part of a male-specific CBC.

Finally, to identify CBC-interacting genes, the GAL4-UAS system was used to over-express full-length and truncated *CBP20* and *CBP80* in non-essential adult fly tissues. The shortest truncations in both cap-binding proteins generated dominant phenotypes. *DmCBP80* dominant phenotype was used in genetic modifier screens to identify CBC-interacting genes. This showed the dominant phenotype was suppressed by an allele of the largest subunit of RNAP II that has a truncated CTD. These findings suggest CBC might interact with the CTD of RNAP II to facilitate the coupling of transcription with pre-mRNA splicing and RNA 3'-end formation.

Acknowledgements

On taking on this PhD I did not appreciate the enormity of the intellectual and emotional demands involved. During this time I have received significant help from friends, family and colleagues. I would like to take this opportunity to sincerely thank them all.

Firstly, I would like to thank Joe Lewis who gave me the opportunity and means to take on & complete my PhD. I want to especially thank my 2nd supervisor, Andrew Jarman who has gone beyond the call of duty. Thank you, for your continued support, encouragement and humour particularly through the difficult times. I emerge a happy, confident, sane person who has no need for therapy!

Many thanks to: Joao Ferreira, whose brief, crazy company I enjoyed enormously. Yiota Kafasla and Rym Houalla, for being so encouraging & returning me to my usual loud, singing, dancing self. Gillian McKay, who teetered round the lab wearing high heels & vest top whilst complaining about the lack of 6th floor eye candy. Truly refreshing. Nina McDougall, for your teachings on the joy of female steriles. I also discovered much joy during our stimulating scientific interchanges.

In addition, I want to thank colleagues who have taken time to help me and created a light-hearted & enjoyable atmosphere for me to work in: Ilan Davis, Hille Tekotte, Gavin Wilkie, Ruth Kirby, Petra zur Lage, Emma Rawlins, Neville Cobbe, Marie-Louise Loupart, Brian McHugh, Sharron Vass, Jane Ilsley, Zoe Popper, Joanna Kufel, Mar Carmena & Earnshaw Lab members, Hiro Okhura, Fiona Cullen, Nicola Wrobel, Anne Abraham, John Findlay, Ian Gilbert, Mary Bownes and Brian Kilbey.

My family have played an incredibly important role during this time. I want to thank Irfan for his patience & understanding even through my mood swings. Imran, for his kind-heartedness & unquestioning support financially when times were hard. Farhanna for her emotional support & humour, which have been invaluable (along with our mutual love of noodles & slimfast!). Shazia who taught me to never question my ability. Through her I have learnt the value of persistence & seizing opportunities. I have been amazed how this time has strengthened our relationship. I now treasure you as highly as my duvet!

Finally, I am greatly indebted to my loving parents. Their belief in me, encouragement & support has been endless. Thank you.

Table of Contents

Table of Figures

Gene Names and Abbreviations

aa	Amino acid
BDGP	Berkeley *Drosophila* genome project
BLAST	Basic local alignment research tool
bp	Base pair
BR	Balbiani ring
CBC	Cap-binding complex
CBP	Cap-binding protein
CBP20R	*CBP20 Related*
CBP80R	*CBP80 Related*
CID	CTD-interacting domain
CF	*Cleavage factor*
CPSF	*Cleavage and polyadenylation specificity factor*
CstF	*Cleavage stimulation factor*
CTD	Carboxyl-terminal domain
DAPI	4, 6-diamidino-2-phenylindole
dH_2O	Distilled water
DIC	Diffraction interference contrast
DIG	Digoxygenin
Dpp	*Decapentaplegic*
DSIF	*DRB inducing sensitivity factor*
eIF	*Eukaryotic initiation factor*
EM	Electron microscopy
EMS	Ethyl methane sulphate
EST	Expressed sequence tag
Ey	*Eyeless*
GFP	Green fluorescent protein
kDa	KiloDalton
kb	Kilobase
NELF	*Negative elongation factor*
NES	Nuclear export signal
NLS	Nuclear localisation signal
NMD	Nonsense mediated decay
NPC	Nuclear pore complex
mRNA	messenger RNA
OD	Optical density
ORF	Open reading frame
PABP2	*Poly(A)-binding protein 2*
PAP	*Poly(A) polymerase*
Ptc	*Patched*
PHAX	*Phosphorylated adaptor for RNA export*
PIC	Pre-initiation complex
P-TEFb	*Positivetranscription elongation factor b*
RBD	RNA-binding domain
RNAP II	*RNA polymerase II*
RNAP II A	*RNA polymerase II with a hypophosphorylated CTD*
RNAP IIO	*RNA polymerase II with a hyperphosporylated CTD*
RRM	RNA recognition motif
rRNA	Ribosomal RNAs
Rp	Ribosomal protein
Sca	*Scabrous*
SCAF	*SR-like CTD associated factors*
SEM	Scanning electron microscopy
SOP	Sensory organ precursor
snoRNA	small nucleolar RNA
STS	Sequence tagged site
TBP	*Tata-binding protein*
TF	Transcription factor
ts	Temperature sensitive
UsnRNA	Uracil-rich small nuclear RNA
UTR	Untranslated region
Vg	*Vestigal*
Xop1	*Exportin 1*
YP	Yolk protein

CHAPTER 1

Introduction

Protein-encoding genes and major spliceosomal U snRNAs with the exception of U6 snRNA, are specifically transcribed by RNA polymerase II (RNAP II). RNAP II transcripts are characterised by a 5'-end modification termed the monomethyl guanosine cap. The cap structure consists of an inverted 7-methyl guanosine, which is linked to the first transcribed nucleotide of RNA through a 5' to 5' triphosphate bridge (Figure 1.1 and Shatkin, 1976; Salditt-Georgieff *et al.*, 1980). The cap was initially characterised as a blocked structure that protects the 5'-end of primary transcripts against 5' to 3' exoribonucleases and as a result promotes RNA stability (Furuichi *et al.*, 1977; Shimotohno *et al.*, 1977; Schwer *et al.*, 1998). Subsequent studies have shown that the cap is functionally important in eukaryotic gene expression where it plays a central role in RNAP II transcript metabolism. In the nucleus, the cap is involved in RNA stability, pre-mRNA splicing (Konarska *et al.*, 1984; Krainer *et al.*, 1984; Ohno *et al.*, 1987), RNA 3'-end formation (Cooke and Alwine, 1996; Gilmartin *et al.*, 1988; Hart *et al.*, 1985), RNA export and nucleolar localisation of U3 and U8 small nucleolar RNAs (Hamm and Mattaj, 1990; Jarmolowski *et al.*, 1994; Jacobson and Pederson, 1998). Whilst in the cytoplasm, the cap is required for cap-dependent translation of mRNAs (Shatkin, 1985; Sachs *et al.*, 1997).

A key observation in the field of gene expression is that transcription, RNA processing and translation appear to be mechanistically linked *in vivo*. This would make biological sense, as it would ensure the most efficient passage for gene expression. However, the molecular mechanisms by which these events are integrated remains to be elucidated. In the last few years, it has become increasingly apparent that RNAP II also contributes to various RNA processing reactions required to synthesis a mature mRNA.

Figure 1.1 Cap Structure

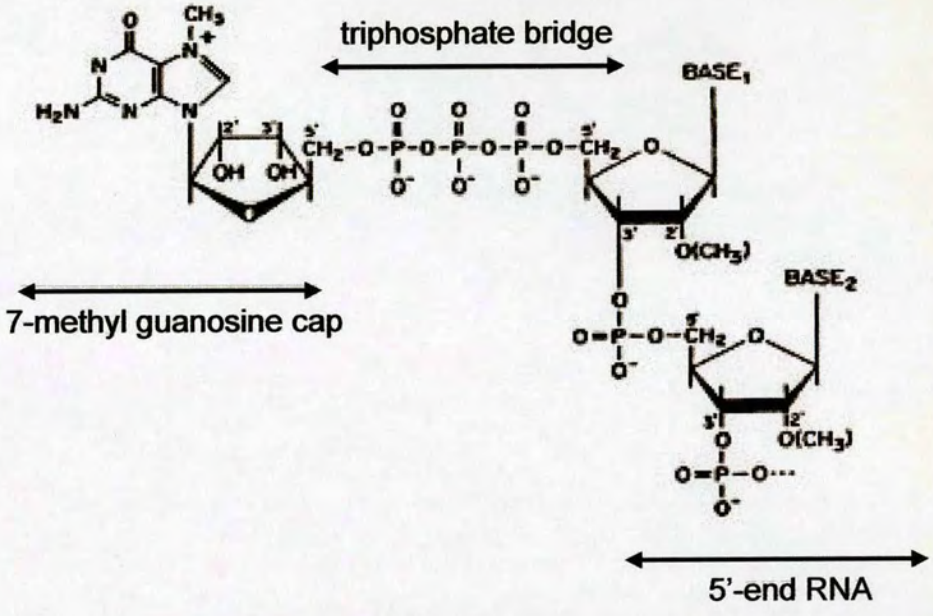


Figure 1.1 Structure of the cap. RNAP II transcripts are characterised by a 5' modification termed the monomethyl guanosine cap. The cap consists of an inverted 7-methyl guanosine linked to the first transcribed nucleotide of RNA through a 5' to 5' triphosphate linkage. Figure is taken from Shatkin, 1976.

In particular, the carboxyl-terminal domain (CTD) of the largest subunit of RNAP II is found to integrate nuclear RNA processing with transcription. The CTD appears to not only recruit RNA processing factors to sites of transcription but also directly functions in RNA cap formation, pre-mRNA splicing and RNA 3'-end formation. The cap acts as a molecular tag to define RNAP II transcripts. As a result, the cap mediates the recognition of RNAP II transcripts by various components of nuclear RNA processing machinery. This suggests that the cap might also interact with the CTD to facilitate the coupling of transcription with nuclear RNA processing.

1.1 The capping reaction

The capping reaction is catalysed by a series of three enzymatic activities; RNA 5'-triphosphatase, RNA guanylyl transferase and RNA (guanine-7)-methyl transferase respectively. Firstly, RNA 5'-triphosphatase hydrolyses the γ-phosphate from the 5'-triphosphate end of the nascent transcript to a 5'-diphosphate. Next, RNA guanylyl transferase transfers GMP derived from GTP to the newly formed 5'-diphosphate end forming a blocked guanosine cap (GpppN). Lastly, RNA (guanine-7)-methyl transferase adds a methyl group to the N7 position of the cap guanosine to generate the m^7 G (5')ppp(5') N cap structure (Figure 1.1 and Shuman, 1995). The capping enzyme in *S. pombe* and *S. cerevisiae* has been characterised and is found to consist of a heterodimer comprised of separate triphosphatase (Cet1) and guanylyl transferase (Ceg1) subunits (Shibagaki, 1992). In contrast, the capping enzyme in metazoans is encoded in a single protein that consists of an amino-terminal triphosphatase domain linked to a carboxyl-terminal guanylyl transferase domain (Yue *et al.*, 1997; Saha *et al.*, 1999).

Evidence from several laboratories has demonstrated in yeast that capping is essential for cell viability. Mutations of the triphosphatase, guanylyl transferase or methyl transferase components of the capping apparatus are lethal *in vivo* (Schwer and Schuman, 1994; Shuman *et al.* 1994; Fresco and Buratowski, 1996; Mao 1996; Wang 1997; Ho *et al.* 1998). This is not unexpected considering the cap plays an important role in RNAP II metabolism events such as splicing, RNA 3'-end

formation, U snRNA export and cap-dependent translation initiation (Konarska *et al.*, 1984; Shatkin, 1985; Jarmolowski *et al.*, 1994; Cooke and Alwine, 1996). In addition, failure to cap pre-mRNAs results in their accelerated decay through the activities of 5' to 3' exoribonucleases (Furuichi *et al.*, 1977; Shimotohno *et al.*, 1977; Schwer *et al.*, 1998).

1.2 The transcription cycle

Cap formation is the first detectable RNAP II processing event that occurs co-transcriptionally. Synthesis of RNAP II transcripts involves a complete round of transcription that consists of pre-initiation complex (PIC) formation, transcription initiation, transition to RNAP II elongation complex and transcription termination (Reines *et al.*, 1999; Buratowski, 2000). The transcription PIC is formed when RNAP II and multiple general transcription factors are recruited to the core promoter. Firstly, the TATA-binding protein (TBP) subunit of TFIID directly binds to the TATA element within the promoter. Next, TFIIB in collaboration with TFIIF bridges TBP and recruits the RNAP II holoenzyme. TFIIE and the ATP-dependent helicase component of TFIIH then unwind the promoter DNA, which allows transcription to be initiated (Berk, 1999; Green, 2000). During transcription initiation, RNAP II synthesises 20-40 nucleotides of the nascent transcript and then pauses. It is within this initiation phase that the capping reaction occurs (Coppola *et al.*, 1983; Jove and Manley, 1984). RNAP II makes the transition to the elongation complex and the holoenzyme then escapes from the promoter and transcribes the remainder of the gene. Finally, transcription is terminated and both RNAP II and the nascent transcript are released from the DNA template.

1.3 Coupling capping with transcription

The capping reaction occurs co-transcriptionally on nascent transcripts that are approximately 20-40 nucleotides in length. At this length, RNAs are sufficiently extruded from the RNA-binding pocket of RNAP II to become accessible to the capping enzymes (Coppola *et al.*, 1983; Hagler and Shuman, 1992; Rasmussen and Lis, 1993). Evidence from several laboratories has shown that RNA cap formation is coupled to transcription. This appears to be facilitated by either the phosphorylated CTD or the DSIF subunit, Spt5 both of which physically interact with the capping enzyme and stimulate its guanylyl transferase activity.

Capping is specifically targeted to RNAP II transcripts through the physical interaction of one or more of the components of the capping apparatus with the phosphorylated CTD. In *S. cerevisiae*, Ceg1 and Abd1 bind directly to the phospho-serine-5 residues of the CTD *in vitro* (Cho *et al.*, 1997; McCracken *et al.*, 1997a). Whilst in mammals, both RNA triphosphatase and guanylyl transferase activities bind to the phosphorylated CTD (McCracken *et al.*, 1997a; Yue *et al.*, 1997; Ho *et al.*, 1998; Ho and Shuman, 1999).

The CTD consists of a heptad repeat with the consensus sequence Tyr-Ser-Pro-Thr-Ser-Pro-Ser, which is remarkably conserved in sequence but not copy number. For example, in *S. cerevisiae* the CTD has 26 repeats, *Drosophila melanogaster* CTD has 42 repeats, whilst the mammalian CTD has 52 repeats (Allison *et al.*, 1985; Corden *et al.*,1985; Allison *et al.*, 1988). The CTD is rich in potential phospho-acceptor amino acid residues and is subject to reversible phosphorylation during the transcription cycle (Dahmus, 1996). The phosphorylation status of the CTD is important since it not only plays an important role in transcription initiation but also recruits factors involved in RNA cap formation, pre-mRNA splicing and RNA 3'-end formation (Carlson, 1997; Rodriguez *et al.*, 2000; Schroeder *et al.*, 2000). Figure 1.2 summarises the dynamic association of RNA capping and processing factors with the differentially phosphorylated forms of RNAP II throughout the transcription cycle.

Figure 1.2 Transcription is Coupled with mRNA Processing

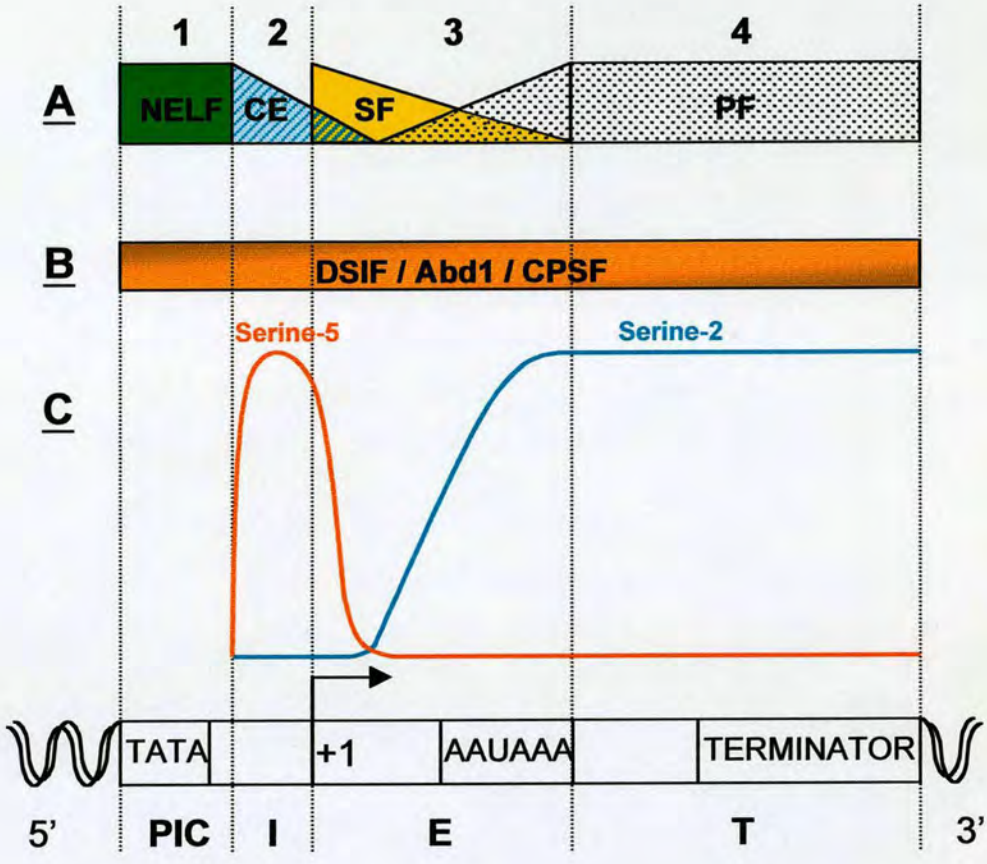


Figure 1.2 Transcription is coupled with mRNA processing. A round of transcription consists of pre-initiation complex formation (PIC), transcription initiation (I), transition to RNAP II elongation complex (E) and transcription termination (T). **A.** Phosphorylation status of CTD affects recruitment of RNA processing factors. **1.** NELF is recruited to RNAP IIA where it represses elongation in association with DSIF. **2.** Capping enzyme (CE) binds serine-5 phosphorylated CTD during initiation. **3.** Splicing factors (SF) associate with elongating RNAP IIO. **4.** 3'-end processing factors (PF) recruited to RNAP IIO during elongation and termination. **B.** RNA processing factors recruited to RNAP IIA at promoter and remain associated with RNAP IIO for entire length of gene. These include transcription regulator DSIF, cap methyl transferase Abd1 and 3' processing factor CPSF. **C.** Phosphorylation status of RNAP II during transcription. **1.** CTD is hypophosphorylated during PIC assembly. **2.** CTD undergoes hyperphosphorylation during transcription initiation when serine-5 is primarily phosphorylated at the promoter region. **3.** Serine-2 phosphorylation is prevalent around transcription start site and throughout transcription elongation and **4.** termination.

Phosphorylation of the RNAP II CTD serves two functions. Firstly, it leads to the release of the PIC from the promoter and secondly, it recruits the capping enzyme and subsequent RNA processing factors to the transcription elongation complex. At the start of transcription, the PIC assembles on the promoter and contains RNAP II with a hypophosphorylated CTD (RNAP IIA) (O'Brien *et al.*, 1994; Weeks *et al.*, 1993). During transcription initiation, the CTD undergoes hyperphosphorylation (RNAP IIO) and consequently activates transition from the initiation complex to a stable elongation complex (Lu *et al.*, 1991; Dahmus, 1994). Phosphorylation of the CTD primarily occurs at serine-2 and serine-5 residues of the heptapeptide consensus sequence (Dahmus, 1996). However, the phosphorylation of these residues is not equivalent but varies along the length of the transcribed gene. The CTD undergoes a cycle of extensive phosphorylation and dephosphorylation of serine-2 and serine-5, which is co-ordinated with the transcription cycle. Serine-5 is predominantly phosphorylated at the promoter region, whereas serine-2 phosphorylation is prevalent around the transcription start site and throughout transcription elongation and termination (Figure 1.2, Panel C 1-4 and Kormarnitsky *et al.*, 2000). This distribution in phospho-serines is thought to result from the activities of different CTD-specific kinases and phosphatases throughout the transcription cycle. Although the identification and timing of such modulators is unclear, in *S. cerevisiae*, serine-5 phosphorylation at the promoter region has been shown to be dependent on the kinase Kin28 (Rodriguez *et al.*, 2000). Whilst in metazoans, serine-2 phosphorylation in the coding region is mediated by the kinase Cdk9, which is a subunit of the positive transcription elongation factor b (P-TEFb) (Zhou *et al.*, 2000).

Genetic studies in *S. cerevisiae* have shown that Kin28 is responsible for phosphorylating serine-5 residues near the gene promoter and recruiting the capping enzyme to the CTD (Figure 1.2, Panel C 2-3 and Rodriguez *et al.*, 2000). Kin28 is a component of the general transcription initiation factor TFIIH that is present in the promoter bound PIC. Phosphorylation of the CTD at serine-5 by Kin28 activates the transition of RNAP IIA to RNAP IIO, thereby positively regulating transcription (Feaver *et al.*, 1994; Svejstrup *et al.*, 1996; Hengartner *et al.*, 1998). In addition, Kin28 mediated phosphorylation directly recruits the capping apparatus to the RNAP

II holoenzyme co-transcriptionally (Rodriguez *et al.*, 2000). As a result, Kin28 appears to be responsible for coupling RNA cap formation with the switch from transcription initiation to elongation. After cap formation, the phosphatase Fcp1 is thought to dephosphorylate phospho-serine-5, which subsequently triggers the dissociation of the capping enzyme from the elongation complex (compare Panels A and C, 2-3 and Chambers and Dahmus, 1994; Cho *et al.*, 1999; Kobor *et al.*, 1999; Schroeder *et al.*, 2000). Although the capping enzyme is released early in elongation, Abd1, the cap methyl transferase appears to remain associated with RNAP IIO until the 3'-end of the gene (compare Panels B and C, 1-4 and Schroeder *et al.*, 2000). Current evidence only supports a role for Abd1 during the capping reaction, however, the association of Abd1 with the elongating polymerase suggests that it might also function in downstream transcription or mRNA processing events.

It is not clear which CTD-specific kinases are responsible for serine-2 phosphorylation at the later stages of transcription. However, in metazoans, a strong candidate is the cyclin-dependent kinase Cdk9, which is a component of the elongation factor P-TEFb (Mancebo *et al.*, 1997; Price, 2000). P-TEFb consists of Cdk9 and several alternative regulatory subunits that include cyclin T1, T2a or T2b, which once associated with Cdk9, increase its kinase activity (Zhu *et al.*, 1997; Peng *et al.*, 1998; Wei *et al.*, 1998). P-TEFb hyperphosphorylates the CTD and stimulates RNAP II transcript elongation *in vivo* (Marshall and Price, 1995; Marshall *et al.*, 1996; Mancebo *et al.*, 1997). In general, RNAP IIO elongation is regulated by a number of positive and negative transcription factors that counteract each others activity. The interplay between these factors shortly after transcription initiation appears to cause a kinetic delay that allows the replacement of initiation factors with elongation and processing factors. As a result, this allows transcription to be coupled with RNA processing (Price, 2000).

Recent data has suggested that in higher eukaryotes, the capping enzymes may be recruited to the PIC independently of the CTD. *In vitro*, the main mechanism that controls P-TEFb stimulated elongation involves DRB inducing sensitivity factor (DSIF) and negative elongation factor (NELF) (Wada *et al.*, 1998a). In mammals,

NELF has five subunits including one called RD, which has a putative RNA-binding domain (Yamaguchi *et al.*, 1999a). Whilst DSIF is a heterodimer comprised of Spt4 and Spt5 subunits, the latter of which interacts with RNAP II (Hartzog *et al.*, 1998; Wada *et al.*, 1998b). Both DSIF and NELF are found in the transcription PIC where they repress the switch from transcription initiation to elongation (Figure 1.2, Panels A and B 1). Hyperphosphorylation of the CTD by P-TEFb disrupts the DSIF-NELF interaction and reverses the negative effect on transcription elongation (Wada *et al.*, 1998a; Yamaguchi *et al.*, 1999a and 1999b). As a result, NELF dissociates from the RNAP IIO elongation complex. In contrast, DSIF remains associated with RNAP IIO to additionally function in targeting cap formation to RNAP II nascent transcripts and facilitating transcription elongation (compare Panels A, B and C 1-4 and Wu-Baer *et al.*, 1998; Wen and Shatkin, 1999; Andrulis *et al.*, 2000; Kaplan *et al.*, 2000; Chiu *et al.*, 2001).

In addition to the phosphorylated CTD, the selective capping of RNAP II transcripts is linked to transcription through the interaction of the capping enzyme with Spt5 (Wen and Shatkin, 1999). Like the phosphorylated CTD, human Spt5 physically interacts with the mammalian capping enzyme and stimulates its guanylyl transferase activity to promote RNA cap formation (Wen and Shatkin, 1999; Chiu *et al.*, 2001). However, it is currently unknown whether Spt5 targets the capping enzyme to RNAP II transcripts prior to, or independent of, the phosphorylated CTD.

Similar to RNA cap formation, pre-mRNA splicing and RNA 3'-end formation are also coupled to transcription through the CTD of RNAP II. Both biochemical and *in vivo* studies have shown that the CTD specifically binds components of the splicing and 3'-end processing machinery to enhance the efficiency of these processes (Hirose and Manley, 2000). The CTD not only recruits processing factors to transcription sites but also appears to function directly with the splicing and 3'-end processing machinery. For example, some members of the SR protein family, that are required for splicing, have been shown to specifically interact with the phosphorylated CTD but not the unphosphorylated CTD (Figure 1.2, Panels A and C 2-3 and Hirose *et al.*, 1999). In contrast binding of 3'-end processing factors occurs

independently of the CTD phosphorylation status (McCracken *et al.*, 1997b; Hirose and Manley, 1998). For example, factors such as cleavage and polyadenylation specificity factor (CPSF) and yeast cleavage factor IB (CFIB/ Hrp1), which is also implicated in mRNA turnover, associate early with RNAP II at the promoter and remain stably associated with the transcribing polymerase for the entire length of the gene (Panels A, B and C 1-4 and Dantonel *et al.*, 1997; Komarnitsky *et al.*, 2000).

1.4 Cap-binding proteins

Many of the effects of the cap in RNAP II transcript metabolism are mediated by proteins that specifically recognise and bind the cap structure. Although several cap-binding proteins (CBPs) have been reported, only two activities are well characterised in eukaryotes, namely the translation initiation factor eIF4E and nuclear cap binding complex (CBC) (Sonenberg, 1978 and 1979; Patzelt *et al.*, 1983; Rozen and Sonenberg, 1987; Izaurralde *et al.*, 1992).

CBC is a heterodimer comprised of CBP20 and CBP80 subunits, both of which are evolutionarily conserved (Ohno *et al.*, 1990; Izaurralde *et al.*, 1994). CBC binds nascent RNAP II transcripts and facilitates the positive effects of the cap in nuclear RNA processing events that include pre-mRNA splicing, RNA 3'-end formation and U snRNA export (Izaurralde *et al.*, 1995; Lewis *et al.*, 1996; Flaherty *et al.*, 1997; Lewis and Izaurralde, 1997). As a result, CBC accompanies pre-mRNA from the site of synthesis until it is processed and exported to the cytoplasm for translation (Gorlich *et al.*, 1996; Visa *et al.*, 1996).

Upon export to the cytoplasm, protein synthesis is initiated by the association of the cap-dependent translation machinery with mRNA. In most cases, ribosomes do not directly bind mRNA but are recruited by a number of eukaryotic initiation factors (eIFs) that position a translationally competent ribosome at the initiation codon. During cap-dependent translation the most important cis-acting sequences are the cap and the poly(A) tail that effectively recruit the translation initiation complex eIF4F. eIF4F is a heterotrimer comprised of eIF4A, eIF4G and eIF4E subunits and serves to direct the translational machinery to the 5'-end of mRNA (Sachs *et al.*, 1997). The

cap-binding component of eIF4F is eIF4E. Whilst eIF4G acts as a molecular bridge between the components of the translational machinery and mRNA. eIF4A is an ATP-dependent RNA helicase, which in conjunction with the associated co-factor eIF4B, unwinds the secondary structures present in the cap-proximal 5' UTR. This promotes the binding of the 40S small ribosomal subunit and associated proteins that scan the mRNA sequence for the initiation codon. After the initiation codon has been recognised, the eIFs dissociate allowing the 60S ribosomal subunit to bind and protein synthesis to be initiated (Kozak, 1989; Pestova *et al.*,2001).

In addition to its function in cap recognition, eIF4F appears to mediate the synergistic effect of the 5' cap and 3' poly(A) tail on cap-dependent translation initiation. eIF4G directly interacts with poly(A) binding protein, which is bound to the poly(A) tail and consequently brings the 5' and 3'-ends into close proximity (Gallie, 1991; Tarun *et al.*, 1997; Wells *et al.*, 1998). This cap-poly(A) synergy is proposed to either facilitate the recycling of ribosomes from the 3'-end back to the 5' cap or enhance the formation of eIF-mRNA complexes (Preiss and Hentze, 1998; Borman *et al.*, 2000).

In order to allow eIF4F to bind the cap, the nuclear CBC must exchange with eIF4E in the cytoplasm. Recent data has shown that in *S. cerevisiae*, eIF4G physically interacts with CBP80 and this association is antagonised by eIF4E (Fortes *et al.*, 2000). In addition, the CBP80-eIF4G interaction appears to be evolutionarily conserved. Mammalian immunoprecipitation experiments have demonstrated a significant nuclear pool of eIF4G *in vivo* that is predominantly associated with CBC (McKendrick *et al.*, 2001). This suggests that in the nucleus, CBC interacts with eIF4G at the 5'-end of mRNA. In the nucleus, eIF4G also interacts directly with poly(A) binding protein, which is bound to the 3' poly(A) tail. As a result, the mRNA is circularized in a manner analogous to that described for mRNA in the cytoplasm. After the mRNP is exported to the cytoplasm, CBC must exchange with eIF4E to initiate cap-dependent translation. The molecular mechanism by which this occurs remains to be elucidated. However, it is thought that the exchange of cap-binding proteins is mediated through eIF4G (McKendrick *et al.*, 2001).

1.5 nuclear Cap-Binding Complex

CBC was initially identified in HeLa cell nuclear extracts as an activity that specifically bound to capped RNAs synthesised *in vitro* (Ohno *et al.*, 1990; Izaurralde *et al.*, 1992). Purification of this cap-binding activity showed that CBC exists as a heterodimer comprising cap-binding proteins of 20 kDa and 80 kDa (Ohno *et al.*, 1990; Izaurralde *et al.*, 1994). Subsequent immunofluorescence studies in HeLa cells have demonstrated that the steady state levels of CBC are predominantly nuclear. In the nucleus, CBC is localised ubiquitously but is excluded from the nucleolus (Kataoka *et al.*, 1994; Izaurralde *et al.*, 1994 and 1995).

Both subunits of the human (Hs) CBC have been cloned and sequenced. The predicted amino acid sequence of HsCBP80 shows that it contains a bipartite nuclear localisation signal (NLS) at the amino-terminus (Mattaj, 1993; Izaurralde *et al.*, 1994). The NLS has been shown to be functional since mutation of this sequence inhibits the nuclear import of HsCBP80. In the absence of HsCBP80, HsCBP20 is not actively transported into the nucleus. This suggests that HsCBP20 is only actively imported into the nucleus in the form of a heterodimer with HsCBP80 (Izaurralde *et al.*, 1995). Recently a multiple position–specific iterated (PSI)-Blast search of the national centre for biotechnology information (NCBI) protein database revealed significant sequence similarities between the middle portion of eIF4G (MIF4G) and domains in NMD2/Upf2 and CBP80 homologues (Ponting, 2000). NMD2/Upf2 is a key component of the surveillance complex that detects premature stop codons as part of the nonsense-mediated decay (NMD) pathway (Czaplinski *et al.*, 1999; Hentze and Kulozik, 1999). Interestingly, recent work in mammalian cells has suggested that both nuclear and cytoplasmic NMD take place in association with CBP80 (Ishigaki *et al.*, 2001). Although the functional significance of the MIF4G domain in RNA degradation is currently unknown, it is thought to mediate protein-protein interactions.

In contrast to HsCBP80, the predicted structure of HsCBP20 essentially consists of a RNA-binding domain (RBD) comprised of two RNP motifs, followed by a carboxyl-tail rich in arginine and glycine (RGG) residues (Izaurralde *et al.*, 1995; Makkerh *et*

al., 1996). Evidence from the X-ray crystal structure of HsCBC together with site-directed mutagenesis has shown that the RBD of HsCBP20 subunit binds capped RNA (Mazza *et al.*, 2001). Despite possessing a RBD, HsCBP20 on its own does not exhibit detectable cap-binding or RNA-binding activity unless complexed with HsCBP80 (Izaurralde *et al.*, 1995). This suggests that the requirement for CBC heterodimerization results from the poor stability of folded CBP20 in the absence of CBP80 (Mazza *et al.*, 2001). The exact role of the carboxyl-terminal auxiliary RGG domain in CBP20 is currently unknown. The RGG domain is predicted to stabilise the interaction between CBC and capped RNA (Birney *et al.*, 1993; Burd and Dreyfuss, 1994). However, there is no direct evidence and the crystal structure of CBC-cap complex will be required to analyse this further.

CBP20 and CBP80 homologues have been identified in a variety of eukaryotes. Comparison of the predicted amino acid sequences has demonstrated that both subunits show a high degree of similarity in human, *Xenopus laevis* and *S. cerevisiae*. This suggests that the function of CBC might also be highly conserved between these species.

1.6 pre-mRNA splicing

After RNAP II nascent transcripts receive a cap to the 5'-end, the pre-mRNAs undergo a number of other processing steps before they are exported to the cytoplasm. These include the generation of a new 3'-end by endonucleolytic cleavage and polyadenylation and the removal of internal non-coding sequences by splicing. For most RNAP II transcribed genes, intron sequences must be removed from the primary transcript before mRNA can be exported for translation. In general, specific mechanisms exist in the nucleus that prevent the export of unspliced RNAs unless they are complexed with proteins that bypass the normal mRNA transport machinery (Nakielny and Dreyfuss, 1997). Consequently, any inappropriately spliced pre-mRNAs are retained in the nucleus and subject to degradation. Recently, a nuclear pathway has been identified in *S. cerevisiae* that rapidly degrades unspliced pre-mRNAs in a 5' to 3' direction by the exonuclease Rat1 and 3' to 5' direction by the exosome complex (Bousquet-Antonelli *et al.*, 2000). Interestingly, CBC has also

been implicated in this pathway, which suggests an additional role for CBC in nuclear RNA degradation (Das *et al.*, 2000).

Introns are recognised and defined by four cis-acting sequences: the 5' splice site, branch point, polypyrimidine tract and 3' splice site. The removal of introns and re-ligation of exons to form mature mRNA occurs within a large dedicated ribonucleoprotein complex termed the spliceosome. The spliceosome consists of five uracil-rich small nuclear ribonucleoprotein particles (U snRNPs) U1, U2, U4/6, U5 that are complexed with 50-100 non-snRNP splicing factors (Graveley, 2001). Spliceosome assembly is a dynamic, ordered process where several stable intermediates have been identified *in vivo* and *in vitro* (Staley and Guthrie, 1998; Reed, 2000). Although the order of spliceosomal assembly is strongly conserved in both human and *S. cerevisiae* systems, the characterised components of these intermediates only show some similarity.

Pre-mRNA splicing can be conceptually divided into three distinct stages. The initial stage encompasses the recognition of conserved intronic sequences located near the 5' splice site and branchpoint region by a subset of splicing factors. This is followed by the assembly of multiple additional splicing factors that form the mature spliceosome. Rearrangements then occur within the spliceosome that activate the two chemical steps of intron removal. Finally, spliced mRNA is released for export to the cytoplasm whilst the intronic RNA is degraded and the splicing factors are recycled (Moore *et al.*, 1993).

Figure 1.3 describes the splicing cycle in more detail. The first defined step in splicing consists of pre-spliceosomal E complex formation in mammals or commitment complex formation in *S. cerevisiae*. The E complex assembles on the hnRNP in an ATP-independent manner and serves to commit pre-mRNA to the splicing pathway. In E complex, U1 snRNP is bound to the 5' splice site and U2AF (U2 snRNP auxiliary factor) is associated with the polypyrimidine tract.

Figure 1.3 Spliceosomal Assembly Cycle

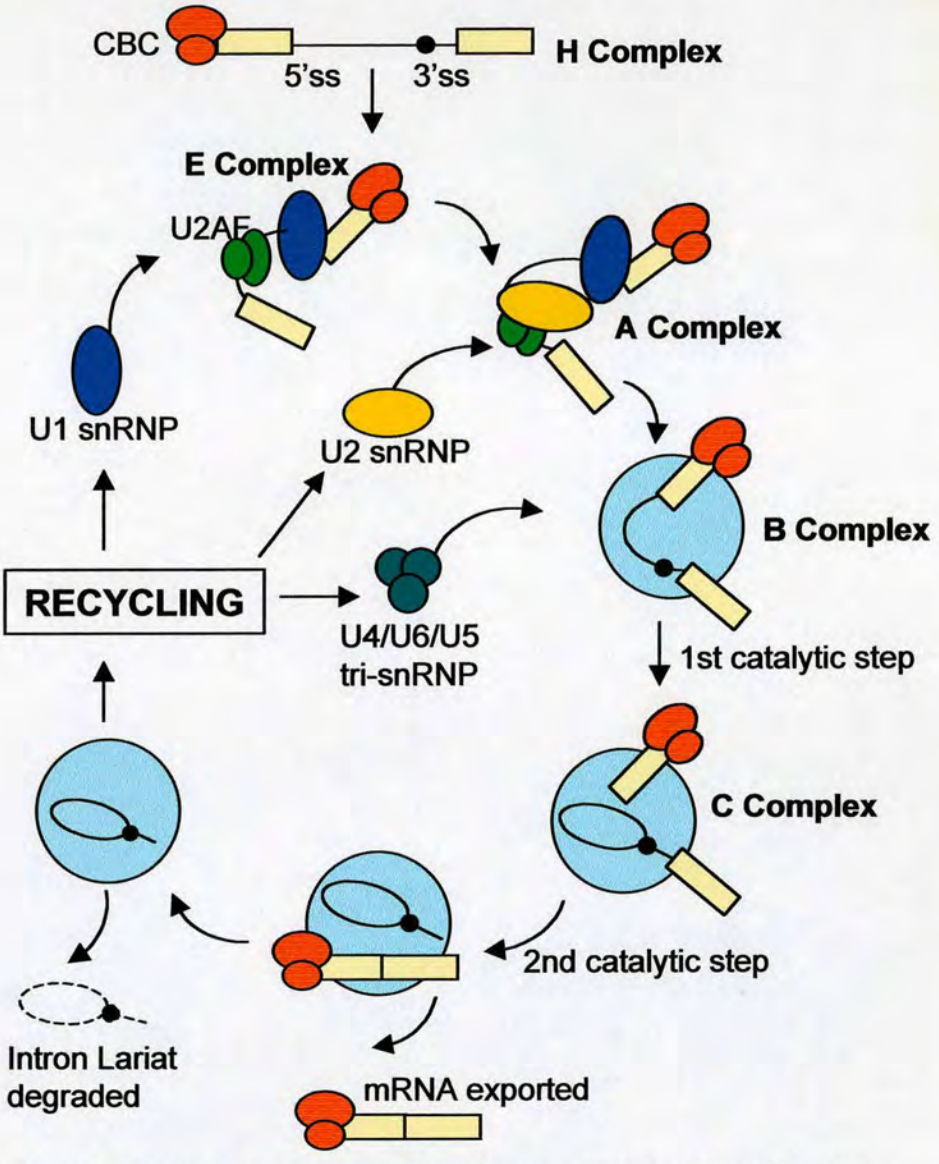


Figure 1.3 Spliceosomal assembly pathway. Simplified description of the splicing cycle. **H complex.** pre-mRNA is associated with RNA-binding proteins including CBC bound to 5' cap. **E complex.** CBC facilitates association of U1 snRNP with 5' splice site (5' ss) and U2 auxiliary factor (U2AF) binds to polypyrimidine tract (black circle) at 3' splice site (3' ss). **A complex.** U2 snRNP binds to the branch point sequence. This step is ATP-dependent as are all of the subsequent steps in the cycle. **B complex.** Addition of U4/U6.U5 tri-snRNP particle forms the active spliceosome. Two sequential transesterification reactions occur within the spliceosome that produce spliced mRNA and intron lariat. mRNA is exported to the cytoplasm for translation and intron lariat is degraded in the nucleus. The spliceosome disassembles and the components are recycled onto another pre-mRNA for subsequent rounds of the splicing cycle.

This promotes the binding of U2 snRNP to the branch point sequence to form the A pre-spliceosomal complex. U2 snRNP addition is ATP-dependent as are all of the subsequent steps in the splicing pathway. The addition of the pre-formed U4/U6.U5 tri-snRNP particle leads to the formation of the active spliceosome or B complex. Within the spliceosome, numerous ATP-dependent rearrangements occur during the course of two sequential transesterification reactions. After the completion of the splicing reaction, spliced mRNA is released and processed further in the nucleus, whilst the intron lariat is debranched and degraded. The spliceosome disassembles and the U snRNPs and non-snRNP proteins are recycled onto another pre-mRNA molecule for subsequent rounds of the splicing cycle.

1.7 Role of CBC in pre-mRNA splicing

A role for the cap in stimulating pre-mRNA splicing was originally shown by *in vitro* studies in mammalian extracts and *in vivo* experiments in *Xenopus* oocytes (Konarska *et al.*, 1984; Krainer *et al.*, 1984). Evidence from several laboratories demonstrated that splicing of a pre-mRNA containing a single intron can be inhibited *in vitro* by cap analogues or *in vivo* by short capped competitor RNAs. Furthermore, this stimulatory effect of the cap in pre-mRNA splicing is restricted to the cap-proximal intron (Konarska *et al.*, 1984; Edery and Sonenberg, 1985; Ohno *et al.*, 1987; Patzelt *et al.*, 1987; Inoue *et al.*, 1989). Subsequent work has shown that CBC is responsible for mediating the positive effects of the cap on splicing. HeLa cell nuclear splicing extracts immunodepleted of CBC exhibit a strong reduction in splicing activity and this can be reversed by the addition of recombinant CBC (Izaurralde *et al.*, 1994). Neither HsCBP20 or HsCBP80 on its own is able to restore splicing to the depleted extract, which demonstrates that only the two subunits in the form of a heterodimer constitute the active complex (Izaurralde *et al.*, 1994; Lewis *et al.*, 1996). Characterisation of the splicing defect showed that in the absence of CBC, U1 snRNP fails to associate efficiently with the cap-proximal 5'splice site to form the E complex (Lewis *et al.*, 1996). Since spliceosome assembly is an ordered stepwise process, inhibition of the earliest step inhibits the formation of the mature spliceosome and subsequent splicing.

Interestingly, CBC is also found stably associated with pre-mRNA in the H and B complexes, suggesting that CBC associates with pre-mRNA before entering the splicing pathway and remains bound to the cap when mRNA leaves the spliceosome (Figure 1.3 and Lewis *et al.*, 1996). Despite remaining bound to capped pre-mRNA throughout spliceosome assembly, current evidence only supports a role for CBC during E complex. However, it is important to note that CBC might also function at later stages in the splicing cycle (O'Mullane and Eperon, 1998).

Collectively these data show that CBC increases the efficiency of U1 snRNP binding to the cap-proximal 5' splice site. This enhances E complex formation and the rate of cap-proximal intron splicing. In vertebrates, the molecular mechanism by which CBC functions during E complex formation is currently unknown. However, in *S. cerevisiae* CBC has been shown to interact directly with components of U1 snRNP.

1.8 CBC's function in splicing is conserved from yeast to humans

As mentioned earlier, the components of the spliceosomal assembly pathway show a high degree of similarity within the *S. cerevisiae* and mammalian systems suggesting that the mechanism of splicing is also conserved (Colot *et al.*, 1996; Lewis *et al.*, 1996; Lewis and Izaurralde, 1997). The yeast equivalent to E complex is called the commitment complex since its formation commits pre-mRNA to the splicing pathway. Like E complex, the commitment complex forms in the absence of ATP after the association of U1 snRNP with the 5' splice site and Mud2p, the yeast homologue of U2AF, with the polypyrimidine tract (Lewis and Izaurralde, 1997).

In *S. cerevisiae* (Sc), *MUD13* and *GCR3* encode CBP20 and CBP80 respectively (Uemura and Jigami, 1992; Colot *et al.*, 1996; Gorlich *et al.*, 1996; Lewis *et al.*, 1996b). *ScCBP20* and *ScCBP80* are not essential genes, which suggests that CBC is not required for vegetative growth. *ScCBP20* was originally identified in a synthetic lethal screen designed to identify genes involved in commitment complex assembly (Colet *et al.*, 1996). Similar to the experiments in the HeLa system, *S. cerevisiae* splicing extracts immunodepleted of endogenous CBC exhibit a strong reduction in

splicing activity that is partially restored by the addition of purified ScCBC (Lewis *et al.*, 1996b). Characterisation of the splicing defect showed that in CBC-depleted extracts the formation of commitment complex is deficient (Lewis *et al.*, 1996b).

Although the molecular mechanism by which CBC functions during E/ commitment complex assembly has not been elucidated, recent work in *S. cerevisiae* has shown that CBC interacts directly with components of U1 snRNP and Mud2 (Fortes *et al.*, 1999a; Fortes *et al.*, 1999b). This is thought to promote the formation of a molecular bridge between the 5' cap and 3' splice site and consequently bring both ends of the intron into close proximity for splicing (Figure 1.4). Both yeast and vertebrate U1 snRNPs consist of a core of three Sm proteins and three U1-specific proteins (U1 70K/ Snp1, U1A/ Mud1, U1C/ yUI-C). In addition, yeast U1 snRNP contains at least six specific proteins that currently have no characterised vertebrate homologues. In *S. cerevisiae*, synthetic lethal interactions have been found between CBC and components of the Sm core, U1-specific and yeast-specific U1 snRNP proteins (Figure 1.4). Moreover, genetic and biochemical analysis has shown that CBC physically interacts with Snu56, a yeast-specific component of U1 snRNP, which currently has no identified vertebrate homologue (Fortes *et al.*, 1999a; Fortes *et al.*, 1999b). As a result, this demonstrates that in *S. cerevisiae*, CBC facilitates U1 snRNP recruitment to the 5' splice site through direct contacts with the components of U1 snRNP.

A genetic and physical interaction has also been identified between CBC and Mud2, which is the yeast homologue of U2AF 65 (Fortes *et al.*, 1999a; Fortes *et al.*, 1999b). Mud2 functions in commitment complex assembly where it binds to the polypyrimidine tract and subsequently promotes the binding of U2 snRNP. Interestingly, Mud2 is not required for facilitating the interaction between U1 snRNP and the pre-mRNA 5' splice site. This suggests that the role of CBC in splicing might extend beyond the U1 snRNP function and CBC might also participate at later stages of the splicing cycle (O'Mullane and Eperon, 1998).

Figure 1.4 Model of Yeast Commitment Complex

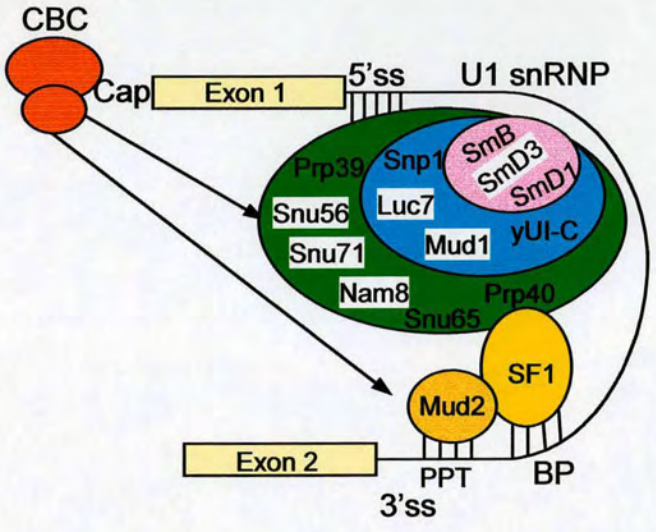


Figure 1.4 Model of yeast commitment complex. A molecular bridge between 5' cap and 3' splice site (3'ss) brings both ends of the intron into close proximity for splicing. CBC-cap facilitate the association of U1 snRNP with 5' splice site (5' ss). Mud2 associates with polypyrimidine tract (PPT) and SF1 binds the branch point sequence (BP). SF1 also physically interacts with Mud2 along with Prp40 component of U1 snRNP. Synthetic lethal interactions have been found between CBC and components of U1 snRNP core Sm proteins, U1-specific proteins and U1-yeast specific proteins (highlighted in white). Direct interactions have been found between CBC and Snu56 and Mud2 (black arrows).

Collectively, these data demonstrate that in splicing, CBC is found to play an analogous role in both human and *S. cerevisiae* systems. Both yeast and mammalian CBC function during the earliest stage of spliceosomal assembly where they increase the efficiency of cap-proximal intron recognition by U1 snRNP. However in *S. cerevisiae*, CBC appears to mediate its effect through a direct interaction with U1 snRNP. This contrasts with the mammalian system, where there is no evidence of a direct interaction and the adaptor molecules that facilitate the effects of CBC in splicing are yet to be elucidated (Lewis *et al.*, 1996).

1.9 Exon definition model: CBC is required for efficient splicing of the cap-proximal intron

The exon definition model states that efficient pre-mRNA splicing relies on the definition of exons as opposed to introns. Exons are defined through co-operative interactions between splicing factors localised on the 3'-end of introns and the downstream 5' splice site across the exon (Robberson *et al.*, 1990). This model is favoured by the small size of mammalian exons, around 137 nucleotides, in addition to the increasing evidence that supports the existence of co-operative interactions across exons (Hoffman and Grabowski, 1992; Zeng and Berget, 2000). However, a problem with the exon definition model is that it fails to explain how the terminal exons are recognised by the components of the spliceosome. An alternative mechanism of intron recognition must exist since the 5' terminal exon has a cap instead of 3' splice site and the 3' terminal exon has a polyadenylation signal instead of 5' splice site.

As mentioned earlier, evidence from mammalian extracts *in vitro* and *Xenopus* oocytes *in vivo* have shown that the cap is only required for efficient splicing of the cap-proximal intron (Ohno *et al.*, 1987; Inoue *et al.*, 1989). This suggests that the first exon might be defined by interactions between the cap and the 5' splice site. However, it does not address why the cap-distal intron exhibits cap-independent splicing. Subsequent experiments that mutated splice site consensus sequences showed that the presence of an upstream polypyrimidine tract is equivalent in

function to the cap for efficient recognition of an internal exon (Lewis *et al.*, 1996). As a result, the cap together with CBC facilitate the association of U1 snRNP with the first functional cap-proximal 5' splice site. Whilst the polypyrimidine tract substitutes for CBC-cap function to define the internal exons. Similarly, the polyadenylation signal in combination with the terminal 3' splice site participates in the definition of the terminal exon.

1.10 Coupling splicing and transcription

Early cytological studies first demonstrated a possible link between splicing and RNAP II transcription. Several laboratories showed that splicing factors are specifically localised to sites of active transcript synthesis (Beyer and Osheim, 1988; Bauren and Wieslander, 1994; Zhang *et al.*, 1994). Subsequent biochemical and *in vivo* studies provided further support for the existence of functional interactions between RNAP II and the splicing apparatus. These included data where the phosphorylated form of RNAP II is found associated with active spliceosomes as well as splicing factors (Chabot *et al.*, 1995; Mortillaro *et al.*, 1996 Yuryev *et al.*, 1996; Kim *et al.*, 1997). Similar to RNA cap formation, experiments using transiently transfected mammalian cells showed that splicing of RNAs transcribed by CTD-truncated RNAP II is inefficient (McCracken *et al.*, 1997b). In addition, over-expression of phosphorylated CTD peptides also inhibits splicing in cultured mammalian cells (Du and Warren, 1997). Collectively, these observations suggest that the hyperphosphorylated CTD is required for targeting splicing factors to transcription sites to ensure efficient splicing.

Recent experiments, in the absence of transcription, have shown that RNAP II also plays a direct and active role in splicing *in vitro*. Purified RNAP IIO strongly activates splicing of several different pre-mRNAs in reconstituted splicing assays (Hirose *et al.*, 1999). RNAP IIO appears to increase the formation of early spliceosomal complexes, most notably A complex. This suggests that RNAP IIO might stimulate splicing by accelerating the rate of the first steps of spliceosome assembly. Taken together, these data indicate that RNAP II not only couples splicing

with transcription by targeting splicing factors to transcription sites but also directly functions with splicing factors to enhance splicing efficiency.

Although it is well established that transcription and splicing are coupled, the molecular mechanisms by which these processes are integrated remains to be elucidated. However, recent work has proposed that SR and SR-like CTD associated factors (SCAFs) might be candidate adaptor molecules that mediate the stimulatory effects of the CTD in splicing. Immunoprecipitates of RNAP IIO are found to contain nuclear matrix components that include members of the serine-arginine (SR) and Sm families of splicing factors (Kim *et al.*, 1997; Mortillaro *et al.*, 1996; Vincent *et al.*, 1996). Although these data support interactions between the phosphorylated CTD and SR proteins, it is unclear whether these interactions occur directly or indirectly.

Direct contacts have been documented between the CTD and the SCAF family of nuclear matrix proteins (Patturajan *et al.*, 1998). SCAFs were initially identified in a yeast two-hybrid screen for CTD-interacting proteins and interact with the CTD via a distinct CTD-interaction domain (Yuryev *et al.*, 1996). Moreover, a subsequent study showed that one SCAF, SCAF8 only interacts with the phosphorylated CTD and is localised to foci that overlap with sites of transcription and processing (Patturajan *et al.*, 1998). Although these results support a role for SCAFs in coupling RNAP II transcription to processing, there is no direct evidence for this and SCAFs do not currently have documented functions in RNA processing events.

1.11 RNA 3'-end formation

Many mRNA precursors are processed at the 3'-end by endonucleolytic cleavage and polyadenylation to form the poly(A) tail. The only known exceptions in metazoan organisms are the major histone mRNAs. Although the histone pre-mRNAs are cleaved, they are not polyadenylated and the factors responsible for the cleavage reaction differ from those acting on all other pre-mRNAs (Dominski and Marzluff, 1999). The poly(A) tail is comprised of a homopolymer of adenosine residues, which after synthesis in the nucleus, is typically 250 nucleotides in length (Minvielle-Sebastia and Keller, 1999). Unlike the 5' cap modification, the polyadenosine tail is a highly dynamic structure that is subject to length alterations during *Xenopus, Drosophila* and mouse development. Changes in poly(A) tail length occur in the cytoplasm and serves to regulate translation (Wickens *et al.*, 1997; Gray and Wicken, 1998).

The poly(A) tail has been proposed to play a role in nearly every aspect of mRNA metabolism. As mentioned earlier, the poly(A) tail and the cap have a synergistic effect on cap-dependent translation initiation. This cap-poly(A) synergy is thought to either facilitate the recycling of ribosomes from the 3'-end back to the 5' cap or enhance the formation of eIF-mRNA complexes (Preiss and Hentze, 1998; Borman *et al.*, 2000). In addition, the poly(A) tail plays a role in maintaining mRNA stability as deadenylation is frequently the first and rate-limiting step of mRNA decay (Beelman and Parker, 1995). Finally, the poly(A) tail is thought to be involved in the export of mRNAs from the nucleus to the cytoplasm for translation (Eckner *et al.*, 1991; Huang and Carmichael, 1996).

New 3'-ends are generated in a two step reaction that is tightly coupled *in vivo*. The pre-mRNA is first cleaved endonucleolytically followed by poly(A) addition to the upstream cleavage fragment (Colgan and Manley, 1997). In higher eukaryotes, RNA 3'-end formation requires two cis-acting sequence elements present in the 3' UTR, which together define the cleavage/ polyadenylation site. The AAUAAA polyadenylation signal is a hexanucleotide repeat that is found 10-30 nucleotides

upstream of the cleavage site. AAUAAA is found in 90% of all sequenced polyadenylation elements and is one of the most highly conserved elements known (Proudfoot, 1991; Wahle and Kuhn, 1997). The second sequence element is the GU-rich motif located 20-40 nucleotides downstream of the cleavage site. In contrast to AAUAAA, the GU-rich element is found in 70% of mammalian pre-mRNAs and is considerably more variable in sequence composition (Wahle and Ruegsegger, 1999). Together, AAUAAA and GU-rich elements and their distance from each other specify the cleavage/ polyadenylation site and determine the strength of the poly(A) signal (MacDonald *et al.*, 1994; Chen *et al.*, 1995).

To date, at least seven trans-acting factors are found to be required for the *in vitro* reconstitution of mammalian RNA 3'-end cleavage and polyadenylation (Wahle and Ruegsegger, 1999; Zhao *et al.*, 1999). The cleavage reaction requires a yet unidentified endonuclease as well as cleavage stimulation factor (CstF), cleavage factors 1 and 2 and the CTD of RNAP II. Poly(A) polymerase (PAP) and the cleavage and polyadenylation specificity factor (CPSF) are required in both cleavage and polyadenylation steps. Whilst, poly(A)-binding protein 2 (PABP2) only functions during polyadenylation.

Although the cis and trans-acting factors required for RNA 3'-end formation are well characterised, the molecular mechanisms involved remain unclear. Figure 1.5 illustrates a current model for the formation of the mammalian RNA 3'-end cleavage complex. Together, the components of the cleavage complex are thought to establish a network of weak co-operative interactions that define the cleavage/ polyadenylation site. CPSF binds directly to the cleavage and polyadenylation signal, AAUAAA and CstF associates with the GU-rich sequence downstream of the cleavage site. The resultant protein-RNA complex then recruits other 3'-end processing factors that include cleavage factors I, II and PAP (Barabino and Keller, 1999).

Figure 1.5 Model of Mammalian RNA 3'-end Cleavage Complex

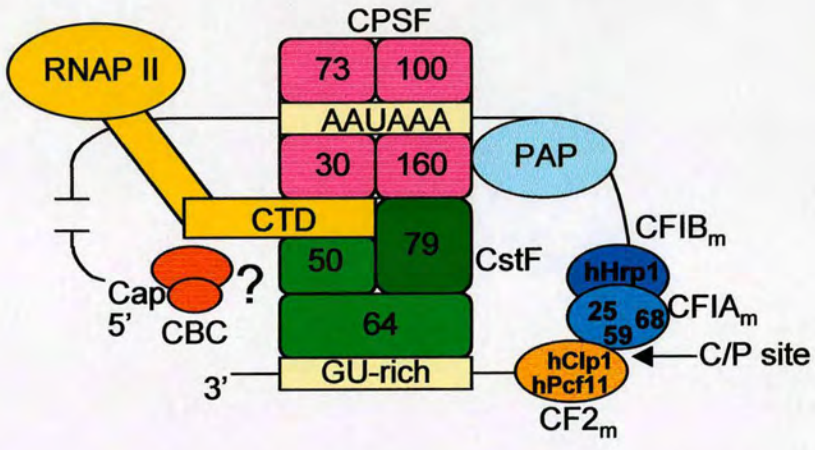


Figure 1.5 Model of mammalian RNA 3'-end cleavage complex. Components of cleavage complex establish a network of weak co-operative interactions that define the cleavage/ polyadenylation site (C/ P site). Cleavage and polyadenylation specificity factor (CPSF) is comprised of four subunits and binds the cleavage and polyadenylation signal, AAUAAA. The heterotrimeric cleavage stimulation factor (CstF) binds the GU-rich signal. This recruits other cleavage factors that include: cleavage factor I ($CFIA_m$ + $CFIB_m$) and 2 ($CF2_m$) along with poly(A) polymerase (PAP). The CTD of RNAP II also participates as a cleavage factor and physically interacts with CPSF and CstF. Both the 5' and 3' ends of RNA contribute to the 3' cleavage reaction. However, the adaptor molecules that mediate the positive effects of CBC-cap are currently unknown.

Recently, the CTD of RNAP II has also been shown to actively participate in RNA cleavage and this role of RNAP II in 3'-end processing will be reviewed separately (Hirose and Manley, 1998). After the cleavage/ polyadenylation site has been defined, the RNA phosphodiester backbone is cleaved by an unknown endonuclease and the polyadenylation complex is recruited. How the cleavage complex makes the transition into the polyadenylation complex is currently not known. However, during polyadenylation CPSF remains associated with the AAUAAA sequence and PAP catalyses the addition of the poly(A) tail. PABP2 subsequently binds the poly(A) and is thought to increase the efficiency of polyadenylation as well as specify the correct length of poly(A) tail (Wahle, 1991; Wahle *et al.*, 1993; Colgan and Manley, 1997) After approximately 250 nucleotides have been synthesised, processive poly(A) elongation is terminated and the mRNA is exported to the cytoplasm for translation.

1.12 Role of CBC in RNA 3'-end formation

The requirement of the cap for efficient 3'-end processing was first demonstrated by *in vitro* studies in HeLa extracts using non-intron containing RNA precursors. The addition of cap analogue to HeLa cell nuclear extracts was found to reduce poly(A) site cleavage (Hart *et al.*, 1985; Cooke and Alwine, 1996). Furthermore, uncapped pre-mRNAs had poorly processed 3'-ends and competed less efficiently for 3'-end processing factors than capped pre-RNAs (Gilmartin *et al.*, 1988; Cooke and Alwine, 1996). The requirement of the cap for efficient 3'-end formation is restricted to non-intron containing genes. This stimulatory effect is abrogated by introducing an intron upstream of the polyadenylation signal (Cooke and Alwine, 1996). Subsequent immunodepletion studies showed directly that the positive effect of the cap in RNA 3'-end formation is mediated by CBC. CBC is required for the cleavage of pre-mRNA but not for the polyadenylation reaction *per se*. (Flaherty *et al.*, 1997).

Taken together, these results suggest that the primary role of CBC in RNA 3'-end processing is to facilitate the cleavage reaction. However, the mechanism by which CBC mediates this function of the cap remains to be determined. Since most of the factors required to reconstitute pre-mRNA 3'-end cleavage *in vitro* have been

identified, a model of the constituents of the cleavage complex can be proposed. Figure 1.5 highlights how both the 5' and 3'-ends of eukaryotic RNAs contribute to RNA 3'-end formation. In order for the 5' and 3'-ends of eukaryotic RNAs to communicate, a network of protein-protein interactions must exist to define the cleavage/ polyadenylation site. Although CBC and the cap have been shown to directly participate in the cleavage reaction, the adaptor molecules that mediate CBC function are yet to be elucidated.

1.13 Coupling RNA 3'-end formation with transcription

Similar to capping and splicing, RNA 3'-end formation is also coupled to RNAP II transcription. Together, biochemical and *in vivo* studies support the existence of functional interactions between the CTD and 3'-end processing machinery. RNAs transcribed by CTD-truncated RNAP II are not efficiently polyadenylated in transiently transfected HeLa cells (McCracken *et al.*, 1997b). In addition, *in vitro* binding experiments show that CPSF and the 50 K subunit of the heterotrimeric CstF physically interact with the CTD. However, unlike capping and splicing, the phosphorylation status of the CTD does not appear to affect the binding of 3'-end processing factors (McCracken *et al.*, 1997b). Interestingly, three of the four subunits of CPSF are also associated with the general transcription factor TFIID that specifically recognises the TATA box of promoter DNA. Furthermore, reconstituted transcription assays show that CPSF is transferred from TFIID to RNAP II concomitantly with transcription initiation (Dantonel *et al.*, 1997). This suggests that in contrast to capping and splicing, 3'-end processing factors can associate early with the transcription PIC. Such factors appear to remain stably associated with the elongating polymerase for the entire length of the gene until they dissociate at the polyadenylation signal to define the cleavage site (Dantonel *et al.*, 1997; McCracken *et al.*, 1997b).

Finally as with capping and splicing, the RNAP II CTD not only recruits processing factors to sites of transcription but also appears to have a direct role in RNA 3'-end formation. In the absence of transcription, the CTD directly activates *in vitro* cleavage reactions. HeLa nuclear extracts immunodepleted of RNAP II exhibit a

reduction in RNA cleavage and this inhibition is restored by the addition of purified enzyme (Hirose and Manley, 1998). Collectively, these results support the hypothesis that the CTD is a component of the RNA 3'-end processing apparatus where it participates in the formation and function of the cleavage complex through direct interactions with other 3'-end processing factors (Figure 1.5).

1.14 RNA nucleocytoplasmic transport

After pre-mRNAs have been capped, spliced and polyadenylated they are exported to the cytoplasm for translation into proteins. Transport in and out of the nucleus occurs through aqueous channels called nuclear pore complexes (NPCs) that penetrate the nuclear envelope. NPCs are approximately 9 nm in diameter and allow the passive diffusion of small molecules and globular proteins less than 60 kDa (Stoffler *et al.*, 1999; Ryan and Wente, 2000). In contrast, the nucleocytoplasmic transport of larger macromolecules such as proteins and RNPs is an active, receptor-mediated process. During active transport, the diameter of the NPC can open to approximately 25 nm indicating that the NPC recognises and reacts to specific transport substrates by undergoing a considerable change in conformation. The transport of molecules across the nuclear envelope requires soluble transport factors. In recent years many of the key receptor and adaptor proteins that control nucleocytoplasmic transport have been identified. As a result, this has lead to a greater understanding of the general mechanisms involved in the nuclear import and export of different classes of RNA.

In order to undergo nucleocytoplasmic transport, different classes of cargo must contain a transport signal that is recognised by a specific receptor protein. In the simplest cases, the cargo is bound by a transport receptor, translocated through the NPC via receptor-mediated interactions and released on the other side. The cargo is released and the empty receptor is recycled to the original compartment to participate in new transport events. Alternatively, the cargo-receptor interaction is not direct but mediated by one or more adaptor proteins. Consequently, both receptor and adaptor proteins have to be recycled after cargo transport and such proteins are found to shuttle continuously between the nuclear and cytoplasmic compartments.

A critical characteristic of nucleocytoplasmic transport is directionality. This ensures cargoes that have completed translocation do not bind receptors and risk being transported to their original compartment. Directionality of all the transport processes studied so far is thought to come from the small GTPase Ran. Although Ran is found in both the nucleus and cytoplasm, regulation of its nucleotide exchange and hydrolysis activity ensures that Ran is asymmetrically distributed within the cell. Ran GTPase activating protein (Ran GAP) is localised to the cytoplasm and Ran GTP exchange factor (Ran GEF) is localised to the nucleus, where it is stably bound to chromatin (Koepp and Silver, 1996; Gorlich, 1997). As a result, nuclear Ran is in the GTP-bound state whereas cytoplasmic Ran is GDP-bound. Import receptors that bind Ran GTP release their cargo. In contrast, export receptors can only bind cargo in the presence of Ran GTP (Azuma and Dasso, 2000). Consequently, Ran GTP releases import cargoes into the nucleus whilst the hydrolysis of GTP promoted by Ran GAP releases export cargoes in the cytoplasm. This Ran GTP gradient across the nuclear envelope provides the directionality necessary for nuclear transport.

The major classes of cellular RNAs that are synthesised in the nucleus must be exported to the cytoplasm to either function in translation (mRNA, rRNA and tRNA) or undergo maturation prior to being re-imported to the nucleus (major spliceosomal U snRNAs). Microinjection studies using *Xenopus* oocytes have shown that the export pathway of each RNA species is uniquely saturable and mediated by class-specific receptors and/ or adaptors (Jarmolowski *et al.*, 1994). Interestingly, the only export receptor known to bind directly to RNA is exportin-t, which is responsible for the export of mature tRNA to the cytoplasm (Arts *et al.*, 1998; Hellmuth *et al.*, 1998; Kutay *et al.*, 1998). In comparison, nuclear export of other classes of RNA appear to be mediated by factors other than export receptors and adaptors.

Key players in mRNA export include TAP and hnRNPs (Reed and Hurt, 2002). TAP is the cellular cofactor required for the nuclear export of unspliced retroviral RNA containing a constitutive transport element. TAP is thought to heterodimerize with p15 to stimulate the export of cellular mRNAs (Wilkie *et al.*, 2001). Nuclear export signals have also been identified in the shuttling hnRNP proteins. One of the best

characterised is the M9 sequence in hnRNPA1. Althought there is no direct evidence, current work strongly suggests the M9 sequence plays a role in mRNA export (Reed and Hurt, 2002).

In addition to trans-acting factors, the cis-acting sequences of RNAs also stimulate RNA export. Important cis-acting sequences include the cap and poly(A) tail that after recognition ensure efficient nuclear export. The following sections describe the contribution of the cap and CBC to the nucleocytoplasmic transport of U snRNAs, snoRNAs and mRNAs.

1.15 U snRNA biogenesis

In order to understand the role of CBC and the cap in U snRNA nuclear export, it is important to describe the export in context of U snRNA biogenesis. The biogenesis of U snRNP complexes requires multiple transport events between the nucleus and cytoplasm. After RNAP II transcription, all the major spliceosomal U snRNAs with the exception of U6 snRNA are capped. The pre-U snRNAs are then exported to the cytoplasm where they interact in an ordered, stepwise manner with seven Sm proteins to form the snRNP Sm core structure. The cap is subsequently hypermethylated and the 3'-ends of U snRNPs are processed (Mattaj, 1986; Chanfreau *et al.*, 1997). After the completion of cytoplasmic processing, newly assembled core U snRNPs are re-imported to the nucleus where they function in splicing. Nuclear import is mediated by a bipartite signal comprised of the U snRNP hypermethylated cap and a yet unidentified Sm core NLS receptor, which is recognised by the snurportin1-importin import receptor complex. In the nucleus, U snRNAs are internally modified at several positions and biogenesis is completed by the association of U snRNP-specific proteins (Fischer *et al.*, 1993; Palacios *et al.*, 1997; Huber *et al.*, 1998). Mature U snRNPs are then available to participate in the spliceosomal assembly pathway where they function in pre-mRNA splicing.

1.16 Role of CBC in U snRNA nuclear export

Early *Xenopus* oocyte studies demonstrated a stimulatory role for the cap in the nuclear export of the major RNAP II transcribed spliceosomal U snRNAs. Nuclear export of capped U snRNAs either microinjected into the oocyte nucleus or transcribed *in vivo* could be efficiently inhibited by an excess of cap analogue (Hamm and Mattaj, 1990; Jarmolowski *et al.*, 1994). Subsequent experiments directly showed that U snRNA nuclear export is inhibited by antibodies directed against CBP20. This inhibition could be specifically reversed by pre-incubating anti-CBP20 antibody with excess recombinant CBP20 (Izaurralde *et al.*, 1995). Although immuno-localisation experiments have demonstrated that the steady state levels of CBC are mainly nuclear, recent evidence has shown that CBC does indeed shuttle between the nuclear and cytoplasmic compartments (Izaurralde *et al.*, 1995; Gorlich *et al.*, 1996; Ohno *et al.*, 2000; Shen *et al.*, 2000). Taken together, these data suggest that the positive effect of the cap in U snRNA export is mediated by CBC.

In addition to the cap, other factors implicated in U snRNA export are Ran GTP, CRM1 or exportin1 (Xop1) and phosphorylated adaptor for RNA export (PHAX). The nuclear export of capped U snRNAs can be inhibited by reducing the nuclear concentration of Ran GTP (Izaurralde *et al.*, 1997). This is expected since Ran GTP has been shown to provide directionality to all transport processes studied so far. Likewise, U snRNA export is also inhibited by inactivating the export receptor Xpo1 with leptomycin B (Fischer *et al.*, 1995; Fornerod *et al.*, 1997). Since Xpo1 specifically binds cargo carrying leucine-rich nuclear export signal (NES), this suggests an adaptor bridges the interaction between Xpo1 and CBC-capped U snRNAs. Recent work has demonstrated the NES-containing adaptor molecule is PHAX (Fornerod 1997; Ohno 2000). Figure 1.6 describes the current model of CBC-dependent U snRNA export as mediated by Xpo1 and PHAX.

Figure 1.6 Model of U snRNA Nuclear Export

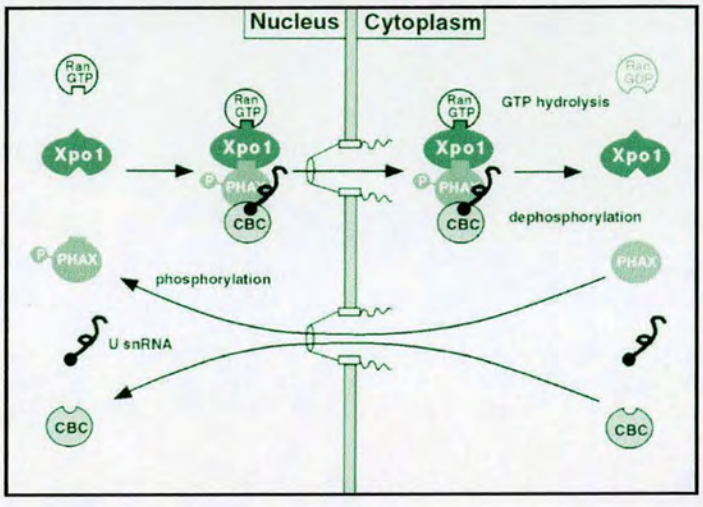


Figure 1.6 Model of U snRNA nuclear export. After RNAP II transcription, CBC associates with capped U snRNA precursors. In the presence of Ran GTP, phosphorylated PHAX mediates the interaction between CBC-capped U snRNA and export receptor Xpo1-Ran GTP This functional export complex is then translocated through the NPC by receptor-mediated processes. After nuclear export, PHAX is dephosphorylated and Ran-GTP is hydrolysed, which leads to U snRNP export complex disassembly. In the cytoplasm, U snRNA precursors undergo cap hypermethylation and RNA 3'-end processing before nuclear re-import to form functional U snRNPs that participate in splicing. PHAX and CBC are recycled to the nucleus for further rounds of export. Figure taken from Ohno *et al.* 2000.

After RNAP II transcription, CBC associates with capped U snRNA precursors in the nucleus. In the presence of Ran GTP, the phosphorylated form of PHAX mediates the interaction between CBC-capped U snRNA and Xpo1-Ran GTP (Figure 1.6). This forms a functional export complex that is translocated through the NPC by receptor-mediated processes. After nuclear export, PHAX is dephosphorylated and Ran-GTP is hydrolysed, which leads to the dissociation of the U snRNP export complex. During U snRNP nuclear export, directionality appears to be conferred by both the asymmetric distribution of Ran GTP and the differential phosphorylation status of PHAX. In the cytoplasm, U snRNA precursors undergo cap hypermethylation and 3'-end processing. Mature U snRNAs are then re-imported to the nucleus to form functional U snRNPs that participate in pre-mRNA splicing (Ohno *et al.*, 2000).

1.17 Nuclear transport of snoRNA

Similar to the major spliceosomal U snRNAs, U3 and U8 small nucleolar (sno) RNAs are also transcribed by RNAP II and their initial cap structure subsequently becomes hypermethylated. U3 and U8 snoRNAs are members of a RNA family that is defined by their nucleolar localisation and association with the nucleolar protein fibrillarin. Both U3 and U8 snoRNAs participate in ribosomal RNA biogenesis where they are essential for pre-rRNA processing (Kass *et al.*, 1990; Peculis and Steitz, 1993; Peculis, 1997). However, unlike the spliceosomal U snRNAs, which are exported to the cytoplasm for cap hypermethylation, U3 and U8 snoRNAs undergo cap hypermethylation within the nucleus. Current evidence suggests that similar to U snRNA export, the cap also contributes to the nucleolar localisation of U3 and U8 snoRNAs.

Microinjection studies using rat kidney epithelial cells have shown that nucleolar localisation of human U3 and U8 snoRNAs is facilitated by the cap. U3 and U8 snoRNAs that contain alternative cap analogues localise less efficiently compared to capped snoRNAs. Furthermore, U3 snoRNA nucleolar localisation is inhibited by an excess of cap analogue. This suggests the stimulatory effects of the cap are facilitated by a cap-binding activity (Jacobson and Pederson, 1998). However, there is no

evidence available to suggest that CBC mediates the positive effect of the cap in U3 and U8 nucleolar localisation.

1.18 Role of CBC in mRNA nuclear export

In comparison to U snRNA and snoRNA export, the role of the cap in the nuclear export of mRNAs remains unclear. Kinetic experiments using *Xenopus* oocytes showed that the cap only increases mRNA export by a factor of two compared to various cap analogues (Izaurralde *et al.*, 1992). In addition, no evidence currently exists to suggest this small stimulatory effect is mediated by CBC. However, immuno-electron microscopy (EM) studies on Balbiani Ring (BR) mRNAs of the insect *Chironomus tentans* have shown that CBC appears to accompany mRNAs from the site of transcription to the cytoplasmic face of the NPC. This suggests that in addition to splicing and RNA 3'-end formation, CBC might function in the nuclear export of mRNAs.

BR mRNAs encode large secretory proteins that are made in larval salivary gland cells. BR transcripts are 30-40 Kb in length and after processing are packaged into 50 nm particles for nuclear export. As a result, BR RNPs are large enough to allow associated RNA-binding proteins to be easily visualised during the successive stages of RNA synthesis, maturation and nuclear export (Daneholt, 1997).

Immuno-EM experiments have shown that CBC binds nascent BR pre-mRNA after transcription initiation, which is consistent with a role in splicing and accompanies the BR mRNP from the site of synthesis to the nuclear envelope. CBC then appears to remain associated with the 5'-end of mRNP during nucleocytoplasmic translocation (Visa *et al.*, 1996). Interestingly, mRNP particles are always translocated with the 5'-end leading, which suggests that the mRNP orientation is significant during transport (Mehlin *et al.*, 1992 and 1995). Since CBC appears to remain bound to the cap during translocation, CBC might also play a role in the early recognition of mRNP particles at the NPC. Alternatively, CBC might participate in the translocation procedure itself. However, there is no current evidence to support this and a role for CBC in mRNA export is purely descriptive.

1.19 Regulation of CBC activity

The available evidence has shown that CBC mediates the positive effects of the cap in three diverse nuclear processes: pre-mRNA splicing, RNA 3'-end formation and U snRNA export. Interestingly, recent work has suggested that the cap-binding activity of CBC is subject to regulation by extracellular stimuli. This raises the question of whether CBC might be the nuclear end-point for certain multiple signalling pathways required to regulate RNAP II transcript metabolism and ultimately cell growth.

Experiments using photoaffinity labelling assays correlated the incorporation of [α-^{32}P] GTP by CBP20 with the ability of CBC to bind capped RNA. As a result, this assay was used in subsequent experiments to show that the cap-binding activity of CBC is stimulated by a broad range of stimuli, which include growth factors and cell stress conditions (Wilson *et al.*, 1999). The cap-binding activity of CBC also appeared to be activated by the expression of a constitutively active form of the GTP-binding protein Ras and Cdc42 (Wilson *et al.*, 1999). This suggests that conditions requiring increased cell growth might promote the activity of CBC through the Ras and Cdc42 signal transduction pathways. Interestingly, HeLa splicing extracts prepared from cells containing the constitutively active form of Cdc42 also appeared to promote cap-dependent splicing, independent of cell growth (Wilson *et al.*, 2000; Erickson and Cerione, 2001). This data implies that the Cdc42 signal transduction pathway stimulates cap-dependent splicing and this effect might be mediated by the activation of CBC. Taken together, these results suggest that in order to maintain cell growth control, Ras and Cdc42 signalling pathways are possible mechanisms by which the activity of CBC and the cap are regulated during RNAP II transcript metabolism.

1.20 CBC-interacting proteins

The current literature has shown that CBC facilitates the stimulatory effects of the cap in at least three diverse aspects of nuclear RNA metabolism, which include pre-mRNA splicing, RNA 3'-end formation and U snRNA export. Despite this, little is known about the proteins that CBC interacts with and the molecular mechanisms involved. Interestingly, genetic studies in *S. cerevisiae* have revealed a number of CBC-interacting proteins that are splicing factors. This provides a possible mechanism by which yeast CBC functions during the splicing of the cap-proximal intron (Fortes *et al.*, 1999a; Fortes *et al.*, 1999b). Although, a number of CBC-interacting proteins have been identified, their roles in the mediating the effects of CBC and the cap in RNAP II transcript metabolism are yet to be elucidated. This section summarises these CBC-interacting proteins and suggests possible CBC-dependent functions during mRNA nuclear processing.

[A] hnRNP F

The CBC-interacting protein, hnRNP F was identified in a three-hybrid screen using HsCBC. This genetic interaction was further supported by *in vitro* binding experiments that showed hnRNP F interacts separately with both HsCBP20 and HsCBP80 subunits. Moreover, hnRNP F preferentially binds to CBC-mRNA complexes compared to RNA alone, which suggests the association between CBC and hnRNP F might have functional consequences (Gamberi *et al.*, 1997).

HnRNPs are amongst the most abundant proteins found in the nucleus and have been proposed to function in nearly every step of mRNA processing (Krecic and Swanson, 1999). HnRNP F has a previously defined role in splicing where it forms part of a complex responsible for activating a splicing event that results in the inclusion of a neuron-specific exon in the *c-src* gene (Black, 1991 and 1992; Min *et al.*, 1995). However, recent evidence has also suggested that hnRNP F has a more general effect on splicing efficiency *in vitro* (Gamberi *et al.*, 1997). In splicing, CBC is required for efficient removal of the cap-proximal intron by facilitating the interaction between U1 snRNP and the 5' splice site (Colot *et al.*, 1996; Lewis *et al.*, 1996 and Lewis and

Izaurralde, 1997). As mentioned earlier, in *S. cerevisiae*, the interaction between CBC and U1 snRNP is direct. However in higher eukaryotes, this interaction appears to be indirect and bridged by adaptor molecules (Lewis *et al.*, 1996b). These results suggest that in vertebrates, hnRNP F might be a candidate adaptor protein that facilitates the positive effects of CBC and the cap during the earliest stages of pre-mRNA splicing.

[B] Npl3

Similar to hnRNP F, genetic studies in *S. cerevisiae* identified another hnRNP, Npl3 as a CBC-interacting protein. Mutations in *NPL3* show synthetic lethal relationships with alleles of both *ScCBP20* and *ScCBP80*. In addition, co-immunoprecipitation experiments demonstrate that Npl3 specifically interacts with ScCBP80. However, this interaction depends on the presence of both ScCBP20 and capped RNA (Shen *et al.*, 2000). It is currently not known whether the interaction between Npl3 and CBC is direct or bridged by adaptor proteins.

Evidence from several laboratories have shown that Npl3 is a shuttling protein that is implicated in mRNA nuclear export (Flach *et al.*, 1994; Singleton *et al.*, 1995; Henry *et al.*, 1996; Lee *et al.*, 1996). Kinetic experiments in *Xenopus* oocytes have shown that the cap has a small stimulatory effect on mRNA export, however, there is no evidence to suggest that this effect is mediated by CBC (Izaurralde *et al.*, 1992). Interestingly, ScCBP80 has been shown to shuttle between the nucleus and cytoplasmic compartments (Shen *et al.*, 2000). In addition, observations from immuno-EM studies show that CBC remains associated with the 5'-end of Balbiani ring mRNPs during nucleocytoplasmic translocation (Visa *et al.*, 1996). Although there is no direct evidence, these data suggest CBC and the cap might function indirectly in mRNA export. A model can be proposed whereby after capping occurs co-transcriptionally, CBC binds nascent transcripts to promote the recruitment of RNA-binding proteins such as Npl3 that are required to function in mRNA nuclear export.

[C] A novel nuclear CBC-dependent degradation pathway?

Recent genetic studies in *S. cerevisiae* have shown that a deletion mutant in *ScCBP80* suppresses a mutation in the *cycl-512* gene. The *cycl-512* mutation generates mRNAs with aberrantly processed 3'-ends. As a result, these transcripts are subsequently degraded and this significantly reduces the level of iso-1-cytochrome *c* protein. However, the *ScCBP80* mutant increases the levels of iso-1-cytochrome *c* protein, which suggests that CBC is involved in mRNA decay (Das *et al.*, 2000). In addition, two-hybrid analysis and co-immunoprecipitation experiments have shown that the riboexonuclease Rrp6 is also a CBC-interacting protein (Das *et al.*, 2000). Rrp6 is a component of the exosome, which is the major nuclear RNA degradation pathway in *S. cerevisiae*. Taken together, these data suggest that CBC and Rrp6 might also be implicated in a novel nuclear decay pathway that degrades aberrantly processed mRNAs retained in the nucleus.

Interestingly, the recently described MIF4G domain that represents significant sequence similarities between the middle portion of eIF4G and domains in NMD2/Upf2 is also conserved in CBP80 homologues (Ponting, 2000). NMD2/Upf2 is a key component of the surveillance complex that detects premature stop codons as part of NMD pathway (Czaplinski *et al.*, 1999; Hentze and Kulozik, 1999). Although the functional significance of the MIF4G domain in CBP80 is currently unknown, it is thought to mediate protein-protein interactions. Recent work has shown that HsCBP80 might also participate in NMD. Both nucleus-associated and cytoplasmic NMD appears to take place in association with HsCBP80 (Ishigaki *et al.*, 2001). During splicing, factors involved in NMD such as Y14 interact with the spliced mRNA around 20 nucleotides upstream of the exon-exon junction to form a stable exon junction complex. Prior to the pioneer round of translation, the components of the exon junction complex remain stably bound and function NMD (Reed and Hurt, 2002). CBC is also bound to the translated mRNA that is surveyed for premature stop codons. If translation termination is sufficiently premature, mRNAs are subject to NMD. However, if the quality of translation is approved, CBC exchanges with

eIF4E and this permits the formation of the steady-state eIF4F translation initiation complex (Fortes *et al.*, 2000; Ishigaki *et al.*, 2001).

[D] eIF4G

A synthetic lethal screen in *S. cerevisiae* identified eIF4G as a CBC-interacting protein. This interaction was further supported by *in vitro* binding experiments that showed CBC physically interacts with eIF4G. Furthermore, the CBC-eIF4G association is antagonised by the cytoplasmic cap-binding protein, eIF4E (Fortes *et al.*, 2000). Biochemical studies in HeLa nuclear extracts have shown that the interaction between CBP80 and eIF4G is evolutionarily conserved. Mammalian immunofluorescence and immunoprecipitation studies have demonstrated a significant nuclear pool of eIF4G *in vivo* that is predominantly associated with CBC (McKendrick *et al.*, 2001). This suggests that in the nucleus, CBC interacts with eIF4G at the 5'-end of mRNA. However, it is currently not known whether this interaction with CBC is indirect and mediated by adaptor proteins (McKendrick *et al.*, 2001).

The function of eIF4G in the nucleus is yet to be elucidated. However, nuclear eIF4G appears to exhibit partial localisation with spliceosomal U snRNPs and splicing factors *in vivo*. Furthermore, eIF4G appears to stably associate with spliceosomes *in vitro*. There is no evidence to suggest that eIF4G participates directly in general splicing. However, these results implicate eIF4G as a candidate adaptor protein that might mediate the positive effects of CBC in cap-dependent splicing.

1.21 *Drosophila melanogaster* as a model system

Current evidence has shown that CBC is a conserved eukaryotic complex that mediates the recognition of capped RNA by various components of the nuclear RNAP II processing machinery. Key roles for CBC have been characterised in pre-mRNA splicing, RNA 3'-end formation and U snRNA export (Lewis and Izaurralde, 1997). In addition, increasing evidence also supports a role for CBC in RNAP II transcript degradation (Das *et al.*, 2000; Ponting, 2000; Ishigaki *et al.*, 2001). CBC function appears to be evolutionarily conserved in metazoans. However, in vertebrates, little is currently known about the proteins CBC interacts with or the mechanisms involved. The overall aim of this thesis is to use the genetic approaches available in *Drosophila melanogaster* to understand the mechanisms by which the cap facilitates multiple aspects of RNAP II transcript metabolism through CBC.

Drosophila provides an excellent system for studying the biology of CBC *in vivo*. Firstly, with only four chromosomes, the fruit-fly is genetically simple and can therefore be easily manipulated in *DmCBP20* and *DmCBP80* gene interaction studies. In addition, the recent publication of the *Drosophila* genome facilitates the identification and mapping of putative CBC-interacting genes (Adams *et al.*, 2000). Secondly, the amenability of the fruit-fly to cell biological and biochemical techniques allows genetic studies to be complimented by molecular and cytological studies. Lastly, the well characterised *Drosophila* developmental biology allows us to investigate whether CBC plays novel roles during spermatogenesis or oogenesis.

As a first step to characterising CBC in *Drosophila*, partial sequences for *DmCBP20* and *DmCBP80* have been identified in Berkeley *Drosophila* Genome Project (BDGP) database and full-length cDNAs have been cloned (Joe Lewis, unpublished data). Overall, both *Drosophila* subunits show high homology to the corresponding human cap-binding proteins. This supports that CBC function might also be conserved between fruit-fly and human species. Furthermore, initial biochemical experiments have shown that DmCBC behaves in a manner analogous to HsCBC. Co-expression of DmCBP20 and DmCBP80 in Sf9 cells forms a stable heterodimeric

CBC complex that specifically binds the cap structure. In addition, either subunit on its own does not exhibit detectable cap-binding activity unless complexed together to form an active CBC (Joe Lewis, unpublished data). This suggests that studying the biology of CBC in *Drosophila* might lead to a greater understanding of its functions and molecular mechanisms in higher eukaryotes.

Previous genetic studies in *S. cerevisiae* have used synthetic lethal screens and yeast two-hybrid systems to identify genes that interact with CBC (Gamberi *et al.*, 1997; Fortes *et al.*, 1999b). However, these experiments have had limited success. For example, in pre-mRNA splicing, the role of CBC has been shown to be evolutionarily conserved. Both yeast and human CBC increase the efficiency of U1 snRNP associating with the cap-proximal 5' splice site. Genetic studies in *S. cerevisiae* have shown that CBC interacts directly with components of U1 snRNP during commitment complex formation. However, in vertebrates, there is no evidence to suggest that a direct interaction exists between CBC and U1 snRNP. Instead, the available data supports that in higher eukaryotes CBC function is mediated through adaptor molecules (Lewis *et al.*, 1996; Fortes *et al.*, 1999b). This strongly suggests that in comparison to yeast, genetic studies in *Drosophila* might be more successful in identifying CBC-interacting proteins and elucidating the mechanism of CBC function in pre-mRNA splicing, as well as other RNA processing events.

Genetic studies in *Drosophila* might also reveal novel functions of CBC during development that are not uncovered using conventional biochemical approaches. The generation of mutants in either *DmCBP20* or *DmCBP80* allows the role of CBC to be investigated during oogenesis or spermatogenesis. Taken together, these data present a strong argument that *Drosophila melanogaster* is model system by which to dissect the biology of CBC *in vivo*.

The research in this thesis is broadly divided into two aims. Firstly, to determine whether CBC has novel functions during fruit-fly development. And secondly, to identify CBC-interacting genes that might elucidate the mechanisms whereby CBC and the cap facilitate multiple aspects of RNAP II transcript metabolism.

CHAPTER 2

Materials and Methods

2.1 MATERIALS

2.1.1 Buffers and solutions

Solution	Components
Buffer A	10 mM Hepes, pH 7.9; 1.5 mM $MgCl_2$; 1 mM KCl; 0.5 mM DTT
Buffer C	20 mM Hepes, pH 7.9; 25% Glycerol; 0.42 M NaCl; 1.5 mM $MgCl_2$; 0.2 mM EDTA; 0.5 mM PMSF; 0.5 mM DTT
Buffer D	20 mM Hepes, pH 7.9; 20% Glycerol; 0.1 M KCl; 0.2 mM EDTA; 0.5 mM PMSF; 0.5 mM DTT
10 x DNA loading buffer	20% (w/v) Ficoll; 1% (w/v) SDS; 0.25% (w/v) Bromophenol blue; 0.25% (w/v) Xylene cyanol; 100 mM EDTA
Grinding buffer	10 mM Tris-HCl, pH 7.5; 10 mM EDTA, pH 8; 60 mM NaCl; 5% (w/v) Sucrose; 5 μg/ml RNase
Hybridisation solution	50% (v/v) deionized Formamide; 5 x SSC; 0.1% (v/v) Tween-20; 100 g/ml *E. coli* tRNA; 50 g/ml Heparin; pH adjusted to 6.5 with HCl
Immunoprecipitation buffer	150 mM NaCl; 50 mM Tris.Cl, pH 7.8; 0.1% (v/v) Triton-X 100; 0.1% (w/v) SDS; 0.1 mM PMSF
Lysis buffer	300 mM Tris, pH 9; 100 mM EDTA, pH 8; 1.25% (w/v) SDS; 5% (w/v) Sucrose
50 x MOPS	1 M MOPS (Na^+ salt); 0.25 M NaOAc; 500 mM EDTA
1 x PBS	137 mM NaOAc; 2.6 mM KCl; 10 mM Na_2HPO_4; 1.8 mM KH_2PO_4; pH adjusted to 7.4 with HCl
PBS-Tw	1 x PBS; 0.1% (v/v) Tween-20
RNA loading buffer	1.6 x MOPS; 64% (v/v) Formamide; 8% (v/v) Formaldehyde; 0.03 mg/ml Ethidium bromide; 0.06% (w/v) Bromophenol blue
20 x SSC	3M NaCl; 0.3 M Sodium citrate; pH adjusted to 7 with HCl
10 x TBE	0.9 M Tris base; 0.9 M Boric acid; 25 mM EDTA (Na^+ salt)
1 x TE	10mM Tris-HCl, pH 7.5; 1 mM EDTA
20 x Transfer buffer	0.2 M Tris base; 1.9 M Glycine
10 x Tris-Tricine	1 M Tris base; 1 M Tricine; 1% (w/v) SDS

2.1.2 Suppliers

Chemicals were purchased from Aldrich, BDH, Boehringer Mannheim, Fisher, Fisons, Flowgen, Gibco BRL and Sigma. Enzymes were obtained from Boehringer Mannheim, New England Biolabs (NEB), Pharmacia, Promega, and Stratagene. Deoxyribonucleotides were supplied by Boehringer Mannheim. Reagents for all growth media were obtained from Difco, Biogene and Sigma. Ampicillin was purchased from Duchefa.

2.1.3 *Escherichia coli* media

All media were autoclaved prior to use and stored at room temperature. For solid media, 2% (w/v) agar was added prior to autoclaving.

Media Type	Components
Luria-Bertani broth	1% (w/v) Bacto-tryptone; 0.5% (w/v) Bacto-yeast extract; 0.5% (w/v) NaCl; pH adjusted to 7.2 with 5 M NaOH
Super broth	0.1% (w/v) Bacto-tryptone; 0.3% (w/v) Bacto-yeast extract; 0.004% (v/v) Glycerol; 0.1 M Potassium phosphate

2.1.4 *Drosophila melanogaster* media

Media Type	Components
Dundee food	1% (w/v) Bacto-tryptone; 0.5% (w/v) Bacto-yeast extract; 0.5% (w/v) NaCl; pH adjusted to 7.2 with 5 M NaOH
Apple juice plates	0.2% (w/v) Bacto-agar; 0.3% (w/v) Sucrose; 0.02% (w/v) Nipagin; 0.1% (v/v) Apple juice concentrate

2.1.5 *Escherichia coli* strains

Strain	Genotype	Source
MC1066	*Δ(lacI POZYA)*74 *galU galK strA*^{r} *leuB*6 *trpC*9830 *pyrF*74::*Tn*5(*Kn*^{r}) *hsdR*$^{-}$	P. Legrain
XL1-Blue	F' *proAB lacI*^{q} *lacZΔM*15 *TN*10 T(t^{r}) *recA*1 *endA*1 *gyrA*96(*Nal*^{r}) *thi*1 *hsdR*17($r_k^- m_k^+$) *sup*E44 *relA*1 *lac*	NEB

2.1.6 *Drosophila melanogaster* strains

A. General Stocks

Stock	Comments	Reference/ Source
w^{1118}	Mutation of the *white* gene.	Bloomington
w; P{w⁺, arm:DmCBP20}	*DmCBP20* ORF over-expressed using the ubiquitous armadillo promoter.	J. Lewis
w; P{w⁺, DmCBP20 genomic}	*DmCBP20* genomic region over-expressed.	This work
ry^{506} *P{ry⁺ᵗ⁷.²=PZ}l(3)0569705697/ TM3,* ry^{RK} Sb^{1} Ser^{1} *Referred in text as: P(lacZ ry⁺) l(3)05697*	Homozygous lethal P-element insertion into the 5' UTR of *DmCBP20* exon 1.	Bloomington
P{w⁺ =lacW}bi ^{X35} Referred in text as: *P(lacW w⁺)bi* ^{X35}	Homozygous viable P-element insertion into *bifid.*	Bloomington
w^{1118}*; P{ry⁺Δ2-3} Sb/ TM6, Ubx, bx, ry, e*	P-element transposase used in *DmCBP80* local hop.	Robertson *et al.*, 1988
Sp/CyO;P{ry⁺Δ2-3} Sb/ TM6, Ubx, bx, ry, e	P-element transposase used in mobilisation of *P(lacZ ry⁺) l(3)05697.*	Robertson *et al.*, 1988
w, FM6	Balancer used in *DmCBP80* local hop.	Bloomington
TM2 Ubx, ry, e/ TM6B, Tb, Hu, ry, e	Balancer used in mobilisation of *P(lacZ ry⁺) l(3)05697.*	Bloomington
Df(3R) P14 sr^{1}*/ TM6B*	Deficiency with breakpoints 90C2-D1; 91A1-2 that uncover *DmCBP20.*	Bloomington
Df(1) bi-D3/ FM7c	Deficiency with breakpoints 4C5-6; 4C7-8 that uncover *DmCBP80.*	Bloomington
Df(3R) e-N19/ TM2	Deficiency with breakpoints 93B; 94 that uncover *DmCBP80R.*	Bloomington

B. GAL4 Drivers

GAL4 Stock	Expression	Reference/ Source
C22 *w; P{w+, C22-GAL4}/ Sm6a*	Ubiquitous throughout embryogenesis.	M. Canning
Decapentaplegic *w; wg^{Sp-1}/ Cyo; P{w+, dpp-GAL4}/ TM6B*	Larval wing, ventral midline of leg, eye morphogenetic furrow and antennal disc.	Staehling-Hampton *et al.*, 1994
Eyeless *w; P{w+, ey-GAL4}*	Larval to late pupal eye disc and subset of CNS.	Eggert *et al.*, 1998b
GMR *w; P{w+, GMR-GAL4}*	Larval to late pupal eye disc, behind the morphogenetic furrow.	Hay *et al.*, 1994
OK-384 *w; P{w+, OK 384-GAL4}/ TM3*	Adult leg, bristles and thoracic muscles.	Connolly *et al.*, 1996
Patched *w; P{w+, ptc-GAL4}/ Sm6a*	Eye cells in embryo; Larval discs, in the stripes between the anterior-posterior boundary.	Speicher *et al.*, 1994
S31 *w; P{w+, 109(2)68-GAL4}/ CyO*	PNS of embryo; Larval eye, leg and wing discs.	Jarman *et al.*, 1998
Scabrous *w; P{w+, sca-GAL4}/ CyO*	PNS of embryo; Larval eye, leg and wing discs.	Hinz *et al.*, 1994
Scalloped *yw^{67}; P{w+, sd-GAL4}*	Pupal wing disc.	F. Sprenger
Vestigal *yw^{67}; P{w+, vg-GAL4}/ CyO; TM2/ TM6*	Pupal wing disc.	F. Sprenger
w; P{w+, UAS-importinβ, GMR-GAL4}/ CyO	Dominant phenotype observed in adult eye.	I. Davis
w; P{w+, UAS-atonal, S31-GAL4}/ CyO	Dominant phenotype observed in scutellar macrochaetae.	Jarman *et al.*, 1998

2.1.7 Oligonucleotides

All oligos used in this work were purchased from Bioline.

A. *DmCBP20* Oligos

Name	Sequence 5' to 3'	Description
S048	TTTCACACAGGAAACAGCTATGAC	*DmCBP20* forward primer used to generate *DmCBP20* C-terminal deletions (Figure 6.2, p135).
S078	GGCTCGAGCTAGAGTGCCATTCACAAAT	*DmCBP20*ΔC $_{1\text{-}95}$ reverse primer with incorporated *Xho* I restriction site, used to amplify 283 bp of *DmCBP20*ΔC $_{1\text{-}95}$ (Figure 6.2, p135).
S079	GGCTCGAGCTAGCACCTGTCCGCCGGTT	*DmCBP20*ΔC $_{1\text{-}125}$ reverse primer with incorporated *Xho* I restriction site, used to amplify 372 bp of *DmCBP20*ΔC $_{1\text{-}125}$ (Figure 6.2, p135).
S264	GGCTCGAGCTAACGCGTGCCGAAAAT	*DmCBP20* genomic forward primer with incorporated *Xho* I restriction site, used to amplify 1053 bp of *DmCBP20* genomic region (Figure 4.3, p85).
S265	GGTCTAGAATGTCTAAGGGGTATGCT	*DmCBP20* genomic reverse primer with incorporated *Xba* I restriction site (Figure 4.3, p85).

B. *DmCBP80* Oligos

Name	Sequence 5' to 3'	Description
S102	GGCTCGAGAAGATGAGTCGCAGGAG	*DmCBP80* forward primer with incorporated *Xho* I restriction site, used to generate *DmCBP80* C-terminal deletions (Figure 6.2, p135).
S103	GGTCTAGACTAGGAGAGCACAGCAAAG A	*DmCBP80*ΔC $_{1\text{-}178}$ reverse primer with incorporated *Xba* I restriction site, used to amplify 537 bp of *DmCBP80*ΔC $_{1\text{-}178}$ (Figure 6.2, p135).
S105	GGTCTAGACTACAGATGATAGGAGAACC	*DmCBP80*ΔC $_{1\text{-}424}$ reverse primer with incorporated *Xba* I restriction site, used to amplify 1275 bp of *DmCBP80*ΔC $_{1\text{-}424}$ (Figure 6.2, p135).
S106	GGTCTAGACTAGCCACCACCAGTTGATG	*DmCBP80*ΔC $_{1\text{-}510}$ reverse primer with incorporated *Xba* I restriction site, used to amplify 1532 bp of *DmCBP80*ΔC $_{1\text{-}510}$ (Figure 6.2, p135).
S107	GGTCTAGACTAATCGTCAGATGGAGAAT	*DmCBP80*CΔ $_{1\text{-}648}$ reverse primer with incorporated *Xba* I restriction site, used to amplify 1946 bp of *DmCBP80*CΔ $_{1\text{-}648}$ (Figure 6.2, p135).

C. *DmCBP80R* Oligos

Name	Sequence 5' to 3'	Description
S471	GGAAGATGAGTCGCAGGAGG	RT-PCR intron 1 forward primer, used to amplify 609 bp across intron 1 of *DmCBP80R* (Figure 5.4, p121 and Figure 5.5, p123).
S472	CCAGAAAGCGCAGCGAGTAC	RT-PCR intron 1 reverse primer (Figure 5.4, p121).
S473	GACATCTGCTACGGCTCCAT	RT-PCR intron 2 forward primer, used to amplify 549 bp across intron 2 of *DmCBP80R* (Figure 5.4, p121).
S474	TCGTTGGCGTACTTGTCATT	RT-PCR intron 2 reverse primer (Figure 5.4, p121 and Figure 5.5, p123).
S299	GGCTCGAGGGAAAGTCTGACGAGGAG	N-T *DmCBP80R* forward primer with incorporated *Xho* I restriction site, used to amplify 1495 bp N-T of *DmCBP80R* (Figure 5.3, p118).
S300	GGTCTAGATTGGTGGTACGAAAGTC	N-T *DmCBP80R* reverse primer with incorporated *Xba* I restriction site (Figure 5.3, p118).

D. *RL19*, *RpA1* and *RpS3* Oligos

Name	Sequence 5' to 3'	Description
S226	GGCTCGAGTTCCTAGCCACGACGAG	Forward primer, used to amplify *RpL19* for northern analysis.
S227	GGTCTAGAGGGATCCAACCAGACCT	*RpL19* reverse primer.
S224	GGCTCGAGCCGTTCACCATTTCTAC	Forward primer, used to amplify *RpA1* for northern analysis.
S225	GGTCTAGAGCACCCCGCCGACTGTT	*RpA1* reverse primer.
S222	GGCTCGAGTTTCCAAGAAACGCAAG	Forward primer, used to amplify *RpS3* for northern analysis.
S223	GGTCTAGACCAAAGGGGGTCACACG	*RpS3* reverse primer.

2.1.8 Plasmids

Plasmid	Description	Source
pΔ2-3	Transposase source.	C. Girdham
pBSK-*DmCBP20*	*DmCBP20* 604 bp ORF cloned into the *Eco*R I and *Xho* I restriction sites of pBSK.	J. Lewis
pBSK-*DmCBP80*	*DmCBP80* 3274 bp ORF cloned into the *Eco*R I and *Xho* I restriction sites of pBSK.	J. Lewis
pBSK-*HsCBP20*	*HsCBP20* 600 bp ORF cloned into the *Eco*R I restriction site of pBSK.	This work
pBSK-*HsCBP80*	*HsCBP80* 2828 bp ORF cloned into the *Eco*R I restriction site of pBSK.	This work
pCasPeR 4–*DmCBP20*	*DmCBP20* genomic amplified from fly DNA using S264 and S265 (Table 2.1.7A) to generate 1053 bp of *DmCBP20* genomic, which was cut with *Xba* I and *Xho* I (restriction sites incorporated into the primers) and cloned between the *Xba* I and *Xho* I sites of pCasPeR 4.	This work
pUAST–*DmCBP20*	pBSK-*DmCBP20* cut with *EcoR* I and *Xho* I to release 604 bp of *DmCBP20* ORF, which was then cloned between the *Eco*R I and *Xho* I sites of pUAST.	This work
pUAST– *DmCBP20*ΔC $_{1\text{-}95}$	*DmCBP20* ORF amplified from pBSK-*DmCBP20* using S048 and S078 (Table 2.1.7A) to generate 283 bp of *DmCBP20*CΔ $_{1\text{-}95}$, which was cut with *Xba* I and *Xho* I (restriction sites incorporated into the primers) and cloned between the *Xba* I and *Xho* I sites of pUAST.	This work
pUAST–*DmCBP20*ΔC $_{1\text{-}125}$	As above, except 372 bp of *DmCBP20*ΔC $_{1\text{-}125}$ amplified with S048 and S079 (Table 2.1.7A).	This work
pUAST–*DmCBP80*	pBSK-*DmCBP80* cut with *Eco*R I and *Xho* I to release 3274 bp of *DmCBP80* ORF, which was then cloned between the *Eco*R I and *Xho* I sites of pUAST.	This work
pUAST–*DmCBP80*ΔC $_{1\text{-}178}$	*DmCBP80* ORF amplified from pBSK-*DmCBP80* using S102 and S103 (Table 2.1.7B) to generate 537 bp of *DmCBP80*ΔC $_{1\text{-}178}$, which was cut with *Xba* I and *Xho* I (restriction sites incorporated into the primers) and cloned between the *Xba* I and *Xho* I sites of pUAST.	This work
pUAST–*DmCBP80*ΔC $_{1\text{-}424}$	As above, except 1275 bp of *DmCBP80*ΔC $_{1\text{-}424}$ amplified with S102 and S105 (Table 2.1.7B).	This work
pUAST–*DmCBP80*ΔC $_{1\text{-}510}$	As above, except 1532 bp of *DmCBP80*ΔC $_{1\text{-}510}$ amplified with S102 and S106 (Table 2.1.7B).	This work
pUAST–*DmCBP80*ΔC $_{1\text{-}648}$	As above, except 2354 bp of *DmCBP80*ΔC $_{1\text{-}648}$ amplified with S102 and S107 (Table 2.1.7B).	This work
pUAST–*HsCBP20*	pBSK-*HsCBP20* cut with *Eco*R I to release 600 bp of *HsCBP20* ORF, which was then cloned between the *Eco*R I site of pUAST.	This work
pUAST–*HsCBP80*	pBSK-*HsCBP80* cut with *Eco*R I to release 2828 bp of *HsCBP80* ORF, which was then cloned between the *Eco*R I site of pUAST.	This work

2.1.9 Antisera

Antibody	Description	Reference/ Source
anti-142[6]	Rabbit polyclonal raised against HsCBP20.	Izaurralde *et al.*, 1994
anti-Actin	Rabbit polyclonal raised against DmActin.	Sigma
anti-DmCBP80R	Rabbit polyclonal raised against DmCBP80R carboxyl-peptide sequence MSEDASDLRPEISKT.	This work
anti-DIG AP	anti-DIG antibody alkaline phosphatase conjugate.	Boehringer Mannheim
anti-mouse IgG-HRP	anti-mouse IgG horse radish peroxidase linked whole antibody generated from sheep.	Amersham
anti-rabbit IgG-HRP	anti-rabbit IgG horse radish peroxidase linked whole antibody generated from donkey.	Amersham
anti-8WG16	Mouse monoclonal raised against non-phosphorylated CTD heptapeptide repeats on largest subunit of wheat germ RNAP II.	Yuryev *et al.*, 1996
anti-YP	Rabbit polyclonal raised against major yolk proteins YP1, YP2 and YP3.	Mary Bownes
anti-Y12	Mouse monoclonal raised against Sm epitope.	Brahms *et al.*, 2000

2.2 METHODS

2.2.1 Growth of bacteria

E. coli strains (Table 2.1.5) were grown at 37°C in Luria-Bertani broth (LB) or Super broth media (Table 2.1.3). To maintain selection for plasmid DNA, transformed bacteria were grown in medium containing ampicillin to a final concentration of 100 μg/ml.

2.2.2 Transformation of competent cells

Competent cells (Table 2.1.5) were prepared using the $CaCl_2$ procedure described in Sambrook *et al.*, 1989. Competent cells were thawed on ice and mixed with the DNA to be transformed. Cells were placed on ice for 45 minutes and then heat-shocked at 42°C for 2 minutes. Immediately after heat-shock, 850 μl of LB (Table 2.1.3) was added and the cells incubated for one hour at 37°C with gentle shaking. The cells were then spread onto LB plates supplemented with 50 μg/ml of ampicillin and incubated at 37°C overnight.

2.2.3 Phenol-chloroform extraction

Nucleic acids were separated from solutions containing residual proteins by adding an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1), vortexing briefly and centrifuging at 14, 000 rpm for 2 minutes. The upper aqueous phase contained the nucleic acids and was transferred to a fresh eppendorf.

2.2.4 Precipitation of nucleic acids

Nucleic acids were precipitated from solution by mixing with 2.5 volumes of absolute ethanol containing 0.1 volumes of 3 M NaOAc, pH 5.2 and freezing for 20 minutes at -70°C. Nucleic acids were pelleted by centrifuging at 14, 000 rpm for 30 minutes at 4°C. The pellet was washed with 70% (v/v) ethanol, dried under vacuum and then resuspended in an appropriate volume of 1 x TE (Table 2.1.1) or dH_2O.

2.2.5 Quantification of nucleic acids

The concentration of DNA or RNA was determined either by spectrophotometry or visual comparison to a known quantity of molecular weight standards by agarose gel electrophoresis. The absorbence of diluted solutions was measured at 260 nm using a Cecil CE 2040 spectrophotometer and quartz cuvette. For double stranded DNA, an OD_{260} value of 1.0 represents a DNA concentration of approximately 50 μg/ml. For single stranded DNA, an OD_{260} value of 1.0 represents a DNA concentration of approximately 33 μg/ml. For single-stranded RNA, an OD_{260} value of 1.0 represents a RNA concentration of approximately 40 μg/ml.

The purity of the DNA or RNA sample was then determined by measuring the absorption at wavelengths of 260 nm and 280 nm. Protein-free preparations of DNA and RNA should give an OD_{260}:OD_{280} ratio close to 1.8 and 2.0 respectively.

2.2.6 Genomic DNA preparation from adult flies

Genomic DNA was extracted from 50 adult flies snap frozen in liquid nitrogen. Flies were homogenised in 400 μl of cold Grinding buffer (Table 2.1.1) using a Polytron

(Kinematika). An equal volume of Lysis buffer (Table 2.1.1) was added, the sample vortexed and placed at 65°C for 30 minutes. Ninety microlitres of 8 M KOAc was then added and the sample incubated on ice for 45 minutes. The tube was centrifuged at 14, 000 rpm for 15 minutes and the supernatant removed and phenol-chloroform extracted (Section 2.2.3) twice. The DNA was precipitated by mixing with 2 volumes of absolute ethanol containing 0.1 volumes of 5 M NaCl and incubated at room temperature for 2 minutes. DNA was pelleted by centrifuging at 14, 000 rpm for 5 minutes at room temperature. The pellet was washed twice with 70% (v/v) ethanol, air dried and resuspended in 50 μl of 1 x TE (Table 2.1.1). Genomic DNA was then stored at -20°C.

2.2.7 Plasmid DNA preparation

Small scale plasmid preparations were purified using the UltraClean Plasmid Miniprep Kit (MoBio), following the manufacturer's instructions. DNA was extracted from 2 ml of *E. coli* culture and resuspended in 50 μl of dH_2O. Larger scale plasmid preparations were isolated using the QIAfilter Maxiprep Kit (Qiagen), following the procedures detailed in the manufacturer's handbook. DNA was extracted from 200 ml of *E. coli* culture, resuspended in 500 μl of 1 x TE (Table 2.1.1) and stored at -20°C.

2.2.8 Agarose gel electrophoresis

DNA fragments were separated on 1-1.5% (w/v) agarose gels. Gels were prepared by melting the appropriate amount of agarose in 1 x TBE (Table 2.1.1) and adding ethidium bromide to a final concentration of 0.1 μg/ml. Prior to loading, DNA samples were mixed with 0.1 volumes of 10 x DNA loading buffer (Table 2.1.1). Electrophoresis was performed in 1 x TBE at 100 V for approximately 40 minutes and the DNA was visualised by UV illumination.

2.2.9 Restriction endonuclease digestion

DNA was digested using the buffer and temperature conditions recommended by the manufacturer. Typically, digests were performed in volumes of 20-200 μl and incubated for a period of 2 hours at 37°C.

2.2.10 Removal of 5' phosphates from DNA

Calf intestinal alkaline phosphatase (CIAP) was used to dephosphorylate linearised vector prior to insert ligation. DNA was incubated with 10 units of CIAP for 30 minutes at 37°C. The enzyme was then deactivated by adding 5 mM EDTA and incubating at 75°C for 10 minutes.

2.2.11 Ligation

Ligations were set up in a final volume of 20 μl containing 0.5-1 μg total DNA, 1 x ligase buffer and 1 unit of T4 DNA Ligase. Generally the vector and insert were present in a 1:3 molar ratio, where the size of the insert was substantially smaller than the vector. For equal size fragments an equimolar 1:1 ratio was used. Ligation reactions were incubated overnight at 16°C and then used to transform competent cells (2.2.2).

2.2.12 Polymerase chain reaction

Specific regions of DNA were amplified using the polymerase chain reaction (PCR). A typical 100 μl reaction mix was set up on ice as follows:

Plasmid DNA	10-30 ng
or Genomic DNA	50 ng
Primers	0.5 pmol
dNTPs	200 μM
Polymerase buffer	1 x
Taq DNA polymerase	5 units
dH_2O	up to 100 μl

Vent polymerase was used for applications that required higher fidelity. Consequently, a $MgCl_2$ titration of 2, 4, and 6 mM was also included.

All PCRs were performed in a Biometra UNO thermoblock (Genetic Research Instrumentation). A typical programme is described below:

Step 1 Denaturation:	94°C for 10 minutes
Step 2 Denaturation:	94°C for 30 seconds
Step 3 Annealing:	55°C for 30 seconds
Step 4 Extension:	74°C for 1 minute
30-40 cycles of Steps 2-4	
Step 5 Final extension:	74°C for 10 minutes

2.2.13 Purification of DNA from agarose gels

DNA was isolated from agarose gel slices using the QIAquick Gel Extraction Kit (Qiagen) following the manufacturer's guidelines. DNA fragments were separated by agarose gel electrophoresis (Section 2.2.8) and the bands visualised on an UV transilluminator. The desired band was excised with a clean razor blade and purified. DNA was typically eluted in 30 µl of dH_2O and stored at -20°C.

2.2.14 Automated DNA sequencing

DNA sequencing reactions were performed using the dRhodamine Terminator Cycle Sequencing Kit (Perkin-Elmer Corporation) in a Biometra UNO thermoblock (Genetic Research Instrumentation). A typical reaction mix was set up as follows:

Plasmid DNA	500 ng
or PCR product	30 ng
Primer	3.2 pmol
Terminator mix	4 µl

Twenty-five cycles as described below were performed:

Step 1:	96°C for 30 seconds
Step 2:	50°C for 15 seconds
Step 3:	60°C for 4 minutes

DNA was precipitated from the reaction mix by adding 25 μl of absolute ethanol and 1 μl of 3 M NaOAc, pH 5.2, incubating on ice for 15 minutes followed by centrifuging at 14, 000 rpm for 30 minutes. The pellet was washed with 250 μl of 70% (v/v) ethanol and dried under vacuum. The extension products were separated on an ABI PRISM 377 DNA sequencer (Perkin-Elmer Corporation) run by Nicola Wrobel (University of Edinburgh) and the sequences analysed using the Gene Jockey II programme.

2.2.15 Radio-labelling DNA fragments by random priming

DNA fragments were labelled with approximately 5000 Ci/mmol [^{32}P]-dCTP (Amersham) using the Megaprime DNA Labelling System (Amersham). DNA labelling reactions were set up according to the manufacturer's protocol.

2.2.16 Removal of unincorporated nucleotides by spin chromatography

Random primed DNA was separated from unincorporated nucleotides using spin chromatography. An one ml syringe column plugged with siliconised glass wool was packed by adding Sephadex G-50 (Amersham) and centrifuging for 3 minutes at 2000 rpm. This procedure was repeated until the gel matrix was packed to a volume of 1 ml. The spin column was equilibrated with 100 μl of 1 x TE (Table 2.1.1) and placed into a 15 ml falcon tube containing a lidless eppendorf. The probe volume was increased to 100 μl with dH_2O, added to the spin column and centrifuged as before. The eluate was collected in the lidless eppendorf and the concentration of labelled probe calculated by scintillation using a Packard Tri-Carb 2100TR Liquid Scintillation Analyser.

2.2.17 Total RNA preparation

Total RNA was prepared using the RNeasy RNA Extraction Kit (Qiagen) following the procedures detailed in the manufacturer's handbook. RNA specifically used in Reverse Transcription (RT)-PCR was extracted using the Absolutely RNA RT-PCR

Kit (Stratagene), according to the manufacturer's instructions. Generally RNA was isolated from approximately 25 mg of fly tissue snap frozen in liquid nitrogen, eluted with 30 µl of RNase-free dH_2O and stored at –20°C for 1 month.

2.2.18 Northern blotting

RNA fragments were separated by electrophoresis in a 1% (w/v) formaldehyde gel. Gels were prepared by dissolving 0.9 g of agarose in 81.6 ml of dH_2O containing 2 ml of 50 x MOPS (Table 2.1.1). The mixture was cooled to 50°C and 16.8 ml of 37% (v/v) formaldehyde was added. Total RNA (Section 2.2.17) was mixed with 0.2 volumes of RNA loading buffer (Table 2.1.1), heated to 95°C for 5 minutes and then chilled on ice immediately prior to loading. Electrophoresis was performed in 1 x MOPS buffer at 100 V for approximately 1.5 hours. A photograph was taken with a ruler placed next to the RNA marker (Promega) lane, using a HeroLab Gel Documentation System. Prior to RNA transfer, the formaldehyde gel was washed twice in 10 x SSC (Table 2.1.1) for 20 minutes.

RNA was transferred from the formaldehyde gel to a Hybond-N membrane (Amersham) by capillary action (Stratagene handbook) in 10 x SSC (Table 2.1.1). After capillary transfer, the membrane was briefly rinsed in 10 x SSC, blotted dry and irradiated in an UV Stratalinker (Stratagene) at 1200 mJ.

Radio-labelled DNA probe (Section 2.2.15) was then hybridised to the Hybond-N membrane using the QuikHyb Hybridisation Solution (Stratagene) following the protocol detailed in the manufacturer's instructions.

2.2.19 RT-PCR

RT-PCR was performed using the Omniscript Reverse Transcriptase Kit (Qiagen) in a two step reaction. First strand cDNA synthesis used 500 ng of total RNA and oligo(dT) priming according to the procedure detailed in the manufacturer's handbook. cDNA product was subsequently amplified using PCR with Taq DNA polymerase (Section 2.2.12).

2.2.20 Synthesis of digoxygenin-labelled RNA

RNA labelled with digoxygenin (DIG)-11-UTP was synthesised using the DIG RNA Labelling Kit (Boehringer Mannheim). Probes were labelled according to the instructions set out in the manufacturer's manual.

2.2.21 mRNA *in situ* hybridisation in wholemount tissues

In situ hybridisation to mRNA in wholemount tissues was performed using the method of Tautz e*t al.*, 1989 with modifications.

In brief, fixed tissue samples (Section 2.2.28) were re-hydrated in 50% methanol in PBS-Tw (Table 2.1.1) for 5 minutes and rinsed in PBS-Tw for 5 minutes. The samples were then post-fixed in 3.7% formaldehyde in PBS-Tw for 20 minutes with gentle shaking. The excess fix was removed by washing the sample 5 x 5 minutes in PBS-Tw. Samples were then washed in 50% Hybridisation solution (Hybe) (Table 2.1.1) in PBS-Tw for 10 minutes followed by a wash in Hybe for a further 10 minutes. Subsequently, the tissues were pre-hybridised in Hybe for 1 hour at 70°C.

DIG labelled RNA probe (Section 2.2.20) was diluted to 5 ng/μl in Hybe (Table 2.1.1), added to the sample and hybridised overnight at 70°C. RNA probe was then removed and the sample washed at 70°C in the following series: Hybe for 30 minutes, 50% Hybe in PBS-Tw (Table 2.1.1) for 30 minutes and 4 x 30 minutes in PBS-Tw. The anti-DIG AP antibody (Table 2.2.9) was diluted 1:2000 in PBS-Tw, added to the samples and incubated at room temperature for 2 hours on a rotating wheel. The excess antibody was then removed by washing 3 x 20 minutes in PBS-Tw.

DIG-labelled RNAs were detected using NBT/BCIP stock solution (Boehringer Mannheim). The colour solution was prepared according to the manufacturer's protocol and the reaction was then terminated by rinsing tissues in 200 mM EDTA in PBS-TW (Table 2.1.1) several times.

2.2.22 Quantification of protein

A. Bradford method

Spectrophotometric quantification of proteins was determined by the Bradford method. A series of bovine serum albumin (BSA) standard solutions were prepared with concentrations ranging from 2.5-30 μg. Two millilitres of Bradford reagent was added to each standard solution, vortexed and allowed to stand for 2 minutes. The OD_{595} of each sample was measured and the values used to plot a graph of BSA concentration versus OD_{595}. The protein of unknown concentration was simultaneously treated in the same manner and the concentration was extrapolated from the standard plot.

B. Sypro-orange staining of SDS-polyacrylamide gels

Quantification of proteins on a SDS-polyacrylamide gel was performed using the sypro-orange (Amersham) detection method. Sypro-orange staining can detect 1-2 ng of protein per band, in contrast to coomassie brilliant blue, which detects 0.3-1 μg protein per band. The SDS-polyacrylamide gel was stained with sypro-orange according to the protocol described in the manufacturer's manual. The sypro-orange stained SDS-polyacrylamide gel was scanned using the Storm 860 Imaging System set on blue fluorescence mode. The area quantitation was then calculated using the ImageQuant program.

2.2.23 SDS-polyacrylamide gel electrophoresis

The Bio-rad mini gel apparatus was assembled according to the manufacturer's instructions. The resolving gel solution was prepared, poured between the glass plates and overlaid with dH_2O. After polymerisation, dH_2O was removed, the stacking gel poured and the comb was inserted in the top of the gel. When set, the comb was taken out and the wells washed with dH_2O. Protein samples were mixed with 0.3 volumes of 3 x SDS loading buffer (NEB), and together with 5 μl pre-stained protein marker (NEB), were heated to 95°C for 5 minutes. Samples were then

briefly centrifuged and loaded onto the gel. Electrophoresis was performed in 1 x Tris-Tricine buffer (Table 2.1.1) at 130 V for approximately 40 minutes.

Resolving gel	8%	10%
ProSieve 50 gel solution (FMC)	1.6 ml	2 ml
1.5 M Tris-HCl, pH 8.8	2.5 ml	2.5 ml
10% SDS	100 µl	100 µl
dH_2O	5.7 ml	5.3 ml
10% APS	100 µl	100 µl
TEMED	4 µl	4 µl

Stacking Gel	5%
ProSieve 50 gel solution (FMC)	500 µl
1 M Tris-HCl, pH 6.8	650 µl
10% SDS	50 µl
dH_2O	3.75 ml
10% APS	50 µl
TEMED	5 µl

2.2.24 Western blotting

Proteins were transferred electrophoretically from the SDS-PAGE gel to a Hybond-ECL membrane (Amersham) using the Bio-rad Transfer System and following the manufacturer's instructions. Electrophoretic transfer was performed in 1 x Transfer buffer (Table 2.1.1) at 400 mA for 1-2 hours. The protein transfer was then confirmed by staining the membrane with Ponceau S.

Hybond-ECL membranes were immunostained according to the following procedure:

The membrane was incubated for 1 hour at room temperature in 5% Blocking solution (5% (w/v) dry milk in PST-Tw (Table 2.1.1) with constant shaking. Primary antiserum (Table 2.1.9) was diluted in 2% Blocking solution (2% (w/v) milk in PST-Tw), applied to the membrane and incubated for 1 hour at room temperature with constant shaking. The membrane was then washed twice for 5 minutes in 2% Blocking solution. Secondary antiserum (Table 2.1.9) was diluted in 2% Blocking solution, applied to the membrane and incubated for 45 minutes at room temperature

with constant shaking. The membrane was then washed as before with an additional 10 minute wash in PBS-Tw followed by a 5 minute rinse in PBS.

Immunodetection of antibodies was performed using enhanced chemiluminescence (ECL). Three millilitres of the ECL developer solution (Amersham) was prepared as detailed in the manufacturer's protocol, applied to the membrane and incubated for 1 minute at room temperature. After incubation the developer solution was removed, the membrane was placed in Saran wrap and exposed to Kodak XAR film. Exposure times were varied according to the strength of the signal. Kodak XAR films were developed in a Konica SRX-101A developer.

2.2.25 Preparation of nuclear extract

HeLa nuclear extract was obtained from 4C Mons Belgium. *Drosophila* nuclear extract was prepared according to the procedure described in Dignam *et al.*, 1983 with modifications.

Dmel2 cells were grown in spinner flasks at 37°C in *Drosophila* Serum Free Media (Sigma) supplemented with 200 mM of L-glutamine. Cells were grown to a density of 5 x 10^9 total cells prior to extract preparation. Dmel2 cells were harvested by centrifuging for 10 minutes at 2000 rpm and resuspended in 5 pellet volumes of cold PBS (Table 2.1.1). Subsequent steps were performed at 4°C. Cells were pelleted by centrifuging for 10 minutes at 1000 g and resuspended in 5 pellet volumes of Buffer A (Table 2.1.1). The cells were then incubated on ice for 10 minutes, centrifuged again at 1000 g for 10 minutes and resuspended in 2 pellet volumes of Buffer A. The suspension was transferred to a Dounce homogeniser and homogenised by 15 slow strokes of the pestle. When approximately 80% of the cells were disrupted, as determined microscopically, the lysate was centrifuged for 10 minutes at 1000 g. The pellet of crude nuclei was resuspended in 3 ml of Buffer C (Table 2.1.1) using 10 slow strokes of the Dounce homogeniser. The suspension was then incubated for 30 minutes and centrifuged at 25 000 g for 30 minutes. The supernatant was dialysed (10 kDa cut-off, Pierce) against 3 changes of 1 litre Buffer D (Table 2.1.1) over a period of 6 hours. The dialysate was centrifuged at 25 000 g for 20 minutes and the

resulting precipitate discarded. The final supernatant, designated the nuclear extract was aliquoted into eppendorfs and stored at -70°C.

2.2.26 Co-immunoprecipitation of proteins

Protein A-Sepharose beads (Amersham) were equilibrated in NT buffer (500 mM NaCl and 50 mM Tris.Cl, pH 7.5) by washing twice for 10 minutes. Washed beads were pelleted by centrifuging at 3000 rpm for 1 minute and resuspended in 300 μl of NT buffer. The appropriate antiserum (Table 2.1.9) was added and allowed to bind the protein A-Sepharose for 1 hour at room temperature on a rotating wheel. The bound antibodies were then washed 3 x 10 minutes in NT buffer before being resuspended in 500 μl of cold Immunoprecipitation (IP) buffer (Table 2.1.1). Nuclear extract was added and the bound antibodies were incubated on a rotating wheel for 2 hours at 4°C. The suspension was then washed 3 x 10 minutes with IP buffer and the beads resuspended in 30 μl of 3 x SDS loading buffer (NEB). The samples were vortexed, heated to 95°C for 5 minutes and the proteins analysed by SDS-polyacrylamide gel electrophoresis and western blotting (Section 2.2.24).

2.2.27 Maintenance of fly stocks

Flies stocks were maintained in vials or bottles containing Dundee Food (Table 2.1.4) at 25°C. All crosses were carried out at 25°C unless otherwise stated.

2.2.28 Embryonic and pupal hatching frequencies

Adult flies containing the homozygous lethal allele over a balancer chromosome were placed in a population cage with a detachable apple juice plate (Table 2.1.4). The caged flies were allowed to mate for 24 hours, after which 100 embryos were carefully removed and aligned on a new apple juice plate. The embryos were then left to develop for a further 48 hours. The number of unhatched embryos were counted and the % hatched determined.

Adult flies containing the homozygous lethal allele over a balancer chromosome were mated and the progeny allowed to develop. At third instar stage, 50 larvae were

carefully transferred to fresh vials containing Dundee Food (Table 2.1.4). The larvae were left for 48 hours to pupate and eclose. The pupal lethals were counted and the % pupated determined.

2.2.29 Fixation of wholemount tissues

Embryos were collected from a large population cage containing a detachable apple juice plate (Table 2.1.4). The plate was removed and the embryos were rinsed into a microsieve with dH_2O using a small paintbrush. The embryos were then dechorionated in 50% bleach for 3 minutes and placed into a scintillation vial containing 50% heptane, 3.7% formaldehyde in PBS (Table 2.1.1). Embryos were fixed for 20 minutes, the formaldehyde removed and 2 volumes of methanol were added. The vial was shaken vigorously and allowed to stand for 30 seconds. The devitellinised embryos settled onto the bottom of the vial and were transferred to an eppendorf, washed twice in methanol and stored in methanol at -20°C.

Third instar larvae were isolated from bottles containing Dundee Food and washed in PBS (Table 2.1.1) to remove excess food. Individual larvae were then transferred to a chilled watchglass containing PBS and the imaginal discs were dissected. Discs were fixed in 3.7 % formaldehyde in PBS for 1 hour with gentle shaking. The excess fix was then removed by rinsing 3 x 5 minutes in PBS-Tw (Table 2.1.1). Larval discs were progressively dehydrated through an ethanol series of 5 minute washes containing 30%, 50%, 70% ethanol in PBS-Tw respectively and stored in absolute ethanol at -20°C.

Newly eclosed flies were fattened for 3 days in bottles containing Dundee Food supplemented with fresh yeast. Ovaries from adult females were dissected in PBS (Table 2.1.1) and fixed in 3.7 % formaldehyde in PBS for 20 minutes. The excess fix was removed by washing 3 x 5 minutes in PBS-Tw (Table 2.1.1) and the ovaries were placed in methanol and stored at -20°C.

2.2.30 DAPI staining of ovaries

Formaldehyde fixed ovaries were washed 3 x 10 minutes in PBS-Tw (Table 2.1.1) and DNA stained with 0.1 μg/ml of 4, 6-diamidino-2-phenylindole (DAPI) for 3 minutes. The ovaries were then mounted in vector shield and analysed by fluorescence microscopy.

2.2.31 Mounting and microscopy of wholemount tissues

Embryos, larval discs, testes and ovaries were mounted in 80% glycerol in PBS-Tw (Table 2.1.1) for imaging. Fly wings were mounted in Hoyer's solution (30% (w/v) gum arabica, 85% (w/v) chloral hydrate, 20% (v/v) glycerol) and dried at 65°C with a lead weight placed on top. DIC and fluorescence images were observed using an Olympus AX-70 'Provis' fluorescence microscope and digital images were taken using a cooled Charge Coupled Device camera with Vysis QUIPS software.

Fly eyes and thoraces were observed under an Olympus SZH10 light microscope and digital images were captured using an Olympus C-2020Z camera.

Flies prepared for scanning electron microscopy (SEM) were desiccated in silica gel and mounted on stubs. The stubs were subsequently coated with gold by John Findlay (University of Edinburgh) and the specimens examined using SEM.

2.2.32 P-element mediated transformation

Three micrograms of recombinant pUAST (Table 2.1.8) was mixed with 1 μg of transposase, pΔ2-3 (Table 2.1.8). The DNA was precipitated by mixing with 2.5 volumes of absolute ethanol containing 0.1 volumes of 3 M NaOAc, pH 5.2 and freezing for 20 minutes at -70°C. DNA was then pelleted by centrifuging at 14, 000 rpm for 10 minutes at 4°C. The pellet was washed with 70% (v/v) ethanol, dried under vacuum for 5 minutes and resuspended in 10 μl of Injection buffer (5 mM KCl and 0.1 mM phosphate buffer, pH 6.8). Prior to loading a femtotip injection needle (Eppendorf), the DNA was briefly centrifuged to remove any debris. The

loaded needle was then carefully mounted in the micro-manipulator apparatus set up in the 18°C constant temperature room.

Freshly eclosed w^{III8} adults were fattened for 3 days in bottles containing Dundee Food supplemented with fresh yeast. Flies were then transferred into a large population cage with a detachable apple juice plate (Table 2.1.4) and placed at 18°C. Newly laid embryos were collected every hour on an apple juice plate, washed with dH_2O and dechorionated in 50% bleach for 3 minutes. After a brief rinse in dH_2O, approximately 50 embryos were lined along the edge of an agar strip with the posterior ends orientated just over the edge of the strip. The line of embryos was then transferred to the edge of a coverslip coated with heptane glue. The coverslip was secured on a glass slide with a drop of halocarbon oil, series 900 (Halocarbon Products Company) and desiccated in silica gel for approximately 3 minutes at room temperature. After this time, the embryos were covered with halocarbon oil, series 700 (Halocarbon Products Company) and mounted on the stage of the micro-manipulator apparatus for injection.

DNA was injected into the posterior pole of pre-blastoderm embryos. Embryos at later stages of development were destroyed by tearing the vitelline membrane. The injected embryos were covered with halocarbon oil, series 900 (Halocarbon Products Company) and allowed to develop at 18°C for 2 days. After hatching, first instar larvae were transferred into vials containing Dundee food (Table 2.1.4) and incubated at 25°C until eclosure. Individual flies were subsequently crossed to w^{III8} males or virgin females and the progeny were screened for transformants containing coloured eyes. As a result, homogenous stocks were established and the transgene insertion was genetically mapped.

CHAPTER 3

Characterisation of DmCBC

3.1 INTRODUCTION

The *Drosophila melanogaster* homologue of CBP20 was initially identified in a TBLASTN search of the BDGP nucleotide sequences using the HsCBP20 peptide sequence. This identified a 330 nucleotide sequence tagged site (STS), *Dm1885* that showed strong homology to the amino-terminal fragment of HsCBP20 (Joe Lewis, unpublished data). Interestingly, *Dm1885* was also found to be associated with the lethal P-element insertion *P(lacZ ry$^+$)l(3)05697*. Chapter 4 subsequently describes how *P(lacZ ry$^+$)l(3)05697* was mobilised to generate excision alleles in *DmCBP20*. A fragment corresponding to *Dm1885* sequence was PCR amplified from embryonic cDNA and used to probe a third instar larval library for full-length cDNAs (Joe Lewis, unpublished data). As a result, the full-length cDNA for *DmCBP20* was isolated and found to be 604 bp in length.

Similar to DmCBP20, the *Drosophila melanogaster* homologue of CBP80 was identified by using the HsCBP80 peptide sequence in a TBLASTN search of the BDGP database. This identified the STS, *Dm84H4T* that showed high homology to amino acids 249-345 of HsCBP80 (Joe Lewis, unpublished data). Likewise, the sequence corresponding to *Dm84H4T* was PCR amplified from adult fly genomic DNA and used to screen a third instar larval library for full-length cDNAs (Joe Lewis, unpublished data). A full-length cDNA for *DmCBP80* was isolated and found to be 3, 224 bp in length.

Searches of FlyBase revealed that apart from one P-insertion in *DmCBP20*, no known mutant alleles for either *DmCBP20* or *DmCBP80* have been described to date. Initial biochemical studies using the *Drosophila* cap-binding proteins showed

that DmCBC behaves in a manner analogous to HsCBC. Co-expression of DmCBP20 and DmCBP80 in Sf9 cells results in a stable heterodimeric CBC that specifically binds the cap structure. In addition, either subunit on its own does not exhibit detectable cap-binding activity unless complexed together to form an active CBC (Izaurralde *et al.*, 1994 and Joe Lewis, unpublished data).

The principal aims of this thesis were to firstly, use *Drosophila melanogaster* as a tool to identify CBC-interacting genes and secondly, determine whether CBC is required for fruit-fly development. In order to achieve this, it is important to commence with an initial characterisation of CBC in *Drosophila melanogaster*. As a result, this Chapter primarily analyses the DmCBC sequence data available along with the expression of DmCBC RNA throughout fruit-fly development.

3.2 RESULTS

3.2.1 Gene structure and mapping of *DmCBP20* and *DmCBP80*

The gene structure for both *DmCBP20* and *DmCBP80* as predicted by Genescan is presented in Figure 3.1. *DmCBP20* consists of three exons, which span 746 bp of genomic DNA with two short introns that are 88 bp and 90 bp in length. The Bloomington stock, *P(lacZ ry$^+$)l(3)05697* contains a P-element inserted into the 5' non-coding region of the first exon, 8 bp upstream of the initiation methionine (Figure 3.1, A). The chromosomal location of *DmCBP20* was mapped to the right arm of chromosome 3 on polytene segment 90E1-2 by *in situ* hybridisation on polytene chromosome squashes using *l(3)05697* as a probe (Alan Spradling). In contrast to *DmCBP20*, the gene structure of *DmCBP80* is relatively complex (Figure 3.1, B). *DmCBP80* contains seven exons and six introns that encompass 8, 299 bp of genomic DNA. To date, no P-elements insertions have been identified within the *DmCBP80* locus. *DmCBP80* has been mapped to 4C7-8 on the X-chromosome by both the BDGP and *in situ* hybridisation on polytene chromosomes using *DmCBP80* cDNA as a probe (Joe Lewis, unpublished data).

Figure 3.1 Gene Structure and Mapping of *DmCBP20* and *DmCBP80*

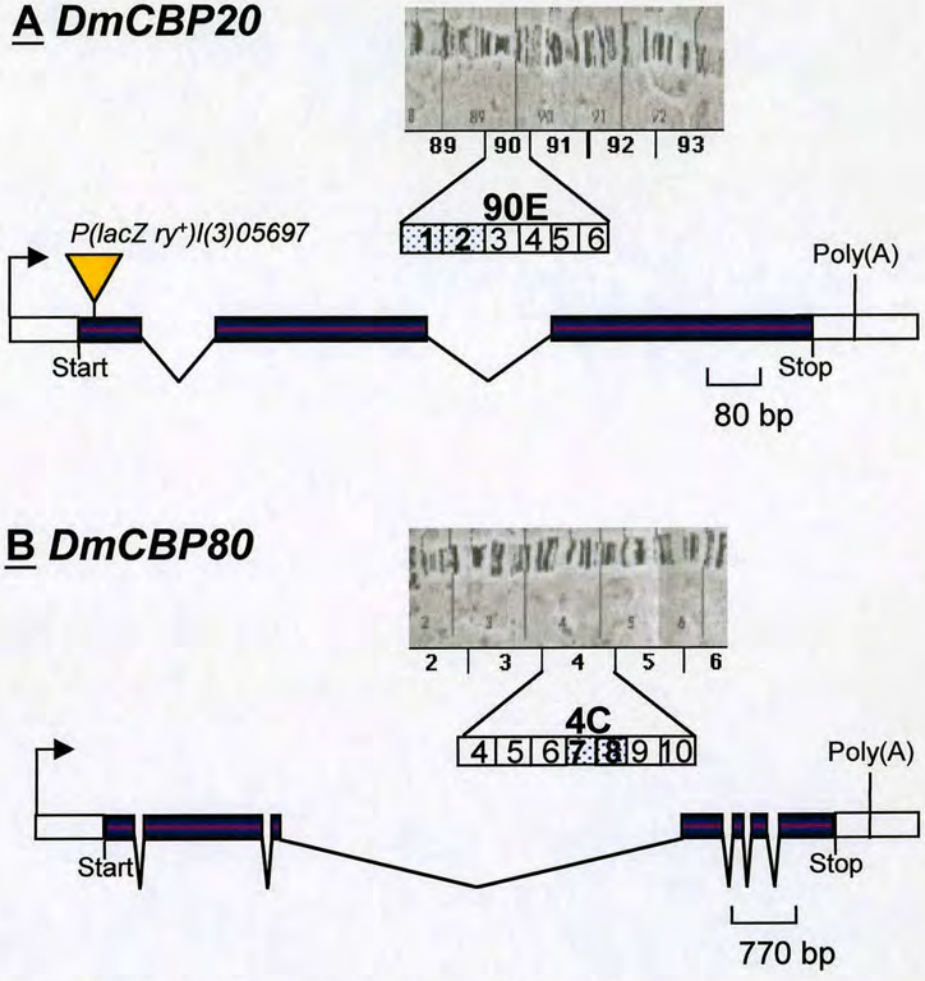


Figure 3.1 Gene structure and chromosomal locations of *DmCBP20* and *DmCBP80*. Schematic representation of gene structures as predicted by Genescan and chromosomal mapped regions. **A.** *DmCBP20* is mapped to right arm of chromosome 3 on polytene segment 90E1-2 by *in situ* hybridisation on polytene chromosomes using *l(3)05697* as a probe (Alan Spradling). *DmCBP20* has three exons depicted by rectangles and two introns represented by black lines, which span a total of 746 bp genomic DNA. P-element *P(lacZ ry+)l(3)05697* is inserted into the 5' non-coding region of the first exon, 8 bp upstream of the initiation methionine. **B.** *DmCBP80* is mapped to 4C7-8 of X-chromosome by both the BDGP and *in situ* hybridisation on polytene chromosomes using *DmCBP80* cDNA as a probe (Joe Lewis, unpublished data). *DmCBP80* has seven exons and six introns that span a total of 8, 299 bp genomic DNA. Start of transcription unit is depicted by black arrow along with poly(A) sites. Translation unit is outlined by start and stop annotations.

3.2.2 *Drosophila melanogaster* and human CBP20 and CBP80 homologues are strongly conserved

To determine whether *Drosophila melanogaster* is a suitable model by which to study the biology of CBC *in vivo*, it is important to compare the predicted amino acid sequences of human and *Drosophila* cap-binding proteins. The conceptual protein product of *DmCBP20* is 154 aa in length compared with 156 aa for HsCBP20. Figure 3.2, A clearly shows that DmCBP20 is highly conserved throughout its length to HsCBP20 exhibiting 78% identity and 88% homology. Extensive homology is particularly apparent within the RNA-binding domain that is comprised of two RNP motifs and the carboxyl-rich RGG tail, which is thought to stabilise the interaction between CBC and the cap (Birney *et al.*, 1993; Burd and Dreyfuss, 1994; Mazza *et al.*, 2001). In contrast, more sequence divergence and length heterogeneity is observed between the amino-terminals of human and *Drosophila* CBP20 proteins.

Similar to DmCBP20, DmCBP80 shows strong conservation to its human homologue. The conceptual translation of *DmCBP80* is predicted to encode a polypeptide that is 800 aa compared with 790 aa for HsCBP80. As can be observed from Figure 3.2, B, there is a high level of identity and similarity between DmCBP80 and HsCBP80 throughout their length at 57% and 75% respectively. Both DmCBP80 and HsCBP80 have a predicted bipartite NLS at the amino-terminus (Mattaj, 1993; Izaurralde *et al.*, 1994). This NLS has been shown to be functional in humans since mutation of this sequence inhibits the nuclear import of HsCBP80 (Izaurralde *et al.*, 1995). In the absence of HsCBP80, HsCBP20 is not actively transported into the nucleus. This suggests that the NLS is required for the active nuclear import of CBC. In addition, DmCBP80 contains the recently described MIF4G domain that represents significant sequence similarities between the middle portion of eIF4G and domains in NMD2/Upf2 and CBP80 (Ponting, 2000; Mazza *et al.*, 2001). Although the functional significance of this domain is currently unknown, the MIF4G domain is thought to mediate protein-protein interactions during NMD (Ishigaki *et al.*, 2001).

Figure 3.2 Sequence Alignments of CBP20 and CBP80 Homologues

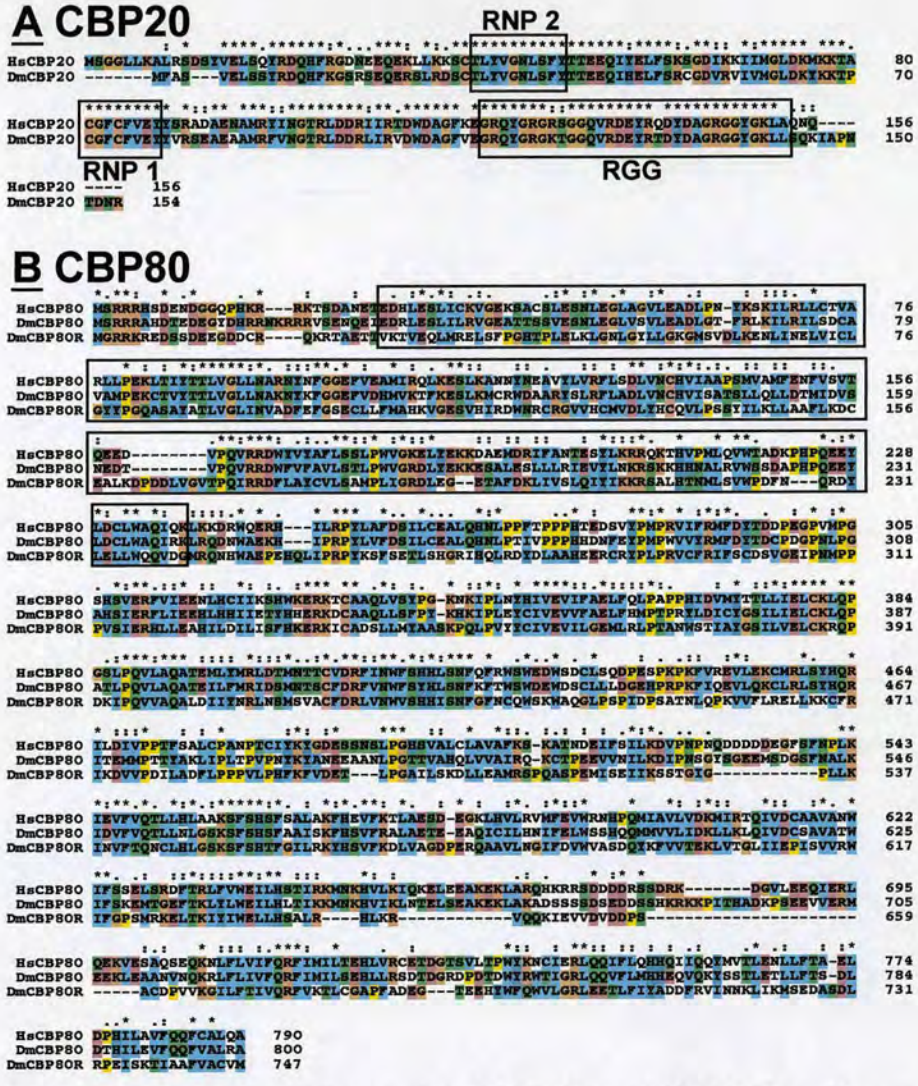


Figure 3.2 Sequence alignments of CBP20 and CBP80 homologues. **A.** Conceptual translation of *DmCBP20* (Q9V3L6) aligned to HsCBP20 (P52298) using Clustal X (Thompson *et al.*, 1998). The highly conserved RNA-binding domain comprised of two RNP motifs is indicated along with the arginine-glycine (RGG) rich carboxyl-domain. **B.** Conceptual translation of *Drosophila melanogaster DmCBP80* (Q9U980) compared to *Homo sapiens* HsCBP80 (Q09161). The conserved MIF4G domain is presented boxed. Identical and related amino acids are illustrated by (*) and (·) respectively. (:) indicates similar amino acids and discontinuities in the alignments are represented by dashes. Blue represents hydrophobic/ aromatic amino acids. Orange depicts basic and glycine residues. Purple shows acidic amino acids. Green and yellow illustrates hydrophilic and proline residues respectively. Significance of DmCBP80R sequence is referred to in Chapter 5.

Collectively, this data shows both DmCBP20 and DmCBP80 are highly homologous to the corresponding human proteins. This strongly suggests that CBC function might also be conserved between *Drosophila* and human species.

3.2.3 DmCBP20 and DmCBP80 mRNAs are ubiquitously expressed throughout development

As a first step to characterising DmCBC, northern analysis was used to determine at which stages in *Drosophila* development DmCBP20 and DmCBP80 transcripts were expressed. Total RNA was extracted from staged embryo collections, three larval instars, adult flies and isolated ovaries. The amounts of total RNA were normalised to the level of ribosomal RNAs (rRNAs) and analysed by northern blotting (Figure 3.3, lowest panel). Full-length *DmCBP20* and *DmCBP80* cDNAs were then used to probe the developmental northern and exposed overnight. As can be observed from Figure 3.3, single transcripts for both DmCBP20 and DmCBP80 were observed at the predicted mRNA size of approximately 600 bp and 3.3 Kb respectively. In addition, DmCBP20 and DmCBP80 mRNAs were found to be expressed throughout *Drosophila* development with both subunits showing similar steady state levels at the same stages. The rRNA doublet bands observed in lanes 8 and 9 were due to the specific cleavage of the large ribosomal subunit RNA (lowest panel).

Figure 3.3 shows that high levels of DmCBP20 and DmCBP80 mRNAs exist in the early embryo (lane 1). The steady state levels of both RNAs decrease as embryogenesis proceeds (lanes 2-5) with the lowest levels being observed in first and second larval instars (lanes 6 and 7). The mRNA steady state levels increase again in third instar larvae, adult male and female stages (lanes 8-10). Most notably, adult females show higher steady state levels for DmCBP20 and DmCBP80 mRNAs than adult males. To determine whether this enrichment of mRNA was due to ovarian expression, ovaries were dissected from mature adult females and the RNA extracted.

Figure 3.3 DmCBP20 and DmCBP80 mRNAs are Expressed Throughout Development

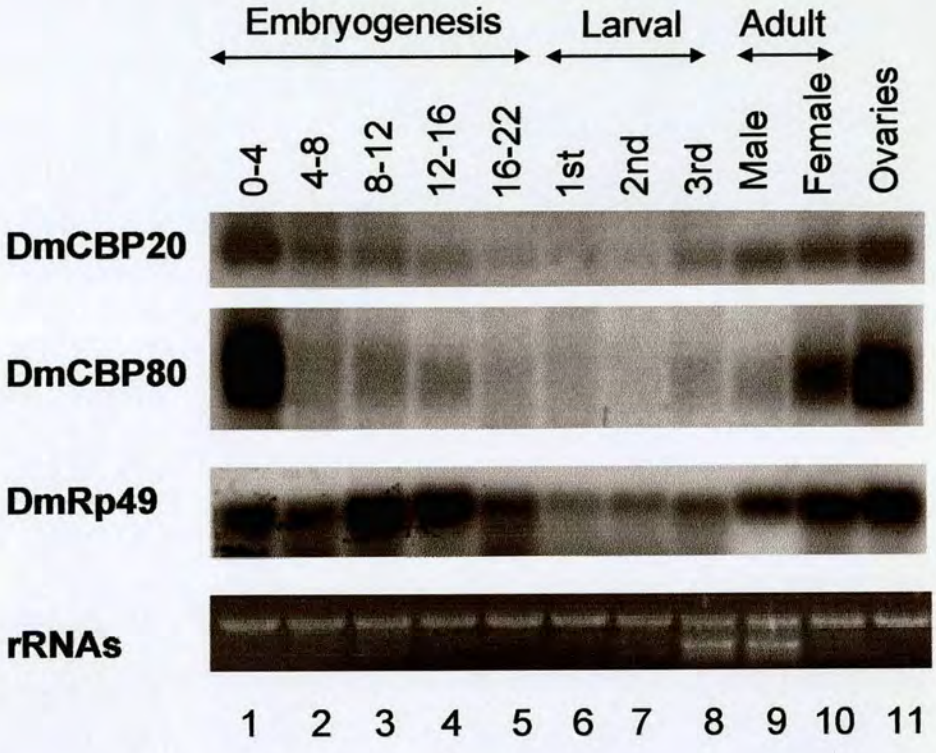


Figure 3.3 DmCBP20 and DmCBP80 mRNAs are expressed throughout development. Developmental northern analysis of DmCBP20 and DmCBP80 transcripts. Staged embryos were collected and matured for the following time points: 0-4 hours, 4-8 hours, 8-12 hours, 12-16 hours, and 16-22 hours along with first, second and third instar larvae, adult males and females and isolated ovaries. Total RNA was extracted under denaturing conditions and resolved by northern blotting. Nitrocellulose filters were probed with either full-length *DmCBP20* or *DmCBP80* cDNA as indicated and exposed. Single transcripts were observed for DmCBP20 and DmCBP80 at approximately 600 bp and 3.3 Kb respectively. 10 μg of RNA per lane was loaded as judged by ethidium bromide staining of ribosomal RNAs (rRNAs). As an additional loading control, the filter was re-probed with full-length ribosomal protein, *DmRp49* cDNA.

Northern analysis showed that the signals for DmCBP20 and DmCBP80 were indeed elevated in ovarian RNA compared to total female RNA suggesting that the components of DmCBC are maternally derived (compare lanes 10 and 11). As an additional loading control, the nitrocellulose membrane was re-probed using full-length cDNA from the ribosomal protein, *DmRp49* that is highly expressed throughout *Drosophila* development (third panel).

The next question to address was whether DmCBP20 and DmCBP80 mRNAs were expressed ubiquitously or restricted to specific cell types/ tissues in fruit-fly development. To test this, RNA *in situ* hybridisation was used to detect DmCBP20 and DmCBP80 RNAs during oogenesis, embryogenesis and selected third instar larval imaginal discs. Full-length *DmCBP20* and *DmCBP80* cDNAs were used to generate DIG-labelled anti-sense RNA that specifically detects mRNA and sense RNA, which serves as a negative control. DmCBP20 and DmCBP80 DIG-labelled RNA probes were hybridised to fixed ovarioles, staged embryo collections and eye-antenna, haltere, brain, leg and wing imaginal discs. As can be observed from Figure 3.4, DmCBP20 and DmCBP80 mRNAs were expressed ubiquitously throughout oogenesis, embryogenesis and larval leg and wing imaginal discs. In keeping with the results from the developmental northern, both cap-binding proteins showed similar expression patterns in the same tissues. No mRNA expression was observed in the controls that used sense DmCBP20 and sense DmCBP80 as probes. This suggested that the staining detected was specific for DmCBP20 and DmCBP80 mRNAs (data not shown).

Figure 3.4 shows that DmCBP20 and DmCBP80 mRNAs were detected in all ovarian tissues (Panel A, [i] and [ii]). Furthermore, the expression levels of RNAs encoding both subunits appears to steadily increase as oogenesis proceeds posteriorly towards the mature egg. Staining is lowest in the germanium and earliest egg chambers staged around 2-4. Detection levels appear to increase as the egg chambers mature with the highest levels being observed in the later stages of oogenesis, around stage 11. This is consistent with the time mRNA is deposited from the nurse cells

into the developing oocyte. These data taken together with the northern analysis strongly support a maternal contribution of both cap-binding proteins.

DmCBP20 and DmCBP80 RNAs were also distributed ubiquitously in early, middle and late staged embryos (Panel B, [i]-[vi]). The intense staining observed in the germ band extension and central nervous tissue of middle and late staged embryos respectively was deemed non-specific since these tissues also preferentially localised sense RNA probe (Panel B, [iii]-[vi] and data not shown). Likewise, RNA *in situ* hybridisation performed on selected imaginal discs of third instar larvae showed that DmCBP20 and DmCBP80 RNAs were not restricted to specific tissues or cells. The ubiquitous staining observed in leg and wing discs was also found to be representative for the other eye-antenna, haltere and brain imaginal discs examined (Panel C, [i]-[iv] and data not shown).

Figure 3.4 DmCBP20 and DmCBP80 mRNAs are Ubiquitously Expressed in Development

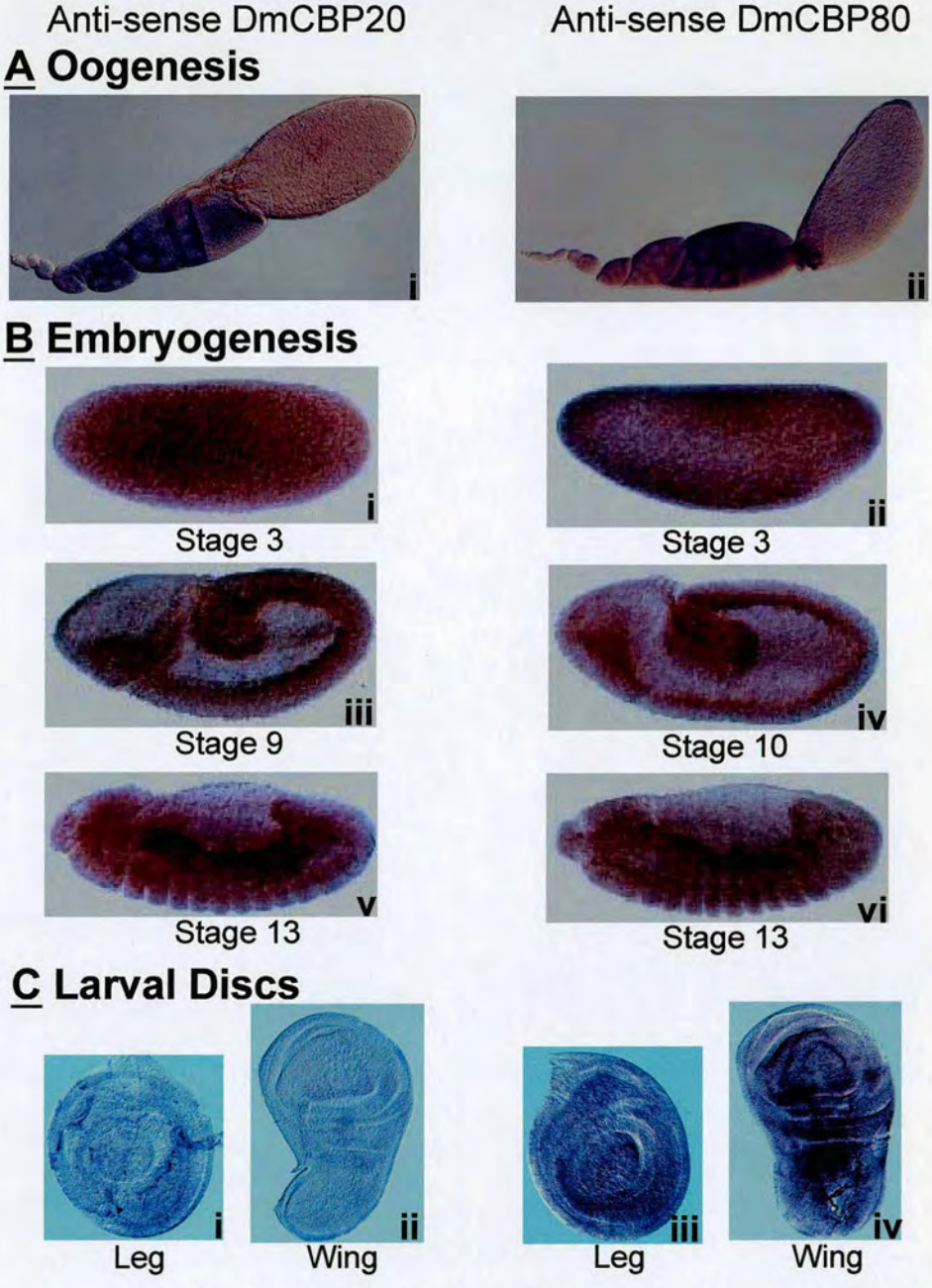


Figure 3.4 DmCBP20 and DmCBP80 mRNAs are ubiquitously expressed in development. RNA *in situ* hybridisation performed on fixed **A**. ovarioles **B.** early **[i-ii]**, middle **[iii-iv]** and late **[v-vi]** staged embryos **C.** third instar larval leg and wing imaginal discs. DmCBP20 and DmCBP80 mRNAs were detected using DIG-labelled anti-sense RNA.

3.3 DISCUSSION

3.3.1 *Drosophila melanogaster* and human CBP20 and CBP80 homologues are strongly conserved

Comparison of the conceptual protein products of *DmCBP20* and *DmCBP80* show they are highly homologous to the corresponding human proteins. CBP80 homologues share a high degree of conservation throughout their length. Both have the bipartite NLS that is functionally required for the nuclear import of HsCBP80, in addition to the recently described MIF4G domain that implicates CBP80 in NMD (Izaurralde *et al.*, 1995; Ponting, 2000; Ishigaki *et al.*, 2001). In comparison to CBP80, *Drosophila* and human homologues of the smaller subunit appear to be more extensively conserved. CBP20 homologues share the conserved RBD along with the carboxyl-terminal RGG tail that is thought to stabilise the interaction between CBC and capped RNA (Birney *et al.*, 1993; Burd and Dreyfuss, 1994). This is consistent with CBP20 having a phylogenetically conserved functional role in binding capped RNA (Mazza *et al.*, 2001).

Taken together, the protein sequence data strongly suggests that CBC function might also be conserved between *Drosophila* and human species. Indeed, the role of CBC in at least pre-mRNA splicing is conserved between yeast and humans. In splicing, CBC increases the efficiency of U1 snRNP binding to the cap-proximal 5' splice site. This enhances commitment/ E complex assembly and consequently the rate of splicing of the cap-proximal intron (Izaurralde *et al.*, 1994; Lewis *et al.*, 1996). However, a striking difference between these two species is that in yeast CBC interacts specifically and directly with the components of U1 snRNP. In contrast, the mechanism of CBC action is currently unknown in mammals. The interaction between HsCBC and U1 snRNP appears to be indirect and mediated by yet unidentified adaptor molecules. There is extensive conservation between the components of the spliceosome in vertebrate and *Drosophila* systems (Mount and Salz, 2000). In addition the complexity of splicing, as well as other RNAP II processing events such as RNA 3'-end formation, appear to be similar between

Drosophila and mammals. This strongly suggests that *Drosophila* would be a model system for dissecting the biology of CBC *in vivo*.

3.3.2 DmCBP20 and DmCBP80 mRNAs are ubiquitously expressed throughout development

Results from the northern analysis indicate that single transcripts exist for DmCBP20 and DmCBP80. As expected from the components of a heterodimer that are only functional as a complex, the expression pattern of both transcripts parallel each other throughout the *Drosophila* developmental stages. In addition, RNA *in situ* analysis shows that DmCBP20 and DmCBP80 mRNAs are distributed ubiquitously and not restricted to specific tissues. This is consistent with CBC playing a key role in RNAP II transcript metabolism of all cells in *Drosophila* development.

Elevated levels of DmCBP20 and DmCBP80 mRNAs exist during oogenesis and early embryogenesis, which suggests both subunits are maternally derived. A maternal contribution is further corroborated by RNA *in situ* analysis, which shows CBC expression to progressively increase as oogenesis proceeds until they are deposited from the nurse cells into the maturing oocyte. The high levels of DmCBP20 and DmCBP80 RNAs during early development might reflect a greater requirement for CBC in general RNA metabolism during the accelerated cell cycles at these stages. This profile matches that of other genes that participate in RNAP II transcript processing such as the splicing factors *sans fille* and *noisette* (Flickinger and Salz, 1994; Meyer *et al.*, 1998). Alternatively, it might indicate that either or both cap-binding subunits have novel functions during early *Drosophila* development.

3.4 CONCLUSIONS

Initial characterisation of DmCBC suggests *Drosophila melanogaster* is a good model organism for understanding the mechanisms and functions of CBC. DmCBP20 and DmCBP80 mRNAs are ubiquitously expressed in all tissues throughout *Drosophila* development. Apart from one P-element in *DmCBP20*, no known alleles have been described for either cap-binding protein to date.

CHAPTER 4

P-element induced mutagenesis of *DmCBP20* and *DmCBP80*

4.1 INTRODUCTION

The results from the previous Chapter have shown that both CBP20 and CBP80 are highly conserved between *Drosophila melanogaster* and human species. This suggests that the function of CBC might also be evolutionarily conserved. In addition, DmCBP20 and DmCBP80 mRNAs are ubiquitously expressed in all tissues throughout fruit-fly development. Elevated levels of both subunits exist in the earliest stages of development, which suggests a greater requirement for CBC during these stages. It is possible that DmCBC plays a general role in RNAP II transcript metabolism or alternatively, either/ both subunits might have novel functions in early development. In order to address this, mutants in the CBC heterodimer can be analysed to study the biology of CBC *in vivo* and determine whether novel functions of CBC exist in *Drosophila* development.

Searches of FlyBase indicated that apart from one P-insertion in *DmCBP20*, no known mutant alleles in *DmCBP20* and *DmCBP80* have been described to date. As a result, various physical and chemical mutagenesis strategies were considered by which to create mutants alleles in *DmCBP20* and *DmCBP80*. Since a P-element already resides within the *DmCBP20* locus, it was reasoned that P-element induced mutagenesis would be a straightforward method of generating mutants in the CBC heterodimer.

In *Drosophila*, the primordial P-element is a transposable element that can catalyse its own excision and insertion events within the genome. However, the activities of the P-element have been separated into two distinct genetic elements that either contain the transposable or the transposase activity. The transposable element is defective in the production of transposase but contains the inverted terminal repeats necessary for its own transposition. In contrast, *P(ry$^+$Δ2-3)* encodes the transposase source required to catalyse movement but is itself stably integrated into the genome (Robertson *et al.*, 1988). In order to mobilise the transposable element within the genome, genetic schemes can be designed whereby both the transposable element and *P(ry$^+$Δ2-3)* are expressed in the male or female germline.

P-element transposition is a useful tool by which to create mutations in a particular gene. P-elements can be mobilised to generate new insertions in a gene of interest by local hopping. Alternatively, an existing P-insertion can be excised to produce additional mutant alleles. Local hopping describes the phenomenon whereby P-elements once mobilised, preferentially re-insert close to or within 200 Kb of their original insertion site (Tower *et al.*, 1993; Zhang and Spradling, 1993). Consequently, local hopping has been successfully used to transpose nearby P-elements into a gene of interest to create P-element induced alleles (Littleton *et al.*, 1993; Segalat *et al.*, 1992). Not only are P-elements transposed to new chromosomal locations but P-element excision can also produce new alleles. Figure 4.3, B illustrates the different excision alleles generated by P-element mobilisation. In summary, precise excision of the P-element results in alleles of the wild-type genomic sequence. Whilst imprecise excision alleles have portions of genomic sequence deleted or alternatively, contain residual P-element sequences.

The aim of this Chapter was to use P-element induced mutagenesis to create mutations in the cap-binding components of the CBC heterodimer. The generation of such mutants might uncover novel functions of CBC in fruit-fly development. P-element induced mutagenesis was used to firstly, create P-insertion alleles of *DmCBP80* by local hopping. And secondly, to generate *DmCBP20* excision alleles by mobilising the P-element residing within the *DmCBP20* locus.

4.2 RESULTS

4.2.1 Generating P-insertions in *DmCBP80*

Since no known mutant alleles have been described for *DmCBP80* to date, P-element induced mutagenesis was used to create lesions in *DmCBP80*. Searches of all the known P-element lines on FlyBase revealed that no P-elements were found in the *DmCBP80* locus. Therefore, the strategy employed was to generate P-insertion alleles of *DmCBP80* by local hopping. In order to isolate P-insertions in *DmCBP80*, a local mutagenesis strategy was designed to transpose a nearby P-element into *DmCBP80*, which is mapped to polytene segment 4C7-8 on the X-chromosome. The starter-element used for the local transposition was the homozygous viable line $P(lacW\ w^{+})bi^{X35}$. $P(lacW\ w^{+})bi^{X35}$ is inserted into the 5' non-coding region of the gene *bifid* (*bi*) that is mapped to 4C2-3. The starter-element is located approximately 220 Kb from *DmCBP80* and therefore has a good chance of re-inserting into the *DmCBP80* locus by local hopping (Tower *et al.*, 1993; Zhang and Spradling, 1993). It is important to note that at the time this P-hop was performed, *DmCBP20* was identified as an essential gene. As a result, the *DmCBP80* P-mutagenesis screen was based on the assumption that like *CBP20*, *CBP80* was also an essential gene in *Drosophila*.

A genetic scheme was subsequently designed to recover homozygous lethal insertion alleles of $P(lacW\ w^{+})bi^{X35}$ (Figure 4.1). $P(lacW\ w^{+})bi^{X35}$ is genetically marked with *white* (*w*) and adults have white eyes that are orange equatorial to the polar tips. To mobilise the P-element, $P(lacW\ w^{+})bi^{X35}$ was crossed with the transposase source $P(ry^{+}\Delta 2\text{-}3)$ and the F1 progeny scored for coloured eyes. New P-insertions were identified by eye colours that differed from $P(lacW\ w^{+})bi^{X35}$ and ranged from light orange to red due to different chromosomal insertion positions. Whilst the F1 progeny that had white eyes were discarded since they represented lines where the P-element had excised and not re-inserted.

Figure 4.1 Generating P-insertions in *DmCBP80* by Local Hopping

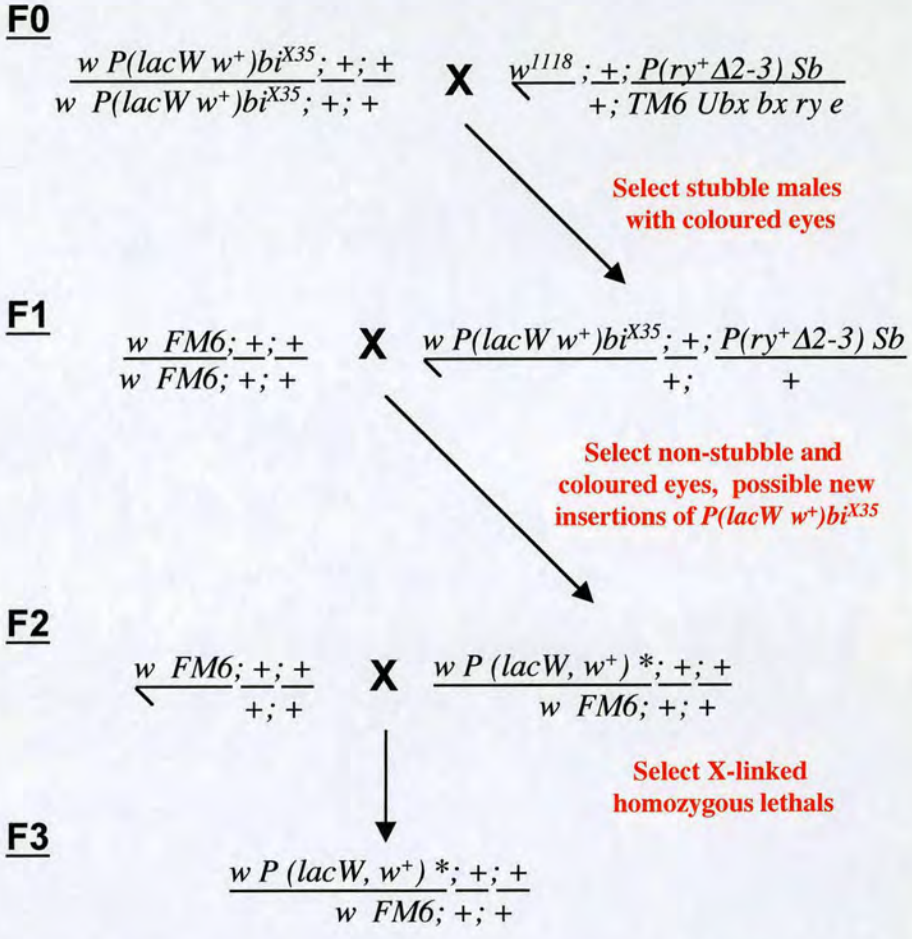


Figure 4.1 Genetic scheme for generating P-insertions in *DmCBP80* by local hopping. Homozygous viable *P(lacW w+)bi^{X35}* was mobilised and new homozygous lethal P-elements on the X chromosome screened for insertions in *DmCBP80*. **F0.** *P(lacW w+)bi^{X35}* is inserted into *bifid* located approximately 220 Kb from *DmCBP80*. *P(lacW w+)bi^{X35}* is marked with *white* and adult have white eyes that are orange equatorial to polar tips. *P(lacW w+)bi^{X35}* virgins are mated to males carrying transposase activity *P(ry+Δ2-3)* marked with stubble. **F1.** Male progeny with coloured eyes and stubble contain both *P(lacW w+)bi^{X35}* and transposase activity. They are mated to homozygous balancer FM6 virgins. **F2.** Progeny carrying bar eyes with different eye colours from *P(lacW w+)bi^{X35}* contain new P-insertions (indicated by asterisk). They are crossed with FM6 balancer flies. **F3.** Progeny scored for hemizygous X-linked lethal P-insertions and heterozygous balanced stocks established. X-linked lethals are then tested for complementation with *Df(1)bi-D3*, which uncovers *DmCBP80* and analysed by southerns.

Upon mobilisation of *P(lacW w^+)bi^{X35}* in a total of 500 independent lines, 280 lines (56%) were found to contain new chromosomal P-insertions. Since *DmCBP80* was assumed to be essential, these lines were subsequently screened for homozygous lethal insertions on the X-chromosome. The remaining 220 lines were found to have white eyes, indicating that the P-element had excised to generate a deletion in the *white* gene. The X-linked homozygous lethal insertions were then tested for lesions in or near the *DmCBP80* locus by complementation with *Df(1)bi-D3*. A total of 10 lines (2%; 10 from 280) failed to complement *Df(1)bi-D3*. These 10 X-linked lethal lines were then recovered in the absence of the transposase activity and the P-insertions characterised by southern analysis. Genomic DNA was extracted from the X-linked lethal balanced lines and digested with *Bam* HI to release *DmCBP80* genomic fragment. DNA was then resolved by southern blotting and the filters probed with full-length *DmCBP80* cDNA and exposed (data not shown). This determined that none of the X-linked homozygous lethal insertions were found to reside within the *DmCBP80* locus.

The local hopping strategy was not successful in generating a P-element induced allele of *DmCBP80*. P-element insertional mutagenesis might have failed either because not enough flies were screened or the assumption that P-insertions in *DmCBP80* would be homozygous lethal was incorrect. Alternatively, *DmCBP80* might not be readily mutable by P-elements. P-element induced mutagenesis is unpredictable, P-elements are found to preferentially insert in some loci and rarely in others. However, what defines certain loci as hot spots for P-insertions and others as not is currently unknown.

4.2.2 Generation of *P(lacZ ry^+)l(3)05697* excision alleles

The amino-terminal fragment of *DmCBP20* was originally identified as the STS, *Dm1885* that is associated with the homozygous lethal P-insertion *P(lacZ ry^+)l(3)05697* (Alan Spradling). Comparison of *Dm1885* with the *DmCBP20* cDNA sequence showed that *P(lacZ ry^+)l(3)05697* is inserted into the 5' non-coding region of the first exon, 8 bp upstream of the initiation methionine (Figure 4.3, A).

Characterisation of *P(lacZ ry$^+$)l(3)05697* using embryonic and pupal hatching frequencies staged the lethality of homozygous mutants to either first or second larval instar stages (data not shown). In order to test whether the P-insertion in *DmCBP20* was the cause of lethality and also generate P-excision alleles, *P(lacZ ry$^+$)l(3)05697* was mobilised. A genetic scheme was subsequently designed to recover both imprecise and precise excision alleles of *P(lacZ ry$^+$) l(3)05697* (Figure 4.2). *P(lacZ ry$^+$) l(3)05697* is genetically marked with *rosy* (*ry$^+$*), therefore the presence of the P-element can be followed by *ry$^+$* eye colour. *P(lacZ ry$^+$) l(3)05697* was transposed by crossing with transgenic flies expressing *P(ry$^+$Δ2-3)* transposase activity. *P(lacZ ry$^+$) l(3)05697* excision alleles were subsequently identified by non-rosy eyes and recovered in the absence of the P-element.

In summary, *P(lacZ ry$^+$) l(3)05697* was transposed in 232 independent lines. This generated 69 homozygous viable excision lines, of which 2 lines (3%) were male and female sterile, 31 lines (45%) were male sterile and 36 lines (52%) were revertants to wild-type *DmCBP20*. The remaining 163 homozygous lethal excision lines were shown to either have residual P-element sequences or deletions in *DmCBP20* gene locus (Figure 4.3, B and C). Finally, no temperature sensitive lines were found at 29°C.

4.2.3 *CBP20* is an essential gene in *Drosophila melanogaster*

Upon the mobilisation of *P(lacZ ry$^+$)l(3)05697* in a total of 232 independent lines, 163 imprecise excision lines (70%) were found to be homozygous lethal (*DmCBP20^{HL}*) (Figure 4.3, C). All 163 *DmCBP20^{HL}* alleles failed to be complemented by either the starter-element *P(lacZ ry$^+$)l(3)05697* or *Df(3R)P14 sr^1*, which uncovers the *DmCBP20* locus (data not shown). These results suggest that all 163 *DmCBP20^{HL}* lines had P-element induced lesions that resided within the *DmCBP20* locus. It was therefore reasoned that these mutations were probably due to either deletions in *DmCBP20* genomic sequence or residual P-element sequences.

Figure 4.2 Generation of *P(lacZ ry⁺)l(3)05697* excision alleles

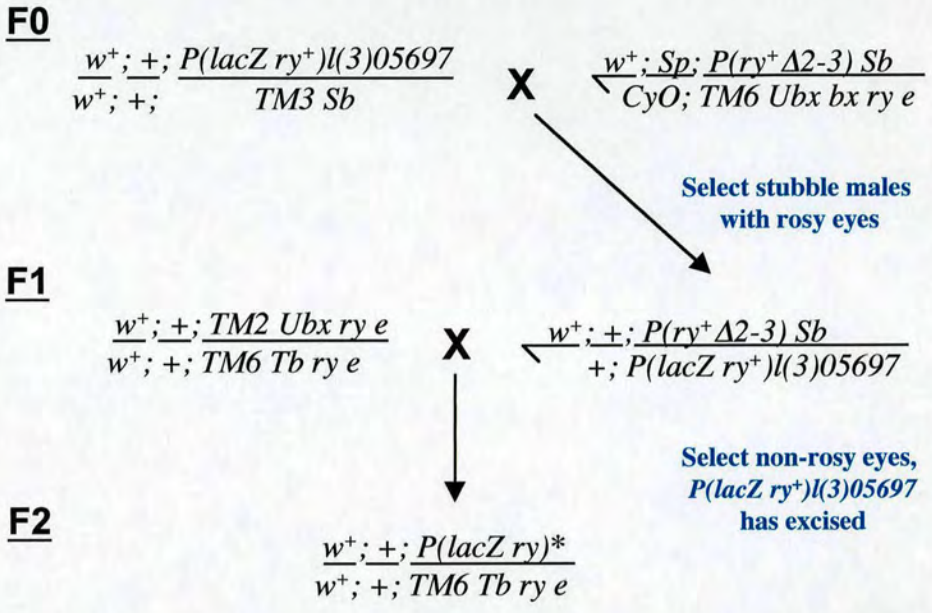
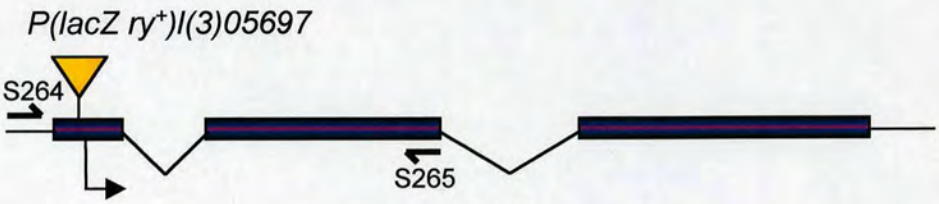


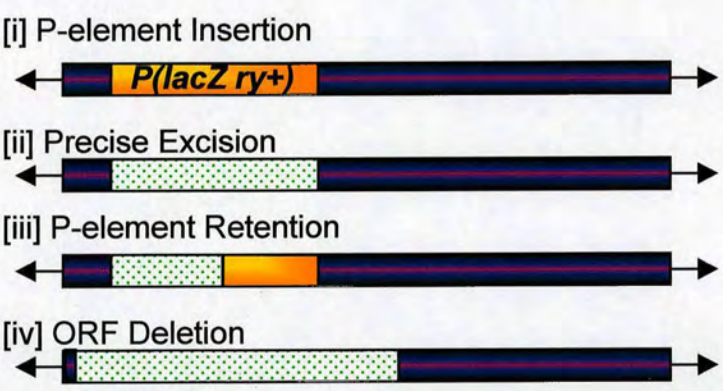
Figure 4.2 Genetic scheme for generating excision alleles of *P(lacZ ry⁺) l(3)05697*. Homozygous lethal *P(lacZ ry⁺)l(3)05697* located in the first exon of *DmCBP20* was mobilised and excision alleles screened for mutations in *DmCBP20*. **F0.** *P(lacZ ry⁺)l(3)05697* virgins are mated to males carrying transposase activity *P(ry+Δ2-3)* marked with stubble. **F1.** Male progeny with rosy eyes and stubble contain both *P(lacZ ry⁺)l(3)05697* and transposase activity. They are mated to double balancer TM2/ TM6 virgins. **F2.** Progeny with non-rosy eyes represent lines where the P-element has excised (indicated by asterisk). The excision lines are then tested for complementation with *Df(3R)P14 sr¹*, which uncovers the *DmCBP20* locus and stable stocks generated.

Figure 4.3 Mobilisation of *P(lacZ ry⁺)l(3)05697*

A



B



C

Excision Alleles	Number of Lines
A. Homozygous viable	69
1. Female sterile	2
2. Male & Female sterile	31
3. Revertants	36
B. Homozygous lethal	163

Figure 4.3 Mobilisation of *P(lacZ ry⁺)l(3)05697*. **A.** Schematic representation of *DmCBP20* gene structure with positions of primers used for characterisation of *P(lacZ ry⁺)l(3)05697* excision lines. **B.** P-element mobilisation generates four distinct classes of mutations: **[i]** P-element re-inserts in a new chromosomal location. **[ii]** P-element excises precisely from insertion site. Imprecise excisions generate lesions within P-insertion site that **[iii]** contain residual P-element sequences or **[iv]** delete portions of the genomic sequence. **C.** Summary of excision lines generated from mobilisation of *P(lacZ ry⁺)l(3)05697*.

To differentiate between deletion and P-element retention mutations, a mass PCR screen was employed to primarily detect *DmCBP20*^{HL} alleles that contained residual P-element sequences. Figure 4.3, A illustrates schematically the primer pair (oligos S264 + S265) designed to amplify across 445 bp of *DmCBP20* around the *P(lacZ ry*$^{+}$*)l(3)05697* insertion site. Genomic DNA was extracted from all 163 *DmCBP20*^{HL}/ TM3 lines and used in a standard Taq PCR reaction. The PCR products were subsequently resolved and analysed on 3% nusieve gels (Appendix A). This allowed the detection of *DmCBP20*^{HL} lines containing residual P-element sequences, since one wild-type band from the balancer chromosome was amplified along with a band from *DmCBP20*^{HL} chromosome. In order to substantiate that the PCR amplified bands did correspond to wild-type and imprecise excision alleles, selected DNA bands were gel purified and sequenced (data not shown).

The results from the PCR screen are presented in Appendix A. In summary, from 163 *DmCBP20*^{HL} lines screened, 51 lines (31%) were identified as containing residual P-element sequences that ranged from approximately 25 to 1, 700 bp in size. It was reasoned that the remaining 112 lines either contained deletions in *DmCBP20* genomic sequence or very small residual P-element sequences that were not detected on the agarose gels. To test this, putative *DmCBP20* deletion lines #1, #2 and #3 were characterised by southern analysis. This confirmed that these *DmCBP20*^{HL} lines indeed contained small deletions in *P(lacZ ry*$^{+}$*)l(3)05697* insertion site, which removed approximately 200 bp of genomic sequence (E. Thompson).

Out of 163 *DmCBP20*^{HL} lines screened, 7 lines were selected for further analysis. They included the three deletion lines described above and P-element retention lines #6, #121, #16 and #65 that contained approximately 25, 100, 400 and 1, 700 bp of residual P-element sequences respectively. To determine the stage of lethality, embryonic and pupal hatching frequencies were performed and the lethal phase found to occur during the first or second larval instars. Moreover, closer examination of homozygous *DmCBP20*^{HL} first instar larvae showed no obvious morphological

defects when compared to wild-type larvae (data not shown).

In order to verify that the lethality observed in homozygous *DmCBP20*^{HL} mutants was due to the imprecise excision of *P(lacZ ry*$^{+}$*)l(3)05697*, rescue crosses were performed on the 7 selected *DmCBP20*^{HL} lines. Two different *DmCBP20* transgenes were constructed and over-expressed in flies. The first transgene expressed *DmCBP20* cDNA under the control of the constitutive armadillo promoter and the second transgene expressed *DmCBP20* genomic fragment with its own promoter sequences. Transgenic lines over-expressing either *DmCBP20* construct were crossed into *DmCBP20*^{HL} mutant background and the adult progeny scored for rescue of homozygous lethality. Both transgenes that expressed either *DmCBP20* cDNA or *DmCBP20* genomic region generated viable adults homozygous for *DmCBP20*^{HL}. Rescued flies expressing either transgene were phenotypically wild-type and fertile, which indicated that the rescue was complete. Furthermore, the rescue was independent of whether one or two copies of the transgene were present (data not shown).

Taken together, these results suggest that *CBP20* is an essential gene in *Drosophila melanogaster*. *P(lacZ ry*$^{+}$*)l(3)05697* imprecise excision alleles either have small deletions or residual P-element sequences in the 5'UTR of *DmCBP20* locus. These alleles appear to be hypomorphs and the mutations might affect the translation of *DmCBP20*. In order to confirm this, western analysis on staged embryos, first and second instar larvae of wild-type and homozygous *DmCBP20*^{HL} mutants needs to be performed. This would be expected to show that in *DmCBP20*^{HL} mutants, DmCBP20 protein declines as development proceeds and the maternal contribution titrates out.

4.2.4 Testing for genetic interaction with *DmCBP20^{HL}* mutants

After the initial characterisation of *DmCBP20^{HL}*, the next question to address was whether these mutants were useful tools to identify CBC-interacting genes. Therefore, a preliminary screen was performed where selected genes and deficiencies were tested for interaction with *DmCBP20^{HL}*. In particular, *DmCBP20^{HL}* #1 and #65, which have 200 bp of genomic sequence deleted and 1, 700 bp residual P-element sequence in the 5' UTR respectively, were selected for genetic analysis. Since *DmCBP20^{HL}* heterozygotes are phenotypically wild-type, they were crossed to selected alleles and F1 adult progeny scored for dominant phenotypes that included lethality, male/ female sterility and male:female ratio. The alleles tested for genetic interaction with *DmCBP20^{HL}* #1 and #65 included mutants in *DmCBP80*, RNAP II and third chromosome deficiency kit.

Considering DmCBP20 and DmCBP80 must heterodimerise to form a functional CBC (Izaurralde *et al.*, 1994), it was reasoned that an obvious gene to test for interaction with *DmCBP20^{HL}* would be its cap-binding partner *DmCBP80.* At the time these genetic interaction crosses were performed, the only available alleles in *DmCBP80* were *Df(1)bi-D3*, which uncovers the *DmCBP80* locus and four transgenic lines that over-express *DmCBP80* using the ubiquitous armadillo promoter. This determined that crossing *DmCBP80* alleles into the genetic background of either *DmCBP20^{HL}* #1 or #65 did not generate observable dominant adult phenotypes.

Many lines of evidence have shown that RNAP II transcription is coupled *in vivo* to various mRNA processing events. Current evidence strongly supports a model whereby the CTD of the large subunit of RNAP II (*RNAP II 215*) not only recruits mRNA processing factors to nascent transcripts but also directly functions in RNA cap formation, pre-mRNA splicing and RNA 3'-end formation (Hirose and Manley, 2000 and Chapter 1). Since CBC mediates the positive effects of the cap in splicing and RNA 3'-end formation, it was reasoned that CBC might also co-operate with the

CTD of RNAP II to facilitate these processes. As a result, *DmCBP20*^{HL} #1 and #65 were crossed to various alleles of the largest subunit of RNAP II listed in Appendix F. Similar to *DmCBP80* alleles tested, the F1 adult progeny were viable and fertile exhibiting no dominant phenotypes. Finally, *DmCBP20*^{HL} #1 and #65 were crossed to a collection of deficiencies on the third chromosome. These deletions collectively uncovered approximately 65% of the third chromosome. However, no *DmCBP20*-interacting deficiencies were identified.

Taken together, the data suggests that *DmCBP20*^{HL} alleles were not ideal tools to use in screens for CBC-interacting genes. *DmCBP20*^{HL} heterozygote adults are phenotypically wild-type in appearance and do not provide a sensitised genetic background. As a result, the identification of CBC-interacting genes is limited by this strategy.

4.2.5 *DmCBP20* contributes to gametogenesis

Mobilisation of *P(lacZ ry*$^{+}$*)l(3)05697* in 232 independent lines generated 69 lines (30%) that were homozygous viable (Figure 4.3, C). These excision lines were subsequently screened for temperature sensitivity at 29°C and male or female sterility. In order to test for temperature sensitivity at 29°C, homozygous viable lines were crossed to *Df(3R)P14 sr* ^{1} and the F1 progeny scored for viability at 25°C and 29°C. This determined that none of the 69 homozygous viable lines were temperature sensitive at 29°C. Male or female sterility was assayed by crossing to wild-type *Ore R* adults and observing whether viable F1 adult progeny were generated. Surprisingly a very high number, approximately 45% (31 from 69 lines) of homozygous lines were found to be male sterile. In contrast, a relatively small number were identified as both male and female sterile (3%; 2 from 69). The remaining 36 homozygous lines (52%) were predicted to be precise excisions and therefore contain wild-type *DmCBP20* sequences. To test this, genomic DNA was extracted from 6 randomly selected putative precise excision lines and full-length *DmCBP20* genomic region was PCR amplified using oligos S264 and S265, gel purified and sequenced (data not shown). This confirmed that *P(lacZ ry*$^{+}$*)l(3)05697*

had precisely excised from the 5' UTR of *DmCBP20* to generate a wild-type genomic sequence.

4.2.6 Characterisation of *DmCBP20^{MFS}* mutants

Out of 69 homozygous viable excision lines, two (3%) were initially identified as female sterile. However, further examination of homozygous males demonstrated that both lines were also male sterile. *DmCBP20* male and female sterile (*DmCBP20^{MFS}*) lines #54 and #84 were found to be lethal when crossed with *Df(3R) P14 sr^{1}*, which uncovers *DmCBP20* locus (data not shown). In addition, both alleles failed to complement the original P-element, *P(lacZ ry^{+})l(3)05697*. As a result, *DmCBP20^{MFS}* #54/ *P(lacZ ry^{+})l(3)05697* progeny were lethal whilst *DmCBP20^{MFS}* #84/ *P(lacZ ry^{+})l(3)05697* were females sterile (data not shown). Taken together, these genetic crosses strongly suggest that lesions in *DmCBP20^{MFS}* lines #54 and #84 occurred within *DmCBP20* locus.

In order to identify the mutations at DNA level, both *DmCBP20^{MFS}* alleles were sequenced. Genomic DNA was extracted from homozygous *DmCBP20^{MFS}* #54 and #84 adults and *DmCBP20* genomic region PCR amplified using the primers illustrated in Figure 4.3, A. DNA bands were subsequently extracted, gel purified and sequenced (data not shown). The results from the DNA sequencing showed that *DmCBP20^{MFS}* #54 and *DmCBP20^{MFS}* #84 contained 65 bp and 30 bp of residual P-element sequences respectively. This data suggests that *DmCBP20^{MFS}* alleles might be hypomorphs where additional sequences in the 5' UTR might affect *DmCBP20* translation and consequently generate the sterility phenotype.

The next question to address was whether gonads from *DmCBP20^{MFS}* adults exhibited any gross morphological abnormalities. As a result, testes were dissected from homozygous #54 and #84 adults, squashed and examined by light microscopy (Mar Carmena and data not shown). This determined that the morphology of external genitalia from both *DmCBP20^{MFS}* lines appeared normal with no obvious defects

compared to wild-type. All the early stages of spermatogenesis appeared to be present from germ stem cells to elongating sperm. Testes from *DmCBP20*^{MFS} #54 and #84 adults contained both pre-meiotic spermatocytes and post-meiotic spermatids, which indicated that meiosis was normal. In addition, cysts that contained elongating sperm were apparent. At the end of spermatid maturation these cysts form compact bundles of mature, individualised sperm (Venables and Eperon, 1999). Although individualised sperm were observed in lines #54 and #84, it was unclear whether these sperm were present at the wild-type number (Mar Carmena and data not shown).

These observations suggest that in *DmCBP20*^{MFS} #54 and #84 homozygous males, the early meiotic stages of spermatogenesis are similar to wild-type. However, it is possible that the defects that induce male sterility arise during the final stages of spermatid differentiation. Consequently, closer examination of the final stages of spermatogenesis is required, for example, to determine whether mature sperm are motile. Whilst EM studies would reveal if prominent structures in the spermatid tail such as the axoneme and mitochondria are present.

In contrast to homozygous males, examination of ovaries dissected from matured homozygous females revealed gross morphological defects. Representative mutant ovarioles from both homozygous *DmCBP20*^{MFS} lines are illustrated in Figure 4.4. As can be observed, the phenotypes of homozygous *DmCBP20*^{MFS} #84 and *DmCBP20*^{MFS} #54 ovarioles were distinct from wild-type (compare Panels A with B and C). Closer examination of homozygous *DmCBP20*^{MFS} #84 females showed they did not lay eggs and had ovaries that were approximately 30% smaller than wild-type (data not shown). A majority of ovarioles (74%) appeared to be mutant and this phenotype was consistent within an ovary (Figure 4.4, B). Generally, the early stages of oogenesis looked wild-type, however, developing oocytes appeared to progressively deteriorate from stage 6.

Figure 4.4 Homozygous *DmCBP20*^{MFS} #54 and #84 Ovarioles have Gross Morphological Defects

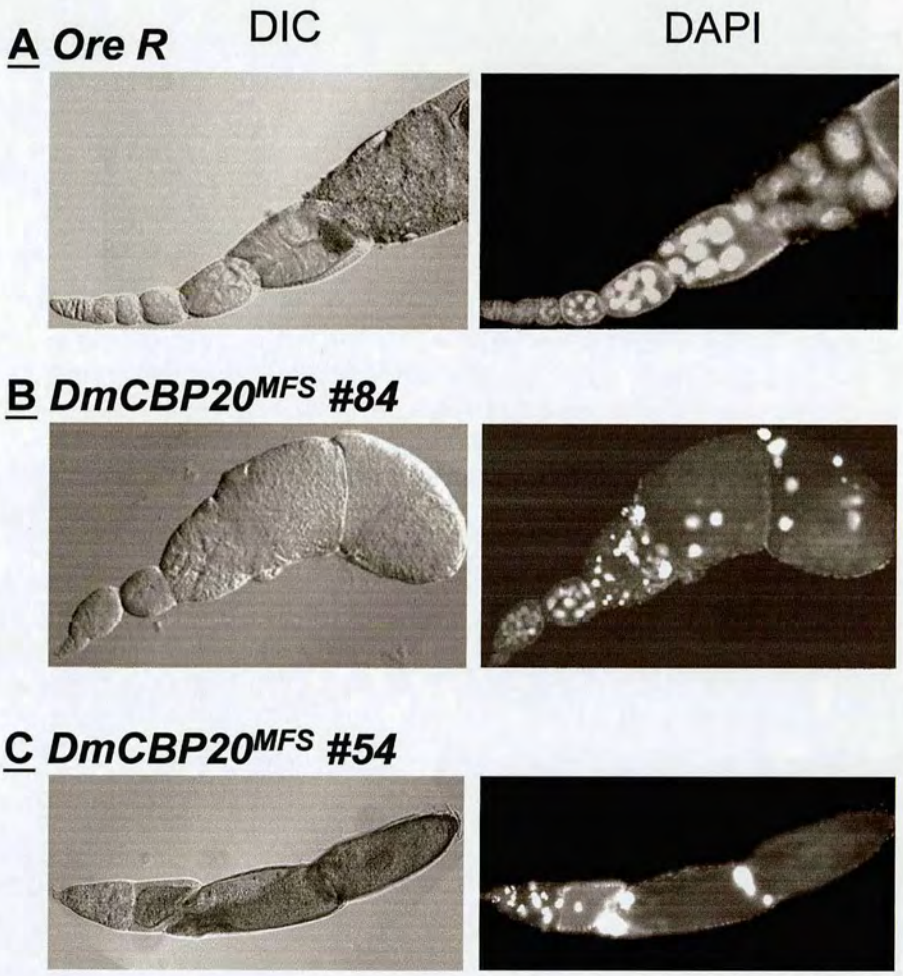


Figure 4.4 Homozygous *DmCBP20*^{MFS} #54 and #84 ovarioles have gross morphological defects. Ovaries were dissected and fixed from matured **A.** wild-type *Ore R*, **B.** homozygous *DmCBP20*^{MFS} *#84* and **C.** homozygous *DmCBP20*^{MFS} *#54* adults. DNA was then stained with DAPI and ovarioles visualised using DIC and fluorescence microscopy.

As can be observed from Figure 4.4 B, DAPI stained homozygous *DmCBP20*^{MFS} #84 ovariole clearly showed discrete DNA bodies from stage 7 onwards. This suggests that the egg chambers might undergo degeneration by apoptosis. Consistent with cell death, the nurse cells also looked abnormal and fragmented around these stages. Furthermore, ovarioles appeared to back-up as each ovariole typically acquired two stage 13-14 egg chambers. If mature egg chambers in backed-up ovarioles do not enter the oviduct they are eventually reabsorbed. This probably accounts for why homozygous *DmCBP20*^{MFS} #84 females do not lay eggs.

Initial observations also showed that deteriorating egg chambers appeared to be fused, which suggests signalling defects might underlie the homozygous *DmCBP20*^{MFS} #84 mutant phenotype. To test this, homozygous *DmCBP20*^{MFS} #84 ovaries were stained with DAPI, examined on an inverted Olympus IX70 microscope and fluorescence images processed by deconvolution analysis (Gillian MacKay). As can be observed from Figure 4.5, each chamber was indeed individualised and surrounded by an intact cell membrane. This was further confirmed by staining homozygous *DmCBP20*^{MFS} #84 ovarioles with rhodamine-phalloidin that specifically detects actin (Gillian MacKay and data not shown). In addition, egg chambers that matured far enough contained mature oocytes that were determined with wild-type follicle cells. Whilst RNA *in situ* analysis of gurken mRNA, which contributes to anterior-posterior axis specification, showed that it was correctly localised in early oocytes in a posterior crescent between the nucleus and follicle cells (Nina MacDougall and data not shown). Taken together, these data suggest that in homozygous *DmCBP20*^{MFS} #84 ovarioles cell signalling and specification of axis formation are wild-type.

Figure 4.5 Homozygous *DmCBP20^{MFS}* #84 Chambers are not Fused

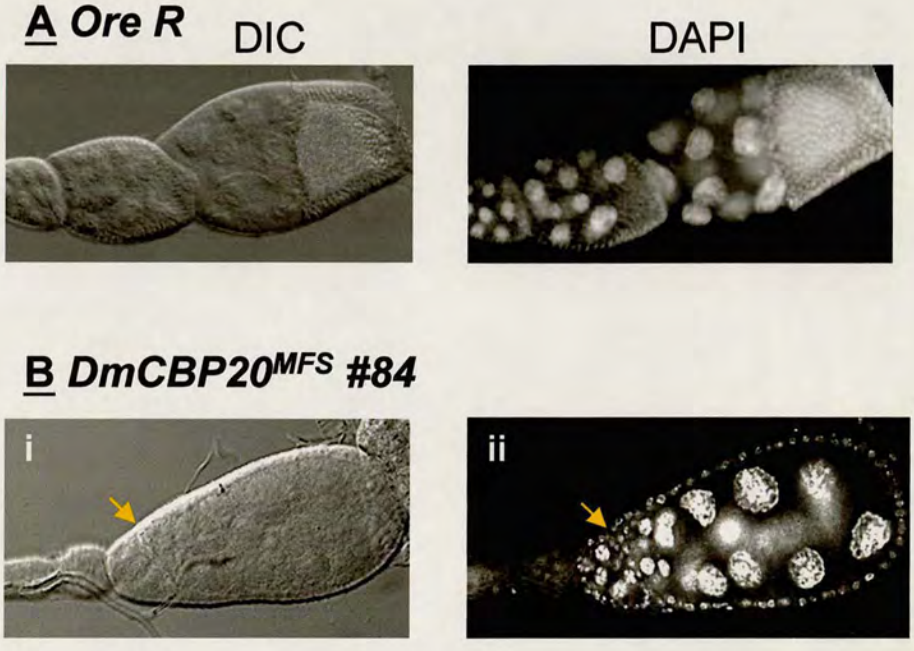


Figure 4.5 Homozygous *DmCBP20^{MFS}* #84 chambers are not fused. Ovaries were dissected and fixed from matured **A.** wild-type *Ore R* and **B.** homozygous *DmCBP20^{MFS} #84* adults. DNA was then stained with DAPI and ovarioles visualised using DIC and fluorescence microscopy on an inverted Olympus IX70 microscope. Fluorescence images were then processed by deconvolution analysis. **[i]** Under DIC, the yellow arrow points to a homozygous *DmCBP20^{MFS} #84* egg chamber that appears to be fused. **[ii]** DAPI image shows the ovariole has two distinct egg chambers.

In contrast to homozygous *DmCBP20*^{MFS} #84, homozygous *DmCBP20*^{MFS} #54 females did lay eggs, some of which were strikingly flaccid in appearance. Further examination of ovaries showed they were approximately 20% smaller than wild-type. A majority of ovarioles looked normal with only 23% exhibiting a mutant phenotype, which was consistent within an ovary (data not shown). A representative homozygous *DmCBP20*^{MFS} #54 mutant ovariole is presented in Figure 4.4 C. Similar to *DmCBP20*^{MFS} #84, *DmCBP20*^{MFS} #54 phenotype exhibited densely stained DNA bodies from stage 7 onwards. As before, this egg chamber degeneration is suggestive of apoptosis.

Interestingly, initial observations showed approximately 30% of mature homozygous *DmCBP20*^{MFS} #54 eggs were flaccid, which suggests defects might exist within the chorion egg shell. To test this, egg chorions were closely examined using dark field microscopy. This determined that the chorions were intact and looked normal. Moreover, prominent features of the chorion such as the two dorsal appendages and micropyle were wild-type (Mary Bownes and data not shown). It was therefore reasoned that the flaccid appearance of homozygous *DmCBP20*^{MFS} #54 eggs might be due to a deficiency in the major yolk proteins. Yolk proteins are expressed in the fat body and follicle cells and transported to the ovaries where they accumulate in large amounts during stages 8 to 10 (Richard *et al.*, 2001). As a result, levels of the three major yolk proteins were analysed by western blotting using the anti-yolk protein (YP) antibody directed against YP1, YP2 and YP3. This demonstrated that all three yolk proteins were present at similar levels to wild-type ovaries (Figure 4.6, B). Collectively, this data shows that the flaccid appearance of homozygous *DmCBP20*^{MFS} #54 mature eggs does not result from defects in major yolk proteins or egg shell formation.

Figure 4.6 DmCBP20 Protein is Lower in Homozygous *DmCBP20^{MFS}* #54 and #84 Ovaries

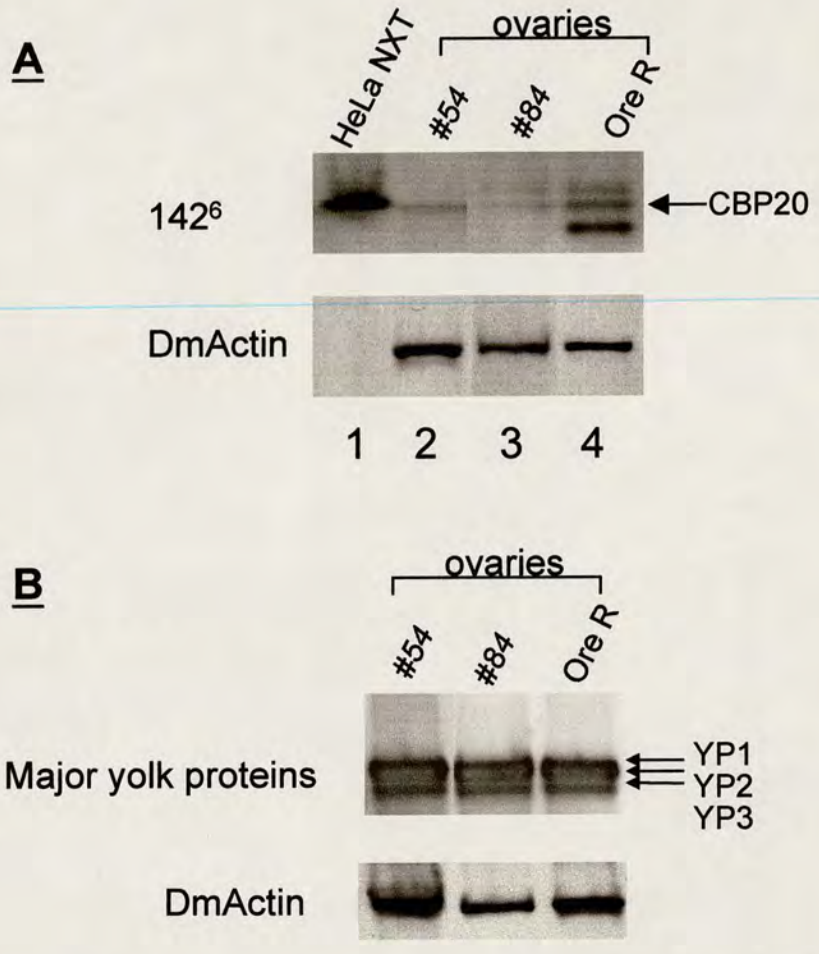


Figure 4.6 DmCBP20 protein is lower in homozygous *DmCBP20^{MFS}* #54 and #84 ovaries. Western blot analysis of DmCBP20 and major yolk proteins in *DmCBP20^{MFS} #54* and *#84* ovaries compared to wild-type. Ovaries were dissected from matured *Ore R* wild-type, homozygous *DmCBP20^{MFS} #54* and homozygous *#84* adults and total protein extracted. Equivalent amounts of total protein were loaded as quantitated by sypro-orange staining using Storm 860 PhosphorImager in blue fluorescence together with ImageQuant software. Proteins were then resolved by SDS-PAGE, western blotted and nitrocellulose membrane probed with **A.** anti-142^6 antibody and **B.** anti-yolk protein (YP) that detects major YP1-3. As an additional loading control, filter was re-probed with anti-DmActin antibody that is ubiquitously expressed throughout oogenesis.

In order to address whether *DmCBP20*^{MFS} #54 and #84 mutations belong to the same complementation group, both alleles were crossed together and the F1 trans-heterozygote progeny scored for sterility. This determined that both male and female *DmCBP20*^{MFS} #84/ *DmCBP20*^{MFS} #54 adults were fertile, which suggests that the two alleles belong to different complementation groups (data not shown). To examine the female trans-heterozygotes more directly, ovaries were dissected from matured adults and stained with DAPI. As can be observed from Figure 4.7, the morphology of *DmCBP20*^{MFS} #84/ *DmCBP20*^{MFS} #54 ovarioles appeared normal when compared to wild-type (compare Panels A and B). The characteristic discrete DNA bodies that were prevalent around stage 7 onwards in both homozygous *DmCBP20*^{MFS} #54 and #84 ovarioles were absent. Furthermore, mature eggs from *DmCBP20*^{MFS} #84/ *DmCBP20*^{MFS} #54 adults were wild-type in appearance. Ideally, western analysis of ovaries dissected from trans-heterozygotes would also be performed to support the genetic data. This would be expected show that DmCBP20 protein level of *DmCBP20*^{MFS} #84/ *DmCBP20*^{MFS} #54 ovaries are higher compared to either homozygous *DmCBP20*^{MFS} allele alone. Taken together, these genetic data suggest that mutations in *DmCBP20*^{MFS} #54 and #84 do complement each other and might therefore induce different molecular defects.

To confirm the sterility observed in homozygous *DmCBP20*^{MFS} #54 and #84 adults was due to the imprecise excision of *P(lacZ ry*$^{+}$*)l(3)05697* from *DmCBP20*, specific rescue crosses were performed. Transgenic lines expressing either *DmCBP20* cDNA expressed under a constitutive promoter or *DmCBP20* genomic region were crossed into the homozygous *DmCBP20*^{MFS} background. The adult progeny were subsequently scored for rescue of male and female homozygous sterility and in both cases, fertile adults were found homozygous for *DmCBP20*^{MFS}.

Figure 4.7 *DmCBP20^{MFS}* #54 Complements *DmCBP20^{MFS}* #84

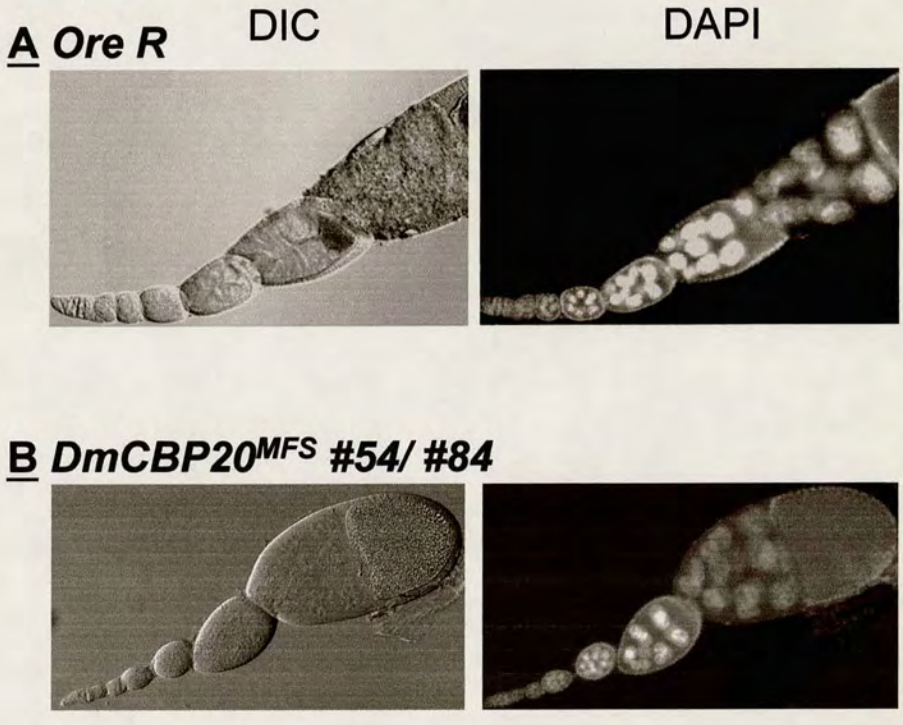


Figure 4.7 *DmCBP20^{MFS}* #54 complements *DmCBP20^{MFS}* #84. Complementation testing of *DmCBP20^{MFS}* alleles. Ovaries were dissected and fixed from matured **A.** wild-type *Ore R* and **B.** *DmCBP20^{MFS} #54/ 84* trans-heterozygotes. DNA was then stained with DAPI and ovarioles visualised using DIC and fluorescence microscopy.

Furthermore, morphological analysis of ovaries dissected from rescued *DmCBP20*^{MFS} #54 and #84 females showed they appeared normal with no obvious defects compared to wild-type ovaries (data not shown). This indicated that the rescue of homozygous *DmCBP20*^{MFS} #54 and #84 females was complete.

The next question to address was whether CBC mRNA expression was disrupted in homozygous *DmCBP20*^{MFS} #54 and #84 lines. RNA *in situ* hybridisation was therefore used to detect the localisation of DmCBP20 and DmCBP80 transcripts in ovarian tissues. Full-length *DmCBP20* and *DmCBP80* cDNAs were used to generate DIG-labelled anti-sense RNA that specifically detects mRNA and sense RNA, which serves as a negative control. The RNA probes were then hybridised to fixed homozygous *DmCBP20*^{MFS} #54, #84 and wild-type ovaries. As can be observed from Figure 4.8, the expression pattern of DmCBP20 and DmCBP80 RNAs were the same in *DmCBP20*^{MFS} and wild-type ovarioles. Transcripts encoding both cap-binding proteins were ubiquitously expressed in all ovarian tissues from *DmCBP20*^{MFS} #54 and #84 mutants. As a result, the localisation of DmCBP20 and DmCBP80 RNAs did not appear to be affected in homozygous *DmCBP20*^{MFS} #54 and #84 adults.

Since RNA *in situ* hybridisation is a qualitative assay, it could not be used to assess the levels of DmCBP20 RNA present in *DmCBP20*^{MFS} #54 and #84 ovaries. Consequently, western blotting was used to examine the amount of DmCBP20 protein in *DmCBP20*^{MFS} mutants. Ovaries were dissected from mature wild-type, homozygous *DmCBP20*^{MFS} #54 and #84 adults and the total protein extracted. Equivalent amounts of protein were loaded as quantitated by sypro-orange fluorescence using the Storm 860 PhosphorImager together with ImageQuant software (data not shown). Proteins were then resolved by SDS-PAGE and analysed by western blotting. CBP20 protein was detected using anti-142[6] antibody that is directed against HsCBP20 but also cross-reacts with DmCBP20 (Izaurralde *et al.*, 1994).

Figure 4.8 DmCBP20 and DmCBP80 mRNAs are Ubiquitously Expressed in Homozygous *DmCBP20^{MFS}* #54 and #84 Ovarioles

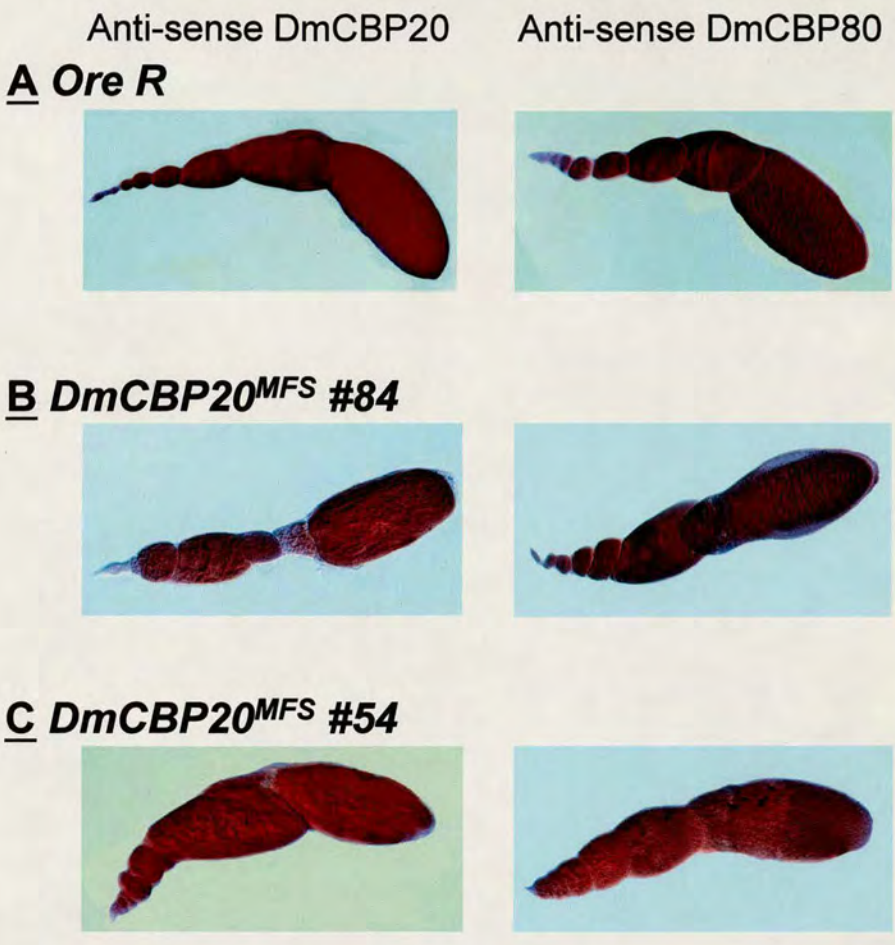


Figure 4.8 DmCBP20 and DmCBP80 mRNAs are ubiquitously expressed in homozygous *DmCBP20^{MFS}* #54 and #84 ovarioles. RNA *in situ* hybridisation performed on fixed matured **A**. wild-type *Ore R*, **B.** homozygous *DmCBP20^{MFS} #84* and **C.** homozygous *DmCBP20^{MFS} #54* ovaries. DmCBP20 and DmCBP80 mRNAs were detected using DIG-labelled anti-sense RNA.

The specificity of anti-142[6] for CBP20 is observed in Figure 4.6, A where the antibody predominantly recognised a band corresponding to CBP20 in both HeLa nuclear extracts and ovaries (upper panel, compare lanes 1 and 4). The western blot shows that the levels of DmCBP20 protein appear to be reduced in homozygous *DmCBP20*^{MFS} #54 and #84 ovaries (upper panel, compare lane 2 and 3). As an additional loading control the western blot was re-probed with anti-DmActin that is expressed ubiquitously throughout oogenesis (lower panel). Collectively, this data suggests that the additional residual P-element sequences in the 5'UTR of *DmCBP20*^{MFS} #54 and #84 might affect *DmCBP20* translation in gonads. This leads to lower levels of DmCBP20 protein, which consequently generates the sterility phenotype.

4.2.7 Analysis of RNA processing in *DmCBP20*^{MFS} ovaries

Current evidence has shown that CBC facilitates the positive effects of the cap in at least three major aspects of nuclear RNA processing: pre-mRNA splicing, RNA 3'-end formation and U snRNA export (Lewis and Izzauralde, 1997). To test whether the molecular defect in *DmCBP20*^{MFS} ovaries was due to splicing, northern blotting was used to assay for aberrant pre-mRNA splicing of endogenous ribosomal protein RNAs. Ribosomal proteins RpL19, RpS3 and RpA1 were specifically selected for analysis since they are ubiquitously expressed throughout oogenesis and have pre-mRNAs that contain two introns, one intron and no introns respectively. Characterisation of CBC has shown that it mediates the stimulatory effect of the cap in the splicing of the cap-proximal intron (Izaurralde *et al.*, 1994; Lewis *et al.*, 1996). Therefore, if the sterility phenotype results from a general splicing defect, increased steady state pre-mRNA to mRNA ratios would be observed in RpS3 that contains one intron. In contrast, *DmCBP20*^{MFS} mutants would be predicted to have no influence on the splicing of RpL19 that contains two introns or the steady state levels of RpA1, which contains no introns and effectively serves as a negative control.

Ovaries were dissected from matured *Ore R* wild-type, homozygous *DmCBP20^{MFS}* #54 and *DmCBP20^{MFS}* #84 adults and the total RNA extracted. Equivalent amounts of RNA were loaded as determined by ethidium stained ribosomal RNAs and analysed by northern blotting. The nitrocellulose filters were probed with full-length *RpL19*, *RpS3* and *RpA1* cDNAs and the steady state ratios of pre-mRNA to mRNA quantitated using the Storm 860 PhosphorImager together with ImageQuant software. This determined that neither *DmCBP20^{MFS}* #54 nor #84 exhibited a significant increase in the steady state pre-mRNA to mRNA ratio in any of the ribosomal proteins (data not shown). It is possible that the northern analysis was not sensitive enough to detect a splicing defect in *DmCBP20^{MFS}* ovaries. Alternatively, defects could result from other functions of CBC such as RNA 3'-end formation. It was therefore reasoned that a more sensitive assay such as RNase protection would be better suited to investigate aberrant pre-mRNA splicing in *DmCBP20^{MFS}* adults. In addition, RNase protection could assay whether a defect in poly(A) tail formation existed in *DmCBP20^{MFS}* #54 and #84 ovaries. Unfortunately, these experiments were not completed due to lack of time.

4.2.8 Characterisation of *DmCBP20^{MS}* male steriles

In contrast to the two *DmCBP20^{MFS}* mutations, mobilisation of *P(lacZ ry^{+})l(3)05697* also generated a very high number of lines (45%; 31 from 69 lines) that were homozygous male sterile (*DmCBP20^{MS}*). Initial genetic crosses showed that all 31 *DmCBP20^{MS}* lines failed to complement either *Df(3R)P14 sr^{1}*, which uncovers *DmCBP20* locus or the original P-element. Flies that possessed the genotype *DmCBP20^{MS}/ Df(3R)P14 sr^{1}* were found to be lethal. Whilst *DmCBP20^{MS}/ P(lacZ ry^{+})l(3)05697* adults were either semi-lethal, male or female sterile (data not shown). Taken together, these genetic crosses suggest that similar to *DmCBP20^{MFS}*, *DmCBP20^{MS}* alleles appear to be hypomorphs with mutations that reside within the *DmCBP20* locus.

In order to determine the mutations at DNA level, *DmCBP20*^{MS} lines #1, #15 and #21 were randomly selected for further analysis. Genomic DNA was extracted from homozygous males and *P(lacZ ry*$^{+}$*)l(3)05697* insertion site in *DmCBP20* PCR amplified using oligos S264+ S265 illustrated in Figure 4.3, A. DNA bands were subsequently extracted, gel purified and sequenced. Similar to *DmCBP20*^{MFS} mutants, the results of the DNA sequencing showed that *DmCBP20*^{MS} alleles #1, #21 and #15 contained 53, 40 and 100 bp of residual P-element sequences respectively (data not shown). This suggests that like *DmCBP20*^{MFS} alleles, additional sequences in the 5' UTR might affect the translation of *DmCBP20* leading to the male sterility phenotype.

To determine whether the homozygous male sterility could be rescued by *DmCBP20* transgenes, *DmCBP20*^{MS} lines #1, #15 and #21 were subjected to the same rescue crosses previously performed on *DmCBP20*^{HL} and *DmCBP20*^{MFS} mutants. Transgenic lines expressing either *DmCBP20* cDNA expressed under a constitutive promoter or a fragment of genomic region encompassing *DmCBP20* were crossed into the homozygous *DmCBP20*^{MS} background. Homozygous *DmCBP20*^{MS} adult male progeny were subsequently scored for fertility by crossing to wild-type females. As a result, homozygous *DmCBP20*^{MS} males, which carried a copy of *DmCBP20* transgene were fertile (data not shown). This indicated that the male sterility observed was specifically due to mutations in *DmCBP20* locus.

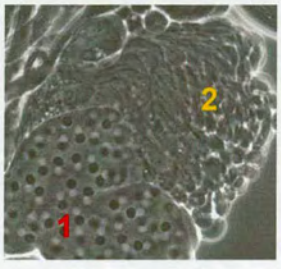
The next question to address was whether testes from *DmCBP20*^{MS} mutants exhibited any gross morphological abnormalities. Consequently, testes were dissected from homozygous #1, #15 and #21 adults, squashed and examined by light microscopy (Mar Carmena). Similar to *DmCBP20*^{MFS} #54 and #84 males, the morphology of external genitalia from homozygous #1, #15 and #21 *DmCBP20*^{MS} lines appeared normal with no obvious defects compared to wild-type. All the stages of early spermatogenesis were present from germ stem cells to elongating sperm. Testes from *DmCBP20*^{MS} males contained wild-type pre-meiotic spermatocytes and

post-meiotic spermatids indicating that meiosis was normal. In addition, cysts containing elongating sperm were apparent in all *DmCBP20*^{MS} adults examined (Figure 4.9 and data not shown). At the end of spermatid maturation these cysts form compact bundles of mature, individualised sperm (Venables and Eperon, 1999). Although mature sperm were present in line #21, none were observed in *DmCBP20*^{MS} lines #1 or #15 (Mar Carmena).

These observations suggest that like *DmCBP20*^{MFS} #54 and #84, the early meiotic stages of spermatogenesis in homozygous *DmCBP20*^{MS} males are wild-type. However, it appears that *DmCBP20*^{MS} mutations might induce defects during the final stages of spermatid differentiation. As a result, this would affect the production of mature, viable sperm and cause the male sterility phenotype.

Figure 4.9 Homozygous *DmCBP20^{MF}#21* Testes do not have Gross Morphological Defects

A ***Ore R***



B ***DmCBP20^{MF} #21***

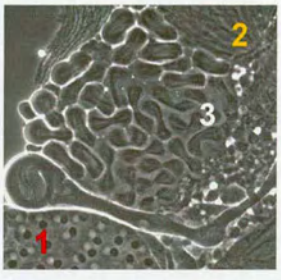


Figure 4.9 Homozygous *DmCBP20^{MF}* #21 testes do not have gross morphological defects. Testes were dissected from matured **A.** wild-type *Ore R* and **B.** homozygous *DmCBP20^{MF} #21* adults. Testes were then visualised using phase contrast microscopy. **1.** Onion-stage spermatid cyst. **2.** Cyst of elongated spermatids. **3.** Cyst of 32 cells at anaphase/ telophase meiosis II.

4.3 DISCUSSION

4.3.1 Generating P-insertions in *DmCBP80*

The P-element mutagenesis was not successful in generating a P-induced allele of *DmCBP80* by local hopping. Although the hop was performed on a small scale, a reasonable number (2%; 10 from 280 lines) of new P-insertions were homozygous lethal and mapped to 4C5-6; 4C7-8 on the X-chromosome near *DmCBP80*, which is located at 4C7-8. This result is encouraging and suggests that it is worthwhile repeating the local hop strategy but with specific modifications. Firstly, the mutagenesis should be scaled up and a genetic scheme designed whereby two or more starter-elements located within 200 Kb of *DmCBP80* are transposed. Secondly, the assumption that P-insertion alleles in *DmCBP80* would be homozygous lethal should be discarded and all transposon insertions mapped to 4C5-6; 4C7-8 on the X-chromosome analysed molecularly. This assumption was based on the observation that *DmCBP20* is an essential gene, together with data describing the characterisation of both subunits from Chapter 3. These results suggest that *DmCBP20* and *DmCBP80* behave similarly, as expected from the constituents of a cap-binding complex that is only functional as a heterodimer. Interestingly, Chapter 5 describes how a second CBP80 homologue has been identified in *Drosophila melanogaster*. In contrast to *DmCBP80*, *DmCBP20* is a unique gene. This suggests that DmCBP20 might have additional functions in conjunction with the second DmCBP80 homologue, which unlike DmCBP80, render it essential. Finally, instead of southern analysis, an alternative molecular screening strategy is required if thousands of mutagenised flies are to be investigated within months. An assay such as PCR-mediated by vectorette is suitable for a large insertional mutagenesis screen (Eggert *et al.*, 1998a). This sensitive assay allows the detection of one fly with the correct P-insertion from a pool of 100 flies containing alternative independent transposon integrations. As a result, the number of single PCR reactions for screening can be reduced along with the time taken for the P-element induced mutagenesis screen.

4.3.2 *CBP20* is an essential gene in *Drosophila melanogaster*

Mobilisation of *P(lacZ ry$^+$)l(3)05697*, which is inserted into the 5' non-coding region of *DmCBP20*, generates a very high number (70%) of imprecise excision alleles that are homozygous lethal. Characterisation of *DmCBP20^{HL}* alleles show they either contain residual P-element sequences that range from 25 bp to 1.7 Kb or small genomic deletions of around 200 bp. Homozygous *DmCBP20^{HL}* mutants die during the first or second larval instar stages with no obvious morphological abnormalities. This lethality is rescued by expressing either *DmCBP20* using its own promoter sequences or the ubiquitously expressed armadillo promoter. As a result, these data show that *DmCBP20* is an essential gene in *Drosophila melanogaster*. This is not surprising considering CBC plays a key role in the metabolism of RNAP II transcripts in all cells.

Although *DmCBP20^{HL}* mutants have not been analysed biochemically, the lesions in the 5' UTR suggest that translation of *DmCBP20* might be affected. Chapter 3 has already shown that DmCBP20 is maternally derived. This suggests that even though the maternal contribution might be sufficient to propel development through embryogenesis, protein from *DmCBP20^{HL}* mutants might be insufficient to sustain *Drosophila* development beyond the first or second larval instar stages.

The molecular defect(s) that lead to the homozygous lethality in *DmCBP20^{HL}* mutants are yet to be elucidated. However, selected *DmCBP20^{HL}* lines have been crossed over GFP-marked balancers to form stable stocks. This will allow homozygous *DmCBP20^{HL}* embryos and larvae to be readily detected and isolated. Therefore, defects in pre-mRNA splicing, RNA 3'-end formation or U snRNA export could be assayed in homozygous *DmCBP20^{HL}* mutants using RNase protection assay and RNA *in situ* hybridisation respectively. It is thought that if DmCBP20 protein levels are lower in *DmCBP20^{HL}* mutants compared to wild-type, the lethality might not result from the disruption of one specific mRNA processing event. Instead, it is probably due to a culmination of defects in splicing, RNA 3'-end formation and U

snRNA export that ultimately disrupt RNAP II transcript metabolism and lead to lethality.

4.3.3 Genetic and molecular analysis of *DmCBP20^{HL}* mutants

Genetic interaction screens that use *DmCBP20^{HL}* deletion line #1 and P-element retention line #65 do not identify any CBC-interacting genes or deficiencies. Adults that contain *DmCBP20^{HL}* mutation in combination with the deficiency that uncovers *DmCBP80* or alleles that over-express *DmCBP80* are viable and fertile with no obvious dominant phenotypes. Similarly, no dominant adult phenotypes are observed when the third chromosome deficiency kit or *RNAP II 215* alleles are crossed into *DmCBP20^{HL}* background.

An interaction would be predicted between *DmCBP20^{HL}* alleles and mutants in the cap-binding partner *DmCBP80*. However it is possible that these interactions are not strong enough to be manifested as observable external adult phenotypes. This suggests that *DmCBP20^{HL}* mutants do not provide a sensitised genetic background suitable for genetic modifier screens. Instead, biochemical analysis of *DmCBP20* and *DmCBP80* double mutants might be more successful in revealing molecular defects in CBC's function during splicing, RNA 3'-end formation or U snRNA export. As a result, Chapters 6 and 7 describe an alternative genetic approach where the GAL4-UAS system is used to identify CBC-interacting genes. GAL4-UAS is used to over-express the components of CBC in non-essential adult tissues to generate dominant phenotypes. These dominant phenotypes can then be used to facilitate genetic modifier screens to identify CBC-interacting genes. Therefore unlike *DmCBP20^{HL}* mutants, both strong and weak genetic interactors of CBC can be identified.

4.3.4 *DmCBP20* contributes to gametogenesis

Transposition of *P(lacZ ry$^+$)l(3)05697* also generates adults that are either male sterile or male and female sterile. Two excision lines (3%), *DmCBP20^{MFS}* #54 and *DmCBP20^{MFS}* #84 are found where both females and males are homozygous sterile. In contrast, a surprisingly high proportion of homozygous excision lines (45%) are male sterile alone. Sequencing of the P-insertion site in *DmCBP20^{MFS}* and *DmCBP20^{MS}* lines #1, #15 and #21 show they retained short residual P-element sequences in the 5' UTR of *DmCBP20*. *DmCBP20^{MFS}* #54 and *DmCBP20^{MFS}* #84 contain 65 bp and 30 bp of residual P-element sequences respectively. Whilst the three *DmCBP20^{MS}* lines retain 40 to 100 nucleotides in the 5' UTR. Rescue crosses further confirm the sterility observed in *DmCBP20^{MFS}* and *DmCBP20^{MS}* lines is due to the imprecise excision of *P(lacZ ry$^+$)l(3)05697*. Consequently, fertility is restored to both homozygous females and males expressing *DmCBP20* transgenes under the control of the armadillo or its own promoter sequences.

RNA *in situ* analysis of *DmCBP20^{MFS}* mutants demonstrate the localisation of DmCBP20 and DmCBP80 transcripts are not affected in ovaries isolated from either homozygous *DmCBP20^{MFS}* #54 or *DmCBP20^{MFS}* #84 adults. Both cap-binding proteins are ubiquitously distributed in all ovarian tissues. However, western analysis shows the levels of DmCBP20 protein are lower in *DmCBP20^{MFS}* #54 and #84 ovaries compared to wild-type.

Taken together, these data implicate a role for *DmCBP20* during gametogenesis. It is interesting to note that 45% of homozygous viable lines are male sterile in comparison to 3% that are both male and female sterile. This strongly suggests that *DmCBP20* makes a greater contribution to male fertility. The additional P-element sequences in the 5' UTR of *DmCBP20^{MFS}* #54 and #84 might affect *DmCBP20* translation. As a result, the lower levels of DmCBP20 protein are insufficient to support either oogenesis or spermatogenesis. Similar to *DmCBP20^{MFS}* mutants, P-

element retention in *DmCBP20*^{MS} lines might also produce lower levels of DmCBP20 protein that cannot sustain spermatogenesis.

4.3.5 Morphological and molecular analysis of *DmCBP20*^{MFS} female steriles

Morphological analysis of ovarioles dissected from mature homozygous *DmCBP20*^{MFS} #54 and *DmCBP20*^{MFS} #84 adults demonstrate gross morphological defects. Generally, both #54 and #84 ovarioles consistently show intense DNA bodies in egg chambers staged 6 onwards. This observation suggests that a checkpoint mechanism might exist around this stage that promotes the degeneration of defective chambers by apoptosis. A high proportion of homozygous *DmCBP20*^{MFS} #84 ovarioles (74%) are mutant in phenotype and adults do not lay eggs. Although initial observations suggest the egg chambers might have signalling defects, further investigations show this is not the case. In comparison to #84, relatively few (23%) homozygous *DmCBP20*^{MFS} #54 ovarioles exhibit a mutant phenotype. As a result adult females do lay eggs, some of which are strikingly flaccid in appearance. This observation suggests that mature eggs might be defective in the chorion egg shell or major yolk proteins. However, further examinations demonstrate that this is not the case.

Although *DmCBP20*^{MFS} #54 and #84 ovaries have lower levels of DmCBP20, complementation testing show both mutations can rescue each other. This suggests that the molecular defect(s) incurred by the lower amounts of DmCBP20 during gametogenesis are different in both *DmCBP20*^{MFS} #54 and #84 lines. Northern analysis does not detect a defect in the splicing of RpS3 ribosomal protein RNAs in homozygous *DmCBP20*^{MFS} #54 or #84 ovaries. RpS3 contains one intron and therefore is subject to CBC and cap-dependent intron removal. As a result, it is reasoned that a more sensitive assay such as RNase protection would be better suited for investigating aberrant splicing or RNA 3'-end formation. Alternatively, microarray analysis could be used to detect a splicing defect in general or a specific

subset of RNAs.

4.3.6 Morphological analysis of *DmCBP20^{MFS}* and *DmCBP20^{MS}* male steriles

In comparison to ovaries from *DmCBP20^{MFS}* adults, testes from homozygous *DmCBP20^{MFS}* and *DmCBP20^{MS}* males exhibit no obvious abnormalities compared to wild-type controls. All the stages of early spermatogenesis are present from germ stem cells to elongating sperm. Furthermore, the morphology of both pre-meiotic spermatocytes and post-meiotic spermatids are normal. However, mature sperm are not observed in *DmCBP20^{MS}* lines #1 and #15. Whilst the number of individualised sperm in *DmCBP20^{MFS}* #54 and #84 do not appear to be the same as wild-type. Taken together, these observations suggests that defects leading to *DmCBP20^{MFS}* and *DmCBP20^{MS}* male sterility probably arise during the final stages of spermatid differentiation.

During fruit-fly spermatogenesis, RNAP II transcription rarely occurs after meiosis. mRNAs are transcribed early and then sequestered in a stable form. Translation is regulated and therefore delayed until proteins are required for function in post-meiotic cells (Standart and Jackson, 1994; Curtis *et al.*, 1995). Current data has characterised roles for CBC in nuclear RNAP II transcript processing events such as splicing and RNA 3'-end formation that are coupled to transcription (Lewis and Izaurralde, 1997; Hirose and Manley, 2000). This suggests that if the mechanism of *DmCBP20* action is the same in somatic and germ cells, *DmCBP20* would only be required during active transcription in pre-meiotic cells. However, in all *DmCBP20^{MFS}* and *DmCBP20^{MS}* males steriles examined, the morphology of pre-meiotic spermatocytes and post-meiotic spermatids are wild-type. Instead, observations suggest the mature sperm are either absent or reduced in number compared to wild-type controls. As a result, *DmCBP20^{MFS}* and *DmCBP20^{MS}* mutations appear to induce defects in the final stages of spermatid differentiation where RNAP II genes are transcriptionally silent. This in turn implicates novel

role(s) for *DmCBP20* outwith nuclear RNA processing, in post-meiotic cells during the later stages of spermatogenesis.

Since transcription is absent in post-meiotic cells, regulation of mRNA translation is an effective mechanism for controlling protein expression. Surprisingly little is known about the strategies and mechanisms involved in translation regulation during spermatogenesis. However, many nuclear RNA-binding proteins have been identified that mediate post-transcriptional regulation and are consequently essential for spermatogenesis (Venables and Eperon, 1999). These proteins are defined by a common RNA-binding domain termed the RNA recognition motif (RRM) or RNP motif (Karsch-Mizrachi and Haynes, 1993; Elliot *et al.*, 1997; Haynes *et al.*, 1997; McGuffin *et al.*, 1998) Interestingly, *CBP20* also has two RNP motifs that comprise a RNA-binding domain at the amino-terminus. In addition, a second CBP20 homologue has recently been identified in humans termed *HsCBP20 Related* or *HsCBP20R*. Current data suggests the expression of *HsCBP20* is predominantly restricted to adult testes (E. Thompson, unpublished data). As a result, these data point to role(s) for *CBP20* during spermatogenesis in both humans and *Drosophila.* Moreover, the function of *DmCBP20* appears to extend beyond the characterised role in nuclear processing of RNAP II transcripts. Instead, it is plausible to consider that *DmCBP20* might also participate in mechanisms that regulate RNAP II translation during spermatogenesis.

4.4 CONCLUSIONS

An existing P-insertion in *DmCBP20* was excised to produce additional *DmCBP20* mutant alleles. This generated a series of *DmCBP20* male and female sterile, male sterile and homozygous lethal lines. *DmCBP20* is an essential gene in *Drosophila melanogaster*. As expected from a key RNA processing factor, DmCBP20 functions in somatic tissues. In addition, *DmCBP20* plays an important role during spermatogenesis and oogenesis. *DmCBP20* appears to make a greater contribution to male fertility where it might have novel roles in RNAP II translation regulation.

CHAPTER 5

Characterisation of *DmCBP80R*

5.1 INTRODUCTION

After the completion of the *Drosophila melanogaster* genome sequencing project, the BDGP database was searched for sequences homologous to *DmCBP20* and *DmCBP80* (Adams *et al.*, 2000). This determined that in *Drosophila melanogaster*, *CBP20* is an unique gene. However, in fruit-flies, a second *DmCBP80* homologue named *CG7907* or *DmCBP80 Related* (*DmCBP80R*) was also identified. Although single-copy genes for both cap-binding proteins are present in lower eukaryotes, humans have two *HsCBP20* homologues and only one *HsCBP80* homologue. Intriguingly, initial characterisation of *HsCBP20 Related* (*HsCBP20R*) has shown it is predominantly expressed in adult testes (E. Thompson, unpublished data). This raises the question of whether functional parallels can be drawn between human *CBP20R* and *Drosophila CBP80R* homologues. As a result, Chapter 5 presents an initial characterisation of *DmCBP80R* by analysing the sequence and RNA data available.

5.2 RESULTS

5.2.1 Identification of a second CBP80 homologue in *Drosophila melanogaster*

In order to determine whether *DmCBP80* is an unique gene, the predicted amino acid sequence of *DmCBP80* was used in a TBLASTN search of the genomic sequences in the BDGP database. This identified a second sequence homologous to DmCBP80 named *CG7907* or *DmCBP80 Related* (*DmCBP80R*). The genome project has mapped *DmCBP80R* to the right arm of chromosome 3 on polytene segment 93E1-2 (inferred) and the predicted gene structure of *DmCBP80R* is presented in Figure 5.1. In comparison to *DmCBP80*, the gene structure of *DmCBP80R* is relatively simple. *DmCBP80R* consists of three exons, which span 2, 420 bp of genomic DNA with two short introns that are 88 bp and 90 bp in length respectively. Searches of FlyBase revealed no known mutant alleles have been described for *DmCBP80R* to date. Interestingly, no ESTs have been found to correspond to *DmCBP80R* suggesting that it might be expressed at low levels, show tissue-restricted expression or alternatively, be a pseudogene.

The conceptual protein product of *DmCBP80R* is 747 aa in length compared with 800 aa of DmCBP80. Overall, both *Drosophila* proteins show 31% identity and 50% homology to each other throughout their length. As can be observed from Figure 5.2, A, high sequence identity is observed at the amino-terminals of DmCBP80 and DmCBP80R proteins. This includes the putative bipartite NLS and recently described MIF4G domain that implicates CBP80 in NMD (Ponting, 2000; Ishigaki *et al.*, 2001). In contrast, greater sequence divergence and length heterogeneity is observed within the carboxyl-terminal regions of DmCBP80 and DmCBP80R.

Figure 5.1 Gene Structure and Chromosomal Location of *DmCBP80R*

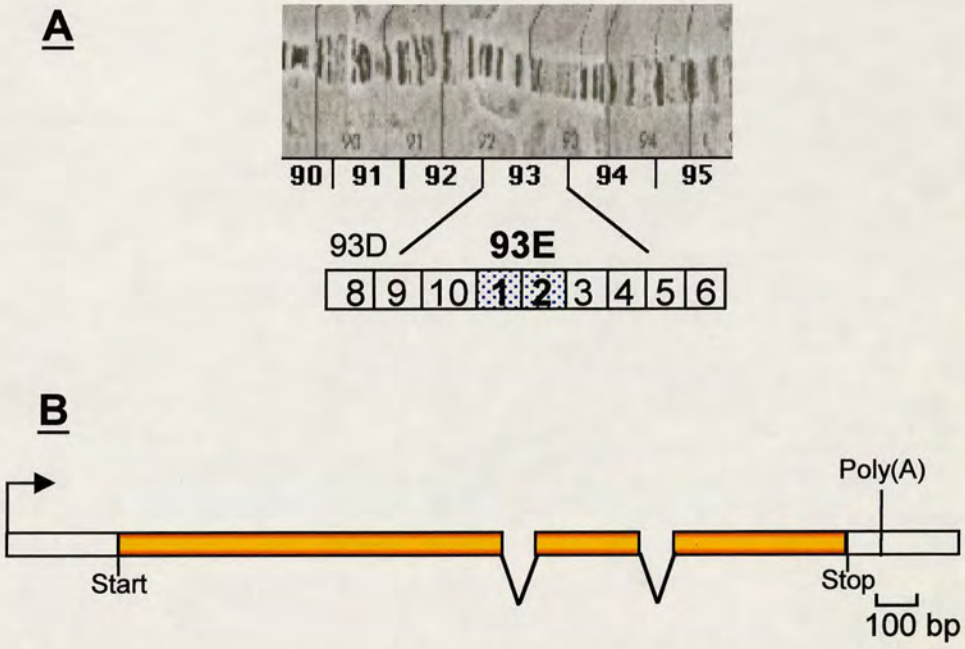


Figure 5.1 Gene structure and chromosomal location of *DmCBP80R*. **A.** Berkeley genome project has mapped *DmCBP80R* (*CG7907*) to right arm of chromosome 3 on polytene segment 93E1-2 (inferred). **B.** Schematic representation of *DmCBP80R* gene structure as predicted by Genescan. *DmCBP80R* has three exons depicted by rectangles and two introns represented by black lines, which span a total of 2, 420 bp genomic DNA. Start of transcription unit is depicted by black arrow along with poly(A) sites. Translation unit is outlined by start and stop annotations.

Figure 5.2 CBP80 Homologues

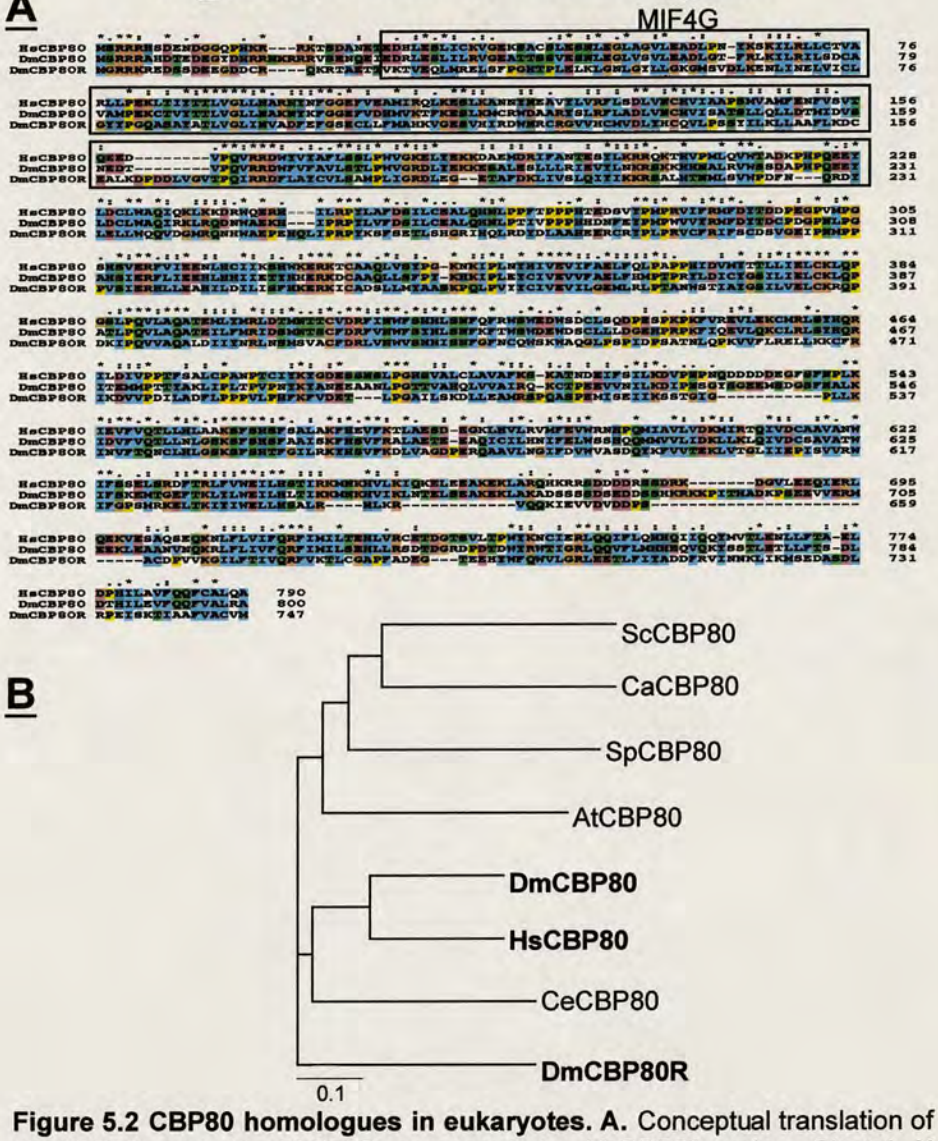


Figure 5.2 CBP80 homologues in eukaryotes. A. Conceptual translation of *DmCBP80* and *DmCBP80R* compared to *HsCBP80* using Clustal X (Thompson *et al.*, 1998). The conserved MIF4G domain is presented boxed. Identical and related amino acids are illustrated by (*) and (·) respectively. (:) indicates similar amino acids and discontinuities in the alignments are represented by dashes. Blue represents hydrophobic/ aromatic amino acids. Orange depicts basic and glycine residues. Purple shows acidic amino acids. Green and yellow illustrates hydrophilic and proline residues respectively. **B.** Phylogenetic tree showing the high divergence of DmCBP80R from DmCBP80 and HsCBP80. The sequences used are: *Saccharomyces cerevisiae* ScCBP80 (P34160), *Candida albicans* CaCBP80 (BAA76378), *Arabidopsis thaliana* AtCBP80 (Q9LKN6), *Caenorhabditis elegans* CeCBP80 (T15197), *Drosophila melanogaster* DmCBP80 (Q9U980), *Homo sapiens* HsCBP80 (Q09161), *Drosophila melanogaster* Related DmCBP80R (Q9VDA4) and *Schizosaccharomyces pombe* SpCBP80 (T39057).

The high level of divergence between the *Drosophila* homologues is more apparent by comparing the phylogenetic relationships of CBP80 from a wide variety of eukaryotes (Figure 5.2, B). The evolutionary tree shows that the duplication of *Drosophila CBP80* was an ancient event in the arthropod lineage prior to the divergence of the different metazoan phyla. As a result, DmCBP80 is more closely related to HsCBP80 whilst DmCBP80R is more distantly related. The high level of divergence between the two *Drosophila* homologues suggests that DmCBP80R might have acquired new functions during evolution. However, in the absence of any functional data this is purely hypothetical.

5.2.2 DmCBP80R mRNA is expressed in 3rd instar larvae and adult males

To address whether *DmCBP80R* is a functional gene, northern analysis was used to assay if DmCBP80R mRNA is expressed in the various *Drosophila* developmental stages. A fragment corresponding to 1, 495 bp amino-terminus of *DmCBP80R* was PCR amplified from adult fly genomic DNA using oligos S299 + S300, gel purified and sequenced (Figure 5.3, A and data not shown). This fragment was subsequently used to re-probe the developmental northern originally used to characterise the expression of DmCBP20 and DmCBP80 mRNAs described in Chapter 3. Total RNA was extracted from staged embryo collections, three larval instars, adult males and females and isolated ovaries. The amounts of total RNA loaded were then normalised to the level of ribosomal RNAs (Figure 5.3, B lowest panel). As an additional loading control, the filter was re-probed with full-length cDNA from the ribosomal protein *DmRp49*, which is ubiquitously expressed throughout fruit-fly development (third panel). No mRNA corresponding to the predicted size of DmCBP80R was observed in any of the staged embryonic collections, three larval instars, adult females or isolated ovaries (upper panel and data not shown). However, a weak signal was reproducibly observed at the predicted mRNA size of approximately 2.5 Kb, in preparations of adult males only (upper panel).

Figure 5.3 Expression of DmCBP80R mRNA

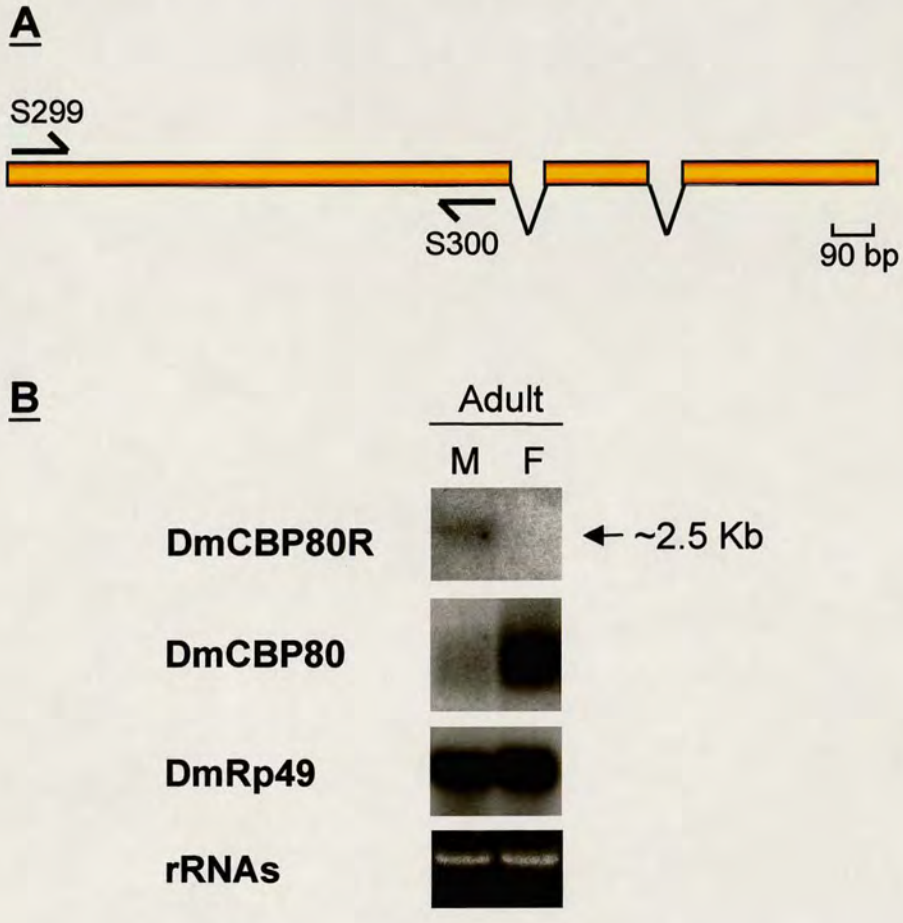


Figure 5.3 Expression of DmCBP80R mRNA. A. Schematic representation of *DmCBP80R* gene structure with positions of the primer pair S299 + S300 used to amplify 1, 495 bp amino-terminal cDNA probe. **B.** Nitrocellulose filters used to characterise DmCBP20 and DmCBP80 transcripts were stripped and re-hybridised with the amino-terminal *DmCBP80R* probe and exposed. A single transcript was observed for DmCBP80R at approximately 2.5 Kb in adult males (M) only. No mRNA was detected in adult females (F). 10 μg of total RNA per lane was loaded as judged by ethidium bromide staining of ribosomal RNAs (rRNAs). As an additional loading control, the filter was re-probed with full-length ribosomal protein *DmRp49* cDNA.

This contrasts to DmCBP80, where steady state levels of mRNA are apparent in both adult males and females (second panel). More notably, adult females show strikingly higher steady state levels of DmCBP80 than adult males. RNA analysis of DmCBP80 from Chapter 3 has already shown that the enrichment of DmCBP80 mRNA in adult females is due to ovarian expression. This suggests that unlike DmCBP80, DmCBP80R transcripts are not maternally derived.

Taking into account the northern data, it was reasoned that DmCBP80R expression might be restricted to specific cell types or tissues in fruit-fly development. In order to examine this more directly, RNA *in situ* hybridisation was used to determine whether DmCBP80R RNA is expressed ubiquitously or in certain cells/ tissues during oogenesis, embryogenesis and selected third instar larval imaginal discs. The amino-terminal *DmCBP80R* fragment used for the northern analysis was also used to generate DIG-labelled anti-sense RNA that specifically detects mRNA and sense RNA, which serves as a negative control. RNA probes were subsequently hybridised to fixed ovarioles, staged embryos and third instar eye-antenna, haltere, brain, leg and wing imaginal discs. However, no specific RNA localisation was observed using DIG-labelled anti-sense DmCBP80R suggesting that DmCBP80R mRNA was not expressed during these stages (data not shown). Again this contrasts with RNA *in situ* analysis of DmCBP80 described in Chapter 3, where DmCBP80 mRNA is found to be ubiquitously expressed in all tissues throughout oogenesis, embryogenesis and third instar imaginal stages.

To investigate whether the weak signal observed in adult males from the northern blotting was due to DmCBP80R mRNA expression, a more sensitive assay was employed. RT-PCR was used to firstly examine whether unspliced or spliced DmCBP80R RNA was expressed throughout the various *Drosophila* developmental stages and secondly, to determine if the splicing of DmCBP80R occurred at the predicted splice sites.

Figure 5.4, A illustrates schematically the primer pairs designed to amplify across intron 1 (oligos S471 + S472) and intron 2 (oligos S473 + S474) of *DmCBP80R*. The correct splicing of intron 1 gives a product of 521 bp and the unspliced form is 609 bp. Whilst the spliced form of intron 2 is 459 bp and the pre-mRNA is 549 bp. Total RNA was extracted from staged embryo collections, three larval instars, adult flies and analysed by RT-PCR. RNase treated RNA preparations were included as a negative control to ensure that the RT products were derived from RNA (data not shown).

Complementing the data from the northern and RNA *in situ* analysis, no RT bands corresponding to spliced CBP80R were observed in RNA isolated from staged embryos, as well as first or second instar larvae (Figure 5.4, Panel B [i] and [ii] lanes 1-7). It is important to note that the RT-PCR gels presented here had annealing temperatures optimised for the PCR of spliced cDNA. Therefore, when the annealing temperature was varied, bands corresponding to unspliced RNA were detected in the embryonic and first and second larval stages (for example [ii] lane 1, 3, 4 and 7 and data not shown). In addition, although certain lanes have high levels of non-specific background, repeat performances of these experiments showed no underlying cDNA bands were obscured (data not shown). In contrast to embryonic, first and second larval stages, RNA samples isolated from adult flies showed that DmCBP80R splicing was exclusive to adult males. Adult females showed no detectable splicing of either intron with only low levels of unspliced pre-mRNA being observed (compare lanes 10-11 from [i] and [ii]). Interestingly, splicing of DmCBP80R introns 1 and 2 was also observed in male and female third instar larvae ([i] and [ii] lanes 8 and 9). Generally, RNAs of proteins predominantly expressed in male testes are only detected in male third instar larvae (Haynes *et al.,* 1997). One possibility is that DmCBP80R mRNAs were detected in both male and female third instar larvae due to mis-sexing. However, this is unlikely since individuals experienced in larval sexing (Petra zur Lage and Nicola Wrobel) confirmed separation of male and female larvae.

Figure 5.4 DmCBP80R mRNA is Expressed in 3rd Instar Larvae and Adult Males

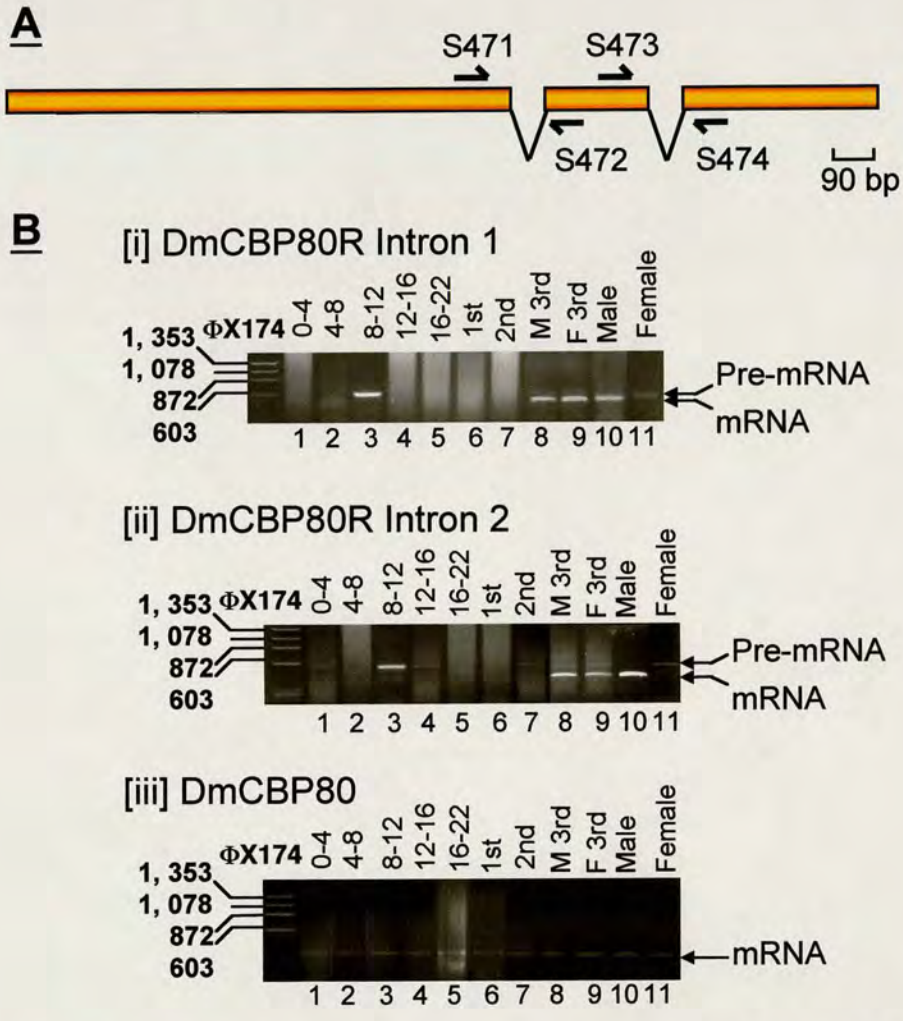


Figure 5.4 DmCBP80R mRNA is expressed in 3rd instar larvae and adult males. RT-PCR analysis of DmCBP80R splicing during *Drosophila* development. **A.** Schematic diagram illustrating predicted *DmCBP80R* gene structure with positions of primer pairs used for RT-PCR. Oligos S471 + S472 and S473 + S474 were used to perform RT-PCR across introns 1 and 2 respectively. **B.** Total RNA was isolated from embryos staged 0-4 hours, 4-8 hours, 8-12 hours, 12-16 hours and 16-22 hours. Along with first, second and third instar male and female larvae, adult males and females. RT-PCR was performed across *DmCBP80R* intron 1 and intron 2. **[i]** Splicing of intron 1 gives a product of 521 bp and unspliced product is 609 bp. **[ii]** Spliced intron 2 is 459 bp and unspliced form is 549 bp. **[iii]** RT-PCR across intron 2 of *DmCBP80* was included as a positive control, spliced intron 2 is 423 bp and unspliced form is 484 bp.

The most likely explanation is that DmCBP80R mRNA expression is tightly regulated. Transcripts appear to be expressed early in pole cell development but are switched off later in females during pupation.

To ensure the RNA preparations used in these experiments were intact and not degraded, a positive control RT-PCR that amplified across intron 2 of *DmCBP80* was also included. Splicing of DmCBP80 intron 2 gives a product of 423 bp and the unspliced form is 484 bp. As expected, DmCBP80 mRNA was spliced throughout embryogenesis ([iii] lanes 1-5), three larval instars ([iii] lanes 6-9) and adult flies ([iii] lanes 10-11). It is important to note that comparisons of the RT gels in Figure 5.4, B show that the RT-PCR was not quantitative. However, it is apparent that RT bands corresponding to DmCBP80 mRNA are detected in adult males and females, which suggests the RNA preparations used were intact and not degraded. This strongly supports the presence of spliced DmCBP80R in males and unspliced CBP80R in females (compare lanes 10-11 from [i]-[iii]). In order to substantiate the cDNA products from the RT-PCR corresponded to spliced and unspliced RNAs, the RT bands were gel purified and sequenced (data not shown). This confirmed that in third instar larvae and adult males, both introns of DmCBP80R were spliced at the correct splice sites and the slower migrating species corresponded to the intron-containing fragment.

5.2.3 DmCBP80R is predominantly expressed in the abdomen of adult males

The RT-PCR experiments have shown that *DmCBP80R* is a *bona fide* gene that is restrictively expressed in *Drosophila* development. DmCBP80R mRNA is first detected in male and female third instar larvae and mRNA expression is subsequently restricted to adult males only. The next question to address was in which tissues is DmCBP80R mRNA expressed. RNA *in situ* hybridisation experiments performed on selected third instar imaginal discs showed no specific mRNA expression in the eye-antenna, haltere, brain, leg or wing discs (data not shown).

Figure 5.5 DmCBP80R mRNA is Predominantly Expressed in the Male Abdomen

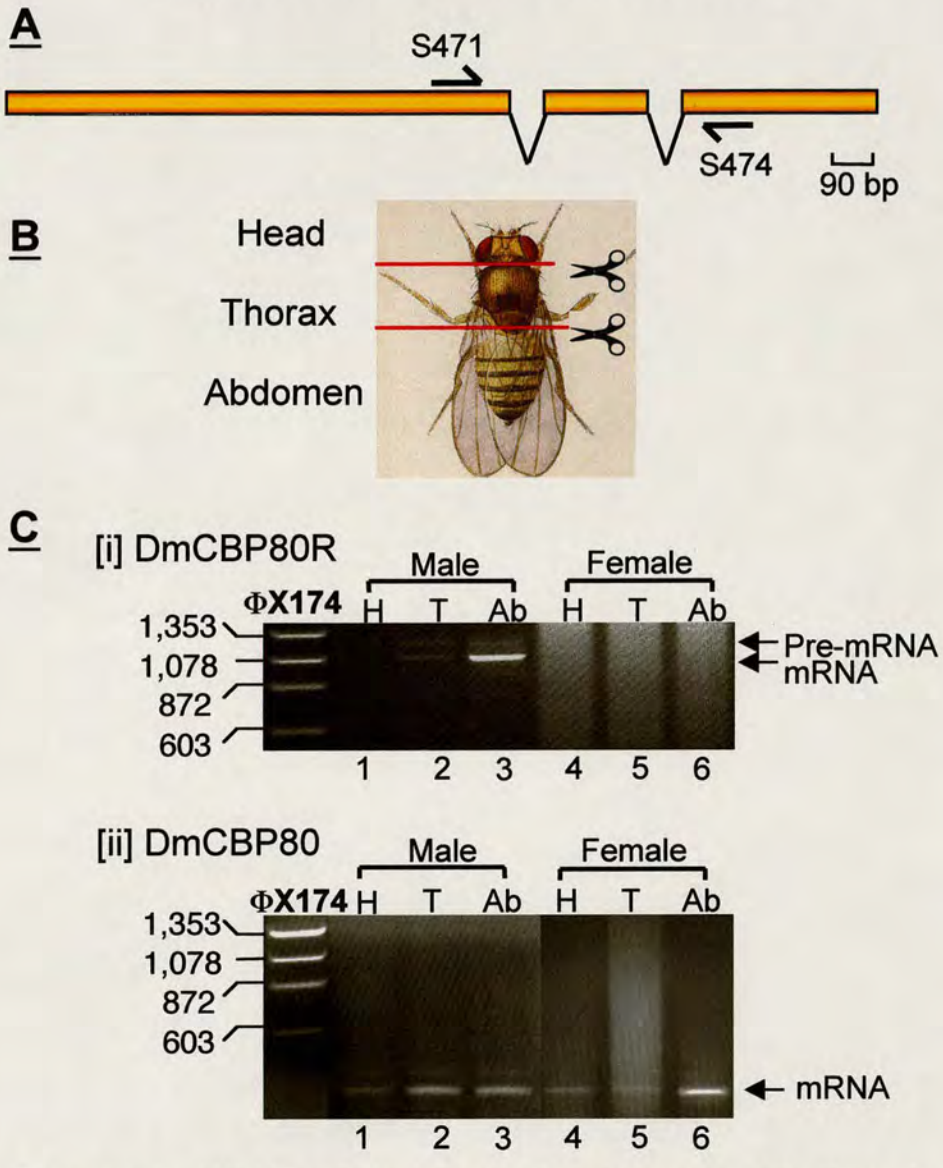


Figure 5.5 DmCBP80R mRNA is predominantly expressed in the male abdomen. RT-PCR analysis of DmCBP80R splicing in adult tissues. **A.** Schematic representation of predicted *DmCBP80R* gene structure with position of primer pair used for RT-PCR. Oligo pair S471 + S474 was used to perform RT-PCR across introns 1 and 2. **B**. Total RNA was extracted from 50 adult males and females cut into three body parts, shown as head (H), thorax (T) and abdomen (Ab). **C. [i]** RT-PCR across introns 1 and 2 gives a spliced product of 1, 021 bp and unspliced form is 1, 196 bp. **[ii]** RT-PCR across intron 2 of *DmCBP80* was included as a positive control, spliced intron 2 is 423 bp and unspliced product is 484 bp.

This implies that DmCBP80R mRNA must be restricted to other third instar larval tissues such as the fat body, salivary glands or gonads. Since adult males are easier to manipulate than larvae, it was decided that DmCBP80R mRNA expression would be further examined in adult male tissues.

As before, RT-PCR was used to determine whether unspliced or spliced DmCBP80R was present in various adult body parts. A total of 50 adult females and adult males were cut into three body parts illustrated in Figure 5.5, B. Total RNA was extracted and used in RT-PCR in combination with the oligo pair S471 + S474, which amplifies across both introns 1 and 2 (Panel C). Correct splicing of DmCBP80R introns 1 and 2 gives a 1, 021 bp product whilst the pre-mRNA is 1, 196 bp. In adult females, no detectable spliced RNA was detected in either the head, thorax or abdomen tissues (Panel C, [i] lanes 4-6). In contrast, spliced DmCBP80R was predominantly observed in the male abdomen with very low detectable levels present in the male thorax ([i] lanes 2 and 3). Although not clearly visible in Figure 5.5 B, low levels of unspliced DmCBP80R RNA were detected in the head, thorax and abdomen tissues ([i] lanes 1-3). Again the positive control, which used primers designed against *DmCBP80* intron 2, demonstrated that RNA preparations used in these experiments were intact and not degraded. As expected, DmCBP80 mRNA is expressed throughout the head, thorax and abdomen tissues of adult males and females ([ii] lanes 1-6). The negative controls using total RNA treated with RNase showed that the RT products were derived from RNA (data not shown). Subsequent preliminary experiments have performed RT-PCR on adult male abdomens with testes removed and isolated testes. This determined that in the male abdomen CBP80R mRNA is predominantly expressed in adult testes (data not shown).

5.2.4 Production of polyclonal antibodies against DmCBP80R

In order to proceed with immunofluorescence and biochemical investigations, it was necessary to raise antibodies against DmCBP80R. Since DmCBP80R was not readily available in a vector to express protein, it was decided to raise peptide antibodies against DmCBP80R. Comparison of DmCBP80R and DmCBP80 protein sequences demonstrated greater sequence divergence and length heterogeneity within the carboxyl-regions. As a result, two rabbits were immunised with a 15 amino acid synthetic peptide (MSEDASDLRPEISKT) designed against the carboxyl-terminus of DmCBP80R. The rabbits were subjected to three bleeds before they were killed and purified polyclonal sera extracted (Abcam). Western blots were prepared that contained total protein extracted from 50 adult male and female abdomens, which effectively served as positive and negative controls. The conceptual translation of *DmCBP80R* gives a protein product of 85 KDa. However, western analysis of the bleeds and purified antibody showed no bands were recognised corresponding to the predicted size of DmCBP80R in male abdomens (data not shown).

5.3 DISCUSSION

5.3.1 *Drosophila melanogaster* has two CBP80 homologues

Searches of the completed *Drosophila* genome indicate *DmCBP20* is a single-copy gene and *DmCBP80* has two homologues. Results from Chapter 3 show *DmCBP80* is mapped to 4C7-8 on the X-chromosome. In contrast, *DmCBP80R* is located to the right arm of chromosome 3 on polytene segment 93E1-2. Comparison of the predicted protein sequences show DmCBP80R is highly divergent from DmCBP80. However, DmCBP80R does share many of the strongly conserved residues found in CBP80 homologues from higher eukaryotes. Indeed, the bipartite NLS and recently described MIF4G domain are strongly conserved in both *Drosophila* proteins (Izaurralde *et al.*, 1995; Ponting, 2000). The functional role of the MIF4G domain is currently not known, however, it is thought to implicate CBP80 in NMD (Mendell *et al.*, 2000; Ishigaki *et al.*, 2001). In comparison to the amino-region, the carboxyl-terminus of DmCBP80R is highly divergent to CBP80. This is particularly apparent in the last 150 amino acids of DmCBP80R, which show very low sequence identity to DmCBP80. In addition, a majority of the size difference between DmCBP80 and DmCBP80R proteins results from deletions of amino acids within this region.

Recently, the X-ray crystal structure of HsCBP80 has been determined and shows the carboxyl-terminal 150 amino acids fold into two alpha helices that form a very long coiled-coil. This coiled-coil protrudes from CBP80 and is thought to provide a distinct surface for protein-protein interactions between CBP80 and RNA processing/ export factors (Mazza *et al.*, 2001). Although the coiled-coil is strongly conserved in DmCBP80, most of the corresponding amino acids in DmCBP80R are deleted or share very low sequence identity to HsCBP80. This suggests that in comparison to DmCBP80 and HsCBP80, the carboxyl-region of DmCBP80R might mediate interactions with different cellular proteins. Taken together, these data suggest the high sequence divergence between the two *Drosophila* homologues might result from DmCBP80R acquiring new functions during evolution.

5.3.2 CBP80R is predominantly expressed in 3rd instar larvae and adult testes

Results from the northern and RT-RCR analysis show DmCBP80R is restrictively expressed throughout *Drosophila* development. Spliced DmCBP80R is initially detected in third instar larvae and subsequently becomes restricted to adult males only. In adult males, DmCBP80R mRNA is restricted to the abdomen and preliminary results show it appears to be predominantly expressed in the testes. Consequently, the RNA analysis strongly suggests that DmCBP80R plays a role in spermatogenesis. In contrast to DmCBP80R, DmCBP80 is ubiquitously expressed in development and appears to be maternally derived (Chapter 3). Although roles for DmCBP80 have been defined in somatic cells, it is unclear whether DmCBP80 also functions during spermatogenesis. Moreover, it remains to be determined whether DmCBP80 and DmCBP80R are functionally redundant. Future RT-PCR experiments will determine whether DmCBP80 and DmCBP80R are co-expressed or mutually excluded in testes. In addition, RNA *in situ* analysis can be used to address if DmCBP80R is expressed ubiquitously or restricted to specific cell types in the testes.

Chapter 4 has already described how the mobilisation of a P-element in *DmCBP20* generates a very high number of homozygous viable male sterile lines. Initial characterisation of these lines suggests the sterility might arise from defects in spermatid differentiation during the final stages of spermatogenesis. It is well established that CBC has a characterised role in the RNA processing events of pre-mRNA splicing and RNA 3'-end formation that are coupled with transcription (Lewis and Izaurralde, 1997; Hirose and Manley, 2000). However, in spermatogenesis, RNAP II transcription rarely occurs after meiosis. Instead, mRNAs required for the later stages of spermatogenesis are transcribed early and sequestered in a stable form ready for translation. Therefore, defects in CBC function would be expected to affect the early pre-meiotic stages of spermatogenesis. Taken together, the *DmCBP20* male sterile mutations suggest *DmCBP20* might have novel functions outside nuclear RNA processing and this might occur with *DmCBP80R*.

During spermatogenesis the translation of certain mRNAs are delayed until the proteins are required in post-meiotic cells (Venables and Eperon, 1999). However, the strategies and mechanisms for regulating translation remain unclear. Many RNA-binding proteins, like *DmCBP20*, appear to be important for spermatogenesis but little is understood about their specific functions (Venables and Eperon, 1999). One possibility is that DmCBP20 could bind DmCBP80R to form a male-specific CBC that affects the translational status of mRNAs transcribed during early spermatogenesis. The male-specific CBC could mask mRNAs from the translation machinery until later stages of spermatogenesis when the proteins are required for function. In order to support this hypothesis, additional work is necessary to initially determine whether DmCBP80R can bind DmCBP20 to form a male-specific CBC.

5.4 CONCLUSIONS

A second divergent CBP80 homologue has been identified in *Drosophila melanogaster*. DmCBP80R mRNA is detected in third instar larvae and adult males where it is predominantly expressed in testes.

CHAPTER 6

Expression of wild-type and mutant CBC protein using the GAL4-UAS system

6.1.1 INTRODUCTION

The available evidence shows that CBC is a key player in the metabolism of RNAP II transcripts. CBC is found to mediate the positive effects of the cap in at least three major aspects of nuclear RNA processing: pre-mRNA splicing, RNA 3'-end formation and U snRNA export (Lewis and Izaurralde, 1997). Despite this, little is known about the proteins CBC interacts with or the molecular mechanisms involved. Genetic studies in *S. cerevisiae* have used synthetic lethal screens and yeast two-hybrid systems to identify CBC-interacting genes (Gamberi *et al.*, 1997; Fortes *et al.*, 1999b). However, these experiments have had limited success and current data strongly suggests that genetic approaches in *Drosophila melanogaster* might be more useful in identifying CBC-interacting genes.

Characterisation of CBC's function in pre-mRNA splicing shows that it promotes the association of U1 snRNP with the cap-proximal splice site. Furthermore, the role of CBC in cap-dependent splicing is found to be conserved from yeast to humans (Colot *et al.*, 1996; Lewis *et al.*, 1996). Although the molecular mechanism by which CBC functions during splicing has not been elucidated, recent genetic studies in *S. cerevisiae* show that CBC appears to mediate the positive effect of the cap through direct interactions with U1 snRNP (Fortes *et al.*, 1999a and 1999b). This contrasts with the mammalian system, where there is no evidence of a direct interaction between CBC and U1 snRNP. Instead, the available data suggests that in higher eukaryotes, CBC function is facilitated through yet unidentified adaptor proteins (Lewis *et al.*, 1996). As a result, the mechanism of CBC action appears to be

different between higher and lower eukaryotes. Given that *Drosophila* is closer to humans than yeast, genetic studies in fruit-flies might be more successful in identifying candidate adaptor molecules that facilitate the function of CBC and the cap in RNAP II transcript metabolism.

Further support for using fruit-flies as a tool for identifying CBC-interacting genes comes from the characterisation of DmCBC. Although mutations in the constituents of *S. cerevisiae* CBC only impair vegetative growth, in *Drosophila*, at least *CBP20* is an essential gene (Chapter 4). This suggests that the requirement for CBC in RNAP II transcript metabolism is greater in higher eukaryotes. In addition, both human and *Drosophila* cap-binding proteins share a very high degree of conservation (Chapter 3). This strongly supports that CBC function might also be conserved between these species. Finally, the recent sequencing of the *Drosophila* genome has revealed extensive conservation between the components of the spliceosome in vertebrate and *Drosophila* systems (Adams *et al.*, 2000; Mount and Salz, 2000). This suggests the complexity of mechanisms such as splicing, which contribute to RNAP II transcript processing are similar in both *Drosophila* and humans. Therefore, the identification of CBC-interacting genes in fruit-flies might lead to a greater understanding of the biology of CBC. Taken together, these data provide a strong case for using genetic approaches in *Drosophila melanogaster* to elucidate the mechanisms by which CBC and the cap facilitate multiple aspects of RNAP II transcript metabolism.

6.1.2 GAL4-UAS system

In order to identify interacting genes in *Drosophila*, one genetic approach is to use flies with a sensitized genetic background in genetic modifier screens. These flies can be generated by creating mutants in the endogenous gene of interest, or alternatively, using the GAL4-UAS system to produce transgenic flies with dominant phenotypes. The GAL4-UAS system is a powerful tool that allows the biology of genes to be studied *in vivo* (Brand and Perrimon, 1993). Classically, GAL4-UAS has been used to investigate the effects of over-expressing wild-type or mutated proteins that cause dominant phenotypes (Kraus and Lis, 1994; Pan and Rubin, 1997; Elefant and Palter, 1999). Targeted gene expression studies have been useful in examining the roles of regulatory proteins such as wingless and EGFR in specifying cell fates (Freeman, 1996; Johnston and Schubiger, 1996). However, GAL4-UAS generated dominant phenotypes have also been successfully used in genetic modifier screens to identify interacting gene products or to establish epistatic relationships (Rorth *et al.*, 1998; Helms *et al.*, 1999; Wan *et al.*, 2000).

GAL4-UAS gene expression is a binary system where the target gene is separated from its transcriptional activator in the form of two distinct transgenic strains, the GAL4 driver and the UAS line (Figure 6.1). The GAL4 driver contains the *S. cerevisiae GAL4* gene that has been randomly inserted into the fly genome and is under the control of a tissue-specific promoter. The UAS-transgenic line is generated by cloning the gene of interest into pUAST, the vector that directs Gal4 dependent transcription. In pUAST, the cloned gene is situated downstream of five Gal4 upstream activation sites (UAS) and upstream of a SV40 transcriptional terminator and polyadenylation site. Consequently, gene transcription is only activated within cells that express Gal4 protein. P-element mediated transformation is used to integrate the UAS-gene into the fly genome and the resultant transgenic lines contain the UAS-gene that is transcriptionally silent.

Figure 6.1 GAL4-UAS System

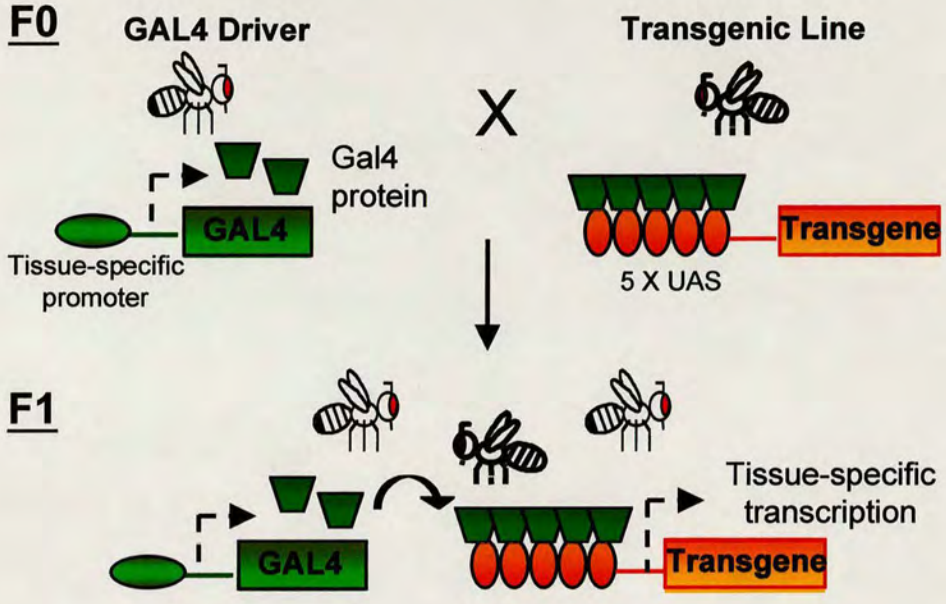


Figure 6.1 The GAL4-UAS system. Targeted gene expression using the GAL4-UAS system. **F0.** GAL4 driver is crossed to the UAS-transgenic line. GAL4 driver contains the yeast *GAL4* gene that has been randomly inserted into the fly genome and is under the control of a tissue-specific promoter. UAS-transgenic line contains the gene of interest, which is cloned behind five Gal4 upstream activating sites (UAS). In the absence of Gal4, the target gene is transcriptionally silent. **F1.** Resultant progeny can activate UAS-transgene transcription in the tissues where Gal4 protein is expressed.

UAS-gene transcription is only activated when a GAL4 driver is crossed with a UAS-transgenic line. Subsequently, the F1 progeny can only activate UAS-gene transcription in specific tissues allowing the phenotypic consequences of targeted expression to be studied.

Complementary to the P-element induced mutagenesis described in Chapter 4, the GAL4-UAS system was used to over-express wild-type and mutant CBP20 and CBP80 proteins in non-essential adult tissues. These transgenic flies were screened for dominant phenotypes, which were subsequently used in genetic modifier screens to identify CBC-interacting genes. This Chapter describes the identification and characterisation of dominant phenotypes in both *CBP20* and *CBP80*.

6.2 RESULTS

6.2.1 GAL4-UAS Crosses

In order to generate dominant adult phenotypes in CBC, the GAL4-UAS system was used to employ two different strategies. Firstly, the system was used to over-express wild-type *Drosophila* and human CBP20 and CBP80 proteins. Secondly, one of the cap-binding proteins in the CBC heterodimer was mutated and over-expressed with the aim of generating putative dominant negative forms of CBC. Full-length *Drosophila UAS-CBP20* and *UAS-CBP80* were cloned into pUAST along with the human transgenes. Despite being highly homologous to DmCBC, the human cap-binding proteins contain notable amino acid changes at conserved regions that might represent different protein-protein interactions. It was therefore reasoned that over-expression of either *HsCBP20* or *HsCBP80* might generate a hybrid CBC that would also produce dominant phenotypes in adult tissues.

The strategy employed in creating a dominant negative form of CBC involved making C-terminal truncations in *DmCBP20* and *DmCBP80* by PCR and cloning these constructs into pUAST. The predicted amino acid sequence of DmCBP80 contains a conserved putative bipartite NLS at the amino-terminus. The NLS has been demonstrated to be functional in humans since mutation of this sequence

inhibits the nuclear import of HsCBP80 (Mattaj, 1993; Izaurralde *et al.*, 1994). As a result, four C-terminal truncations were made outside this domain, at regions believed to be conserved boundaries from amino acid sequence alignments (Figure 6.2, A). Interestingly, the recently described MIF4G domain that represents significant sequence similarities between the middle portion of eIF4G, domains in NMD2/Upf2 and CBP80 homologues was also left intact in all four truncations (Ponting, 2000; Mazza *et al.*, 2001). As mentioned earlier, the MIF4G motif implicates CBC in NMD. Although the functional significance of this domain is currently unknown, it is thought to mediate protein-protein interactions (Ishigaki *et al.*, 2001).

In contrast to DmCBP80, the amino acid structure of DmCBP20 consists essentially of a RBD comprised of two RNP motifs, followed by a C-terminal tail rich in arginine and glycine residues (Makkerh 1996; Izaurralde *et al.*, 1995). The RBD of CBP20 is thought to have a phylogenetically conserved functional role in binding capped RNA (Mazza *et al.*, 2001). Since the RBD appears to be essential for RNA or cap-binding activity, it was left intact. Instead truncations were made in the auxiliary RGG tail, whose function in CBC is currently unknown, but probably serves to stabilise the interaction between CBC and capped RNA (Birney *et al.*, 1993; Burd and Dreyfuss, 1994). As a result, two truncations were made in *DmCBP20*, the first in the middle of the RGG domain and the second deleted this domain altogether (Figure 6.2, B).

Figure 6.2 UAS-Constructs Transformed into Fruit-Flies

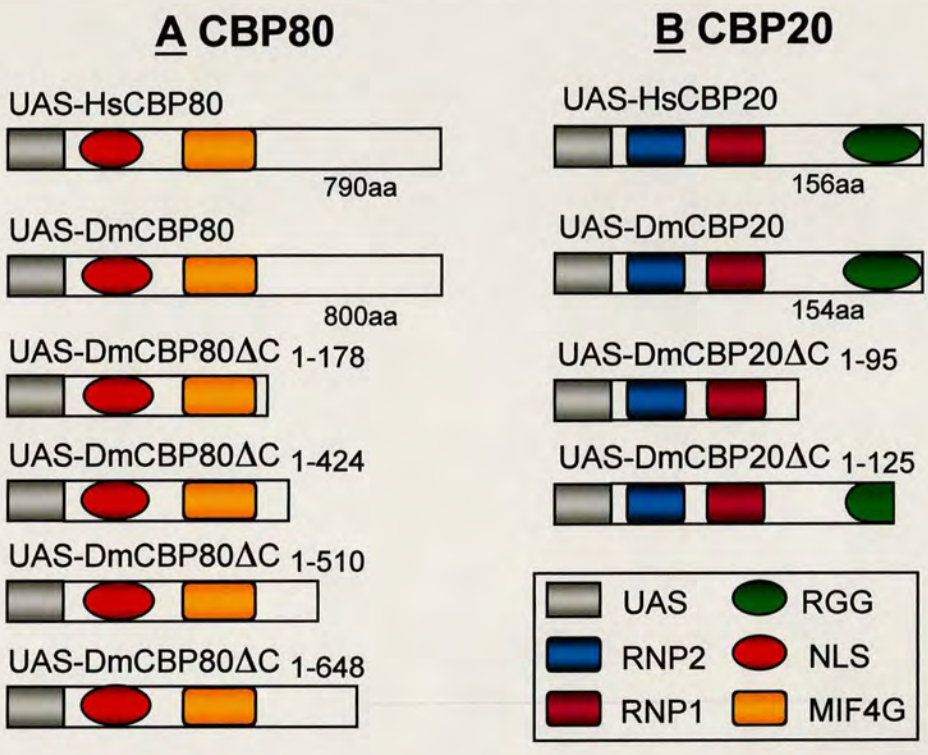


Figure 6.2 *UAS-CBP20* and *UAS-CBP80* constructs transformed into flies. Schematic diagram representing **A.** UAS-CBP80 and **B.** UAS-CBP20 constructs. Full-length human and *Drosophila* ORFs together with C-terminal truncations were cloned into pUAST. UAS-transgenic lines were then generated by P-element mediated transformation.

Both full-length and truncated *Drosophila* and human *CBP20* and *CBP80* constructs were cloned into pUAST and sequenced. P-element mediated transformation was then used to generate UAS-transgenic lines, which were characterised according to the eye colour of female adults and the mapping of transgene insertions. Each UAS-construct generated lines with independent insertions as demonstrated by the range of *white* marker expression, which might suggest varying strengths of transgene expression and the transgenes mapping to different chromosomes (Appendix B). In order to check that the UAS-transgenic lines were transcriptionally silent, adult flies were examined under a dissecting light microscope. This determined that in the absence of Gal4, UAS-transgenic flies did not exhibit any visible phenotypes.

6.2.2 Screen for *CBP20* and *CBP80* dominant adult phenotypes

To test whether *CBP20* and *CBP80* UAS-transgenic strains generated dominant phenotypes once transcriptionally activated, a screen was performed crossing five independent lines from each UAS-transgene to six different GAL4 drivers. The UAS-transgenic lines were selected on account of varying eye colour and transgene insertion on different chromosomes. The GAL4 drivers chosen for the screen included *patched*, *decapentaplegic*, *scabrous*, *S31*, *vestigal*, *scalloped*, *GMR*, *eyeless* and *OK-384*, which collectively express Gal4 in the larval eye, wing and leg imaginal discs. Since Gal4 activity is temperature dependent, crosses were set up at 29°C, 25°C and 18°C and the F1 progeny were screened for dominant adult phenotypes. As a control, GAL4 lines were crossed to w^{1118} flies to observe the effects of expressing Gal4 alone. The results from these mass GAL4-UAS crosses are presented in Appendix C. In summary, over-expression of full-length *Drosophila* and human CBP20 and CBP80 proteins did not produce any obvious defects in adult tissues. However, the targeted expression of the smallest truncations in both *DmCBP20* and *DmCBP80*, which remove 29 and 152 amino acids from the C-terminus respectively, did generate dominant adult phenotypes.

6.2.3 Identification of the dominant truncation *UAS-DmCBP20ΔC $_{1-125}$*

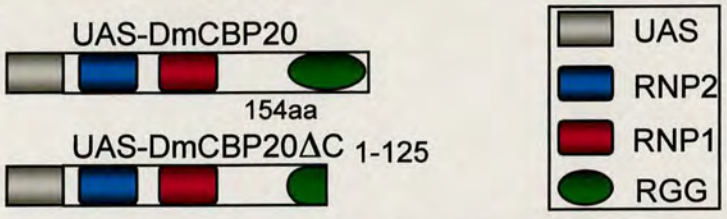
A total of two C-terminal truncations were generated in *DmCBP20*, the first was made in the middle of the RGG domain and the second deleted this domain all together. Only the truncation in the middle of the RGG domain, *UAS-DmCBP20ΔC $_{1-125}$* (*20ΔC $_{1-125}$*) generated visible phenotypes when crossed with *patched* (*ptc*) GAL4 alone (Appendix C). Although the function of the RGG domain in *CBP20* is currently not known, it is thought to stabilise the interaction between CBC and capped RNA (Birney *et al.*, 1993; Burd and Dreyfuss, 1994; Izaurralde *et al.*, 1995).

The transcriptional activation of *20ΔC $_{1-125}$* with *ptc-GAL4* produced dominant phenotypes in the eye, wing and thoracic tissues of adult flies. This is consistent with *ptc* being a segment polarity gene and directing Gal4 expression in the stripes between the anterior-posterior boundary of larval imaginal discs (Speicher *et al.*, 1994). Although, a total of three independent transgenic lines were generated containing the *20ΔC $_{1-125}$* transgene, only line #1 produced notable phenotypes in combination with *ptc-GAL* (Figure 6.3, A). As a result, the dominant phenotypes observed in *20ΔC $_{1-125}$ #1/ ptc-GAL4* adult eye, wing and thoracic tissues are described below.

The thorax of an adult fly can be broadly divided into two regions, the upper notum and the lower scutellum. In a wild-type fly, the scutellum is characterised by four long bristles termed macrochaetae and these are highlighted in the control fly that expresses Gal4 alone (Figure 6.4, Panel A[i] and B[iii]). The transcriptional activation of *20ΔC $_{1-125}$ #1* in combination with *ptc-GAL4* caused the scutellar macrochaetae located proximally on the thorax to over-produce. As can be observed from Figure 6.4, at 25°C the two proximal scutellar macrochaetae were multiplied nearly 5 fold. The macrochaetae increased from the wild-type number of two, to 9 macrochaetae in the *20ΔC $_{1-125}$#1/ ptc-GAL4* mutant (compare Panel B [iii] with [iv]).

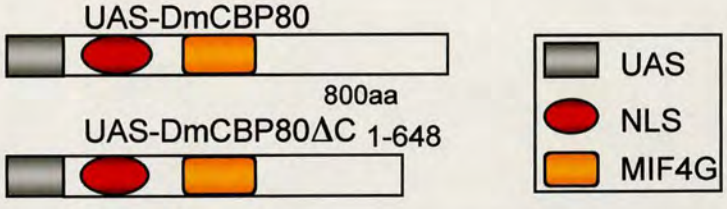
Figure 6.3 Identification of Dominant Truncations in CBC

A CBP20



Line Number	Eye Colour	Chromosome	GAL4 Driver
1	Red	2	*ptc*
2	Orange	3	None
3	Orange	3	None

B CBP80



Line Number	Eye Colour	Chromosome	GAL4 Driver
6	Dark Red	2	*sca, dpp, ey, ptc, 384*
4	Red	2	*sca, dpp, ey, ptc, 384*
2	Orange	X	*ey, ptc, 384*
7	Orange	3	*ey, ptc, 384*

Figure 6.3 Identification of dominant truncations *UAS-DmCBP20ΔC $_{1-125}$* and *UAS-DmCBP80ΔC $_{1-648}$*. Summary of UAS-transgenic lines and GAL4 drivers that generate dominant phenotypes. **A.** Out of three transgenic lines containing *UAS-DmCBP20ΔC $_{1-125}$* transgene, line #1 in combination with *patched-GAL4* (*ptc*) produced observable phenotypes in eye, wing and thoracic tissues. **B.** Eight transgenic lines contained the *UAS-DmCBP80ΔC $_{1-648}$* transgene of which, four lines generated visible phenotypes when crossed with *scabrous* (*sca*), *decapentaplegic* (*dpp*), *eyeless* (*ey*), *ptc* or *OK-384* GAL4 drivers. UAS-lines are listed in order of relative strength, where line #6 is the strongest and line #7 is the weakest.

Figure 6.4 *UAS-DmCBP20ΔC $_{1-125}$ #1/ ptc-GAL4* Dominant Bristle Phenotype

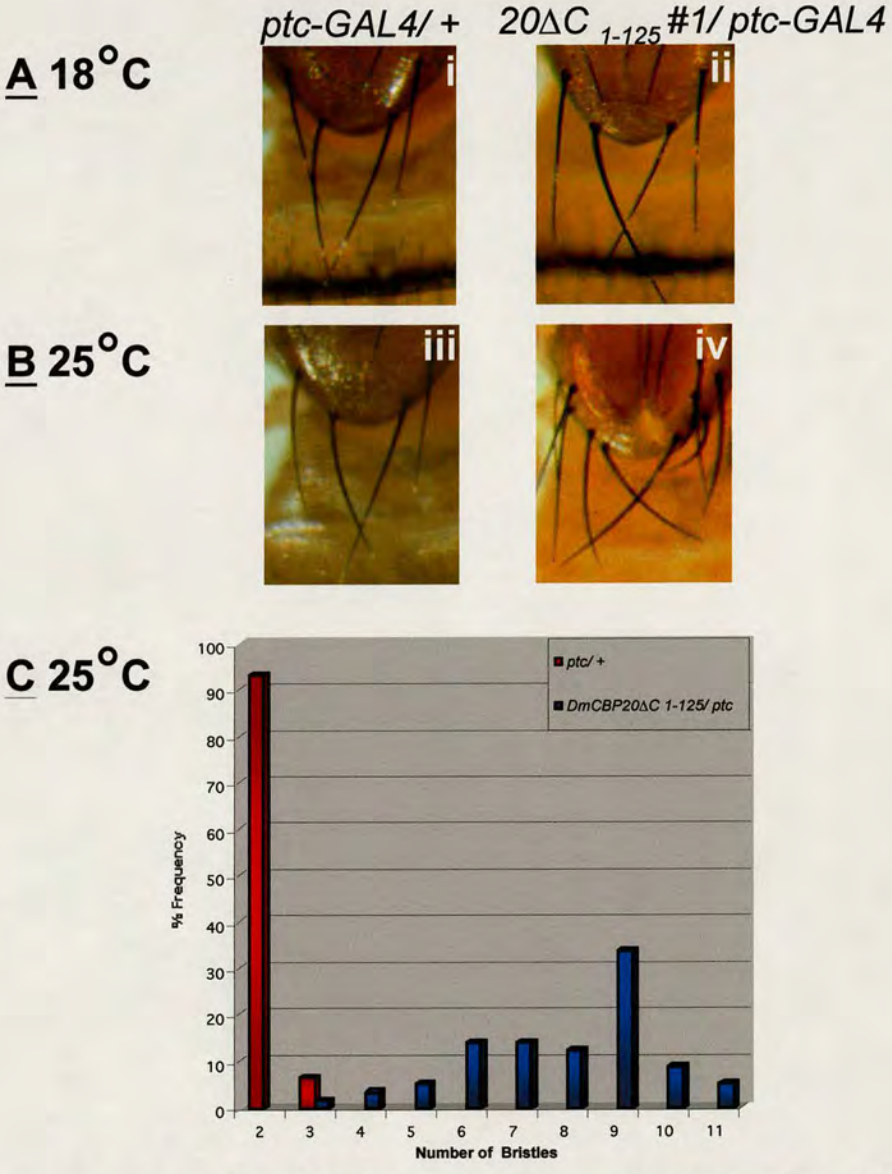


Figure 6.4 *UAS-DmCBP20ΔC $_{1-125}$ #1/ ptc-GAL4* dominant bristle phenotype. At 25°C, over-expression of *UAS-DmCBP20ΔC $_{1-125}$ #1* under *ptc-GAL4* control leads to over-production of the two proximal scutellar macrochaetae. Scutellum of *ptc-GAL4/ +* control flies is characterised by four macrochaetae at **A. [i]** 18°C and **B. [iii]** 25°C. In *UAS-DmCBP20ΔC $_{1-125}$ #1/ ptc-GAL4* adults, proximal set of scutellar macrochaetae are **B. [iv]** multiplied at 25°C and **A. [ii]** wild-type at 18°C. **C.** Graph illustrating frequency of proximal scutellar macrochaetae within population of *UAS-DmCBP20ΔC $_{1-125}$ #1/ ptc-GAL4* and control flies at 25°C.

However, this increase in proximal scutellar macrochaetae was not consistent within the F1 population. The graph in Panel C illustrates the range in the number of proximal scutellar macrochaetae in *20ΔC $_{1-125}$#1/ ptc-GAL4* and GAL4 controls. In *20ΔC $_{1-125}$ #1/ ptc-GAL4* adults, the macrochaetae ranged from 3 to 11 bristles, where 34% of the population exhibited 9 bristles. In contrast, 94% of GAL4 control flies had the wild-type number of 2 bristles and the remaining 6% of control flies had 3 proximal scutellar macrochaetae. Since the GAL4-UAS system is temperature dependent, repeating the cross at a lower temperature would be expected to decrease the severity of the dominant phenotype. Indeed, this was observed at 18°C where the proximal scutellar macrochaetae reverted to the wild-type number of two (compare Panels A [i] with [ii]). Conversely, increasing the Gal4 activity by raising the temperature to 29°C resulted in the more severe phenotype of pupal lethality (data not shown). Taken together, these observations suggest that the dominant macrochaetae phenotype is due to *20ΔC $_{1-125}$#1* transgene expression.

In addition to the over-production of scutellar macrochaetae, the transcriptional activation of *20ΔC $_{1-125}$#1* under the control of *ptc-GAL4* generated dominant eye and wing phenotypes at 25°C. Similar to the macrochaetae phenotype, the aberrant eye phenotype was not fully penetrant within the F1 population. As can be observed from Figure 6.5, eyes from the control fly that expressed Gal4 alone were wild-type (Panel A [i]). In comparison, *20ΔC $_{1-125}$#1/ ptc-GAL4* eyes ranged from mildly rough and reduced to wild-type in appearance (Panels A compare [ii and iii] with [iv and v]). This dominant eye phenotype also had a low penetrance within the F1 population with only 25% of adults exhibiting aberrant eyes (Panels A [ii] to [iv]). Whilst the remaining *20ΔC $_{1-125}$#1/ ptc-GAL4* flies had eyes that were wild-type (Panel A [v]). Decreasing the Gal4 activity by reducing the temperature to 18°C produced *20ΔC $_{1-125}$#1/ ptc-GAL4* eyes that were wild-type in appearance. Conversely, increasing the temperature to 29°C resulted in pupal lethality (data not shown). Taken together, these observations suggest that the dominant eye phenotypes generated at 25°C are due to the over-expression of *20ΔC $_{1-125}$#1* truncation.

Figure 6.5 *UAS-DmCBP20ΔC $_{1-125}$ #1/ ptc-GAL4* Dominant Eye and Wing Phenotype

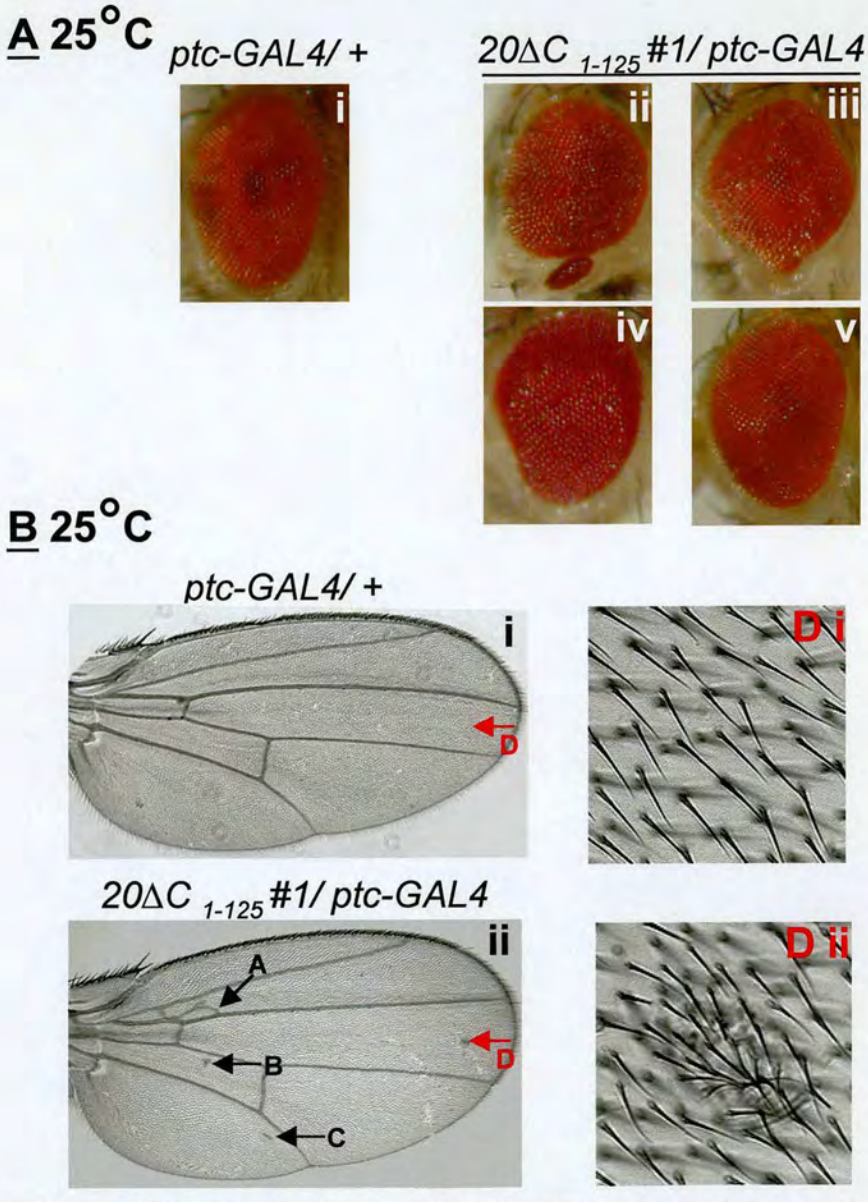


Figure 6.5 *UAS-DmCBP20ΔC $_{1-125}$ #1/ ptc-GAL4* dominant wing and eye phenotype. At 25°C, over-expression of *UAS-DmCBP20ΔC $_{1-125}$ #1* directed by *ptc-GAL4* generates aberrant phenotypes in adult eye and wing tissues. **A. [i]** Eye from *ptc-GAL4/ +* control is wild-type. **[ii-v]** Eyes from *UAS-DmCBP20ΔC $_{1-125}$ #1/ ptc-GAL4* range from mildly rough and reduced to wild-type **B. [i]** Wing from *ptc-GAL4/ +* control is wild-type. **[ii]** Wing from *DmCBP20ΔC $_{1-125}$ #1/ ptc-GAL4* has additional short veins indicated as A-D. **D. [i]** *ptc-GAL4/ +* control **[ii]** Additional short veins result from over-production of wing bristles.

Over-expression of *20ΔC $_{1-125}$#1* under the *ptc* promoter also elicited mild growth defects in the developing wing. At 25°C, additional veins were found in *20ΔC $_{1-125}$ #1/ ptc-GAL4* wings that were not present in the control flies that expressed GAL4 alone (Figure 6.5, Panels B compare [i] with [ii]). In *20ΔC $_{1-125}$#1/ ptc-GAL4* flies, the wing bristles were over-produced at specific regions and this generated additional short veins (Panel B compare [i] D with [ii] D). In comparison to the dominant eye phenotype, the wing phenotype was more penetrant within the F1 population. Interestingly, extra short veins were consistently found in the first posterior cell and distal tip of the wing in 75% and 49% of *20ΔC $_{1-125}$#1/ ptc-GAL4* adults respectively (Panel B [ii], A and D). Whilst two other short veins were also present in 62% of mutant flies but these were randomly located on the adult wing (for example, Panel B [ii], B and C).

Although, *20ΔC $_{1-125}$#1/ ptc-GAL4* flies exhibited dominant phenotypes in the eye, wing and thoracic tissues, only the scutellar macrochaetae phenotype was considered suitable to facilitate a genetic screen. The macrochaetae phenotype was easily detected under a dissecting light microscope and in addition, was consistent within a population at 21°C and 25°C (data not shown). To proceed with the genetic analysis, it was necessary to recombine *ptc-GAL4* and *20ΔC $_{1-125}$#1* transgenes onto the same chromosome. Consequently, large scale recombination crosses were performed at 21°C and 25°C. In over 1000 crosses, no viable recombination events were found at either temperature that included both UAS and GAL4 transgenes. This suggests that recombining *20ΔC $_{1-125}$ #1* and *ptc-GAL4* might be lethal or alternatively, both transgenes are inserted too close to allow recombination. As a result, work in the following sections concentrate on the *DmCBP80* dominant truncation, *UAS-DmCBP80ΔC $_{1-648}$*.

6.2.4 Identification of the dominant truncation *UAS-DmCBP80ΔC$_{1-648}$*

A total of four C-terminal truncations were generated in *DmCBP80* at regions believed to be conserved domain boundaries from multiple sequence alignments. As before, transgenic lines containing these truncations were transcriptionally activated by crossing with several GAL4 drivers. Out of the four truncations, only *UAS-DmCBP80ΔC $_{1-648}$* (*80ΔC $_{1-648}$*) transgene induced significant phenotypic effects when over-expressed with *ptc*, *decapentaplegic*, *scabrous*, *eyeless* and *OK-384* GAL4 drivers (Appendix C). *80ΔC $_{1-648}$* is the shortest truncation in *DmCBP80* that removes 152 amino acids from the C-terminus. However, the functional significance of this deletion is currently unknown. A total of eight independent lines were generated containing the *80ΔC $_{1-648}$* transgene, of which four lines produced dominant phenotypes. Figure 6.3, B presents these four lines placed in order of their relative strengths of expression, as judged by the range of phenotypes observed with various GAL4 drivers at different temperatures. The dominant phenotypes observed in combination with *eyeless*, *OK-384* and *scabrous* GAL4 drivers are described below.

Transgenic line *UAS-DmCBP80ΔC $_{1-648}$#4* (*80ΔC $_{1-648}$ #4*), which was mapped to the second chromosome and the second strongest transgene, produced dominant adult phenotypes under the control of *scabrous*, *decapentaplegic*, *OK-384*, *ptc* and *eyeless* GAL4 drivers (Figure 6.3, B). Transcriptional activation of *80ΔC $_{1-648}$ #4* in combination with *eyeless* (*ey*)-GAL4 generated dominant phenotypes in the adult eye. This is consistent with *eye* directing Gal4 expression to cells anterior of the morphogenetic furrow in larval and late pupal eye discs (Eggert *et al.*, 1998b). At 25°C, over-expression of *80ΔC $_{1-648}$ #4* generated smaller eyes compared to the control that expressed GAL4 alone (Figure 6.6, B). Examination of the adult eyes by scanning electron microscopy (SEM) clearly showed that the basic unit of the eye, the ommatidia clusters, were reduced in number over gross regions (Panels B, compare [i] with [ii-iv]). There was no consistent pattern of ommatidia loss with approximately 15% to 80% of ommatidia being absent in adult eyes.

Figure 6.6 *UAS-DmCBP80ΔC $_{1-648}$ #4; ey-GAL4* Dominant Eye Phenotype

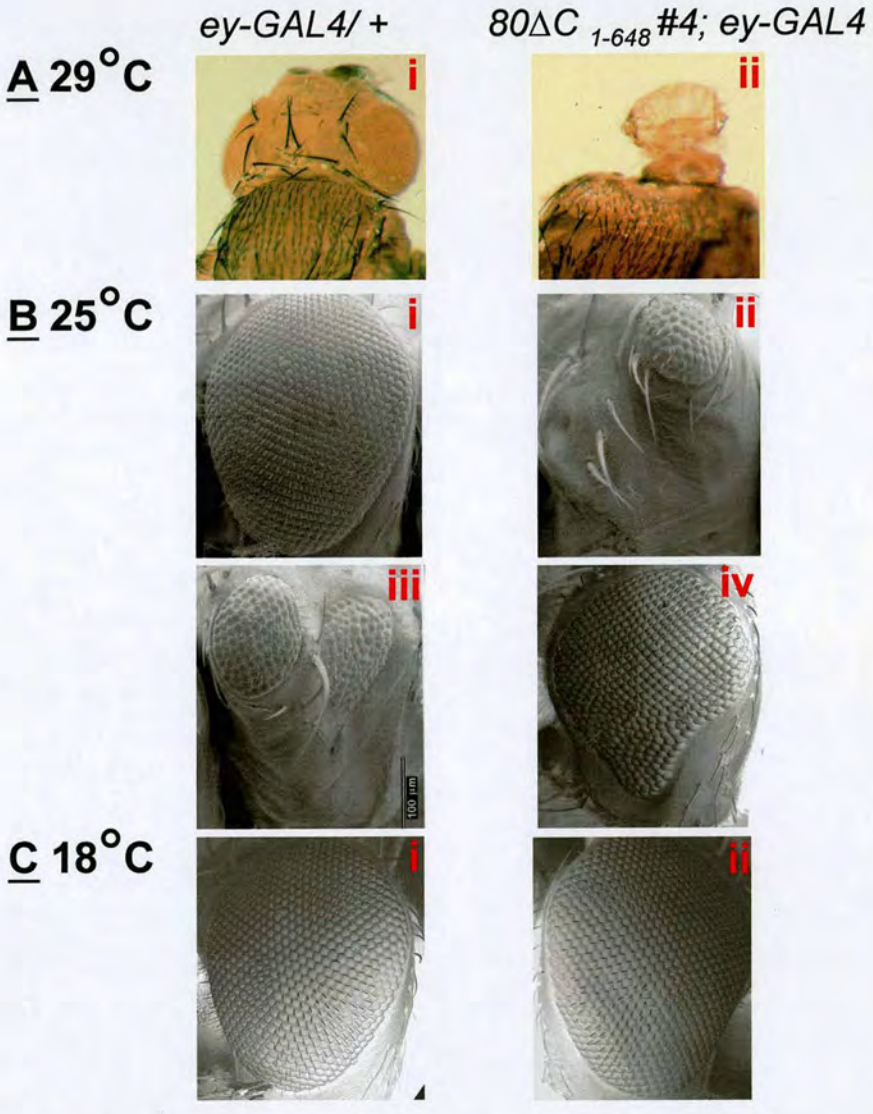


Figure 6.6 *UAS-DmCBP80ΔC $_{1-648}$ #4; ey-GAL4* dominant eye phenotype. Over-expression of *UAS-DmCBP80ΔC $_{1-648}$ #4* under control of *ey-GAL4*, which expresses in the head and eye,results in loss of eye tissue at 25°C and smaller head at 29°C. **A.** At 29°C, eyes from **[i]** *ey-GAL4/ +* control pharate at late pupal stage are wild-type. **[ii]** In *80ΔC $_{1-648}$ #4; ey-GAL4* pharate of same stage, both eyes are absent and the head is smaller. Scanning electron micrographs of adult eyes. **B. [i]** At 25°C, eye from *ey-GAL4/ +* control flies is wild-type. **[ii-iv]** *80ΔC $_{1-648}$ #4; ey-GAL4* eyes are reduced to varying degrees. Measurements based on images suggest approximately 15% to 80% of the eye is absent. **C. [i]** and **[ii]** At 18°C, *ey-GAL4/ +* and *80ΔC $_{1-648}$ #4; ey-GAL4* eyes are wild-type.

This suggests that the dominant phenotype is caused by growth defects in ommatidial assembly. Since the GAL4-UAS system is temperature dependent, increasing the Gal4 activity by raising the temperature to 29°C generated the more severe phenotype of pupal lethality. On closer examination, eyes from *80ΔC $_{1-648}$ #4; ey-GAL4* pharates were found to be completely absent (Panels A, compare [i] and [ii]). This significantly reduced the size of the head, which suggests the lethality might result from the inability of pharates to push through the pupal case and eclose. Conversely, decreasing the temperature to 18°C produced *80ΔC $_{1-648}$ #4; ey-GAL4* adults with eyes that were wild-type in comparison to the GAL4 control (Panels C, compare [i] and [ii]). Taken together, these observations suggest that the dominant eye phenotypes produced at higher temperatures are specifically due to *80ΔC $_{1-648}$ #4* transgene expression.

Although the dominant phenotype observed in *80ΔC $_{1-648}$ #4; ey-GAL4* flies obviously arises from defects in eye development, it is unclear which stages of development are being affected. Ommatidial assembly is an ordered process of cell recruitment where the photoreceptor cells form a cluster along with cone cells, pigment cells and bristle cells respectively (Chang *et al.*, 1994). Eyes from *80ΔC $_{1-648}$ #4; ey-GAL4* adults clearly show that the ommatidia are reduced in number. This suggests that the dominant eye phenotype might arise from defects in cell growth. Disruptions in cell proliferation, cell death or cell fate decisions would affect ommatidial assembly and decrease the number of ommatidial clusters. Interestingly, *GMR-GAL4* failed to generate aberrant eye phenotypes when crossed with *80ΔC $_{1-648}$* transgene at 18°C, 25°C or 29°C. *GMR-GAL4* expresses Gal4 during the later stages of eye development, in cells posterior to the morphogenetic furrow. Collectively, this suggests that the dominant eye phenotype might arise due to defects in cell growth during the earliest stages of eye development. This is consistent with the time cells undergo rapid cell proliferation.

In contrast to the restricted expression of *ey-GAL4*, *OK-384-GAL4* drives Gal4 expression in most adult tissues including the wing, eye, leg and thorax (Connolly *et al.*, 1996). It is therefore not surprising that over-expression of *80ΔC* $_{1\text{-}648}$ under the control of *OK-384* resulted in lethality. The transgenic line *UAS-DmCBP80ΔC* $_{1\text{-}648}$ *#7* (*80ΔC* $_{1\text{-}648}$ *#7*), which was mapped to the third chromosome and the weakest transgene, produced dominant phenotypes under the control of *ey*, *ptc* and *OK-384* GAL4 drivers (Figure 6.3, B). The transcriptional activation of *80ΔC* $_{1\text{-}648}$ *#7* in combination with *OK-384-GAL4* resulted in pupal lethality at 29°C. Whilst the control flies that expressed Gal4 alone were viable and wild-type in appearance at this temperature. Decreasing the temperature and consequently Gal4 activity to 25°C or 18°C produced viable *80ΔC* $_{1\text{-}648}$ *#7; OK-384-GAL4* adults that did not exhibit any visible phenotype (data not shown). Taken together, these data suggest that the pupal lethality at 29°C is specifically due to the over-expression of *80ΔC* $_{1\text{-}648}$ truncation.

The transgenic line *UAS-DmCBP80ΔC* $_{1\text{-}648}$ *#6* (*80ΔC* $_{1\text{-}648}$ *#6*), which was mapped to the second chromosome and the strongest transgene, produced dominant phenotypes under the control of *decapentaplegic*, *ey*, *ptc*, *OK-384* and *sca* GAL4 drivers (Figure 6.3, B). *Scabrous* (*sca*) specifically directs Gal4 expression in a subset of cells called the proneural cluster and sensory organ precursor (SOP), which are involved in the development of the external sense organs (Hinz *et al.*, 1994). Defects in this developmental pathway are easily visualised in adults as aberrant bristles on the head, thorax and wing tissues. When *80ΔC* $_{1\text{-}648}$ *#6* was crossed with *sca-GAL4* at 25°C, *80ΔC* $_{1\text{-}648}$ *#6/ sca-GAL4* progeny exhibited a severe loss of macrochaetae. To illustrate this point, the phenotype observed in the thoracic tissues will be described in comparison to the control fly that expressed Gal4 alone. Both the dorso-central region of the notum and scutellum are defined by four macrochaetae that are indicated in Figure 6.7, Panel A [i]. At 25°C, activation of *80ΔC* $_{1\text{-}648}$ *#6* transcription generated a loss of these macrochaetae and this was 100% penetrant within the F1 population.

Figure 6.7 *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4* Dominant Bristle Phenotype

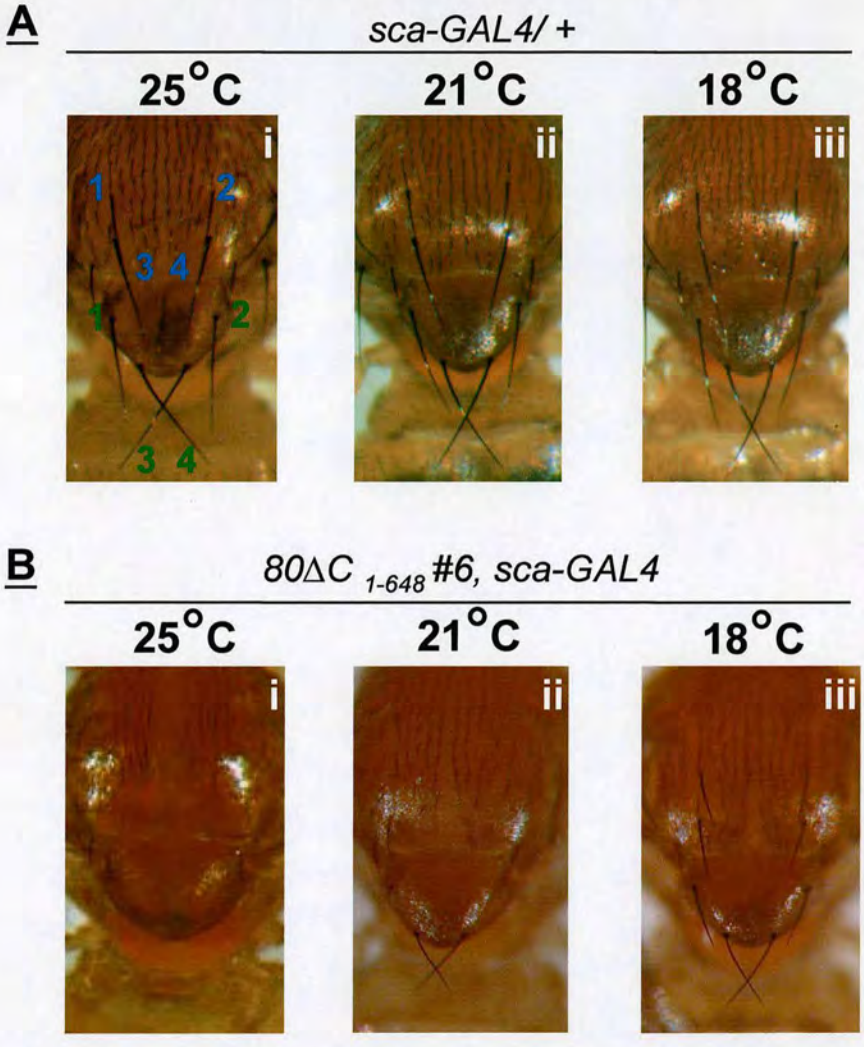


Figure 6.7 *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4* dominant bristle phenotype. Over-expression of *UAS-DmCBP80ΔC $_{1-648}$ #6* under control of *sca-GAL4* produces aberrant macrochaetae on thorax. **A. [i]** Central region of notum and scutellum are each defined by four macrochaetae, depicted 1-4 in blue and green respectively. **[i]-[iii]** In *sca-GAL4/ +* adults, macrochaetae on thorax are wild-type at 18°C, 21°C and 25°C. **B. [i]** In *80ΔC $_{1-648}$ #6, sca-GAL4* adults, macrochaetae on notum and scutellum are absent at 25°C. **[ii]** At 21°C, two distal scutellar macrochaetae are present, whilst macrochaetae on notum are still absent. Measurements based on images suggest scutellar macrochaetae are 45% shorter and thinner than GAL4 controls. **[iii]** At 18°C, *80ΔC $_{1-648}$ #6, sca-GAL4* adults have wild-type number of macrochaetae on both notum and scutellum. These macrochaetae are approximately 35% shorter and thinner than GAL4 controls.

In addition, there was a loss of smaller bristles or microchaetae on the notum compared with the Gal4 control (compare Panels A [i] with B [i]). In *80ΔC $_{1-648}$ #6/ sca-GAL4* adults, this 'bald' phenotype was reduced in severity at lower temperatures, which is consistent with a decrease in Gal4 activity. At 21°C, two out of four scutellar macrochaetae were present within 65% of the F1 population. The remaining 35% of *80ΔC $_{1-648}$ #6/ sca-GAL4* adults had an additional macrochaeta found on the central region of the notum. These macrochaetae were approximately 45% shorter and thinner than the GAL4 controls. Also associated with the reduced temperature was an increase in microchaetae towards the wild-type number (compare Panels A [ii] with B [ii]). At the lowest temperature, the wild-type number of both macrochaetae and microchaetae were present. However, the macrochaetae were approximately 35% shorter in length and thinner than the GAL4 controls (compare Panels A [iii] with B [iii]). Conversely, increasing the temperature to 29°C resulted in the more severe phenotype of pupal lethality (data not shown).

Taken together, these observations show that *80ΔC $_{1-648}$* dominant phenotypes described in the eye and thorax are independent of the chromosomal integration site and reproducible with different GAL4 drivers. As a result, the dominant adult phenotypes are specifically due to the over-expression of *80ΔC $_{1-648}$* truncation.

Out of the four transgenic lines containing *80ΔC $_{1-648}$* transgene, *80ΔC $_{1-648}$ #6/ sca-GAL4* dominant bristle phenotype was considered the best suited for genetic modifier screens. This macrochaetae phenotype was easily detected under a dissecting light microscope at either 25°C or 21°C. Furthermore, the dominant bristle phenotype was highly penetrant within the F1 population at 25°C and 21°C, with little variation within the population (Figure 6.7 and data not shown). To further facilitate the genetic analysis, it was necessary to recombine *sca-GAL4* and *80ΔC $_{1-648}$ #6* transgenes onto the same chromosome. Consequently, mass recombination crosses were performed at 25°C and 21°C. This generated *80ΔC $_{1-648}$ #6, sca-GAL4* recombinants on the second chromosome at both temperatures. Unfortunately, *80ΔC $_{1-648}$ #6, sca-GAL4* adults were sterile at 25°C. However, at 21°C, *80ΔC $_{1-648}$ #6, sca-GAL4* adults were fertile and the macrochaetae phenotype was tested for

reproducibility. This confirmed that the recombinant chromosome produced a dominant macrochaetae phenotype indistinguishable from when the transgenes resided on homologous chromosomes (data not shown).

The primary aim of this Chapter was to generate a dominant adult phenotype in either *CBP20* or *CBP80* that could be used in a screen for genetic interactors. Therefore, time was not spent examining the developmental defects that underlie the *80ΔC* $_{1\text{-}648}$ dominant phenotypes. Like *80ΔC* $_{1\text{-}648}$ *#4; ey-GAL4*, it remains to be determined which specific stages of sense organ development are being disrupted in *sca-GAL4, 80ΔC* $_{1\text{-}648}$ *#6* adults. In brief, sense organ development involves the selection of the SOP precursor from the proneural cluster. The SOP then divides to form four components of the external sense organ that are the macrochaeta, socket, neuron and sheath (Jan and Jan, 1993). In *sca-GAL4, 80ΔC* $_{1\text{-}648}$ *#6* adults, the macrochaetae were strikingly thinner and shorter than GAL4 controls at 21°C and 18°C. Whilst at 25°C, the macrochaetae were absent with very few sockets being observed. This suggests that similar to *80ΔC* $_{1\text{-}648}$ *#4; ey-GAL4* eyes, the macrochaetae phenotype might arise from defects in cell growth. The same factors are probably responsible for generating the range of dominant phenotypes observed in all four *80ΔC* $_{1\text{-}648}$ transgenic lines. It has not been investigated whether the neuron and sheath components of the external sense organ are also formed in *80ΔC* $_{1\text{-}648}$ *#6, sca-GAL4* adults at 25°C. As a result, defects in cell proliferation or cell fate decisions are possible candidates for inducing *80ΔC* $_{1\text{-}648}$ dominant phenotypes during eye and sense organ development.

In comparison to the developmental defects that underlie *80ΔC* $_{1\text{-}648}$ dominant phenotypes, it was considered more important to investigate the molecular defects generated by *80ΔC* $_{1\text{-}648}$ truncation. This would reveal the significance of the deleted domain in *80ΔC* $_{1\text{-}648}$ mutants and its contribution to CBC function. Out of the four transgenic lines, *80ΔC* $_{1\text{-}648}$ *#6* had the strongest relative strength of expression. As a result, the next section describes how *80ΔC* $_{1\text{-}648}$ *#6* mutants were analysed biochemically for defects in RNAP II transcript processing.

6.2.5 Analysis of RNA processing in *UAS-DmCBP80ΔC $_{1-648}$ #6* mutants

Characterisation of CBC has shown that it mediates the positive effects of the cap in pre-mRNA splicing, RNA 3'-end formation and U snRNA export (Lewis and Izaurralde, 1997). To test whether *80ΔC $_{1-648}$ #6* mutants exhibit a splicing defect, the same northern blotting assay described in the characterisation of *DmCBP20^{MFS}* in Chapter 4 was used. Whilst ovaries from *DmCBP20^{MFS}* adults were easily dissected for analysis, *80ΔC $_{1-648}$ #6* also needed to be ubiquitously expressed in tissues that were easily isolated and collected. Consequently, the GAL4 driver *C22* was chosen to direct *80ΔC $_{1-648}$ #6* expression in embryos. In order to characterise *C22-GAL4* expression, it was used to transcriptionally activate *UAS-GFP* and the resultant *UAS-GFP/ C22-GAL4* embryos were examined by fluorescence microscopy (data not shown). This determined that *UAS-GFP* was expressed ubiquitously throughout embryogenesis. As a result, *80ΔC $_{1-648}$ #6* was over-expressed in combination with *C22-GAL4* and this generated *80ΔC $_{1-648}$ #6/ C22-GAL4* embryos that were lethal at 25°C and viable at 18°C. Closer examination of *80ΔC $_{1-648}$ #6/ C22-GAL4* embryos staged the lethality towards the end of embryogenesis, around stage 14. In contrast, *C22/ +* control embryos that expressed Gal4 alone were viable at both 25°C and 18°C (data not shown). This confirmed that the embryonic lethality observed was specifically due to the effects of *80ΔC $_{1-648}$ #6* transgene expression and not Gal4 alone.

To determine whether the lethality of *80ΔC $_{1-648}$ #6/ C22-GAL4* embryos resulted from defects in pre-mRNA splicing, northern analysis was used to examine the splicing of endogenous ribosomal protein mRNAs during embryogenesis. Ribosomal proteins RpL19, RpS3 and RpA1 were specifically selected for analysis since they are ubiquitously expressed throughout embryogenesis and have pre-mRNAs that contain two introns, one intron and no introns respectively. Characterisation of CBC has shown that it mediates the stimulatory effects of the cap in the splicing of the cap-proximal intron alone (Izaurralde *et al.*, 1994; Lewis *et al.*, 1996). Therefore, if *80ΔC $_{1-648}$ #6* truncation generates a splicing defect, increased steady state pre-mRNA

to mRNA ratios would be observed in RpS3 pre-mRNAs that contain one intron. In contrast, over-expression of *80ΔC $_{1-648}$ #6* would be predicted to have no influence on the splicing of RpL19 that contains two introns and RpA1, which contains no introns and effectively serves as a negative control.

Embryos were collected and matured from *80ΔC $_{1-648}$ #6/ C22-GAL4* and *C22-GAL4/ +* at 18°C and 25°C for the following time points: 0-4, 4-8, 8-12, 12-16 and 16-22 hours. Total RNA was then isolated from staged embryos and equivalent amounts of RNA were analysed by northern blotting. The nitrocellulose filters were subsequently probed with full-length *RpL19*, *RpS3* and *RpA1* cDNAs and the steady state ratios of pre-mRNA to mRNA were quantitated using Storm 860 PhosphorImager together with ImageQuant software (data not shown). However, no significant increase was observed in the steady state pre-mRNA to mRNA ratio of RpS3 ribosomal protein pre-mRNAs throughout embryogenesis (data not shown). This suggested that the northern analysis might not be sensitive enough to detect a splicing defect in *80ΔC $_{1-648}$ #6/ C22-GAL4* embryos. Alternatively, embryonic lethality might be induced by defects in other functions of CBC such as RNA 3'-end formation (Flaherty *et al.*, 1997). It was therefore reasoned that a more sensitive assay, such as RNase protection, would be better suited to address whether over-expression of *80ΔC $_{1-648}$ #6* generates aberrant pre-mRNA splicing. RNase protection assay would also address whether *80ΔC $_{1-648}$ #6/ C22-GAL4* embryos had defects in poly(A) tail formation. Unfortunately, the RNase experiments were not completed due to lack of time.

6.2.6 *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4* is not a true dominant negative

Before *80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae phenotype can be used in a genetic modifier screen, it needs to be characterised genetically. The original aim of this Chapter was to truncate either *DmCBP20* or *DmCBP80* and identify a dominant negative form of CBC, which gives a visible phenotype in adult tissues. To determine whether *80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae phenotype was due to a

dominant negative effect, specific test crosses were performed at 21°C. If a truncation acts as a dominant negative, simultaneous removal of one wild-type copy of the gene would be predicted to increase the severity of the dominant phenotype. Conversely, flies that over-express full-length transgenes, as well as the truncation, would be predicted to suppress the dominant phenotype. As a result, flies were generated that in addition to over-expressing *80ΔC $_{1-648}$ #6, sca-GAL4*, also contained *Df(1) bi-D3* that uncovers the *DmCBP80* locus. This determined that the macrochaetae phenotype was not enhanced by removing one wild-type copy of *DmCBP80* (data not shown). However, the dominant phenotype was partially suppressed by introducing transgenes that over-expressed full-length *UAS-DmCBP80 #1*. As can be observed from Figure 6.8, all four macrochaetae were present on the central notum and scutellum of *80ΔC $_{1-648}$ #6, sca-GAL4; UAS-DmCBP80 #1* adults. However, these macrochaetae were approximately 35% shorter and thinner than GAL4 controls (compare [i], [ii] and [iii]). It was reasoned that the suppression observed might not be caused by *UAS-DmCBP80* but by a lack of Gal4 available to activate both *UAS-DmCBP80* and *80ΔC $_{1-648}$ #6* transgenes. To test this, full-length *UAS-DmCBP20* and truncated *UAS-DmCBP80* constructs were crossed into the *80ΔC $_{1-648}$ #6, sca-GAL4* background. These constructs did not modify the dominant macrochaetae phenotype, which suggests that the partial suppression observed was specific to over-expressing full-length *UAS-DmCBP80*. Taken together, this data shows that *80ΔC $_{1-648}$ #6* truncation is not a true dominant negative. Instead, *80ΔC $_{1-648}$ #6* appears to be a gain-of function mutant.

Figure 6.8 *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4* Dominant Bristle Phenotype is Partially Suppressed by *UAS-DmCBP80 #1* and *Df(3R) P14 sr^{1}* that uncovers *DmCBP20*

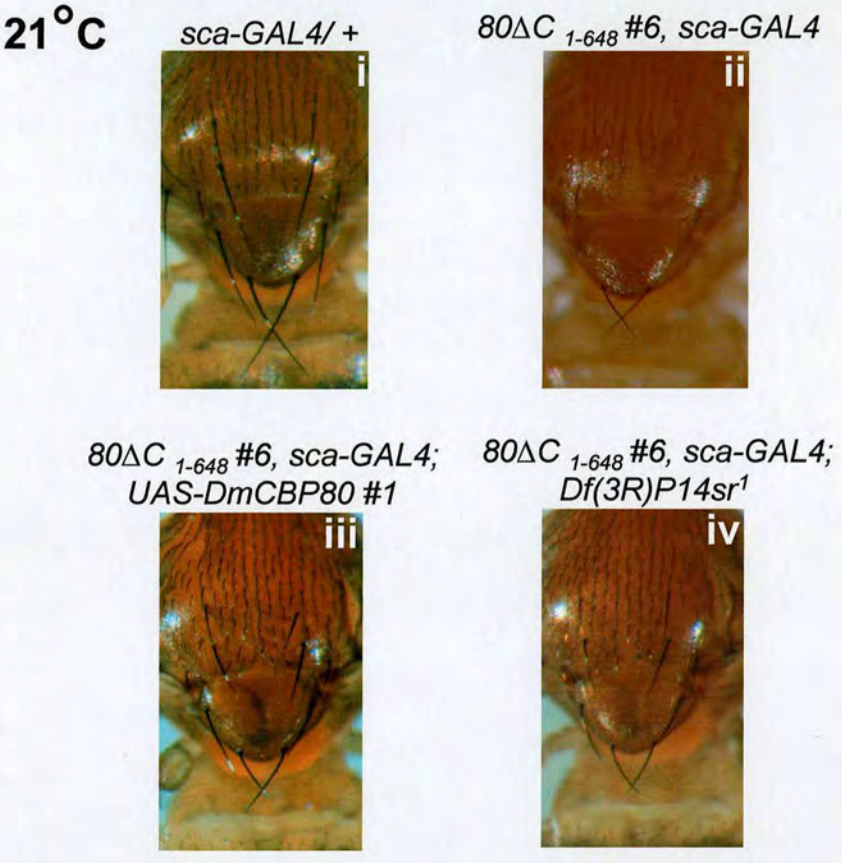


Figure 6.8 *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4* bristle phenotype is partially suppressed by *UAS-DmCBP80 #1* and *Df(3R) P14 sr^{1}* that uncovers *DmCBP20*. [i] At 21°C, *sca-GAL4/ +* control has four wild-type macrochaetae on central region of notum and scutellum. **[ii]** In *80ΔC $_{1-648}$ #6, sca-GAL4* adults, two distal scutellar macrochaetae are present, whilst macrochaetae on notum are absent. **[iii]** Over-expression of *UAS-DmCBP80 #1* in *80ΔC $_{1-648}$ #6, sca-GAL4* background. All four macrochaetae are present on central notum and scutellum. Measurements based on images suggest macrochaetae are 35% shorter and thinner than control. **[iv]** Dominant phenotype is also partially suppressed by crossing *Df(3R) P14 sr^{1}* into *80ΔC $_{1-648}$ #6, sca-GAL4* background. All four macrochaetae are present on central notum and scutellum and are approximately 55% shorter and thinner than control.

6.2.7 *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4* interacts with *DmCBP20* deficiency

Considering DmCBP20 and DmCBP80 must heterodimerise to form a functional CBC (Izaurralde *et al.*, 1994), it was reasoned that the dominant macrochaetae phenotype might be modified by alleles of *DmCBP20*. In order to test this, various *DmCBP20* alleles were crossed into *80ΔC $_{1-648}$ #6, sca-GAL4* background at 21°C. These alleles included transgenes ubiquitously expressing *DmCBP20* genomic region and ORF; *Df(3R) P14 sr^{1}* that uncovers the *DmCBP20* locus; together with *DmCBP20^{HL}* and *DmCBP20^{MFS}* alleles generated from the P-hop described in Chapter 4. This determined that the macrochaetae phenotype was not modified by transgenes that over-expressed *DmCBP20*. However, the dominant phenotype was partially suppressed by crossing in *Df(3R) P14 sr^{1}* and *DmCBP20* deletion alleles #1 and #2. As can be observed from Figure 6.8, all four macrochaetae were present on central notum and scutellum of *80ΔC $_{1-648}$ #6, sca-GAL4*; *Df(3R) P14 sr^{1}* adults. However, these macrochaetae were approximately 55% shorter and thinner than GAL4 controls (compare [i], [ii] and [iv]). Collectively, this genetic data suggests that increasing the level of DmCBP20 protein appears to have no effect on 80ΔC $_{1-648}$ #6 activity. Whereas reducing endogenous DmCBP20, by the introduction of a deficiency or *DmCBP20* deletion alleles, does partially suppress the dominant macrochaetae phenotype.

6.2.8 *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4* does not interact with *DmCBP80R* deficiency

Since *Drosophila* has a second DmCBP80 homologue, an obvious question to address was whether mutant alleles in *DmCBP80R* could modify *80ΔC $_{1-648}$ #6, sca-GAL4* dominant phenotype. To date, no known mutant alleles have been described for *DmCBP80R*. However, a BDGP search identified *Df(3R) e-N19* that uncovers *DmCBP80R* locus on polytene segment 93E1-2. *Df(3R) e-N19* was crossed into *80ΔC $_{1-648}$ #6, sca-GAL4* background at 21°C. This determined that *Df(3R) e-N19* did not modify the macrochaetae phenotype (data not shown). Characterisation of

DmCBP80R from Chapter 5 shows that DmCBP80R mRNA is restricted to third instar larvae and adult males, where it is predominantly expressed in the testes. Therefore, a deficiency uncovering *DmCBP80R* would not be predicted to interact with *80ΔC $_{1-648}$ #6* truncation that is over-expressed in somatic tissues. Indeed, this is consistent with the result observed here.

6.3 DISCUSSION

6.3.1 Identification of the dominant truncation *UAS-DmCBP20Δ $_{1-125}$*

Results from the GAL4-UAS crosses show that over-expression of full-length *Drosophila* and human *UAS-CBP20* and *UAS-CBP80* do not produce visible phenotypes in adult flies. However, the targeted over-expression of the shortest truncations in both *DmCBP20* and *DmCBP80*, which remove 29 and 152 amino acids from the C-terminus respectively, do generate dominant adult phenotypes. A total of three transgenic lines contained the *20ΔC $_{1-125}$* transgene, of which line #1 generated phenotypes in combination with *ptc-GAL4* alone. Aberrant phenotypes were observed in the eye, wing and thoracic regions of adult flies, which is consistent with *ptc* expression. However, since only one *20ΔC $_{1-125}$* transgenic line produces dominant phenotypes, it cannot be assumed that the phenotypes observed were independent of the chromosomal integration site and specific to *20ΔC $_{1-125}$* transgene expression.

The *20ΔC $_{1-125}$* transgene is predicted to encode a protein that is truncated in the middle of the C-terminal RGG tail. The function of the RGG domain in DmCBP20 is currently unknown, however, RGG tails often serve as an auxiliary domain in RNA-binding (Birney *et al.*, 1993; Burd and Dreyfuss, 1994). It is therefore thought that the RGG domain in *DmCBP20* might act to stabilise the interaction between CBC and capped RNA.

In contrast to *20ΔC $_{1-125}$*, the transgene *20ΔC $_{1-95}$* does not produce visible adult phenotypes when crossed with the range of GAL4 drivers used in this study. The conceptual protein product of *20ΔC $_{1-95}$* deletes the RGG domain altogether. One possibility is that flies over-expressing *20ΔC $_{1-95}$* might form mutant CBC that does not stably bind capped RNA. Instead, endogenous CBC preferentially binds and consequently *20ΔC $_{1-95}$* expression does not generate aberrant phenotypes. In comparison, over-expression of *20ΔC $_{1-125}$* would form mutant CBC containing a partially deleted RGG domain. This mutant CBC might be relatively more stable in binding capped RNA. However, the mutant CBC might dissociate prematurely during RNAP II transcript processing. This would destabilise protein-protein interactions and induce molecular defects in cap-dependent RNA metabolism and the dominant phenotypes observed in adult tissues.

6.3.2 Identification of the dominant truncation *UAS-DmCBP80ΔC$_{1-648}$*

Four truncations were made in *DmCBP80*, however, only the over-expression of *80ΔC $_{1-648}$* generated dominant phenotypes in adult tissues. A total of eight transgenic lines contained *80ΔC $_{1-648}$* transgene. Out of these, four independent lines generate dominant adult phenotypes in combination with several GAL4 drivers. Over-expression of *80ΔC $_{1-648}$* creates a number of adult phenotypes ranging from visible tissue defects to lethality. This demonstrates that the phenotypic effects are independent of the chromosome integration site and reproducible with different GAL4 drivers. Taken together, these data support that the dominant adult phenotypes observed are specifically due to *80ΔC $_{1-648}$* transgene over-expression.

Site-directed mutagenesis of HsCBP80 has identified a number of key residues that interact with HsCBP20. Since only CBP20 binds the cap, CBP80 must heterodimerise with CBP20 to form a functional CBC (Mazza *et al.*, 2001). Interestingly, the key residues in CBP80 that interact with CBP20 are extensively conserved and reside within the middle and carboxyl-regions of CBP80. The three longest UAS-truncations in *DmCBP80* delete these conserved residues. Therefore,

these *UAS-DmCBP80* truncations would not be predicted to form a stable heterodimer with endogenous DmCBP20. As a result, no mutant CBC is formed and no dominant phenotypes are observed in combination with various GAL4 drivers.

80ΔC $_{1-648}$ is the shortest truncation made in *DmCBP80* and removes 152 amino acids from the C-terminus. This truncation was made at a region believed to be a conserved boundary from amino acid alignments. Recently, the crystal structure of HsCBC has been determined (Mazza *et al.*, 2001). A distinct feature of HsCBP80 carboxyl-terminus is a very long coiled-coil that is comprised of two alpha helices. The coiled-coil protrudes from HsCBP80 and might provide a distinct surface for protein-protein interactions. Interestingly, this coiled-coil is highly conserved in DmCBP80. Furthermore, *80ΔC $_{1-648}$* is predicted to encode a protein where the coiled-coil is nearly deleted altogether. Although the functional significance of this domain is currently not known, truncation of HsCBP80 coiled-coil does not appear to affect HsCBP20 or cap-binding activities (Mazza *et al.*, 2001). This suggests that the coiled-coil might mediate interactions between CBP80 and RNAP II processing/ export factors. Therefore, over-expression of *80ΔC $_{1-648}$* might generate mutant CBC that is defective in mRNA processing. It is possible that mutant CBC containing *80ΔC $_{1-648}$* truncation generates disruptions in a number of cap-dependent mRNA processing events. A culmination of molecular defects in pre-mRNA splicing, RNA 3'-end formation and U snRNA export might ultimately disrupt RNAP II transcript metabolism and produce the dominant adult phenotypes observed here. Northern analysis did not detect a defect in the splicing of RpS3 RNAs in embryos that over-expressed *80ΔC $_{1-648}$* transgene. However, further biochemical analysis of *80ΔC $_{1-648}$* mutants using more sensitive assays such as RNase protection or alternatively, microarray analysis might be more successful in determining the functional significance of the deleted domain.

6.3.3 CBC dominant phenotypes might arise from defects in cell growth

Over-expression of *20ΔC* $_{1\text{-}125}$ and *80ΔC* $_{1\text{-}648}$ generate a range of aberrant adult phenotypes in the wing, eye and thoracic tissues. Since no evidence suggests CBC has specific functions in either wing, eye or sense organ development, the defects that underlie these dominant phenotypes have not been investigated. Interestingly, recent work has suggested that the cap-binding activity of CBC is subject to regulation by extracellular stimuli (Wilson *et al.*, 1999 and 2000). This raises the question of whether CBC might be the nuclear end-point for multiple signalling pathways required to regulate RNAP II transcript metabolism and ultimately cell growth. It can therefore be reasoned that defects in CBC activity and subsequently cap-dependent RNA metabolism might in turn deregulate cell growth control. As a result, the molecular defects induced by *80ΔC* $_{1\text{-}648}$ and *20ΔC* $_{1\text{-}95}$ transgene over-expression might ultimately affect cell growth control and account for the dominant phenotypes observed.

Disruptions in cell growth can encompass a range of possible defects in processes such as cell proliferation, cell death or cell fate decisions. However, the dominant adult phenotypes arising from the transcriptional activation of *80ΔC* $_{1\text{-}648}$ probably result from a common defect. In *UAS-DmCBP80ΔC* $_{1\text{-}648}$ *#6, sca-GAL4* flies, the macrochaetae on the central notum and scutellum are absent with very few sockets present at 25°C. Whilst at 21°C and 18°C, the macrochaetae on the thorax are notably shorter and thinner than the GAL4 controls. This suggests that disruptions in cell proliferation or cell differentiation might be responsible for these phenotypes. Similarly, in *UAS-DmCBP80ΔC* $_{1\text{-}648}$ *#4; ey-GAL4* adults, the ommatidia are strikingly reduced in number. No dominant eye phenotypes are observed when *80ΔC* $_{1\text{-}648}$ is transcriptionally activated with *GMR-GAL4* that expresses later in eye development. This suggests that CBC-related defects occur early in eye development, at times consistent with cell proliferation. Alternatively, it is possible that checkpoint mechanisms exist during eye and sense organ development that detect general defects in cell growth. As a result, disruptions in cell growth might

induce cell death by apoptosis and generate the dominant phenotypes observed. However, in the absence of any experimental data, this is purely hypothetical.

6.3.4 *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4* is a gain-of function mutant

UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4 dominant phenotype is not enhanced by a deficiency that uncovers *DmCBP80* locus. However, macrochaetae are partially suppressed by simultaneously over-expressing *UAS-DmCBP80* in *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4* background. This partial suppression is not due to a lack of Gal4 to activate both *80ΔC $_{1-648}$* and *UAS-DmCBP80* transgenes. Interestingly, over-expression of full-length *UAS-DmCBP80* in combination with the GAL4 drivers used in this study does not generate dominant adult phenotypes. This suggests *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae phenotype observed here is specific to the over-expression of *80ΔC $_{1-648}$* truncation. Therefore, *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4* is not a true dominant negative but rather a gain-of function mutant.

The data suggests the partial suppression observed from over-expressing both full-length *UAS-DmCBP80* and truncated *80ΔC $_{1-648}$* in the same tissues might increase the competition for endogenous DmCBP20. As a result, the activity of mutant CBC containing *80ΔC $_{1-648}$* truncation would decrease and cause the macrochaetae phenotype to be partially suppressed. However, reducing the endogenous levels of *DmCBP80* in *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4* background does not affect the activity of *80ΔC $_{1-648}$*. Therefore, the dominant macrochaetae phenotype is not modified.

6.3.5 *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4* interacts with *DmCBP20* deficiency

UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4 dominant phenotype was not modified by *DmCBP20^{MFS}* mutants or transgenes that ubiquitously express *DmCBP20* genomic region and ORF. However, macrochaetae are partially suppressed by a deficiency that uncovers *DmCBP20* locus and *DmCBP20* deletion alleles # 1 and #2 described in Chapter 4. An interaction between *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4* and *DmCBP20* mutants is not surprising, since DmCBP20 and DmCBP80 are only functional as constituents of a heterodimeric CBC. This data provides further support that the macrochaetae phenotype observed in *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4* adults is specific to *80ΔC $_{1-648}$* transgene over-expression. It appears that reducing the endogenous DmCBP20 protein available decreases the activity of 80ΔC $_{1-648}$ and partially suppresses the macrochaetae phenotype.

6.3.6 *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4* does not interact with *DmCBP80R* deficiency

UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4 dominant phenotype is not modified by a deficiency that uncovers *DmCBP80R* locus. This negative result is not surprising since decreasing the endogenous amounts of DmCBP80R would have no effect on somatic tissues that over-express *80ΔC $_{1-648}$*. Instead, this deficiency would be predicted to reduce the endogenous levels of DmCBP80R present in adult testes. However, an interesting question to address is whether DmCBP80 and DmCBP80R are functionally redundant. Therefore, to determine if over-expressing *DmCBP80R* in somatic tissues could rescue mutants in *DmCBP80*. Unfortunately, the genetic tools required to test such a hypothesis are currently not available.

Although DmCBP80 and DmCBP80R share high conservation, much of the sequence divergence lies within the carboxyl-terminal 200 amino acids. In HsCBP80 and DmCBP80, this region is predicted to encode a prominent coiled-coil that may mediate protein-protein interactions between CBP80 and mRNA processing/ export

factors. The carboxyl-terminal 200 amino acids of DmCBP80R shares very low sequence identity to either DmCBP80 or HsCBP80. This suggests that in DmCBP80R, the carboxyl-region might mediate different protein-protein interactions. Intriguingly, since the coiled-coil region is deleted in *80ΔC* $_{1\text{-}648,}$ both DmCBP80R and *80ΔC* $_{1\text{-}648}$ are predicted to be structurally similar. One possibility is that the dominant phenotypes observed in *UAS-DmCBP80ΔC* $_{1\text{-}648}$ *#6, sca-GAL4* adults is caused by *80ΔC* $_{1\text{-}648}$ behaving in a manner analogous to *DmCBP80R* in somatic cells. One might predict that the misexpression of DmCBP80R in somatic cells would give the same phenotypes as the over-expression of the *80ΔC* $_{1\text{-}648}$ truncation.

6.4 CONCLUSIONS

The targeted over-expression of *UAS-DmCBP20ΔC* $_{1\text{-}125}$ and *UAS-DmCBP80ΔC* $_{1\text{-}648}$ using the GAL4-UAS system generated dominant adult phenotypes. Characterisation of *UAS-DmCBP80ΔC* $_{1\text{-}648}$ *#6, sca-GAL4* shows it is a gain-of function mutant whose dominant adult macrochaetae phenotype is suitable for genetic modifier screens.

CHAPTER 7

Screen for CBC-Interacting Genes

7.1 INTRODUCTION

Characterisation of the cap and CBC show they act as a molecular tag to define RNAP II transcripts for nuclear processing. Key roles for CBC have been described in pre-mRNA splicing, RNA 3'-end formation and U snRNA export (Lewis and Izaurralde, 1997). Despite this, little is currently known about the proteins that CBC interacts with or the molecular mechanisms involved. In metazoans, both cap-binding proteins appear to be evolutionarily conserved. Indeed, results from Chapter 3 have shown that DmCBP20 and DmCBP80 are highly homologous to the corresponding human proteins. This suggests that CBC function might also be conserved between *Drosophila melanogaster* and human species.

The completed *Drosophila melanogaster* genome shows that components of the RNAP II transcript processing machinery are highly conserved (Adams *et al.*, 2000). In general, *Drosophila* proteins are more closely related to their vertebrate counterparts than *S. cerevisiae*. This implies the complexity of mechanisms that contribute to RNAP II transcript processing are similar between *Drosophila* and vertebrates. Therefore, genetic studies in fruit-flies rather than yeast might be more successful in identifying candidate adaptor molecules that facilitate the function of CBC and the cap.

The overall aim of the GAL4-UAS study was to generate a dominant adult phenotype in either *DmCBP20* or *DmCBP80* that could be used in genetic modifier screens to identify CBC-interacting genes. Chapter 6 has described the identification and initial characterisation of *UAS-DmCBP80ΔC* $_{1\text{-}648}$ *#6, sca-GAL4* dominant macrochaetae phenotype. However, before a dominant adult phenotype can be used

in a genetic modifier screen, it needs to fulfil three main criteria. Firstly, the dominant phenotype must be easily detected under a dissecting microscope. Indeed, the previous Chapter has already shown that *UAS-DmCBP80ΔC $_{1-648}$#6, sca-GAL4* recombinants exhibit a macrochaetae phenotype that is clearly observed at 21°C. At this temperature, adults have no macrochaetae on the central notum and only two out of four scutellar macrochaetae. This produces an intermediate phenotype that allows both enhancers and suppressers to be readily scored. Secondly, the dominant adult phenotype needs to be consistent within a population. As mentioned earlier, *UAS-DmCBP80ΔC $_{1-648}$#6, sca-GAL4* macrochaetae phenotype is highly penetrant within a population at 21°C. Approximately 65% of *UAS-DmCBP80ΔC $_{1-648}$#6, sca-GAL4* recombinants exhibit the phenotype described above. The remaining 35% of adults have an additional macrochaeta on the central region of the notum. Finally, the dominant adult phenotype must be sensitive to modification.

In order to address the last criteria, Chapter 7 describes how *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4* adults were crossed to a collection of deficiencies and candidate CBC-interacting genes. This determined whether *UAS-DmCBP80ΔC $_{1-648}$#6, sca-GAL4* dominant macrochaetae phenotype would be a useful tool that could be used in future large-scale genetic modifier screens.

7.2 RESULTS

7.2.1 Deficiency screen using *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4*

To test whether *80ΔC $_{1-648}$ #6, sca-GAL4* dominant phenotype could be modified and useful for identifying *CBP80* interactors, a screen was performed using chromosomal deficiencies. *80ΔC $_{1-648}$ #6, sca-GAL4* adults were crossed to a deficiency collection, which uncovered approximately 65% of the second chromosome and 50% of the third chromosome. Crosses were set up at 21°C and the F1 progeny were screened for modification of *80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae phenotype. To confirm the genetic interaction was specific to *80ΔC $_{1-648}$* truncation, the modifying deficiencies were subjected to three control crosses. Firstly, the interacting deficiency was crossed to *sca-GAL4* alone at 21°C. This ensured that the modification observed did not result from an interaction between the deficiency and Gal4 protein. Secondly, the deficiency was crossed with an unrelated UAS-GAL4 transgenic line and the F1 progeny screened for modification. The selected line was *UAS-atonal, S31-GAL4* that exhibits a dominant scutellar macrochaetae phenotype at 25°C. In contrast to CBC, atonal is a RNAP II transcription factor with specified roles in sense organ development. *S31-GAL4* driver expresses Gal4 in the same tissues as *sca-GAL4*. Similar to *80ΔC $_{1-648}$ #6, sca-GAL4* dominant phenotype, *UAS-atonal, S31-GAL4* adults lack the two distal macrochaetae present on the scutellum at 25°C (Jarman *et al.*, 1998). This determined whether the enhancement or suppression was specific to CBC. The final control involved further mapping the interacting genetic regions using smaller overlapping deficiencies. This eliminated the possibility of second site modifiers causing the modification of *80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae. Once the minimal interacting breakpoints had been defined, the polytene regions were analysed for candidate CBC-interacting genes using the cytosearch tool in Flybase. Since CBC facilitates the positive effects of the cap in

splicing, RNA 3'-end formation and U snRNA export, genes involved in these processes and other aspects of RNAP II transcript metabolism were investigated.

The results from the deficiency screen are presented in Figure 7.1. In summary, *80ΔC $_{1-648}$ #6, sca-GAL4* dominant phenotype was both enhanced and suppressed by deficiencies on the second and third chromosome. Out of 100 deficiencies screened, 5 independent deficiencies were found to modify the macrochaetae phenotype at 21°C. A total of 70 deficiencies were tested on the second chromosome, of which four deficiencies interacted. Whilst one interacting deficiency was identified from 30 deficiencies residing on the third chromosome. The following sections describe these interacting genetic regions in more detail along with the candidate CBC-interacting genes uncovered.

Figure 7.1 Summary of Deficiencies Modifying *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4* Macrochaetae Phenotype

Interacting Deficiency	Mapped Region	Modification
Df(2L)Dwee-delta5	27C1-27E	Suppressor
Df(2R)ST1	42B3-42C	Enhancer
Df(2R)knSA3	51C1-51D8	Enhancer
Df(2R)Pu-D17	57D12-58D1	Enhancer
Df(3L)th102	71F-72D1	Suppressor

Figure 7.1 Summary of deficiencies modifying *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae phenotype. *80ΔC $_{1-648}$ #6, sca-GAL4* recombinants were crossed to 70 deficiencies on the second chromosome and 30 deficiencies on the third chromosome. Crosses were set up at 21°C and the F1 progeny were scored for suppression or enhancement compared to *80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae phenotype and *sca-GAL4* adults alone.

7.2.2 Genetic region 27C1-27E suppresses *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae phenotype

Df(2L)Dwee-delta5 was initially found to strongly suppress *80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae phenotype at 21°C. Introducing *Df(2L)Dwee-delta5* into *80ΔC $_{1-648}$ #6, sca-GAL4* background restored the macrochaetae to the wild-type number and this was fully penetrant within the F1 population. As can be observed from Figure 7.2, A, *Df(2L)Dwee-delta5/ 80ΔC $_{1-648}$ #6, sca-GAL4* adults have all four macrochaetae present on the central notum and scutellum (compare [i] and [iii]). However, the distal pair of macrochaetae on the notum and scutellum were approximately 10% shorter and thinner than the GAL4 control (compare [ii] and [iii]).

Df(2L)Dwee-delta5 uncovers polytene segments 27A; 28A on the second chromosome. To determine whether this interacting region could be further defined, smaller overlapping deficiencies listed in Appendix D were tested for interaction with *80ΔC $_{1-648}$ #6, sca-GAL4* at 21°C. This demonstrated that like *Df(2L)Dwee-delta5*, *Df(2L)J-H* and *Df(2L)spd^{i2}* also suppressed the dominant macrochaetae phenotype. As a result, the interacting genetic region was further mapped to 27C1-27E (Figure 7.2, B). These deficiencies did not enhance or suppress *sca-GAL4* at 21°C. In addition, the deficiencies did not modify *UAS-atonal, S31-GAL4* dominant scutellar macrochaetae phenotype at 25°C (data not shown). This suggests that the modification observed was due to a specific interaction between the genetic interacting region 27C1-27E and *80ΔC $_{1-648}$*.

At the time, cytosearches in Flybase revealed a number of genes that resided within the interacting region 27C1-27E. To determine whether the suppression could be reproduced and the putative CBC-interacting gene identified, all the Bloomington stocks available of known genes along with randomly selected P-insertions in unknown genes were crossed with *80ΔC $_{1-648}$ #6, sca-GAL4* at 21°C (Appendix E).

Figure 7.2 Genetic Region 27C1-27E Suppresses *UAS-DmCBP80ΔC $_{1-648}$ #6, sca* Phenotype

Figure 7.2 Genetic region 27C1-27E suppresses *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae phenotype. A. [i] At 21°C, *80ΔC $_{1-648}$ #6, sca-GAL4* adult has two out of four distal scutellar macrochaetae whilst three out of four macrochaetae on central notum are absent. **[ii]** In *sca-GAL4/ +* control, four macrochaetae on central notum and scutellum are present. For ease of comparison the same control images [i] and [ii] are used in Figures 7.2 to 7.7. **[iii]** Crossing *Df(2L)Dwee-delta5* into *80ΔC $_{1-648}$ #6, sca-GAL4* background restores all four macrochaetae on central notum and scutellum. **B.** *80ΔC $_{1-648}$ #6, sca-GAL4* interacting region was further mapped to 27C1-27E. Red rectangles indicate deficiencies that suppressed *80ΔC $_{1-648}$ #6, sca-GAL4* phenotype. White rectangles depict deficiencies that did not modify *80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae. Black lines indicate possible deficiency breakpoints.

The genetic region 27C1-27E uncovered 5 genes with roles implicated in RNAP II transcript metabolism; *snRNP70K*, *Hrb27C*, *x16*, *nop5* and *CC1.3*. *snRNP70K* encodes a U1-specific protein, *Hrb27C* is a hnRNP whilst *x16* and *CC1.3* are predicted splicing factors. Characterisation of CBC has shown it facilitates the interaction between U1 snRNP and the cap-proximal 5' splice site (Lewis *et al.*, 1996). In higher eukaryotes, the molecules that mediate CBC function during splicing are currently unknown. Therefore, *snRNP70K*, *Hrb27C*, *x16* or *CC1.3* might be possible adaptor proteins that facilitate the effects of CBC and the cap in spliceosomal assembly. *Nop5* encodes a snoRNA involved in rRNA processing. Although the cap contributes to the export of RNAP II transcribed snoRNAs, the cap-binding activity that mediates the cap's effect has not been identified (Jacobson and Pederson, 1998). It is therefore possible that this effect is mediated by CBC. As a result, mutant alleles were obtained for each of these gene except *CC1.3* and tested for interaction with *80ΔC $_{1-648}$ #6, sca-GAL4* at 21°C. However, none of these alleles were found to modify *80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae phenotype (Appendix E and data not shown).

7.2.3 Genetic region 42B3-42C enhances *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae phenotype

In contrast to 27C1-27E , *Df(2R)ST1* was identified as enhancing *80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae phenotype at 21°C. Furthermore, the enhancement was fully penetrant within the F1 population. Figure 7.3, A shows that adults containing *Df(2R)ST1* in combination with *80ΔC $_{1-648}$ #6, sca-GAL4* have severely reduced distal macrochaetae on the scutellum (compare [i], [ii] and [iii]). *Df(2R)ST1* uncovers polytene segments 42B3-5; 42E15-18 on chromosome 2. Further mapping of this genetic region using the deficiencies listed in Appendix D narrowed the interacting region to 42B3-42C (Figure 7.3, B). *Df(2R)ST1* did not modify *sca-GAL4* alone at 21°C, nor *UAS-atonal, S31-GAL4* dominant scutellar macrochaetae phenotype at 25°C (data not shown).

Figure 7.3 Genetic Region 42B3-42C Enhances *UAS-DmCBP80ΔC $_{1-648}$ #6, sca* Phenotype

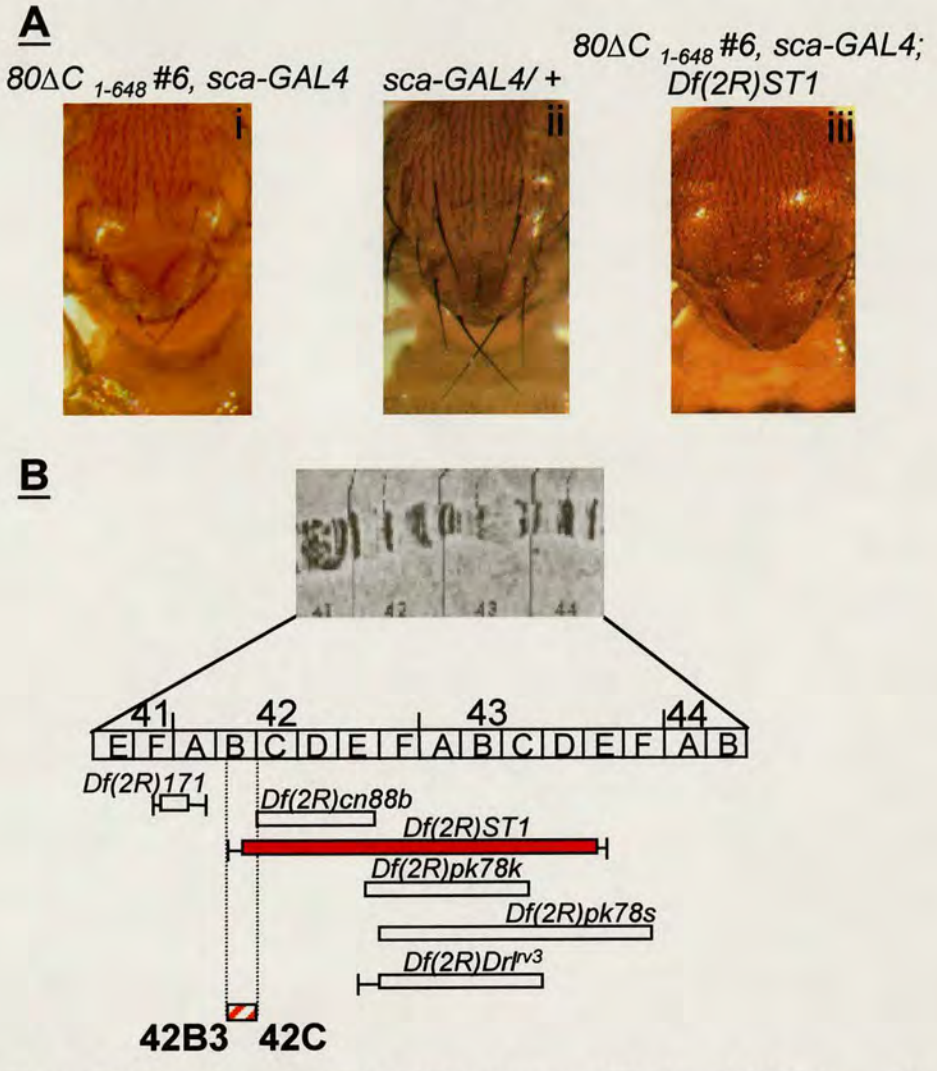


Figure 7.3 Genetic region 42B3-42C enhances *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae phenotype. A. [i] At 21°C, *80ΔC $_{1-648}$ #6, sca-GAL4* adult has two out of four distal scutellar macrochaetae whilst three out of four macrochaetae on central notum are absent. **[ii]** In *sca-GAL4/ +* control, four macrochaetae on central notum and scutellum are present. For ease of comparison the same control images [i] and [ii] are used in Figures 7.2 to 7.7. **[iii]** Crossing *Df(2R)ST1* into *80ΔC $_{1-648}$ #6, sca-GAL4* background severely reduced the two distal scutellar macrochaetae. **B.** *80ΔC $_{1-648}$ #6, sca-GAL4* interacting region was further mapped to 42B3-42C. Red rectangles indicate deficiencies that suppressed *80ΔC $_{1-648}$ #6, sca-GAL4* phenotype. White rectangles depict deficiencies that did not modify *80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae. Black lines indicate possible deficiency breakpoints.

This suggests that the enhancement of *80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae phenotype resulted from a specific interaction between *80ΔC $_{1-648}$* and the genetic region 42B3-42C.

Cytosearches of the genetic region 42B3-42C showed no obvious putative CBC-interacting genes with functions in RNAP II transcript processing. Two Bloomington P-insertion stocks were available for the genes *jing* and *Adf1* that are predicted to encode RNAP II transcription factors. As a result, these alleles along with 3 randomly selected Bloomington P-insertions in yet uncharacterised genes were obtained and tested for interaction with *80ΔC $_{1-648}$ #6, sca-GAL4* at 21°C (Appendix E). However, no genetic interaction was observed (data not shown).

7.2.4 Genetic region 51C1-51D8 enhances *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae phenotype

Df(2R)knSA3 was found to strongly enhance *80ΔC $_{1-648}$ #6, sca-GAL4* dominant macrochaetae at 21°C. As can be observed from Figure 7.4, A, crossing *Df(2R)knSA3* into *80ΔC $_{1-648}$ #6, sca-GAL4* background increased the severity of the dominant phenotype. In *Df(2R)knSA3/ 80ΔC $_{1-648}$ #6, sca-GAL4* adults, both the distal scutellar macrochaetae and sockets were notably absent at 21°C (compare [i], [ii] and [iii]). *Df(2R)knSA3* uncovers polytene segments 51B5-11; 51D7-E2 on the second chromosome. The interacting genetic region was further mapped to 51C1-51D8 using surrounding deficiencies listed in Appendix D (Figure 7.4, B). *Df(2R)knSA3* did not modify *sca-GAL4* alone at 21°C, nor *UAS-atonal, S31-GAL4* dominant scutellar macrochaetae phenotype at 25°C (data not shown). This suggests that the enhancement of *80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae resulted from a specific interaction between *80ΔC $_{1-648}$* and the genetic region 51C1-51D8.

Figure 7.4 Genetic Region 51C1-51D8 Enhances *UAS-DmCBP80ΔC $_{1-648}$ #6, sca* Phenotype

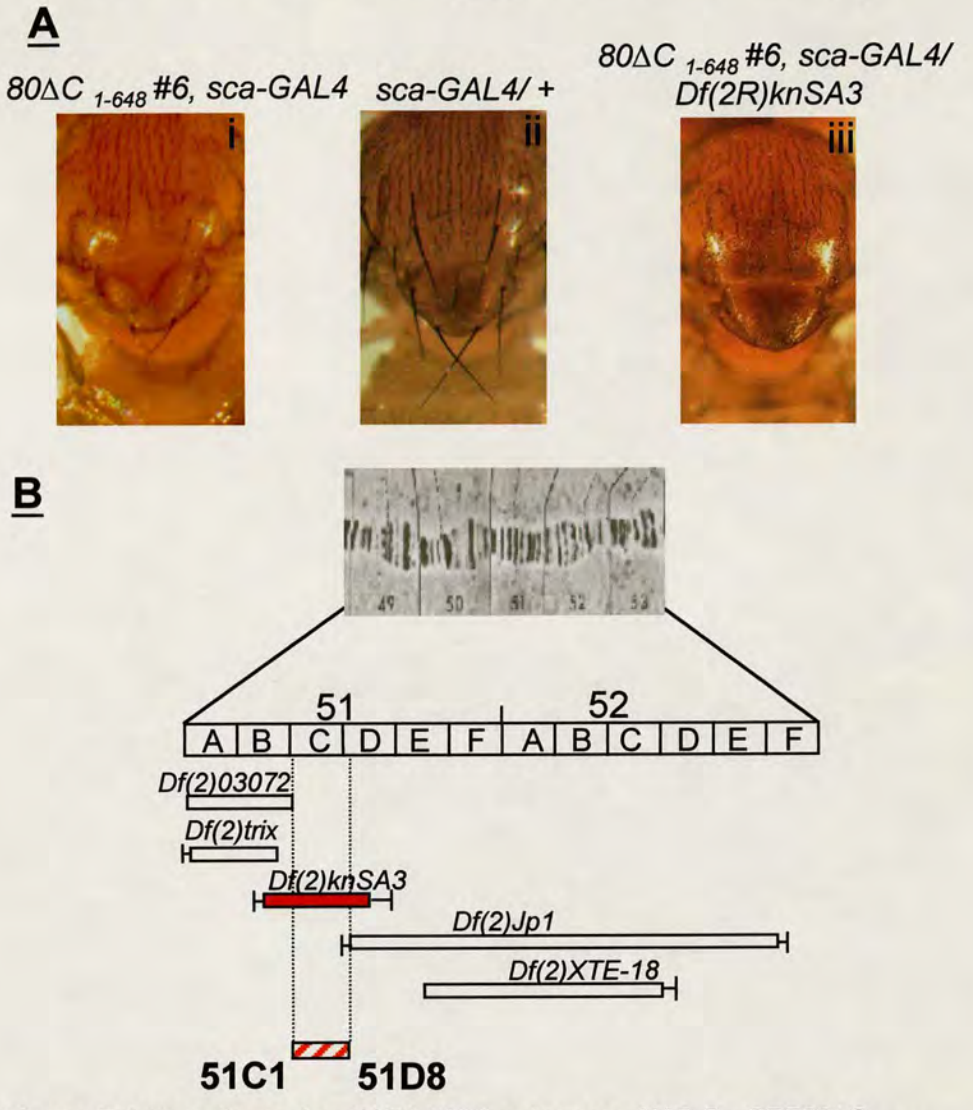


Figure 7.4 Genetic region 51C1-51D8 enhances *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae phenotype. A. [i] At 21°C, *80ΔC $_{1-648}$ #6, sca-GAL4* adult has two out of four distal scutellar macrochaetae whilst three out of four macrochaetae on central notum are absent. **[ii]** In *sca-GAL4/ +* control, four macrochaetae on central notum and scutellum are present. For ease of comparison the same control images [i] and [ii] are used in Figures 7.2 to 7.7. **[iii]** Crossing *Df(2R)knSA3* into *80ΔC $_{1-648}$ #6, sca-GAL4* background results in the absence of the two distal scutellar macrochaetae. **B.** *80ΔC $_{1-648}$ #6, sca-GAL4* interacting region was further mapped to 51C1-51D8. Red rectangles indicate deficiencies that suppressed *80ΔC $_{1-648}$ #6, sca-GAL4* phenotype. White rectangles depict deficiencies that did not modify *80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae. Black lines indicate possible deficiency breakpoints.

Cytosearches in Flybase revealed a number of genes that resided within 51C1-51D8. The most interesting candidate CBC-interacting gene was *cleavage factor IA* that is predicted to function in the cleavage step of RNA 3'-end formation. Given the role of CBC in RNA 3'-end processing, CBC might interact with *cleavage factor IA* to facilitate RNA cleavage (Flaherty *et al.*, 1997). However, no mutant alleles were available in *cleavage factor IA*. As a result, the available alleles of genes that mapped to 51C1-51D8 were crossed with *80ΔC $_{1-648}$ #6, sca-GAL4* at 21°C and the F1 progeny screened for modification (Appendix E). These included RNAP II transcription factor *Kn*, endopeptidase regulator *Rpn6*, *sf* and *boc* that do not currently have predicted functions. However, no genetic interaction with *80ΔC $_{1-648}$ #6, sca-GAL4* was observed (data not shown).

7.2.5 Genetic region 57D12-58D1 enhances *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae phenotype

Df(2R)Pu-D17 was also identified as enhancing *80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae phenotype at 21°C. Figure 7.5, A shows that adults containing *Df(2R)Pu-D17* in combination with *80ΔC $_{1-648}$ #6, sca-GAL4* lack the two distal scutellar macrochaetae and sockets at 21°C (compare [i], [ii] and [iii]). *Df(2R)Pu-D17* uncovers polytene segments 57B4; 58B on chromosome 2. To determine whether this interacting region could be further defined, smaller overlapping deficiencies listed in Appendix D were tested for interaction with *80ΔC $_{1-648}$ #6, sca-GAL4* at 21°C. This demonstrated that like *Df(2R)Pu-D17*, *Df(2R)PK1* and *Df(2R)Egfr5* also enhanced the dominant macrochaetae phenotype (Figure 7.5, B and data not shown). As a result, the interacting genetic region was mapped to 57D12-58D1. These deficiencies did not enhance or suppress *sca-GAL4* at 21°C. In addition, the deficiencies did not modify *UAS-atonal, S31-GAL4* dominant scutellar macrochaetae phenotype at 25°C (data not shown). This suggests that the modification observed was due to a specific interaction between the genetic region 57D12-58D1 and *80ΔC $_{1-648}$*.

Figure 7.5 Genetic Region 57D12-58D1 Enhances *UAS-DmCBP80ΔC $_{1-648}$ #6, sca* Phenotype

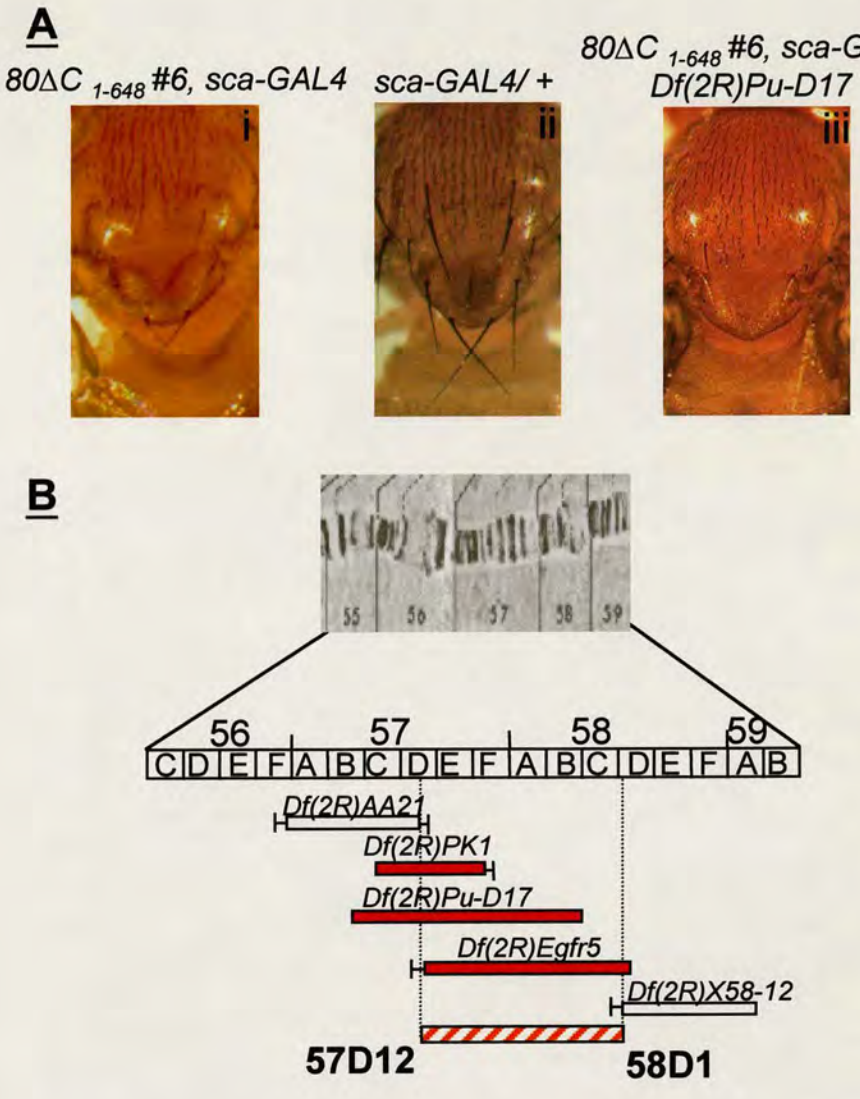


Figure 7.5 Genetic region 57D12-58D1 enhances *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae phenotype. A. [i] At 21°C, *80ΔC $_{1-648}$ #6, sca-GAL4* adult has two out of four distal scutellar macrochaetae whilst three out of four macrochaetae on central notum are absent. **[ii]** In *sca-GAL4/ +* control, four macrochaetae on central notum and scutellum are present. For ease of comparison the same control images [i] and [ii] are used in Figures 7.2 to 7.7. **[iii]** Crossing *Df(2R)Pu-D17* into *80ΔC $_{1-648}$ #6, sca-GAL4* background results in the absence of the two distal scutellar macrochaetae. **B.** *80ΔC $_{1-648}$ #6, sca-GAL4* interacting region was further mapped to 57D12-58D1. Red rectangles indicate deficiencies that suppressed *80ΔC $_{1-648}$ #6, sca-GAL4* phenotype. White rectangles depict deficiencies that did not modify *80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae. Black lines indicate possible deficiency breakpoints.

The interacting genetic region 57D12-58D1 was used in cytosearches and this revealed no genes with predicted roles in RNAP II transcript processing. However, the transcription factor *dve* along with four P-insertions in yet uncharacterised genes did reside within 57D12-58D1 and were crossed into *80ΔC $_{1-648}$ #6, sca-GAL4* background at 21°C (Appendix E). This showed that none of these alleles reproduced the enhancement of *80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae observed with *Df(2R)Pu-D17*.

7.2.6 Genetic region 71F-72D1 suppresses *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae phenotype

Df(3L)th102 was the only deficiency on the third chromosome that modified *80ΔC $_{1-648}$ #6, sca-GAL4* phenotype at 21°C. Introducing *Df(3L)th102* into *80ΔC $_{1-648}$ #6, sca-GAL4* restored the macrochatae to the wild-type number and this was fully penetrant within the F1 population. As can be observed from Figure 7.6, A, *Df(3L)th102/ 80ΔC $_{1-648}$ #6, sca-GAL4* adults have all four macrochaetae present on the central notum and scutellum (compare [i] and [iii]). However, the distal pair of macrochaetae on the notum and scutellum were approximately 20% shorter and thinner than the GAL4 control (compare [ii] and [iii]). *Df(3L)th102* uncovers polytene segments 71F3-5; 72D12. To determine whether this interacting region could be further defined, overlapping deficiencies listed in Appendix D were tested for interaction with *80ΔC $_{1-648}$ #6, sca-GAL4* at 21°C. This demonstrated that like *Df(3L)th102*, *Df(3L)brm11* also suppressed the dominant macrochaetae phenotype (Figure 7.6, B and data not shown). As a result, the interacting genetic region was further mapped to 71F-72D1. *Df(3L)th102* and *Df(3L)brm11* did not modify *sca-GAL4* alone at 21°C, nor *UAS-atonal, S31-GAL4* dominant scutellar macrochaetae phenotype at 25°C (data not shown). This suggests that the suppression of *80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae phenotype resulted from a specific interaction between *80ΔC $_{1-648}$* and the genetic region 71F-72D1.

Figure 7.6 Genetic Region 71F-72D1 Suppresses *UAS-DmCBP80ΔC $_{1-648}$ #6, sca* Phenotype

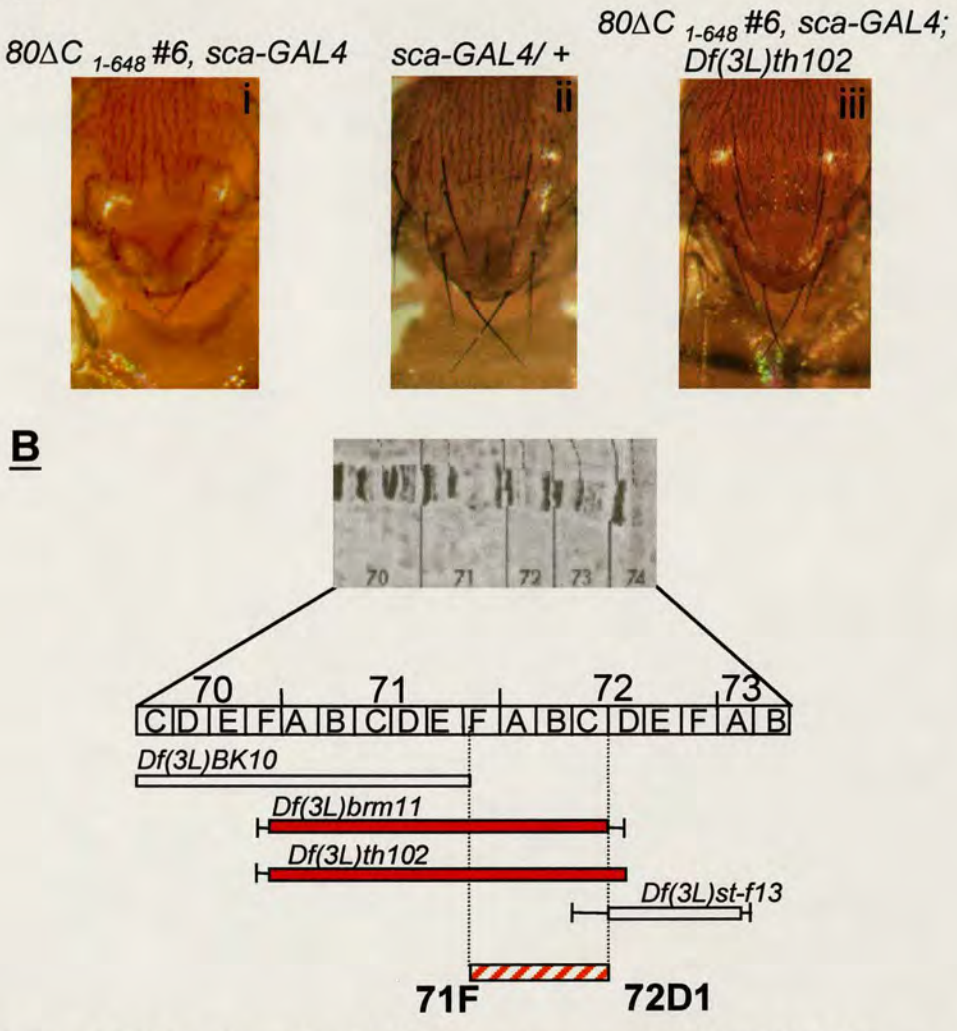


Figure 7.6 Genetic region 71F-72D1 suppresses *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae phenotype. A. [i] At 21°C, *80ΔC $_{1-648}$ #6, sca-GAL4* adult has two out of four distal scutellar macrochaetae whilst three out of four macrochaetae on central notum are absent. **[ii]** In *sca-GAL4/ +* control, four macrochaetae on central notum and scutellum are present. For ease of comparison the same control images [i] and [ii] are used in Figures 7.2 to 7.7. **[iii]** Crossing *Df(3L)th 102* into *80ΔC $_{1-648}$ #6, sca-GAL4* background restores all four macrochaetae on central notum and scutellum. **B.** *80ΔC $_{1-648}$ #6, sca-GAL4* interacting region was further mapped to 71F-72D1. Red rectangles indicate deficiencies that suppressed *80ΔC $_{1-648}$ #6, sca-GAL4* phenotype. White rectangles depict deficiencies that did not modify *80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae. Black lines indicate possible deficiency breakpoints.

Analysis of genetic region 71F-72D1 by cytosearch revealed no obvious candidate CBC-interacting genes. Stocks were available for the RNAP II transcription factor *Arf72A* and P-insertions in yet uncharacterised genes located in 71F-72D1. As a result, these alleles were tested for interaction with *80ΔC $_{1-648}$ #6, sca-GAL4* at 21°C (Appendix E). However, no modification of the macrochaetae phenotype was observed (data not shown).

In summary, the CBC-interacting genes residing in the five interacting deficiencies remain to be determined. However, the results from the deficiency screen do show that *80ΔC $_{1-648}$ #6, sca-GAL4* provides a sensitised genetic background for genetic modifier screens.

7.2.7 Targeted genetic modifier screen using *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4*

In parallel to the deficiency screen, genes that might interact with CBC were screened for interaction with *80ΔC $_{1-648}$ #6, sca-GAL4*. Together, the cap and CBC act as a molecular beacon to facilitate the recognition of capped RNA by various components of RNAP II transcript processing machinery. In particular, CBC mediates the stimulatory effects of the cap in pre-mRNA splicing, RNA 3'-end formation and U snRNA export (Lewis and Izaurralde, 1997). As a result, genes involved in these processes, as well as other aspects of RNA metabolism such as transcription and cap-dependent translation, were screened for interaction with *80ΔC $_{1-648}$ #6, sca-GAL4* (Figure 7.7 and Appendix F).

Although the completed *Drosophila* genome allows the identification of fruit-fly homologues, the availability of mutant alleles in the targeted genes was limited. Consequently, most of the mutant alleles used in the genetic modifier screen constituted deficiencies and P-element insertions. Alleles of selected genes were crossed into *80ΔC $_{1-648}$ #6, sca-GAL4* background at 21°C. The resultant F1 progeny were then scored for suppression or enhancement compared to *80ΔC $_{1-648}$ #6, sca-GAL4* and *sca-GAL4* adults alone. The following sections describes the genes tested for interaction with *80ΔC $_{1-648}$ #6, sca-GAL4* and the reasons why they were targeted.

Figure 7.7 Summary of Proteins Targeted for Interaction with *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4*

Protein	Function
RNAP II 215	Transcription, largest subunit of RNAP II.
RNAP II 140	Transcription, 2nd largest subunit of RNAP II.
CTD-binding proteins	Facilitate the coupling of transcription with nuclear RNA processing?
eIF4E	Cap-binding protein of eIF4F translation initiation complex.
eIF4E-binding protein	Translation, regulate eIF4E activity.
eIF4E-related proteins	Translation ?
eIF4G	Component of eIF4F cap-dependent translation initiation complex.
eIF4A	RNA helicase component of eIF4F cap-dependent translation initiation complex.
CPSF	RNA 3'-end formation, heterotetrameric complex comprised of 160, 100, 73 and 30 kDa subunits. Participates in cleavage and polyadenylation reactions.
CstF	RNA 3'-end formation, heterotrimeric complex comprised of 79, 50 and 64 kDa subunits. Participates in cleavage reaction.
U1 snRNAs	Splicing, U1 snRNA non-coding proteins.
SmD3	Splicing factor, U snRNA core Sm protein.
Sans fille	Splicing factor, homologue of both U1 snRNP-U1A and U2 snRNP-U2B".
Luc7-like	Splicing factor, homologue of yeast U1 snRNP-specific protein Luc7. Participates in commitment complex assembly.
Luc7-related	Splicing factor, homologue of yeast U1 snRNP-specific protein Luc7. Participates in commitment complex assembly.
U2AF 50	Splicing factor, homologue of human U2AF65. Binds polypyrimidine tract to recruit U2 snRNP during A complex.
U2AF 38	Splicing factor, homologue of human U2AF35. Binds polypyrimidine tract to recruit U2 snRNP during A complex.
Noisette	Splicing factor, homologue of human 60 kDa subunit of heterotrimeric SF3a.
B52	Splicing factor, homologue of human SRp55.
Hrb 87F	hnRNP RNA-binding protein.
Pendulin	Nuclear protein transport, importin α homologue of yeast srp1.
Karyopherin α1	Nuclear protein transport, importin α homologue of yeast srp1.
Crm1	U snRNA export receptor, homologue of human Xpo1.
Diminished-7	Ran-binding protein.
Bx34	Nucleoporin, homologue of human TPR.

Figure 7.7 Summary of proteins targeted for interaction with *UAS-DmCBP80ΔC $_{1-648}$, #6, sca-GAL4.* Proteins involved in RNAP II transcript metabolism were selected for the genetic modifier screen using *80ΔC $_{1-648}$ #6, sca-GAL4*. These included proteins with functions in transcription, pre-mRNA splicing, RNA 3'-end formation, RNA export and cap-dependent translation. See text for more detail. Crosses where set up at 21°C and the F1 progeny were scored for for suppression or enhancement compared to *80ΔC $_{1-648}$ #6, sca-GAL4* and *sca-GAL4* adults alone.

CBC's function in pre-mRNA splicing has been shown to be evolutionarily conserved between humans and yeast. CBC increases the efficiency of U1 snRNP binding to the cap-proximal 5' splice site. This enhances the recognition of the 5' splice site and subsequent commitment or E complex assembly. As a result, CBC facilitates efficient pre-mRNA splicing (Colot *et al.*, 1996; Lewis *et al.*, 1996). In vertebrates, little is currently known about the mechanism of CBC action. However, genetic studies in *S. cerevisiae* have identified a number of splicing factors, most of which are components of U1 snRNP, that interact with CBC (Fortes *et al.*, 1999a and 1999b). Therefore, *Drosophila* homologues of both U1 snRNAs and U1 snRNP-specific proteins were targeted for interaction with *80ΔC $_{1-648}$ #6, sca-GAL4* at 21°C.

The *Drosophila* genome contains five genes that encode U1 snRNAs. No mutant alleles were available for these genes. Consequently, small deficiencies that uncover the predicted gene loci of the five *U1 snRNAs* were crossed with *80ΔC $_{1-648}$ #6, sca-GAL4* at 21°C. F1 progeny that carried both alleles were then scored for suppression or enhancement compared to *80ΔC $_{1-648}$ #6, sca-GAL4* and *sca-GAL4* adults alone. This showed that *80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae phenotype was not modified by deficiencies that uncovered *U1 snRNA* genes (Appendix F and data not shown).

Each U snRNP contains a set of Sm core proteins shared with other snRNPs and a set of U snRNP-specific proteins. Genetic studies in *S. cerevisiae* have demonstrated synthetic lethal interactions between CBC and the Sm core protein SmD3, the U1 snRNP-specific proteins Luc7, Mud1, Snu56, Snu71 and Nam8. Moreover, Snu56 has been shown to interact directly with CBC in yeast (Fortes *et al.*, 1999a and 1999b). *Drosophila* homologues have been identified for SmD3, Luc7 and Mud1/ U1A proteins. Interestingly, Luc7 appears to have two homologues in fruit-flies, Luc7 and Luc7-related. Whilst the *Drosophila* protein sans fille corresponds to both vertebrate U1 snRNP-U1A and U2 snRNP-U2B'' proteins (Polycarpou-Schwarz *et al.*, 1996). In contrast, Snu56, Snu71 and Nam8 U1 snRNP-specific proteins do not appear to have clear counterparts in the *Drosophila* database. As a result,

deficiencies that uncovered *SmD3*, *Luc7*, *Luc7-related* and mutant alleles in *sans fille* were crossed with *80ΔC $_{1-648}$ #6, sca-GAL4* at 21°C. F1 progeny were then scored for modification compared to *80ΔC $_{1-648}$ #6, sca-GAL4* and *sca-GAL4* adults alone. This determined that *80ΔC $_{1-648}$ #6, sca-GAL4* dominant phenotype was not modified by the U1 snRNP-specific mutants tested (Appendix F and data not shown).

During commitment or E complex assembly, U1 snRNP associates with the 5' splice site and Mud2/U2AF65-35 binds to the polypyrimidine tract. In *S. cerevisiae*, a genetic and physical interaction has been identified between CBC and Mud2 (Fortes *et al.*, 1999a and 1999b). The *Drosophila* homologue of Mud2 is U2AF50, which forms a heterodimer with U2AF38. P-insertion and deletion alleles were obtained for the components of U2AF heterodimer and tested for interaction with *80ΔC $_{1-648}$ #6, sca-GAL4* at 21°C. However, none of these alleles modified the dominant macrochaetae phenotype in comparison to *80ΔC $_{1-648}$ #6, sca-GAL4* and *sca-GAL4* controls alone (Appendix F and data not shown).

Interestingly, Mud2 is not required for facilitating the interaction between U1 snRNP and 5' splice site but promotes the binding of U2 snRNP for A complex assembly. This suggests that the role of CBC in splicing might extend beyond the U1 snRNP function and CBC might participate at later stages in the splicing cycle (O'Mullane and Eperon, 1998). Therefore, the U2 snRNP-specific protein, noisette and splicing factor, B52 both of which function during the early stages of spliceosome assembly were also tested for interaction with *80ΔC $_{1-648}$ #6, sca-GAL4* at 21°C. This showed the dominant macrochaetae phenotype was not enhanced or suppressed by mutants in *noisette* or *B52* (Appendix F and data not shown).

A role for CBC has also been described in RNA 3'-end formation. During 3'-end processing, CBC facilitates the positive effect of the cap in RNA cleavage (Flaherty *et al.*, 1997). However, the molecular mechanism by which CBC mediates this effect has not been elucidated. To determine whether CBC interacts with components of the cleavage complex, *80ΔC $_{1-648}$ #6, sca-GAL4* was crossed with various alleles of CPSF and CStF at 21°C (Appendix F). In vertebrates, CPSF is a heterotetrameric

complex that binds the cleavage and polyadenylation sequence and functions in both RNA cleavage and polyadenylation reactions. Whilst CStF is a heterotrimer that associates with the GU-rich motif and participates in RNA cleavage only. P-insertion and deletion mutants were available for the largest subunits of both CPSF and CstF complexes (CPSP-160 and CStF-77), along with deficiencies that uncovered the predicted genetic loci of the remaining subunits. However, none of the mutants in CPSF or CstF modified the macrochaetae phenotype in comparison to *80ΔC $_{1\text{-}648}$ #6, sca-GAL4* and *sca-GAL4* controls (Appendix F and data not shown).

Characterisation of CBC has shown that it also mediates the cap's effect in U snRNA export. CBC and the adaptor molecule PHAX are both necessary and sufficient for Xpo1 mediated U snRNA export (Fornerod *et al.*, 1997; Ohno *et al.*, 2000). CBC is then thought to be recycled back to the nucleus via the importin pathway. The NLS in CBP80 appears to bind the adaptor srp1/ importin α, which in combination with importin β, transports cytoplasmic CBC to the nucleus (Gorlich *et al.*, 1996). The *Drosophila* homologue of Xpo1 is Crm1 and homologues of the importin α family include karyopherin α1 and pendulin. Since no mutant alleles were available for these genes, small deficiencies that uncovered the predicted loci were crossed to *80ΔC $_{1\text{-}648}$ #6, sca-GAL4* at 21°C. F1 progeny that carried both alleles were then scored for suppression or enhancement compared to *80ΔC $_{1\text{-}648}$ #6, sca-GAL4* and *sca-GAL4* adults alone. This showed the macrochaetae phenotype was not modified by deficiencies that uncovered *Crm1*, *karyopherin α1* or *pendulin* genes (Appendix F and data not shown).

After RNAP II transcripts have been processed in the nucleus, they are exported to the cytoplasm for translation. In order to initiate cap-dependent translation, CBC must exchange with the cytoplasmic cap-binding protein eIF4E. This allows the heterotrimeric translation initiation complex, eIF4F, to form at the cap. Although the mechanism for the exchange of cap-binding proteins is unknown, it is thought to be mediated through eIF4G (Fortes *et al.*, 2000; McKendrick *et al.*, 2001). As a result, the components of eIF4F were screened for genetic modification with *80ΔC $_{1\text{-}648}$ #6, sca-GAL4*. In addition, proteins that regulate cap-dependent translation such as

eIF4E-binding protein were targeted. eIF4E-binding proteins specifically control the initiation of cap-dependent translation through binding to eIF4E. Although there is a family of eIF4E-binding proteins in vertebrates, the *Drosophila* genome appears to only contain a single gene encoding one eIF4E-binding protein homologue. P-insertion and deletion mutants were obtained for these translation factors and crossed with *80ΔC $_{1-648}$#6, sca-GAL4* at 21°C. F1 progeny were then scored for modification compared to *80ΔC $_{1-648}$#6, sca-GAL4* and *sca-GAL4* adults alone. This determined that *80ΔC $_{1-648}$#6, sca-GAL4* dominant phenotype was not modified by the translation mutants tested (Appendix F and data not shown).

Taken together, the results show that none of the mutants in pre-mRNA splicing, RNA 3'-end formation, RNA export or cap-dependent translation factors modified *80ΔC $_{1-648}$#6, sca-GAL4* dominant phenotype at 21°C. However, the following sections describe how the macrochaetae phenotype was suppressed by an allele of the largest subunit of RNAP II and a deficiency that uncovered the CTD-binding protein, cleavage factor 1A.

7.2.8 *RNAP II 215^{W8}* suppresses *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae phenotype

Many lines of evidence have shown that RNAP II transcription is intimately coupled *in vivo* to nuclear mRNA processing events such as capping, pre-mRNA splicing and RNA 3'-end formation (Hirose and Manley, 2000). Given the role CBC plays in pre-mRNA splicing and RNA 3'-end formation, it was reasoned that CBC might also interact with the CTD to facilitate these processes (see Chapter 1). Therefore, genes tested for interaction with *80ΔC $_{1-648}$#6, sca-GAL4* included *RNAP II 215* and *RNAP II 140* that encode the two largest subunits of RNAP II. In particular, *RNAP II 215* was targeted since its CTD specifically couples transcription with RNA processing. A number of mutants were available in *RNAP II 215* and *RNAP II 140* that included deletion and EMS-induced alleles. These alleles were crossed with *80ΔC $_{1-648}$#6,*

sca-GAL4 and the F1 progeny screened for genetic modification at 21°C (Appendix F).

Surprisingly, one allele of the largest RNAP II subunit, *RNAP II 215* ^{W8} was found to strongly suppress *80ΔC* $_{1-648}$*#6, sca-GAL4* phenotype at 21°C. *RNAP II 215* ^{W8} is a recessive lethal caused by a 631 bp insertion of residual P-element sequences into CTD heptad repeat #20. This generates a RNAP II 215 subunit with a premature stop codon. As a result, the RNAP II 215 subunit contains a partially truncated CTD with only 19 of the normal 42 CTD heptad repeats (Brickey and Greenleaf, 1995). Crossing *RNAP II 215* ^{W8} into *80ΔC* $_{1-648}$ *#6, sca-GAL4* background strongly suppressed the macrochaetae phenotype to wild-type. As can be observed from Figure 7.8, A, all four macrochaetae on the central notum and scutellum were restored. However, these marcrochaetae appeared to be approximately 16% shorter and thinner than the GAL4 control (compare panels [i]-[iii]). In contrast, introducing other *RNAP II 215* alleles only resulted in a mild suppression of the dominant phenotype. This is illustrated in Figure 7.8, A where crossing a null allele of *RNAP II 215* (*RNAP 215* ^{NULL}) restored one of the four macrochaetae on the central notum along with three out of four scutellar macrochaetae. However these macrochaetae were approximately 55% shorter than the GAL4 control (compare panels [i]-[iv]).

To determine whether the suppression observed was specific to *80ΔC* $_{1-648}$ *#6, sca-GAL4, RNAP II 215* alleles were crossed with an UAS-GAL4 control. It was decided that *UAS-atonal, S31-GAL4* would not be an ideal control since *atonal* is a RNAP II transcription factor. Therefore, *RNAP II 215* alleles might modify *UAS-atonal, S31-GAL4* dominant macrochaetae phenotype for legitimate reasons. As expected, crossing the *RNAP II 215* alleles into *UAS-atonal, S31-GAL4* background mildly suppressed the scutellar macrochaetae phenotype at 25°C (Appendix F and data not shown). Instead, *UAS-importin β, GMR-GAL4* was selected as an alternative UAS-GAL4 control. *Importin β* encodes a protein import receptor that binds cargo carrying a classical NLS through the adaptor, importin α. Unlike *atonal*, *importin β* would not be expected to interact with *RNAP II 215*.

Figure 7.8 *RNAP II 215^{W8}* suppresses *UAS-DmCBP80ΔC $_{1-648}$ #6, sca* phenotype

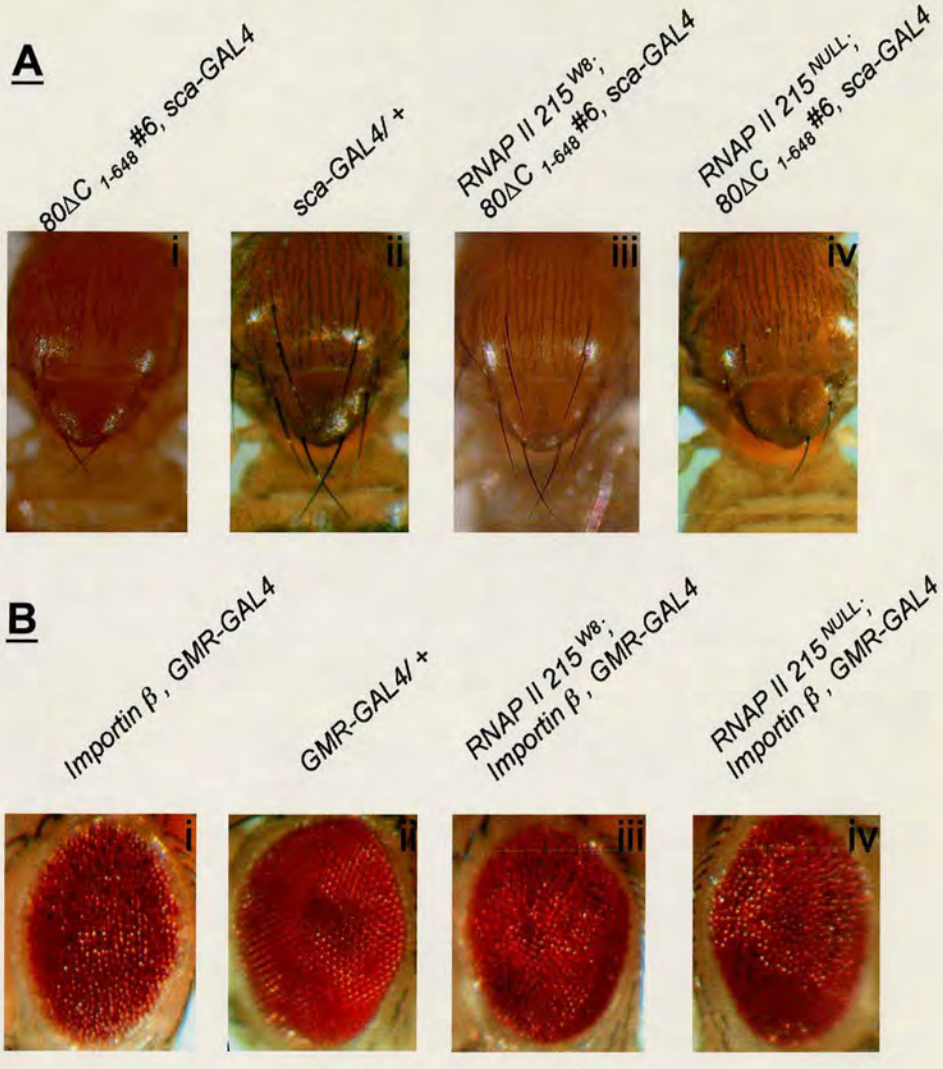


Figure 7.8 *RNAP II 215^{W8}* suppresses *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae phenotype. A. [i] At 21°C, *80ΔC $_{1-648}$ #6, sca-GAL4* adult has two out of four distal scutellar macrochaetae whilst four macrochaetae on central notum are absent. **[ii]** In *sca-GAL4/ +* control, four macrochaetae on central notum and scutellum are present. **[iii]** Crossing *RNAP II 215^{W8}* into *80ΔC $_{1-648}$ #6, sca-GAL4* background restores all four macrochaetae on central notum and scutellum. **[iv]** Crossing *RNAP II 215 ^{NULL}* into *80ΔC $_{1-648}$ #6, sca-GAL4* background partially suppresses *80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae. **B. [i]** At 25°C, *importin β , GMR-GAL4* eye is rough and blistered. **[ii]** *GMR-GAL4/ +* eye. Crossing **[iii]** *RNAP II 215^{W8}* or **[iv]** *RNAP II 215 ^{NULL}* into *importin β , GMR-GAL4* background results in mild suppression.

Figure 7.8, B illustrates the dominant eye phenotype observed in *UAS-importin β, GMR-GAL4* adults at 25°C. Typically, *UAS-importin β, GMR-GAL4* eye appears to be severely rough and blistered at 25°C (compare [i] with [ii]). Introducing various alleles of *RNAP II 215* into *UAS-importin β, GMR-GAL4* background mildly suppressed the dominant eye phenotype at 25°C. As can be observed in Figure 7.8, B *RNAP 215* ^{W8}; *UAS-importin β, GMR-GAL4* eye appears to be rough in comparison to the GAL4 control. Likewise, crossing *RNAP 215* ^{NULL} into *UAS-importin β, GMR-GAL4* background also generated a mildly rough eye. Both *RNAP II 215* ^{NULL} and *RNAP II 215* ^{W8} affect transcription, therefore the observed mild suppression might not be surprising. One possibility is that the *RNAP II 215* mutants might have reduced transcription and consequently produce lower levels of Gal4 protein. The reduction in Gal4 availability might in turn mildly suppress both *UAS-atonal, S31-GAL4* and *UAS-importin β, GMR-GAL4* dominant phenotypes. Similarly, the mild suppression observed in *RNAP 215* ^{NULL}; *80ΔC* $_{1-648}$ *#6, sca-GAL4* macrochaetae was also deemed non-specific. In comparison, the strong suppression observed in *RNAP 215* ^{W8}; *80ΔC* $_{1-648}$ *#6, sca-GAL4* adults was considered significant. Collectively, these data suggest *80ΔC* $_{1-648}$ *#6, sca-GAL4* macrochaetae phenotype is modified by *RNAP 215* ^{W8}. Moreover, this genetic interaction is specific to the CTD of RNAP II 215.

To test whether CBC interacts (directly or indirectly) with RNAP II biochemically, immunoprecipitation experiments were performed on *Drosophila* and HeLa nuclear extracts. RNAP II complexes from both extracts were immunoprecipitated with the monoclonal antibody 8WG16, which recognises non-phosphorylated epitopes on the CTD (Yuryev *et al.*, 1996). RNAP II immunoprecipitates were separated by SDS-PAGE and then analysed by western blotting. The presence of CBC was detected using anti-142[6] antibody that is directed against HsCBP20 but also cross-reacts with DmCBP20 (Izaurralde *et al.*, 1994 and Chapter 4). As a positive control, the western blot was probed with monoclonal antibody Y12 that recognises epitopes on the Sm protein family (Brahms *et al.*, 2000). However, CBC was not found to co-immunoprecipitate with RNAP II complexes in *Drosophila* nuclear extracts (data not

shown). In contrast, immunoprecipitation experiments using larger quantities (approximately 2 mgs) of HeLa nuclear extracts showed CBC was enriched in RNAP II immunoprecipitates (E. Thompson, unpublished data). This suggests a weak association between CBC and the CTD of RNAP II that might be indirect and mediated by adaptor proteins. Alternatively, the interaction between CBC and RNAP II might be direct and enhanced by CTD phosphorylation. Therefore, further work using alternative antibodies directed against different phospho-epitopes on the CTD, such as H5 and H14 that detect serine-2 and serine-5 respectively, might prove more successful in determining a biochemical interaction between CBC and RNAP II CTD.

7.2.9 Deficiency uncovering CTD-interacting protein, cleavage factor 1 suppresses *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae phenotype

Although the CTD of RNAP II couples transcription with mRNA processing, very little is currently known about the molecules that facilitate the effects of the CTD or the mechanisms involved. Yeast two-hybrid screens have identified a family of proteins that specifically interact with the CTD (Yuryev *et al.*, 1996; Bourquin *et al.*, 1997; Tanner *et al.*, 1997). These proteins resemble the SR splicing factors and have been designated SR-like CTD associated factors (SCAFs). SCAFs interact with the CTD through a distinct CTD-interacting domain (CID). To date, two types of CID have been characterised. Type 1 CID is comprised of approximately 80 conserved amino acids located at the carboxyl-terminus. In comparison, type 2 CID has around 120 amino acid conserved region at the amino-terminus. Although both sets of CID are unrelated in amino acid sequence, they are probably related structurally.

In order to identify *Drosophila* proteins that contain a CID, the peptide sequences of type 1 and type 2 CID were used in BLASTP searches of the BDGP database (Joe Lewis). This identified five putative CID genes that are listed in Appendix F. No mutant alleles were available for any of the CID genes. As a result, small deficiencies that uncovered these genes were crossed into *80ΔC $_{1-648}$ #6, sca-GAL4*

background at 21°C. Interestingly, *Df(2)knSA3* that uncovers the CID gene *CG10228*, was already identified as a *80ΔC* $_{1\text{-}648}$ *#6, sca-GAL4* interacting deficiency. *Df(2)knSA3* was found to strongly suppress *80ΔC* $_{1\text{-}648}$ *#6, sca-GAL4* macrochaetae phenotype at 21°C and the genetic interacting region was mapped to 51C1-51D8 (Figure 7.4). *CG10228* is predicted to encode cleavage factor 1A, which is the homologue of *S. cerevisiae* pcf11. *Cleavage factor 1A* is located to 51D1-2 and like CBC, is predicted to function in the cleavage step of RNA 3'-end formation (Barilla *et al.*, 2001; Flaherty *et.al.,* 1997; Gross and Moore, 2001). Taken together, these data suggest that *cleavage factor 1A* is a strong candidate CBC-interacting gene that resides in genetic interacting region 51C1-51D8.

7.3 DISCUSSION

7.3.1 *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4* is a useful tool for genetic modifier screens

The deficiency screen showed that *80ΔC $_{1-648}$ #6, sca-GAL4* dominant phenotype is sensitive to modification. The macrochaetae phenotype is both enhanced or suppressed by a total of five deficiencies from the second and third chromosomes. Consequently, *80ΔC $_{1-648}$ #6, sca-GAL4* dominant phenotype provides a sensitised genetic background for performing genetic modifier screens. Together with being consistent within a population and easily detected at 21°C, the macrochaetae phenotype fulfils all three criteria required for use in genetic modifier screens. This suggests *80ΔC $_{1-648}$ #6, sca-GAL4* is an useful tool by which to identify CBC-interacting genes and could be applied to large-scale genetic screens using P-element collections.

Although Flybase cytosearches identified candidate genes with predicted functions in RNA processing within the interacting deficiencies, very few mutant alleles were available. It was therefore not possible to directly test these putative CBC-interacting genes for interaction with *80ΔC $_{1-648}$ #6, sca-GAL4*. For example, genetic interacting region 51C1-51D8 uncovers *cleavage factor 1A*, which is the *Drosophila* homologue of *S. cerevisiae* pcf11. Characterisation of pcf11 has shown it participates in the cleavage step of RNA 3'-end formation. Since CBC also participates in RNA 3'-end cleavage, it can be reasoned that CBC might facilitate the stimulatory effect of the cap through interactions with cleavage factor 1A. However, in the absence of any direct genetic data this is purely hypothetical. As a result, the identity of the *80ΔC $_{1-648}$ #6, sca-GAL4* interacting genes that reside within the five deficiency mapped regions remains to be determined.

7.3.2 *RNAPII 215* ^{W8} suppresses *UAS-DmCBP80ΔC* $_{1-648}$ *#6, sca-GAL4* macrochaetae phenotype

Chapter 6 has already discussed how *80ΔC* $_{1-648}$ truncation removes a significant portion of a prominent carboxyl-terminal coiled-coil that protrudes from CBP80. In CBC-cap bound mRNA, CBP80 is thought to serve as a platform for the assembly of various components of the RNAP II transcript processing/ export machinery (Mazza *et al.*, 2001). Key roles for CBC have been characterised in cap-dependent splicing, RNA 3'-end formation and U snRNA export (Lewis and Izaurralde, 1997). In addition, recent work in mammalian cells has also suggested that CBC functions during a pioneer round of cap-dependent translation where newly synthesised proteins are scanned for premature stop codons and therefore subject to NMD (Fortes *et al.*, 2000; Ishigaki *et al.*, 2001). One possibility is that the over-expression of *80ΔC* $_{1-648}$ truncation removes the distinct coiled-coil surface required for interactions with different RNA processing/ export/ pioneer round of cap-dependent translation factors. As a result, RNAP II transcripts are not subject to efficient cap-dependent splicing, RNA 3'-end formation or NMD. RNAs are inefficiently processed and translated, which generates the dominant macrochaetae phenotype observed in *80ΔC* $_{1-648}$ *#6, sca-GAL4* adults.

Results from the targeted genetic modifier screen show *80ΔC* $_{1-648}$ *#6, sca-GAL4* macrochaetae phenotype is strongly suppressed by an allele of RNAP II 215 subunit. *RNAP II 215* ^{W8} is a recessive embryonic lethal of the largest subunit of RNAP II. *RNAP II 215* ^{W8} contains a partially truncated CTD caused by a 631 bp insertion of residual P-element sequences into CTD heptad repeat #20, which generates a premature stop codon. As a result, RNAP II 215 subunit has only 19 of the normal 42 CTD heptad repeats (Brickey and Greenleaf, 1995). Interestingly, *80ΔC* $_{1-648}$ *#6, sca-GAL4* macrochaetae phenotype is not significantly modified by *RNAP II 215* EMS or deletion alleles that mutate regions upstream of the CTD. Nor is the dominant phenotype modified by a null allele in the second largest RNAP II subunit, *140*. Collectively, these data suggest a genetic interaction exists between CBC and the

largest subunit of RNAP II. Furthermore, this interaction appears to be specific to the CTD of RNAP II 215.

Characterisation of *RNAP II 215* ^{W8}*/ +* adults show the mRNA steady state levels from the *RNAP II 215* ^{W8} allele are much lower than wild-type. As a result, *RNAP II 215* ^{W8} protein subunit is under-represented and the majority of RNAP II in *RNAP II 215* ^{W8}*/ +* adults contain a wild-type CTD (Brickey and Greenleaf, 1995). *RNAP II 215* ^{W8} contains a premature stop codon approximately 500 nucleotides upstream of the normal termination site. This strongly suggests that the low steady state levels of RNAP II 215 ^{W8} mRNA result from transcript degradation by NMD (Brickey and Greenleaf, 1995). One possibility is that in *80ΔC* $_{1\text{-}648}$*#6, sca-GAL4* adults, mRNAs might undergo defective pioneer translation initiation and therefore are not subjected to efficient NMD. As a result, in *RNAP II 215* ^{W8}*; 80ΔC* $_{1\text{-}648}$*#6, sca-GAL4* adults, there is inefficient detection and degradation of the premature stop codon in RNAP II 215 ^{W8} transcripts. This leads to increased steady state levels of RNAP II 215 ^{W8} mRNAs that are translated and incorporated into RNAP II holoenzymes. Therefore, *RNAP II 215* ^{W8}*; 80ΔC* $_{1\text{-}648}$*#6, sca-GAL4* flies might have catalytically active RNAP II that contains a truncated CTD. Although purely speculative, this hypothesis could be initially tested by using northern analysis to compare the steady state levels of RNAP II 215 ^{W8} transcripts in *RNAP II 215* ^{W8}*/ +* and *RNAP II 215* ^{W8}*; 80ΔC* $_{1\text{-}648}$*#6, sca-GAL4* mutants.

The CTD of RNAP II has been shown to couple transcription with capping, pre-mRNA splicing and RNA 3'-end formation *in vivo*. Not only does the CTD independently stimulate each of these RNA processing events but different regions of the CTD appear to perform distinct functions in mRNA processing (Hirose and Manley, 2000; Fong and Bentley, 2001). Recent work has shown the carboxyl-region of mammalian CTD can support capping, splicing and 3'-end processing. However, the amino-region of the CTD can only support capping (Fong and Bentley, 2001). *RNAP II 215* ^{W8} is predicted to encode a carboxyl-truncated CTD with 19 out of the 42 normal heptad repeats (Brickey and Greenleaf, 1995). One possibility is that *RNAP II 215* ^{W8}*; 80ΔC* $_{1\text{-}648}$*#6, sca-GAL4* adults express RNAP II whose truncated

CTD supports capping but does not carry out efficient transcription coupled splicing and/ or RNA 3'-end formation.

It can therefore be reasoned that in *RNAP II 215* ^{W8}*; 80ΔC* $_{1-648}$ *#6, sca-GAL4* mutants, mRNAs undergo a compounded effect. Firstly, RNAP II transcripts are predicted to undergo inefficient transcription coupled splicing/ RNA 3'-end formation due to the activity of RNAP II with a truncated CTD. Secondly, mRNAs might also undergo aberrant CBC-cap-dependent splicing and 3'-end formation resulting from the over-expression of *80ΔC* $_{1-648}$ truncation. In addition, CBC might co-operate with the CTD of RNAP II to facilitate these mRNA processing events. Before mRNAs are released from transcription sites, they are monitored for correct splicing and 3'-end formation by checkpoints such as the nuclear exosome (Bousquet-Antonelli *et al.*, 2000; Hilleren *et al.*, 2001). Aberrantly processed mRNAs that undergo this compounded effect in *RNAP II 215* ^{W8}*; 80ΔC* $_{1-648}$ *#6, sca-GAL4* mutants might not escape such quality control systems and be retained in the nucleus and degraded. As a result, the inefficiently processed mRNAs in *80ΔC* $_{1-648}$ *#6, sca-GAL4* adults are not translated and the dominant macrochaetae phenotype is suppressed.

7.3.3 Cleavage factor 1A is a CTD-interacting protein uncovered by *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4* interacting region 51C1-51D8

BLASTP searches of the BDGP database identified five *Drosophila* proteins that contained CIDs. Apart from deficiencies, no mutant alleles are available for any of these putative CTD-interacting genes. Interestingly, *Df(2R)knSA3* that uncovers putative CTD-interacting gene *CG10228*, was already identified as a *80ΔC $_{1-648}$ #6, sca-GAL4* interacting deficiency. *Df(2R)knSA3* strongly suppresses *80ΔC $_{1-648}$ #6, sca-GAL4* macrochaetae phenotype at 21°C. *CG10228* is located to 51D1-2 and resides within the genetic interacting region 51C1-51D8. *CG10228* contains a type 2 amino-terminal CID and is predicted to encode cleavage factor 1A, the *Drosophila* homologue of *S. cerevisiae* pcf11. Characterisation of pcf11 shows it is a subunit of the heterotetrameric CF1A, which functions in RNA 3'-end formation and transcription termination. Moreover, recent work has confirmed that pcf11 directly interacts with the CTD of RNAP II and this interaction is stimulated by CTD phosphorylation (Barilla *et al.*, 2001; Gross and Moore 2001). This suggests that pcf11 is an adaptor molecule that couples transcription with RNA 3'-end formation by facilitating the stimulatory effects of the CTD during 3'-end processing.

Taken together, these data suggest that *cleavage factor 1A* is a strong candidate for the CBC-interacting gene that resides within 51C1-51D8. However, until mutant alleles of *cleavage factor 1A* become available to cross with *80ΔC $_{1-648}$ #6, sca-GAL4*, this remains purely speculative. Alternatively, biochemical approaches such as immunoprecipitations could be used to explore a possible interaction between CBC and *cleavage factor 1A*.

Roles in the cleavage step of RNA 3'-end formation have been described for *Drosophila* cleavage factor 1A, CBC and the CTD of RNAP II. One possibility is that during RNA 3'-end formation, the CTD might facilitate the coupling of transcription with RNA 3'-end formation by recruiting 3'-end processing factors to

the cleavage complex. As a result, CTD-interacting adaptor molecules such as cleavage factor 1A might interact with CBC in the cleavage complex to facilitate the positive effects of the cap in RNA 3'-end formation.

7.4 CONCLUSION

Genetic modifier screens using deficiencies and targeted genes have shown *UAS-DmCBP80ΔC* $_{1\text{-}648}$ *#6, sca-GAL4* is a useful tool by which to identify CBC-interacting genes. An allele of the largest subunit of RNAP II that has a carboxyl-truncated CTD suppresses *UAS-DmCBP80ΔC* $_{1\text{-}648}$ *#6, sca-GAL4* macrochaetae phenotype. This suggests an indirect/ direct interaction between CBC and RNAP II that is mediated through the CTD.

CHAPTER 8

Final Summary

The original objectives of this thesis were to firstly, determine whether CBC has novel functions during fruit-fly development. And secondly, identify CBC-interacting genes to further understand the mechanisms by which CBC and the cap facilitate multiple aspects of RNAP II transcript metabolism. Indeed, the research in this thesis presents three main findings. Firstly, novel function(s) are implicated for *DmCBP20* outside RNA processing during spermatogenesis. Next, a second CBP80 homologue has been identified in *Drosophila melanogaster* that is predominantly expressed in third instar larvae and adult testes. Finally, a genetic interaction is identified between CBC and the CTD of RNAP II. The following sections summarise these results and the working hypotheses that can be tested in the future.

Initial characterisation of CBC in *Drosophila* shows that both cap-binding proteins are highly homologous to the corresponding human proteins. DmCBC behaves in a manner analogous to HsCBC and both subunits are required to heterodimerise to form a functional cap-binding complex (Chapter 3). Taken together, these findings suggest that CBC function is conserved between *Drosophila* and human species. This presents a strong argument for using *Drosophila* as a model system to dissect the biology of CBC *in vivo*.

As a first step to characterising CBC function in *Drosophila*, an existing P-insertion in *DmCBP20* was mobilised to generate a battery of *DmCBP20* mutants. Results from the P-hop determined *DmCBP20* is an essential gene that is required for both somatogenesis and gametogenesis (Chapter 4). CBC RNA is expressed throughout *Drosophila* development and appears to be maternally derived (Chapter 3). This profile is typical for a key player in RNAP II transcript metabolism that is important for the viability of all cells in *Drosophila* development.

Surprisingly, *DmCBP20* makes a greater contribution to male fertility than female fertility. Analysis of the testes from *DmCBP20* male sterile lines indicate that the early pre-meiotic stages, which correlate to times of active RNAP II transcription coupled RNA processing, appear to be wild-type. Instead, preliminary observations suggest the defects underlying the male sterility might arise during the final stages of spermatogenesis, at times consistent with tight post-transcriptional regulation (Chapter 4). One possibility is that *DmCBP20* plays novel roles during spermatogenesis in post-transcriptional regulation. Additional work will have to be done to determine the exact stage and cause of sterility during spermatogenesis.

Although *CBP20* is present as a single-copy gene in *Drosophila*, humans have two CBP20 homologues. Initial characterisation of *HsCBP20R* has shown that it is predominantly expressed in the adult testes (E. Thompson). This suggests a role for CBP20 in spermatogenesis is conserved amongst higher eukaryotes. Intriguingly, *Drosophila* has two CBP80 homologues whilst humans only have one. Similar to HsCBP20R, DmCBP80R is predominantly expressed in the adult testes with little expression in the somatic tissues (Chapter 5). The functional significance of HsCBP20R or DmCBP80R is currently unknown. One possibility is that DmCBP80R, together with DmCBP20, might have acquired novel functions in spermatogenesis during evolution. DmCBP80R might bind DmCBP20 to form a male-specific CBC that functions in translation regulation. The male-specific CBC might mask mRNAs transcribed early from the translation machinery until the proteins are required during the later stages of spermatogenesis. Future experiments will be directed to whether DmCBP80R can bind DmCBP20 to form a male-specific CBC. In addition, it remains to be determined whether DmCBP80 and DmCBP80R are expressed in the same cell types or if the expression is mutually exclusive.

Finally, over-expression of the *UAS-DmCBP80ΔC* $_{1\text{-}648}$ truncation using the GAL4-UAS system generates dominant adult phenotypes in various non-essential tissues. Characterisation of the *UAS-DmCBP80ΔC* $_{1\text{-}648}$ *#6, sca-GAL4* dominant macrochaetae phenotype shows it is suitable for large-scale genetic modifier screens to identify *CBP80*-interacting genes (Chapter 6 and 7). Interestingly, the dominant

macrochaetae phenotype was suppressed by an allele of the largest subunit of RNAP II that has a truncated CTD (Chapter 7). This result suggests CBC interacts with the largest subunit of RNAP II and this interaction is facilitated through the CTD.

CBC and the cap act as a molecular tag to define RNAP II transcripts during processing. It is well established that transcription is coupled with RNA processing and this is facilitated through the CTD of the largest subunit of RNAP II. However, very little is known about the molecules that mediate the interaction between the CTD and various RNA processing factors. The strong suppression of *UAS-DmCBP80ΔC* $_{1\text{-}648}$ *#6, sca-GAL4* suggests a connection between the CTD of RNAP II and CBC. One hypothesis is that the cap-CBC might interact with the CTD of RNAP II to facilitate efficient coupling of transcription with processing. Such an interaction would help target transcripts to the CTD for pre-mRNA splicing and RNA 3'-end formation. Alternatively, CBC might play a novel role in RNAP II transcription, for example during transcription initiation. Future efforts will be directed to elucidating whether CBC functions in the early steps of transcription initiation and/ or the coupling of splicing and 3'-end processing to this process.

CHAPTER 9

Appendices

Appendix A: Summary of *DmCBP20^{HL}* PCR Screen

DmCBP20^{HL}/ TM3			
Line	**Mutation**	**Line**	**Mutation**
1	~200 bp deletion	42	Putative deletion
2	~200 bp deletion	43	Putative deletion
3	~200 bp deletion	44	Putative deletion
4	Putative deletion	45	Putative deletion
5	Putative deletion	46	Putative deletion
6	~25 bp residual P-element	47	~50 bp residual P-element
7	Putative deletion	48	Putative deletion
8	Putative deletion	49	Putative deletion
9	~40 bp residual P-element	50	Putative deletion
10	~100 bp residual P-element	51	~600 bp residual P-element
11	Putative deletion	52	~400 bp residual P-element
12	Putative deletion	53	Putative deletion
13	Putative deletion	54	Putative deletion
14	Putative deletion	55	~1, 700 bp residual P-element
15	Putative deletion	56	~1, 200 bp residual P-element
16	~400 bp residual P-element	57	Putative deletion
17	Putative deletion	58	~800 bp residual P-element
18	~1, 200 bp residual P-element	59	Putative deletion
19	Putative deletion	60	~25 bp residual P-element
20	Putative deletion	61	Putative deletion
21	Putative deletion	62	Putative deletion
22	~900 bp residual P-element	63	~100 bp residual P-element
23	Putative deletion	64	Putative deletion
24	Putative deletion	65	~1, 700 bp residual P-element
25	Putative deletion	66	Putative deletion
26	Putative deletion	67	Putative deletion
27	~60 bp residual P-element	68	Putative deletion
28	Putative deletion	69	Putative deletion
29	Putative deletion	70	Putative deletion
30	~50 bp residual P-element	71	Putative deletion
31	Putative deletion	72	~40 bp residual P-element
32	Putative deletion	73	Putative deletion
33	Putative deletion	74	Putative deletion
34	Putative deletion	75	Putative deletion
35	Putative deletion	76	~25 bp residual P-element
36	Putative deletion	77	~600 bp residual P-element
37	~300 bp residual P-element	78	Putative deletion
38	~100 bp residual P-element	79	Putative deletion
39	Putative deletion	80	Putative deletion
40	Putative deletion	81	Putative deletion
41	Putative deletion	82	Putative deletion

DmCBP20^{HL}/ TM3			
Line	**Mutation**	**Line**	**Mutation**
83	Putative deletion	124	Putative deletion
84	~40 bp residual P-element	125	Putative deletion
85	~1, 000 bp residual P-element	126	Putative deletion
86	~600 bp residual P-element	127	~30 bp residual P-element
87	~800 bp residual P-element	128	Putative deletion
88	~30 bp residual P-element	129	~200 bp residual P-element
89	Putative deletion	130	Putative deletion
90	Putative deletion	131	Putative deletion
91	Putative deletion	132	Putative deletion
92	Putative deletion	133	Putative deletion
93	Putative deletion	134	Putative deletion
94	~500 bp residual P-element	135	~1, 500 bp residual P-element
95	Putative deletion	136	~40 bp residual P-element
96	Putative deletion	137	Putative deletion
97	~200 bp residual P-element	138	Putative deletion
98	~25 bp residual P-element	139	Putative deletion
99	Putative deletion	140	Putative deletion
100	Putative deletion	141	~100 bp residual P-element
101	~400 bp residual P-element	142	Putative deletion
102	Putative deletion	143	Putative deletion
103	~100 bp residual P-element	144	~40 bp residual P-element
104	Putative deletion	145	~60 bp residual P-element
105	Putative deletion	146	Putative deletion
106	~100 bp residual P-element	147	Putative deletion
107	Putative deletion	148	Putative deletion
108	Putative deletion	149	~500 bp residual P-element
109	Putative deletion	150	~40 bp residual P-element
110	~200 bp residual P-element	151	Putative deletion
111	Putative deletion	152	Putative deletion
112	Putative deletion	153	~30 bp residual P-element
113	Putative deletion	154	Putative deletion
114	Putative deletion	155	Putative deletion
115	~1, 100 bp residual P-element	156	Putative deletion
116	~50 bp residual P-element	157	~30 bp residual P-element
117	Putative deletion	158	Putative deletion
118	~60 bp residual P-element	159	Putative deletion
119	Putative deletion	160	Putative deletion
120	Putative deletion	161	Putative deletion
121	~100 bp residual P-element	162	Putative deletion
122	~500 bp residual P-element	163	Putative deletion
123	~200 bp residual P-element		

Appendix B: UAS-Transgenic Lines

UAS-HsCBP20			
Line	**Eye Colour**	**Chromosome**	**Homozygous Viable?**
1	Yellow	2	Yes
2	Orange	3	Yes
3	Light Orange	3	Yes
4	Dark Orange	X	Yes
5	Orange	3	No
6	Red	3	No
7	Yellow	3	Yes
8	Light Orange	3	Yes
9	Dark Orange	2	Yes
10	Dark Orange	X	Yes
11	Light Orange	2	No
12	Red	3	Yes
13	Orange	3	Yes
14	Orange	3	Yes
15	Dark Orange	2	Yes
16	Dark Orange	2	Yes
17	Yellow	2	Yes
18	Red	2	Yes
UAS-DmCBP20			
Line	**Eye Colour**	**Chromosome**	**Homozygous Viable?**
1	Light Orange	3	Yes
2	Dark Orange	2	Yes
3	Red	X	Yes
4	Orange	3	Yes
5	Dark Orange	X	Yes
6	Orange	2	No
UAS-DmCBP20ΔC$_{1-95}$			
Line	**Eye Colour**	**Chromosome**	**Homozygous Viable?**
1	Light Orange	2	No
2	Orange	X	Yes
3	Red	2	No
4	Orange	3	Yes
5	Red	3	Yes
6	Yellow	3	Yes
7	Orange	3	Yes
8	Light Orange	2	No
9	Orange	X	Yes
10	Light Orange	3	Yes
11	Orange	3	Yes
12	Orange	2	No
13	Orange	3	No
14	Red	2	No
15	Red	3	Yes
16	Orange	X	Yes
17	Dark Orange	2	No
18	Orange	3	No
19	Red	3	Yes

20	Dark Orange	2	Yes
21	Yellow	3	Yes
22	Orange	2	No
UAS-DmCBP20ΔC$_{1-125}$			
Line	**Eye Colour**	**Chromosome**	**Homozygous Viable?**
1	Red	2	Yes
2	Orange	3	Yes
3	Orange	3	Yes
UAS-HsCBP80			
Line	**Eye Colour**	**Chromosome**	**Homozygous Viable?**
1	Orange	2	Yes
2	Dark Orange	2	Yes
3	Orange	3	Yes
4	Light Orange	2	Yes
5	Yellow	2	Yes
6	Orange	X	Yes
UAS-DmCBP80			
Line	**Eye Colour**	**Chromosome**	**Homozygous Viable?**
1	Red	2	Yes
2	Orange	3	Yes
3	Yellow	3	No
UAS-DmCBP80ΔC$_{1-178}$			
Line	**Eye Colour**	**Chromosome**	**Homozygous Viable?**
1	Orange	X	Yes
2	Dark Orange	3	Yes
3	Orange	2	Yes
4	Dark Orange	3	Yes
5	Orange	X	Yes
6	Orange	X	Yes
7	Orange	3	No
8	Dark Orange	3	Yes
9	Red	3	Yes
10	Light Orange	2	No
11	Red	3	Yes
12	Orange	3	No
13	Light Orange	X	Yes
14	Orange	X	Yes
15	Yellow	2	Yes
UAS-DmCBP80ΔC$_{1-424}$			
Line	**Eye Colour**	**Chromosome**	**Homozygous Viable?**
1	Orange	3	Yes
2	Yellow	2	Yes
3	Light Orange	3	Yes
4	Orange	X	Yes
5	Orange	3	Yes
6	Light Orange	3	Yes
7	Orange	3	No
8	Dark Orange	2	Yes
9	Orange	2	Yes
10	Dark Orange	X	No
11	Orange	3	Yes

12	Yellow	2	Yes
13	Red	2	Yes
14	Dark Orange	2	Yes
UAS-DmCBP80ΔC$_{1-510}$			
Line	**Eye Colour**	**Chromosome**	**Homozygous Viable?**
1	Red	3	Yes
2	Orange	3	Yes
3	Dark Orange	2	Yes
4	Orange	2	Yes
5	Dark Orange	3	Yes
6	Orange	X	Yes
7	Orange	3	No
UAS-DmCBP80ΔC$_{1-648}$			
Line	**Eye Colour**	**Chromosome**	**Homozygous Viable?**
1	Light Orange	2	No
2	Orange	X	Yes
3	Yellow	3	Yes
4	Red	2	Yes
5	Orange	3	Yes
6	Dark Red	2	Yes
7	Orange	3	Yes
8	Dark Orange	3	No

Appendix C: GAL4-UAS Crosses

Transgenic Line			GAL4 Driver				
Line Number	**Chromosome**	**Eye Colour**	***GMR***	***sca***	***vg***	***ptc***	***384***
UAS-HsCBP20							
1	2	Yellow	X	X	X	X	X
2	3	Orange	X	X	X	X	X
3	3	Light Orange	X	X	X	X	X
4	X	Dark Orange	X	X	X	X	X
7	3	Yellow	X	X	X	X	X
12	3	Red	X	X	X	X	X
UAS-DmCBP20							
1	3	Light Orange	X	X	X	X	X
2	2	Dark Orange	X	X	X	X	X
3	X	Red	X	X	X	X	X
4	3	Orange	X	X	X	X	X
5	X	Dark Orange	X	X	X	X	X
6	2	Orange	X	X	X	X	X
UAS-DmCBP20ΔC$_{1-95}$							
2	X	Orange	X	X	X	X	X
4	3	Orange	X	X	X	X	X
14	2	Red	X	X	X	X	X
15	3	Red	X	X	X	X	X
16	X	Orange	X	X	X	X	X
19	3	Red	X	X	X	X	X
UAS-DmCBP20ΔC$_{1-125}$							
1	2	Red	X	X	X	●	X
2	3	Orange	X	X	X	X	X
3	3	Orange	X	X	X	X	X
UAS-DmCBP80							
1	2	Orange	X	X	X	X	X
2	2	Dark Orange	X	X	X	X	X
3	3	Orange	X	X	X	X	X
4	2	Light Orange	X	X	X	X	X
5	2	Yellow	X	X	X	X	X
6	X	Orange	X	X	X	X	X

X = No dominant phenotype observed in F1 progeny at 18°C, 25°C or 29°C.

● = Dominant phenotype observed in F1 progeny at 25°C and 29°C.

Transgenic Line			GAL4 Driver				
Line Number	**Chromosome**	**Eye Colour**	***GMR***	***sca***	***vg***	***ptc***	***384***
UAS-DmCBP80							
1	2	Red	X	X	X	X	X
2	3	Orange	X	X	X	X	X
3	2	Yellow	X	X	X	X	X
UAS-DmCBP80ΔC$_{1-178}$							
1	X	Orange	X	X	X	X	X
3	2	Orange	X	X	X	X	X
8	3	Dark Orange	X	X	X	X	X
9	3	Red	X	X	X	X	X
13	X	Light Orange	X	X	X	X	X
15	2	Yellow	X	X	X	X	X
UAS-DmCBP80ΔC$_{1-424}$							
1	3	Orange	X	X	X	X	X
2	2	Yellow	X	X	X	X	X
4	X	Orange	X	X	X	X	X
6	3	Light Orange	X	X	X	X	X
7	3	Orange	X	X	X	X	X
13	2	Red	X	X	X	X	X
UAS-DmCBP80ΔC$_{1-510}$							
1	3	Red	X	X	X	X	X
2	3	Orange	X	X	X	X	X
3	2	Dark Orange	X	X	X	X	X
4	2	Orange	X	X	X	X	X
5	3	Dark Orange	X	X	X	X	X
6	X	Orange	X	X	X	X	X
UAS-DmCBP80ΔC$_{1-648}$							
1	2	Light Orange	X	X	X	X	X
2	X	Orange	X	X	X	●	●
3	3	Yellow	X	X	X	X	X
4	2	Red	X	X	X	●	●
5	3	Orange	X	X	X	X	X
6	2	Dark Red	X	X	X	●	●
7	3	Orange	X	X	X	●	●
8	3	Dark Orange	X	X	X	X	X

X = No dominant phenotype observed in F1 progeny at 18°C, 25°C or 29°C.

● = Dominant phenotype observed in F1 progeny at 25°C and 29°C.

Appendix D: Deficiencies used for Mapping *UAS-DmCBP80ΔC* $_{1-648}$ *#6, sca-GAL4* Interacting Regions

Deficiency	Breakpoints	Source
Interacting region 27C1-27E		
Df(2L)spd, al^{1} dp^{ov1}	27D-E; 28C	Bloomington
Df(2L)XE-2750	28B2; 28D3	Bloomington
In(1)w^{m4}; Df(2L)E110	25F3;26A1; 26D3-11	Bloomington
Df(2L)Dwee-delta5	27A; 28A	Bloomington
Df(2L)J-H	27C2-9; 28B3-4	Bloomington
w; Df(2L)spd^{j2}, wg^{spd-j2}	27C1-2; 28A	Bloomington
Interacting region 42B3-42C		
Df(2R)17l, cn^{1}	41F3-4; 42A3-9	Bloomington
*Df(2R)cn88b, cn**	42C; 42E	Bloomington
*Df(2R)ST1, Adh^{n5} pr^{1} cn**	42B3-5; 43E15-18	Bloomington
Df(2R)pk78k, Tp(2;2)CA30, sp^{1}	42E3; 43C3	Bloomington
Df(2R)pk78s	42F; 43F8	Bloomington
Df(2R)Drl^{rv3}, bw^{1}	42E1-4; 43C3	Bloomington
Interacting region 51C1-51D3		
Df(2R)03072	51A5; 51C1	Bloomington
Df(2R)trix	51A1-2; 51B6	Bloomington
w^{a} N^{fa-g}; Df(2R)Jp1	51D3-8; 52F5-9	Bloomington
y1 w1/Dp(1;Y)y+; Df(2R)XTE-18	51E3; 52C9-D1	Bloomington
w^{1}/Dp(1;Y)y^{+}; Df(2R)knSA3, Tp(1;2)TE21F22A	51B5-11; 51D7-E2	
Interacting region 57D12-58D1		
Df(2R)AA21, c^{1} px^{1} sp^{1}	56F9-17; 57D11-12	Bloomington
Df(2R)PK1, c^{1} px^{1} sp^{1}	57C5; 57F5-6	Bloomington
Df(2R)Pu-D17, cn^{1} bw^{1} sp^{1}	57B4; 58B	Bloomington
Df(2R)Egfr5, b^{1} pr^{1} cn^{1} sca^{1}	57D2-8; 58D1	Bloomington
Dp(1;Y)y^{+}/y^{1}; Df(2R)X58-12	58D1-2; 59A	Bloomington
Interacting region 71F-72D1		
Df(3L)BK10, ru^{1} Ly^{1} red^{1} cv-c^{1} Sb^{sbd-1} sr^{1} e^{1}	70C; 71F	Bloomington
Df(3L)brm11	71F1-4; 72D1-10	Bloomington
Df(3L)th102, h^{1} kni^{ri-1} e^{s}	71F3-5; 72D12	Bloomington
Df(3L)st-f13, Ki^{1} roe^{1} p^{p}	72C1-D1; 73A3-4	Bloomington

Appendix E: Genes Tested in *UAS-DmCBP80ΔC* $_{1-648}$ *#6, sca-GAL4* Interacting Regions

1. Interacting region 27C1-27E

Allele	Comments	Reference/ Source
snRNP 70K	***27C7-8***	
$P(ry^{+t7.2}\ lacZ)snRNP70K^{02107}cn^{1}$	P-insertion in *snRNP 70K* 5' UTR	Cline *et al.*, 1999
Nop5	***27C3-4***	
$y^{1}\ w^{67c23};\ P(w^{+mC}lacW)nop5^{k00230}$	P-insertion	Vorbruggen *et al.*, 2000
x16	***27C3-4***	
$y^{1}\ w^{67c23};\ P(w^{+mC}lacW)x16^{k00230}$	P-insertion	Vorbruggen *et al.*, 2000
Hrb27C	***27C3-4***	
$P\ (ry^{+t7.2}\ lacZ)Hrb27C^{02647}$	Lethal P-insertion 3.3 Kb from *Hrb 27C* coding sequence	Hammond *et al.*, 1997
$y^{1}w^{67c23};\ P(w^{+mC}lacW)Hrb27C^{k02814}$	Lethal P-insertion in *Hrb 27C*	Bloomington
ee	***25A3-27F2***	
$Dp(1;Y)y^{+};\ ee^{1}b^{1}pr^{1}$	P-insertion ?	Bloomington
$ee^{1}b^{1}pr^{1}$	Recessive viable	Bloomington
cup	***26F3-6***	
$cup^{1}cn^{1}bw^{1}sp^{1}$	Female sterile	Bloomington
$P(ry^{+t7.2}\ lacZ)cup^{01355}cn^{1}$	P-insertion	Bloomington
Rca1	***27B4-27C1***	
$P(ry^{+t7.2}\ lacZ)Rca1^{03300}cn^{1}$	P-insertion	Bloomington
ms(2)		
$P(ry^{+t7.2}\ lacZ)ms(2)27CD^{06491}cn^{1}$	P-insertion	Bloomington
milton	***27D5-6***	
$y^{1}w^{67c23};\ P(w^{+mC}\ lacW)milton^{k04704}$	P-insertion	Bloomington
l(2)k09923	***26D1-2***	
$P(lacW)l(2)k09923^{\ k09923}$	P-insertion	Bloomington
l(2)k14206	***26F3-5***	
$P(lacW)l(2)k14206^{\ k14206}$	P-insertion	Bloomington

2. Interacting region 42B3-42C

Allele	Comments	Reference/ Source
Adf1	***42B3-4***	
$P(ry^{+t7.2}\ lacZ)Adf1^{01349}cn^{1}$	P-insertion	Bloomington
jing	***42B2***	
$P(ry^{+t7.2}\ lacZ)jing^{01094}cn^{1}$	P-insertion	Bloomington
$y^{1}\ w^{67c23};\ P(w^{+mC}lacW)jing^{k03404}$	P-insertion	Bloomington
l(2)k09848	***42A8-12***	
$y^{1}w^{67c23};\ P(w^{+mC}lacW)l(2)k09848^{k09848}$	P-insertion	Bloomington
l(2)k03204	***42B1-3***	
$y^{1}w^{67c23};\ P(w^{+mC}lacW)l(2)k03204^{k03204}$	P-insertion	Bloomington
l(2)04535b	***42C1-2***	
$P(lacZ)l(2)04535b^{04535b}cn^{1}$	P-insertion	Bloomington

3. Interacting region 51C1-51D3

Allele	Comments	Reference/ Source
sf	***51C1-E2***	
sf^2	Recessive viable	Bloomington
$b^1 sf^1$	Recessive viable	Bloomington
Rpn6	***51C2***	
$y^1\ w^{67c23};\ P(w^{+mC}lacW)Rpn6^{k00103}$	P-insertion	Bloomington
Kn	***51C2-5***	
kn^1	Recessive viable	Bloomington
boc	***51D1-12***	
$cn^1\ P(ry^{+t7.2}\ lacZ)boc^1$	P-insertion	Bloomington
l(2)k08015	***51D3-5***	
$y^1\ w^{67c23};\ P(w^{+mC}lacW)l(2)k08015^{k08015}$	P-insertion	Bloomington

4. Interacting region 57D12-58D1

Allele	Comments	Reference/ Source
dve	***58D1-2***	
$y^1\ w^{67c23};\ P(w^{+mC}lacW)dve^{k06515}$	P-insertion	Bloomington
$cn^1\ P(ry^{+t7.2}\ lacZ)dve^{01738}$	P-insertion	Bloomington
l(2)10608	***57E1-2***	
$P(lacZ)l(2)10608^{10608}$	P-insertion	Bloomington
l(2)03605	***57F8-10***	
$P(lacZ)l(2)03605^{03605}$	P-insertion	Bloomington
l(2)k10317	***57E6-7***	
$y^1 w^{67c23};\ P(w^{+mC}lacW)l(2)k10317^{k10317}$	P-insertion	Bloomington
l(2)07837	***58A3-4***	
$P(ry^{+t7.2}\ lacZ)l(2)07837^{07837}cn^1$	P-insertion	Bloomington

5. Interacting region 71F-72D1

Allele	Comments	Reference/ Source
Arf72A	***72C1***	
$Arf72AN6\ P(w^{+m})71F$	P-insertion	Bloomington
l(3)72Ab	***72A3; 72C1***	
$l(3)72AbI24\ P(w^{+m})71F$	P-insertion	Bloomington
l(3)72CD	***72C1-72D2***	
$l(3)72CDaJ12\ P(w^{+m})71F$	P-insertion	Bloomington

Appendix F: Genes Targeted for Interaction with *UAS-DmCBP80ΔC $_{1-648}$ #6, sca-GAL4*

Transcription

Allele	Comments	Reference/ Source
RNAP II 140		
RNAPII 140^{A5}	Null in *140*	Mark Mortin
Df(3R)red^{1}	*140* located to 88A10-12, deficiency breakpoints 88A2; 88C1-D1	Bloomington
RNAP II 215		
RNAPII 215 ^{NULL}	Null in *215*	Mark Mortin
RNAPII 215 ^{K1}	*215* ts allele	Mark Mortin
RNAPII 215 ^{W8}	*215* with truncated CTD, contains 19 out of 42 heptad repeats	Brickey and Greenleaf, 1995
v^{1}RNAPII 215 ^{8}	EMS-induced ts allele in *215*, aa replacement L-878-F	Chen *et al.*, 1993
v^{1}RNAPII 215 ^{12}	EMS-induced ts allele in *215*, aa replacement A-377-T	Chen *et al.*, 1993
v^{1}RNAPII 215 ^{4}	EMS-induced ts allele in *215*, aa replacement R-741-H	Chen *et al.*, 1993
y^{2}ras^{1}v^{1}RNAPII 215 shi^{1}	Deletion in 5' upstream coding region of *215*	Bloomington
v^{1}RNAPII 215 ^{3}	EMS-induced ts allele in *215*	Bloomington
Df(1)GA112	*215* located to 10C2, deficiency breakpoints 10B1-2; 10C1-2	Bloomington
Df(1)N71	*215* located to 10C2, deficiency breakpoints 10B5; 10D4	Biggs *et al.*, 1985; Bloomington
Putative CTD-binding proteins		
Df(2L)VA16	*CG15160* located to 37A2, deficiency breakpoints 36F7-9; 37D1-2	Bloomington
Df(3L)R-G7	*CG9018* located to 62D3, deficiency breakpoints 62B9; 62E7	Bloomington
Df(2L)J39	*CG6521* located to 32B4, deficiency breakpoints 31C; 32D1-E5	Bloomington
Df(2R)knSA3	*CG10228* located to 51D2, deficiency breakpoints 51B5-11; 51D7-E2	Bloomington
Df(2R)Pu-D17	*CG9346* located to 57B16, deficiency breakpoints 57B4; 58B	Bloomington
Df(2R)AA21	*CG4266* located to 57C4-5, deficiency breakpoints 56F9-17; 57D11-12	Bloomington

pre-mRNA splicing

Allele	Comments	Reference/ Source
U1 snRNA		
Df(2L)S2	U1a-21D located to 21E, deficiency breakpoints 21C6-D1; 24D3-4	Mount and Saltz, 2000; Bloomington
Df(3R)110	U1a-82Eb and U1a-82Ec located to 82E, deficiency breakpoints 82C4; 82F3	Bloomington
Df(3R)mbc-R1	U1a-95Ca, U1a-95Cb and U1a-95Cc located to 95C, deficiency breakpoints 95A5-7; 95D6-11	Bloomington
Sans fille		
yw snf^{4210}	Null in *snf*	Flickinger and Salz, 1994
Df(1)JC70	*snf* located to 4F1-2, deficiency breakpoints 4C15-16; 5A1-2	Salz *et al.*, 1994
snf^{1}v^{24}	aa replacement H-49-R in amino terminal RRM of *snf*	Salz and Flickinger, 1996
SmD3		
Df(2R)CB21	*SmD3* located to 48F3, deficiency breakpoints 48E; 49A	Mount and Salz, 2000; Bloomington
Luc7-like		
Df(3L)81K19	*Luc7-like* (*CG7564*) located to 74C1, deficiency breakpoints 73A3; 74F	Mount and Salz, 2000; Bloomington
Related to Luc7		
Tp(1;3)sn^{13a1}	Related to Luc7 (*CG3198*) located to 6C8, deficiency breakpoints 6C; 7C9-7D1	Mount and Salz, 2000; Bloomington
U2af 50		
Df(1)4b18, y cv v f car	*U2af50* located to 14C1 , deficiency breakpoints 14B8; 14C1	Bloomington
U2af50$^{\Delta RS}$	Deletion of RS domain amino acids 1-34 that leaves a single RS dipeptide	Rudner *et al.*, 1998
U2af50$^{\Delta RS.extreme}$	Deletion of amino-terminal acids 1-46	Rudner *et al.*, 1998
U2af 38		
Df(2L)PMF	*U2af 38* located to 21B7-8 , deficiency breakpoints 21A1; 21B7-8	Rudner *et al.*, 1996; Bloomington
cn^{1}P(ry $^{+17.2}$lac Z) U2af38^{06751}	Lethal P-insertion 200bp upstream of *U2af 38* coding region	Rudner *et al.*, 1996
y^{1}w^{67c23}; P(w^{+mC}lacW) U2af38^{k14504}	Lethal P-insertion in *U2af 38*	Bloomington
Noisette		
P(A92)gor^{1}	P-insertion in *Noisette*	Meyer *et al.*, 1998
B52		
wa; B52^{R2}	25 bp deletion in 1stRRM of *B52*	Peng and Mount, 1995
B52^{7}	Deletion in *B52*	Peng and Mount, 1995
w^{1118}; Hrb87F^{2}B52 $^{129.1}$	Lethal P-insertion in *B52*	John Lis
w^{1118}; Hrb87F^{2}B52$^{402.1}$	Lethal P-insertion in *B52*	John Lis
B52^{1}red e	EMS-induced allele in *B52*	Susan Haynes
B52 ^{17}red e	EMS-induced allele in *B52*	Susan Haynes
Df(3R)ry^{85}	*B52* located to 87F7-8, deficiency breakpoints 87B15-C1; 87F15-88A1	Bloomington
Hrb 87F		
Df(3R)Hrb87F	Null in *Hrb87F* and partial deletion in *Tsr*	Bloomington
Hrb87F^{7}	*Hrb87F* deletion allele	Peng and Mount, 1995

RNA 3'-end formation

Allele	Comments	Reference/ Source
Cleavage and polyadenylation specificity factor		
$y^1 w^{67c23}$; *l(2)k16805*	P-insertion in *CPSF-160* 5' UTR.	Salinas *et al.*, 1998
Df(2R)trix	*CPSF-160* located to 51B, deficiency breakpoints 51A1-2; 51B6	Bloomington
$CPSF^V$	500 bp deletion in *CPSF-160*	Salinas *et al.*, 1998
$CPSF\ IV^{II-49}$	Deletion in *CPSF-160* and *Asx*	Salinas *et al.*, 1998
Df(3R)3450	*CPSF-100* located to 98F, deficiency breakpoints 98E3; 99A8	Bloomington
Df(3R)ChaM7	*CPSF-73* located to 91D, deficiency breakpoints 91A; 91F5	Bloomington
Df(2L)al	*CPSF-30* located to 21D, deficiency breakpoints 21B8-C1; 21C8-D1	Bloomington
Df(2L)ast2	*CPSF-30* located to 21D, deficiency breakpoints 21D1-2; 22B2-3	Bloomington
Cleavage stimulation factor		
Df(1)X-1, Df(1)y-ac, $sc^1 su(f)^5$	150 bp deletion in *CstF-77* at positions +2.2 to +4.3	Bloomington
$y^1 v^1 su(f)^6$	1 Kb deletion in *CstF-77* at positions –0.3 to +1.3	Bloomington
$y^2 sc^1 su(s)^2 lz^1 f^1 su(f)^8$	*CstF-77* ts lethal	Bloomington
Df(1)A209	*CstF-77* located to 20E, deficiency breakpoints 20A; 20F	Bloomington
Df(3R)ChaM7	*CstF-64* located to 91D, deficiency breakpoints 91A; 91F5	Bloomington
Df(3R)faf-BP	*CstF-50* located to 100D, deficiency breakpoints 100D; 100F5	Bloomington

RNA transport

Allele	Comments	Reference/ Source
Pendulin		
$P(w^{+mC} lacW)Pen^{k14401}$	Lethal P-insertion in 5'UTR of *Pendulin*	Torok *et al.*, 1995
Df(2L)MdhA	*Pendulin* located to 31A, deficiency breakpoints 30D-30F; 31F	Bloomington
Xpo1		
Df(2L)N22-14	*Xpo1* located to 29C1-2, deficiency breakpoints 29C1-2; 30C8-9	Bloomington
Diminished-7		
Df(3L)pbl-X1	*Diminished*-7 located to 66B8-B10, deficiency breakpoints 65F3; 66B10	Bloomington
Bx34		
Df(2R)en30	*Bx34* located to 48C3, deficiency breakpoints 48A3-4; 48C6-8	Bloomington
Nup154		
Df(2L)J39	*Nup154* located to 32D1-2, deficiency breakpoints 31C-D; 32D-E	Bloomington
Karyopherin α1		
Df(3L)VW3	*Karyopherin α1* located to 76A-F, deficiency breakpoints 76A3; 76B2	Bloomington

Translation

Allele	Comments	Reference/ Source
eIF4E		
Df(3L)AC1	*eIF4E* located to 67B2, deficiency breakpoints 67A2; 67D13	Hernandez *et al.*, 1997; Bloomington
w; eIF4E P589.11	Lethal P-insertion in 1st intron of *eIF4E*	Pascal Lachance
eIF4E-binding protein		
yw; thor	Null in *eIF4E-binding protein*	Pascal Lachance
$Df(2L)tim^{02}$	*eIF4E-binding protein* located to 23F3, deficiency breakpoints 23F2-3; 23F6-24A1	Bernal and Kimbrell, 2000
eIF4E-related proteins		
Df(3L)66C-G28	*CG8023* located to 66C1, deficiency breakpoints 66B8-9; 66C9-10	Lasko, 2000
Df(3L)W5.4	*CG10124* located to 65C1, deficiency breakpoints 65A; 65E1	Lasko, 2000; Bloomington
Df(3R)crb-F89-4	*eIF4E-homologous protein* located to 95D8-10, deficiency breakpoints 95D7-11; 95F15	Bloomington
eIF4A		
$y^1\ w^{67c23}$; $P(w^{+mC}lacW)eIF4A^{k01501}$	Lethal P-insertion in *eIF4A*	Bloomington
$P(ry^{+t7.2}PZ)eIF4A^{02439}$ cn^1	Lethal P-insertion in *eIF4A*	Galloni *et al.*, 1999
$yw; eIF4A^{162\text{-}5}$	700 bp deletion in *eIF4A* at positon +2	Dorn *et al.*, 1993
$yw; eIF4A^{1069}$	Lethal P-insertion into 1st intron of *eIF4A*	Galloni *et al.*, 1999
$yw; eIF4A^{1013}$	Lethal P-insertion into 1st intron of *eIF4A*	Galloni *et al.*, 1999
$yw; eIF4A^{1006}$	Lethal P-insertion into 2nd intron of *eIF4A*	Galloni *et al.*, 1999
Df(2L)E110	*eIF4A* located to 26A9-B1, deficiency breakpoints 25F3-26A1; 26D3-11	Bloomington
eIF4G		
Df(4)G	*eIF4G* located to 102E, deficiency breakpoints 102E2; E10	Hernandez *et al.*, 1998; Bloomington
Df(4)M101-62f	Deficiency breakpoints 101E; 102B10-17	Bloomington

CHAPTER 10

References

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., George, R. A., Lewis, S. E., Richards, S., Ashburner, M., Henderson, S. N., Sutton, G. G., Wortman, J. R., Yandell, M. D., Zhang, Q., Chen, L. X., Brandon, R. C., Rogers, Y. H., Blazej, R. G., Champe, M., Pfeiffer, B. D., Wan, K. H., Doyle, C., Baxter, E. G., Helt, G., Nelson, C. R., Gabor, G. L., Abril, J. F., Agbayani, A., An, H. J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R. M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E. M., Beeson, K. Y., Benos, P. V., Berman, B. P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M. R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K. C., Busam, D. A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J. M., Cawley, S., Dahlke, C., Davenport, L. B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A. D., Dew, I., Dietz, S. M., Dodson, K., Doup, L. E., Downes, M., Dugan-Rocha, S., Dunkov, B. C., Dunn, P., Durbin, K. J., Evangelista, C. C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A. E., Garg, N. S., Gelbart, W. M., Glasser, K., Glodek, A., Gong, F., Gorrell, J. H., Gu, Z., Guan, P., Harris, M., Harris, N. L., Harvey, D., Heiman, T. J., Hernandez, J. R., Houck, J., Hostin, D., Houston, K. A., Howland, T. J., Wei, M. H., Ibegwam, C., et al. (2000). The genome sequence of Drosophila melanogaster. Science *287*, 2185-95.

Allison, L. A., Moyle, M., Shales, M., and Ingles, C. J. (1985). Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. Cell *42*, 599-610.

Allison, L. A., Wong, J. K., Fitzpatrick, V. D., Moyle, M., and Ingles, C. J. (1988). The C-terminal domain of the largest subunit of RNA polymerase II of Saccharomyces cerevisiae, Drosophila melanogaster, and mammals: a conserved structure with an essential function. Mol Cell Biol *8*, 321-9.

Andrulis, E. D., Guzman, E., Doring, P., Werner, J., and Lis, J. T. (2000). High-resolution localization of Drosophila Spt5 and Spt6 at heat shock genes in vivo: roles in promoter proximal pausing and transcription elongation. Genes Dev *14*, 2635-49.

Arts, G. J., Kuersten, S., Romby, P., Ehresmann, B., and Mattaj, I. W. (1998). The role of exportin-t in selective nuclear export of mature tRNAs. Embo J *17*, 7430-41.

Azuma, Y., and Dasso, M. (2000). The role of Ran in nuclear function. Curr Opin Cell Biol *12*, 302-7.

Barabino, S. M., and Keller, W. (1999). Last but not least: regulated poly(A) tail formation. Cell *99*, 9-11.

Barilla, D., Lee, B. A., and Proudfoot, N. J. (2001). Cleavage/polyadenylation factor IA associates with the carboxyl- terminal domain of RNA polymerase II in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A *98*, 445-50.

Bauren, G., and Wieslander, L. (1994). Splicing of Balbiani ring 1 gene pre-mRNA occurs simultaneously with transcription. Cell *76*, 183-92.

Beelman, C. A., and Parker, R. (1995). Degradation of mRNA in eukaryotes. Cell *81*, 179-83.

Berk, A. J. (1999). Activation of RNA polymerase II transcription. Curr Opin Cell Biol *11*, 330-5.

Bernal, A., and Kimbrell, D. A. (2000). Drosophila Thor participates in host immune defense and connects a translational regulator with innate immunity. Proc Natl Acad Sci U S A *97*, 6019-24.

Beyer, A., L., and Osheim, Y., N. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. Genes and Development *2*, 754-765.

Biggs, J., Searles, L. L., and Greenleaf, A. L. (1985). Structure of the eukaryotic transcription apparatus: features of the gene for the largest subunit of Drosophila RNA polymerase II. Cell *42*, 611-21.

Birney, E., Kumar, S., and Krainer, A. R. (1993). Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. Nucleic Acids Res *21*, 5803-16.

Black, D. L. (1992). Activation of c-src neuron-specific splicing by an unusual RNA element in vivo and in vitro. Cell *69*, 795-807.

Black, D. L. (1991). Does steric interference between splice sites block the splicing of a short c-src neuron-specific exon in non-neuronal cells? Genes Dev *5*, 389-402.

Borman, A. M., Michel, Y. M., and Kean, K. M. (2000). Biochemical characterisation of cap-poly(A) synergy in rabbit reticulocyte lysates: the eIF4G-PABP interaction increases the functional affinity of eIF4E for the capped mRNA 5'-end. Nucleic Acids Res *28*, 4068-75.

Bourquin, J. P., Stagljar, I., Meier, P., Moosmann, P., Silke, J., Baechi, T., Georgiev, O., and Schaffner, W. (1997). A serine/arginine-rich nuclear matrix cyclophilin interacts with the C- terminal domain of RNA polymerase II. Nucleic Acids Res *25*, 2055-61.

Bousquet-Antonelli, C., Presutti, C., and Tollervey, D. (2000). Identification of a regulated pathway for nuclear pre-mRNA turnover. Cell *102*, 765-75.

Brahms, H., Raymackers, J., Union, A., de Keyser, F., Meheus, L., and Luhrmann, R. (2000). The C-terminal RG dipeptide repeats of the spliceosomal Sm proteins D1 and D3 contain symmetrical dimethylarginines, which form a major B-cell epitope for anti-Sm autoantibodies. J Biol Chem *275*, 17122-9.

Brand, A., H., and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development *118*, 401-415.

Brickey, W. J., and Greenleaf, A. L. (1995). Functional studies of the carboxy-terminal repeat domain of Drosophila RNA polymerase II in vivo. Genetics *140*, 599-613.

Buratowski, S. (2000). Snapshots of RNA polymerase II transcription initiation. Curr Opin Cell Biol *12*, 320-5.

Burd, C. G., and Dreyfuss, G. (1994). Conserved structures and diversity of functions of RNA-binding proteins. Science *265*, 615-21.

Carlson, M. (1997). Genetics of transcriptional regulation in yeast: connections to the RNA polymerase II CTD. Annu Rev Cell Dev Biol *13*, 1-23.

Chabot, B., Bisotto, S., and Vincent, M. (1995). The nuclear matrix phosphoprotein p255 associates with splicing complexes as part of the [U4/U6.U5] tri-snRNP particle. Nucleic Acids Res *23*, 3206-13.

Chambers, R. S., and Dahmus, M. E. (1994). Purification and characterization of a phosphatase from HeLa cells which dephosphorylates the C-terminal domain of RNA polymerase II. J Biol Chem *269*, 26243-8.

Chanfreau, G., Elela, S. A., Ares, M., Jr., and Guthrie, C. (1997). Alternative 3'-end processing of U5 snRNA by RNase III. Genes Dev *11*, 2741-51.

Chang, H. C., Karim, F. D., O'Neill, E. M., Rebay, I., Solomon, N. M., Therrien, M., Wassarman, D. A., Wolff, T., and Rubin, G. M. (1994). Ras signal transduction pathway in Drosophila eye development. Cold Spring Harb Symp Quant Biol *59*, 147-53.

Chen, F., MacDonald, C. C., and Wilusz, J. (1995). Cleavage site determinants in the mammalian polyadenylation signal. Nucleic Acids Res *23*, 2614-20.

Chen, Y., Weeks, J., Mortin, M. A., and Greenleaf, A. L. (1993). Mapping mutations in genes encoding the two large subunits of Drosophila RNA polymerase II defines domains essential for basic transcription functions and for proper expression of developmental genes. Mol Cell Biol *13*, 4214-22.

Chiu, Y. L., Coronel, E., Ho, C. K., Shuman, S., and Rana, T. M. (2001). HIV-1 Tat protein interacts with mammalian capping enzyme and stimulates capping of TAR RNA. J Biol Chem *276*, 12959-66.

Cho, E. J., Takagi, T., Moore, C. R., and Buratowski, S. (1997). mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. Genes Dev *11*, 3319-26.

Cho, H., Kim, T. K., Mancebo, H., Lane, W. S., Flores, O., and Reinberg, D. (1999). A protein phosphatase functions to recycle RNA polymerase II. Genes Dev *13*, 1540-52.

Cline, T. W., Rudner, D. Z., Barbash, D. A., Bell, M., and Vutien, R. (1999). Functioning of the Drosophila integral U1/U2 protein Snf independent of U1 and U2 small nuclear ribonucleoprotein particles is revealed by snf(+) gene dose effects. Proc Natl Acad Sci U S A *96*, 14451-8.

Colgan, D. F., and Manley, J. L. (1997). Mechanism and regulation of mRNA polyadenylation. Genes Dev *11*, 2755-66.

Colot, H. V., Stutz, F., and Rosbash, M. (1996). The yeast splicing factor Mud13p is a commitment complex component and corresponds to CBP20, the small subunit of the nuclear cap-binding complex. Genes Dev *10*, 1699-708.

Connolly, J. B., Roberts, I. J., Armstrong, J. D., Kaiser, K., Forte, M., Tully, T., and O'Kane, C. J. (1996). Associative learning disrupted by impaired Gs signaling in Drosophila mushroom bodies. Science *274*, 2104-7.

Cooke, C., and Alwine, J. C. (1996). The cap and the 3' splice site similarly affect polyadenylation efficiency. Mol Cell Biol *16*, 2579-84.

Coppola, J. A., Field, A. S., and Luse, D. S. (1983). Promoter-proximal pausing by RNA polymerase II in vitro: transcripts shorter than 20 nucleotides are not capped. Proc Natl Acad Sci U S A *80*, 1251-5.

Corden, J. L., Cadena, D. L., Ahearn, J. M., Jr., and Dahmus, M. E. (1985). A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. Proc Natl Acad Sci U S A *82*, 7934-8.

Curtis, D., Lehmann, R., and Zamore, P. D. (1995). Translational regulation in development. Cell *81*, 171-8.

Czaplinski, K., Ruiz-Echevarria, M. J., Gonzalez, C. I., and Peltz, S. W. (1999). Should we kill the messenger? The role of the surveillance complex in translation termination and mRNA turnover. Bioessays *21*, 685-96.

Dahmus, M. E. (1996). Reversible phosphorylation of the C-terminal domain of RNA polymerase II. J Biol Chem *271*, 19009-12.

Dahmus, M. E. (1994). The role of multisite phosphorylation in the regulation of RNA polymerase II activity. Prog Nucleic Acid Res Mol Biol *48*, 143-79.

Daneholt, B. (1997). A look at messenger RNP moving through the nuclear pore. Cell *88*, 585-8.

Dantonel, J. C., Murthy, K. G., Manley, J. L., and Tora, L. (1997). Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA. Nature *389*, 399-402.

Das, B., Guo, Z., Russo, P., Chartrand, P., and Sherman, F. (2000). The role of nuclear cap binding protein Cbc1p of yeast in mRNA termination and degradation. Mol Cell Biol *20*, 2827-38.

Dignam, J. D., Lebovitz, R. M., and Roeder, R. G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res *11*, 1475-89.

Dominski, Z., and Marzluff, W. F. (1999). Formation of the 3' end of histone mRNA. Gene *239*, 1-14.

Dorn, R., Morawietz, H., Reuter, G., and Saumweber, H. (1993). Identification of an essential Drosophila gene that is homologous to the translation initiation factor eIF-4A of yeast and mouse. Mol Gen Genet *237*, 233-40.

Du, L., and Warren, S. L. (1997). A functional interaction between the carboxy-terminal domain of RNA polymerase II and pre-mRNA splicing. J Cell Biol *136*, 5-18.

Eckner, R., Ellmeier, W., and Birnstiel, M. L. (1991). Mature mRNA 3' end formation stimulates RNA export from the nucleus. Embo J *10*, 3513-22.

Edery, I., and Sonenberg, N. (1985). Cap-dependent RNA splicing in a HeLa nuclear extract. Proc Natl Acad Sci U S A *82*, 7590-4.

Eggert, H., Bergemann, K., and Saumweber, H. (1998a). Molecular screening for P-element insertions in a large genomic region of Drosophila melanogaster using polymerase chain reaction mediated by the vectorette. Genetics *149*, 1427-34.

Eggert, T., Hauck, B., Hildebrandt, N., Gehring, W. J., and Walldorf, U. (1998b). Isolation of a Drosophila homolog of the vertebrate homeobox gene Rx and its possible role in brain and eye development. Proc Natl Acad Sci U S A *95*, 2343-8.

Elefant, F., and Palter, K., B. (1999). Tissue-specific expression of dominant negative mutant Drosophila HSC70 causes developmental defects and lethality. Molecular Biology of the Cell *10*, 2101-2117.

Elliott, D. J., Millar, M. R., Oghene, K., Ross, A., Kiesewetter, F., Pryor, J., McIntyre, M., Hargreave, T. B., Saunders, P. T., Vogt, P. H., Chandley, A. C., and Cooke, H. (1997). Expression of RBM in the nuclei of human germ cells is dependent on a critical region of the Y chromosome long arm. Proc Natl Acad Sci U S A *94*, 3848-53.

Erickson, J. W., and Cerione, R. A. (2001). Multiple roles for Cdc42 in cell regulation. Curr Opin Cell Biol *13*, 153-7.

Feaver, W. J., Svejstrup, J. Q., Henry, N. L., and Kornberg, R. D. (1994). Relationship of CDK-activating kinase and RNA polymerase II CTD kinase TFIIH/TFIIK. Cell *79*, 1103-9.

Fischer, U., Huber, J., Boelens, W. C., Mattaj, I. W., and Luhrmann, R. (1995). The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. Cell *82*, 475-83.

Fischer, U., Sumpter, V., Sekine, M., Satoh, T., and Luhrmann, R. (1993). Nucleo-cytoplasmic transport of U snRNPs: definition of a nuclear location signal in the Sm core domain that binds a transport receptor independently of the m3G cap. Embo J *12*, 573-83.

Flach, J., Bossie, M., Vogel, J., Corbett, A., Jinks, T., Willins, D. A., and Silver, P. A. (1994). A yeast RNA-binding protein shuttles between the nucleus and the cytoplasm. Mol Cell Biol *14*, 8399-407.

Flaherty, S. M., Fortes, P., Izaurralde, E., Mattaj, I. W., and Gilmartin, G. M. (1997). Participation of the nuclear cap binding complex in pre-mRNA 3' processing. Proc Natl Acad Sci U S A *94*, 11893-8.

Flickinger, T. W., and Salz, H. K. (1994). The Drosophila sex determination gene snf encodes a nuclear protein with sequence and functional similarity to the mammalian U1A snRNP protein. Genes Dev *8*, 914-25.

Fong, N., and Bentley, D. L. (2001). Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. Genes Dev *15*, 1783-95.

Fornerod, M., Ohno, M., Yoshida, M., and Mattaj, I. W. (1997). CRM1 is an export receptor for leucine-rich nuclear export signals. Cell *90*, 1051-60.

Fortes, P., Bilbao-Cortes, D., Fornerod, M., Rigaut, G., Raymond, W., Seraphin, B., and Mattaj, I. W. (1999a). Luc7p, a novel yeast U1 snRNP protein with a role in 5' splice site recognition. Genes Dev *13*, 2425-38.

Fortes, P., Inada, T., Preiss, T., Hentze, M. W., Mattaj, I. W., and Sachs, A. B. (2000). The yeast nuclear cap binding complex can interact with translation factor eIF4G and mediate translation initiation. Mol Cell *6*, 191-6.

Fortes, P., Kufel, J., Fornerod, M., Polycarpou-Schwarz, M., Lafontaine, D., Tollervey, D., and Mattaj, I. W. (1999b). Genetic and physical interactions involving the yeast nuclear cap- binding complex. Mol Cell Biol *19*, 6543-53.

Freeman, M. (1996). Reiterative use of the EGF receptor triggers differentiation of all cell types in the Drosophila eye. Cell *87*, 651-660.

Fresco, L. D., and Buratowski, S. (1996). Conditional mutants of the yeast mRNA capping enzyme show that the cap enhances, but is not required for, mRNA splicing. RNA *2*, 584-96.

Furuichi, Y., LaFiandra, A., and Shatkin, A. J. (1977). 5'-Terminal structure and mRNA stability. Nature *266*, 235-9.

Gallie, D. R. (1991). The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. Genes Dev *5*, 2108-16.

Galloni, M., and Edgar, B. A. (1999). Cell-autonomous and non-autonomous growth-defective mutants of Drosophila melanogaster. Development *126*, 2365-75.

Gamberi, C., Izaurralde, E., Beisel, C., and Mattaj, I. W. (1997). Interaction between the human nuclear cap-binding protein complex and hnRNP F. Mol Cell Biol *17*, 2587-97.

Gilmartin, G. M., McDevitt, M. A., and Nevins, J. R. (1988). Multiple factors are required for specific RNA cleavage at a poly(A) addition site. Genes Dev *2*, 578-87.

Gorlich, D., Dabrowski, M., Bischoff, F. R., Kutay, U., Bork, P., Hartmann, E., Prehn, S., and Izaurralde, E. (1997). A novel class of RanGTP binding proteins. J Cell Biol *138*, 65-80.

Gorlich, D., Kraft, R., Kostka, S., Vogel, F., Hartmann, E., Laskey, R. A., Mattaj, I. W., and Izaurraide, E. (1996). Importin provides a link between nuclear protein import and U snRNA export. Cell *87*, 21-32.

Graveley, B. R. (2000). Sorting out the complexity of SR protein functions. RNA *6*, 1197-211.

Gray, N. K., and Wickens, M. (1998). Control of translation initiation in animals. Annu Rev Cell Dev Biol *14*, 399-458.

Green, M. R. (2000). TBP-associated factors (TAFIIs): multiple, selective transcriptional mediators in common complexes. Trends Biochem Sci *25*, 59-63.

Gross, S., and Moore, C. (2001). Five subunits are required for reconstitution of the cleavage and polyadenylation activities of Saccharomyces cerevisiae cleavage factor I. Proc Natl Acad Sci U S A *98*, 6080-5.

Hagler, J., and Shuman, S. (1992). Stability of ternary transcription complexes of vaccinia virus RNA polymerase at promoter-proximal positions. J Biol Chem *267*, 7644-54.

Hamm, J., and Mattaj, I. W. (1990). Monomethylated cap structures facilitate RNA export from the nucleus. Cell *63*, 109-18.

Hammond, L. E., Rudner, D. Z., Kanaar, R., and Rio, D. C. (1997). Mutations in the hrp48 gene, which encodes a Drosophila heterogeneous nuclear ribonucleoprotein particle protein, cause lethality and developmental defects and affect P-element third-intron splicing in vivo. Mol Cell Biol *17*, 7260-7.

Hart, R. P., McDevitt, M. A., and Nevins, J. R. (1985). Poly(A) site cleavage in a HeLa nuclear extract is dependent on downstream sequences. Cell *43*, 677-83.

Hartzog, G. A., Wada, T., Handa, H., and Winston, F. (1998). Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in Saccharomyces cerevisiae. Genes Dev *12*, 357-69.

Hay, B. A., Wolff, T., and Rubin, G. M. (1994). Expression of baculovirus P35 prevents cell death in Drosophila. Development *120*, 2121-9.

Haynes, S. R., Cooper, M. T., Pype, S., and Stolow, D. T. (1997). Involvement of a tissue-specific RNA recognition motif protein in Drosophila spermatogenesis. Mol Cell Biol *17*, 2708-15.

Hellmuth, K., Lau, D. M., Bischoff, F. R., Kunzler, M., Hurt, E., and Simos, G. (1998). Yeast Los1p has properties of an exportin-like nucleocytoplasmic transport factor for tRNA. Mol Cell Biol *18*, 6374-86.

Helms, W., Lee, H., Ammerman, M., Parks, A. L., Muskavitch, M. A., and Yedvobnick, B. (1999). Engineered truncations in the Drosophila mastermind protein disrupt Notch pathway function. Dev Biol *215*, 358-74.

Hengartner, C. J., Myer, V. E., Liao, S. M., Wilson, C. J., Koh, S. S., and Young, R. A. (1998). Temporal regulation of RNA polymerase II by Srb10 and Kin28 cyclin- dependent kinases. Mol Cell *2*, 43-53.

Henry, M., Borland, C. Z., Bossie, M., and Silver, P. A. (1996). Potential RNA binding proteins in Saccharomyces cerevisiae identified as suppressors of temperature-sensitive mutations in NPL3. Genetics *142*, 103-15.

Hentze, M. W., and Kulozik, A. E. (1999). A perfect message: RNA surveillance and nonsense-mediated decay. Cell *96*, 307-10.

Hernandez, G., del Mar Castellano, M., Agudo, M., and Sierra, J. M. (1998). Isolation and characterization of the cDNA and the gene for eukaryotic translation initiation factor 4G from Drosophila melanogaster. Eur J Biochem *253*, 27-35.

Hernandez, G., Diez del Corral, R., Santoyo, J., Campuzano, S., and Sierra, J. M. (1997). Localization, structure and expression of the gene for translation initiation factor eIF-4E from Drosophila melanogaster. Mol Gen Genet *253*, 624-33.

Hilleren, P., McCarthy, T., Rosbash, M., Parker, R., and Jensen, T. H. (2001). Quality control of mRNA 3'-end processing is linked to the nuclear exosome. Nature *413*, 538-42.

Hinz, U., Giebel, B., and Campos-Ortega, J. A. (1994). The basic-helix-loop-helix domain of Drosophila lethal of scute protein is sufficient for proneural function and activates neurogenic genes. Cell *76*, 77-87.

Hirose, Y., and Manley, J. L. (2000). RNA polymerase II and the integration of nuclear events. Genes Dev *14*, 1415-29.

Hirose, Y., and Manley, J. L. (1998). RNA polymerase II is an essential mRNA polyadenylation factor. Nature *395*, 93-6.

Hirose, Y., Tacke, R., and Manley, J. L. (1999). Phosphorylated RNA polymerase II stimulates pre-mRNA splicing. Genes Dev *13*, 1234-9.

Ho, C. K., Schwer, B., and Shuman, S. (1998). Genetic, physical, and functional interactions between the triphosphatase and guanylyltransferase components of the yeast mRNA capping apparatus. Mol Cell Biol *18*, 5189-98.

Ho, C. K., and 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*, 405-11.

Hoffman, B. E., and Grabowski, P. J. (1992). U1 snRNP targets an essential splicing factor, U2AF65, to the 3' splice site by a network of interactions spanning the exon. Genes Dev *6*, 2554-68.

Huang, Y., and Carmichael, G. C. (1996). Role of polyadenylation in nucleocytoplasmic transport of mRNA. Mol Cell Biol *16*, 1534-42.

Huber, J., Cronshagen, U., Kadokura, M., Marshallsay, C., Wada, T., Sekine, M., and Luhrmann, R. (1998). Snurportin1, an m3G-cap-specific nuclear import receptor with a novel domain structure. Embo J *17*, 4114-26.

Inoue, K., Ohno, M., Sakamoto, H., and Shimura, Y. (1989). Effect of the cap structure on pre-mRNA splicing in Xenopus oocyte nuclei. Genes Dev *3*, 1472-9.

Ishigaki, Y., Li, X., Serin, G., and Maquat, L. E. (2001). Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. Cell *106*, 607-17.

Izaurralde, E., Kutay, U., von Kobbe, C., Mattaj, I. W., and Gorlich, D. (1997). The asymmetric distribution of the constituents of the Ran system is essential for transport into and out of the nucleus. Embo J *16*, 6535-47.

Izaurralde, E., Lewis, J., Gamberi, C., Jarmolowski, A., McGuigan, C., and Mattaj, I. W. (1995). A cap-binding protein complex mediating U snRNA export. Nature *376*, 709-12.

Izaurralde, E., Lewis, J., McGuigan, C., Jankowska, M., Darzynkiewicz, E., and Mattaj, I. W. (1994). A nuclear cap binding protein complex involved in pre-mRNA splicing. Cell *78*, 657-68.

Izaurralde, E., Stepinski, J., Darzynkiewicz, E., and Mattaj, I. W. (1992). A cap binding protein that may mediate nuclear export of RNA polymerase II-transcribed RNAs. J Cell Biol *118*, 1287-95.

Jacobson, M. R., and Pederson, T. (1998). A 7-methylguanosine cap commits U3 and U8 small nuclear RNAs to the nucleolar localization pathway. Nucleic Acids Res *26*, 756-60.

Jan, Y. N., and Jan, L. Y. (1993). HLH proteins, fly neurogenesis, and vertebrate myogenesis. Cell *75*, 827-30.

Jarman, A. P., and Ahmed, I. (1998). The specificity of proneural genes in determining Drosophila sense organ identity. Mech Dev *76*, 117-25.

Jarmolowski, A., Boelens, W. C., Izaurralde, E., and Mattaj, I. W. (1994). Nuclear export of different classes of RNA is mediated by specific factors. J Cell Biol *124*, 627-35.

Johnston, L. A., and Schubiger, G. (1996). Ectopic expression of wingless in imaginal discs interferes with decapentaplegic expression and alters cell determination. Development *122*, 3519-29.

Jove, R., and Manley, J. L. (1984). In vitro transcription from the adenovirus 2 major late promoter utilizing templates truncated at promoter-proximal sites. J Biol Chem *259*, 8513-21.

Kaplan, C. D., Morris, J. R., Wu, C., and Winston, F. (2000). Spt5 and spt6 are associated with active transcription and have characteristics of general elongation factors in D. melanogaster. Genes Dev *14*, 2623-34.

Karsch-Mizrachi, I., and Haynes, S. R. (1993). The Rb97D gene encodes a potential RNA-binding protein required for spermatogenesis in Drosophila. Nucleic Acids Res *21*, 2229-35.

Kass, S., Tyc, K., Steitz, J. A., and Sollner-Webb, B. (1990). The U3 small nucleolar ribonucleoprotein functions in the first step of preribosomal RNA processing. Cell *60*, 897-908.

Kataoka, N., Ohno, M., Kangawa, K., Tokoro, Y., and Shimura, Y. (1994). Cloning of a complementary DNA encoding an 80 kilodalton nuclear cap binding protein. Nucleic Acids Res *22*, 3861-5.

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

Kobor, M. S., Archambault, J., Lester, W., Holstege, F. C., Gileadi, O., Jansma, D. B., Jennings, E. G., Kouyoumdjian, F., Davidson, A. R., Young, R. A., and Greenblatt, J. (1999). An unusual eukaryotic protein phosphatase required for transcription by RNA polymerase II and CTD dephosphorylation in S. cerevisiae. Mol Cell *4*, 55-62.

Koepp, D. M., and Silver, P. A. (1996). A GTPase controlling nuclear trafficking: running the right way or walking RANdomly? Cell *87*, 1-4.

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

Konarska, M. M., Padgett, R. A., and Sharp, P. A. (1984). Recognition of cap structure in splicing in vitro of mRNA precursors. Cell *38*, 731-6.

Kozak, M. (1989). The scanning model for translation: an update. J Cell Biol *108*, 229-41.

Krainer, A. R., Maniatis, T., Ruskin, B., and Green, M. R. (1984). Normal and mutant human beta-globin pre-mRNAs are faithfully and efficiently spliced in vitro. Cell *36*, 993-1005.

Kraus, M. E., and Lis, J. T. (1994). The concentration of B52, an essential splicing factor and regulator of splice site choice in vitro, is critical for Drosophila development. Mol Cell Biol *14*, 5360-70.

Krecic, A. M., and Swanson, M. S. (1999). hnRNP complexes: composition, structure, and function. Curr Opin Cell Biol *11*, 363-71.

Kutay, U., Lipowsky, G., Izaurralde, E., Bischoff, F. R., Schwarzmaier, P., Hartmann, E., and Gorlich, D. (1998). Identification of a tRNA-specific nuclear export receptor. Mol Cell *1*, 359-69.

Lasko, P. (2000). The drosophila melanogaster genome: translation factors and RNA binding proteins. J Cell Biol *150*, F51-6.

Lee, M. S., Henry, M., and Silver, P. A. (1996). A protein that shuttles between the nucleus and the cytoplasm is an important mediator of RNA export. Genes Dev *10*, 1233-46.

Lewis, J. D., and Izaurralde, E. (1997). The role of the cap structure in RNA processing and nuclear export. Eur J Biochem *247*, 461-9.

Lewis, J. D., Izaurralde, E., Jarmolowski, A., McGuigan, C., and Mattaj, I. W. (1996). A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. Genes Dev *10*, 1683-98.

Littleton, J. T., Stern, M., Schulze, K., Perin, M., and Bellen, H. J. (1993). Mutational analysis of Drosophila synaptotagmin demonstrates its essential role in Ca(2+)-activated neurotransmitter release. Cell *74*, 1125-34.

Lu, H., Flores, O., Weinmann, R., and Reinberg, D. (1991). The nonphosphorylated form of RNA polymerase II preferentially associates with the preinitiation complex. Proc Natl Acad Sci U S A *88*, 10004-8.

MacDonald, C. C., Wilusz, J., and Shenk, T. (1994). The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location. Mol Cell Biol *14*, 6647-54.

Makkerh, J. P. S., Dingwall, C., and Laskey, R. A. (1996). Comparative mutagenesis of nuclear localization signals reveals the importance of neutral and acidic amino acids. Curr Biol *6*, 1025-7.

Mancebo, H. S., Lee, G., Flygare, J., Tomassini, J., Luu, P., Zhu, Y., Peng, J., Blau, C., Hazuda, D., Price, D., and Flores, O. (1997). P-TEFb kinase is required for HIV Tat transcriptional activation in vivo and in vitro. Genes Dev *11*, 2633-44.

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

Marshall, N. F., Peng, J., Xie, Z., and Price, D. H. (1996). Control of RNA polymerase II elongation potential by a novel carboxyl- terminal domain kinase. J Biol Chem *271*, 27176-83.

Marshall, N. F., and Price, D. H. (1995). Purification of P-TEFb, a transcription factor required for the transition into productive elongation. J Biol Chem *270*, 12335-8.

Mattaj, I. W. (1986). Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding. Cell *46*, 905-11.

Mattaj, I. W. (1993). RNA recognition: a family matter? Cell *73*, 837-40.

Mazza, C., Ohno, M., Segref, A., Mattaj, I. W., and Cusack, S. (2001). Crystal structure of the human nuclear cap binding complex. Mol Cell *8*, 383-96.

McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Shuman, S., and Bentley, D. L. (1997a). 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. Genes Dev *11*, 3306-18.

McGuffin, M. E., Chandler, D., Somaiya, D., Dauwalder, B., and Mattox, W. (1998). Autoregulation of transformer-2 alternative splicing is necessary for normal male fertility in Drosophila. Genetics *149*, 1477-86.

McKendrick, L., Thompson, E., Ferreira, J., Morley, S. J., and Lewis, J. D. (2001). Interaction of eukaryotic translation initiation factor 4G with the nuclear cap-binding complex provides a link between nuclear and cytoplasmic functions of the m(7) guanosine cap. Mol Cell Biol *21*, 3632-41.

Mehlin, H., Daneholt, B., and Skoglund, U. (1995). Structural interaction between the nuclear pore complex and a specific translocating RNP particle. J Cell Biol *129*, 1205-16.

Mehlin, H., Daneholt, B., and Skoglund, U. (1992). Translocation of a specific premessenger ribonucleoprotein particle through the nuclear pore studied with electron microscope tomography. Cell *69*, 605-13.

Mendell, J. T., Medghalchi, S. M., Lake, R. G., Noensie, E. N., and Dietz, H. C. (2000). Novel Upf2p orthologues suggest a functional link between translation initiation and nonsense surveillance complexes. Mol Cell Biol *20*, 8944-57.

Meyer, V., Oliver, B., and Pauli, D. (1998). Multiple developmental requirements of noisette, the Drosophila homolog of the U2 snRNP-associated polypeptide SP3a60. Mol Cell Biol *18*, 1835-43.

Min, H., Chan, R. C., and Black, D. L. (1995). The generally expressed hnRNP F is involved in a neural-specific pre- mRNA splicing event. Genes Dev *9*, 2659-71.

Minvielle-Sebastia, L., and Keller, W. (1999). mRNA polyadenylation and its coupling to other RNA processing reactions and to transcription. Curr Opin Cell Biol *11*, 352-7.

Moore, M. J., and Sharp, P. A. (1993). Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. Nature *365*, 364-8.

Mortillaro, M. J., Blencowe, B. J., Wei, X., Nakayasu, H., Du, L., Warren, S. L., Sharp, P. A., and Berezney, R. (1996). A hyperphosphorylated form of the large subunit of RNA polymerase II is associated with splicing complexes and the nuclear matrix. Proc Natl Acad Sci U S A *93*, 8253-7.

Mount, S. M., and Salz, H. K. (2000). Pre-messenger RNA processing factors in the Drosophila genome. J Cell Biol *150*, F37-44.

Nakielny, S., and Dreyfuss, G. (1997). Nuclear export of proteins and RNAs. Curr Opin Cell Biol *9*, 420-9.

O'Brien, T., Hardin, S., Greenleaf, A., and Lis, J. T. (1994). Phosphorylation of RNA polymerase II C-terminal domain and transcriptional elongation. Nature *370*, 75-7.

O'Mullane, L., and Eperon, I. C. (1998). The pre-mRNA 5' cap determines whether U6 small nuclear RNA succeeds U1 small nuclear ribonucleoprotein particle at 5' splice sites. Mol Cell Biol *18*, 7510-20.

Ohno, M., Kataoka, N., and Shimura, Y. (1990). A nuclear cap binding protein from HeLa cells. Nucleic Acids Res *18*, 6989-95.

Ohno, M., Sakamoto, H., and Shimura, Y. (1987). Preferential excision of the 5' proximal intron from mRNA precursors with two introns as mediated by the cap structure. Proc Natl Acad Sci U S A *84*, 5187-91.

Ohno, M., Segref, A., Bachi, A., Wilm, M., and Mattaj, I. W. (2000). PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation. Cell *101*, 187-98.

Palacios, I., Hetzer, M., Adam, S. A., and Mattaj, I. W. (1997). Nuclear import of U snRNPs requires importin beta. Embo J *16*, 6783-92.

Pan, D., and Rubin, G., M. (1997). Kuzbanian controls proteolytic processing of notch and mediates lateral inhibition during Drosophila and Vertebrate neurogenesis. Cell *90*, 271-280.

Patturajan, M., Wei, X., Berezney, R., and Corden, J. L. (1998). A nuclear matrix protein interacts with the phosphorylated C-terminal domain of RNA polymerase II. Mol Cell Biol *18*, 2406-15.

Patzelt, E., Blaas, D., and Kuechler, E. (1983). CAP binding proteins associated with the nucleus. Nucleic Acids Res *11*, 5821-35.

Patzelt, E., Thalmann, E., Hartmuth, K., Blaas, D., and Kuechler, E. (1987). Assembly of pre-mRNA splicing complex is cap dependent. Nucleic Acids Res *15*, 1387-99.

Peculis, B. A. (1997). The sequence of the 5' end of the U8 small nucleolar RNA is critical for 5.8S and 28S rRNA maturation. Mol Cell Biol *17*, 3702-13.

Peculis, B. A., and Steitz, J. A. (1993). Disruption of U8 nucleolar snRNA inhibits 5.8S and 28S rRNA processing in the Xenopus oocyte. Cell *73*, 1233-45.

Peng, J., Zhu, Y., Milton, J. T., and Price, D. H. (1998). Identification of multiple cyclin subunits of human P-TEFb. Genes Dev *12*, 755-62.

Peng, X., and Mount, S. M. (1995). Genetic enhancement of RNA-processing defects by a dominant mutation in B52, the Drosophila gene for an SR protein splicing factor. Mol Cell Biol *15*, 6273-82.

Pestova, T. V., Kolupaeva, V. G., Lomakin, I. B., Pilipenko, E. V., Shatsky, I. N., Agol, V. I., and Hellen, C. U. (2001). Molecular mechanisms of translation initiation in eukaryotes. Proc Natl Acad Sci U S A *98*, 7029-36.

Polycarpou-Schwarz, M., Gunderson, S. I., Kandels-Lewis, S., Seraphin, B., and Mattaj, I. W. (1996). Drosophila SNF/D25 combines the functions of the two snRNP proteins U1A and U2B' that are encoded separately in human, potato, and yeast. RNA *2*, 11-23.

Ponting, C. P. (2000). Novel eIF4G domain homologues linking mRNA translation with nonsense-mediated mRNA decay. Trends Biochem Sci *25*, 423-6.

Preiss, T., and Hentze, M. W. (1998). Dual function of the messenger RNA cap structure in poly(A)-tail- promoted translation in yeast. Nature *392*, 516-20.

Price, D. H. (2000). P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II. Mol Cell Biol *20*, 2629-34.

Proudfoot, N. (1991). Poly(A) signals. Cell *64*, 671-4.

Rasmussen, E. B., and Lis, J. T. (1993). In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. Proc Natl Acad Sci U S A *90*, 7923-7.

Reed, R. (2000). Mechanisms of fidelity in pre-mRNA splicing. Curr Opin Cell Biol *12*, 340-5.

Reed, R and Hurt, E. (2002). A conserved mRNA export machinery coupled to pre-mRNA splicing. Cell *108*, 523-31.

Reines, D., Conaway, R. C., and Conaway, J. W. (1999). Mechanism and regulation of transcriptional elongation by RNA polymerase II. Curr Opin Cell Biol *11*, 342-6.

Richard, D. S., Gilbert, M., Crum, B., Hollinshead, D. M., Schelble, S., and Scheswohl, D. (2001). Yolk protein endocytosis by oocytes in Drosophila melanogaster: immunofluorescent localization of clathrin, adaptin and the yolk protein receptor. J Insect Physiol *47*, 715-723.

Robberson, B. L., Cote, G. J., and Berget, S. M. (1990). Exon definition may facilitate splice site selection in RNAs with multiple exons. Mol Cell Biol *10*, 84-94.

Robertson, H. M., Preston, C. R., Phillis, R. W., Johnson-Schlitz, D. M., Benz, W. K., and Engels, W. R. (1988). A stable genomic source of P element transposase in Drosophila melanogaster. Genetics *118*, 461-70.

Rodriguez, C. R., Cho, E. J., Keogh, M. C., Moore, C. L., Greenleaf, A. L., and Buratowski, S. (2000). Kin28, the TFIIH-associated carboxy-terminal domain kinase, facilitates the recruitment of mRNA processing machinery to RNA polymerase II. Mol Cell Biol *20*, 104-12.

Rorth, P., Szabo, K., Bailey, A., Laverty, T., Rehm, J., Rubin, G. M., Weigmann, K., Milan, M., Benes, V., Ansorge, W., and Cohen, S. M. (1998). Systematic gain-of-function genetics in Drosophila. Development *125*, 1049-57.

Rozen, F., and Sonenberg, N. (1987). Identification of nuclear cap specific proteins in HeLa cells. Nucleic Acids Res *15*, 6489-500.

Rudner, D. Z., Breger, K. S., Kanaar, R., Adams, M. D., and Rio, D. C. (1998). RNA binding activity of heterodimeric splicing factor U2AF: at least one RS domain is required for high-affinity binding. Mol Cell Biol *18*, 4004-11.

Rudner, D. Z., Kanaar, R., Breger, K. S., and Rio, D. C. (1996). Mutations in the small subunit of the Drosophila U2AF splicing factor cause lethality and developmental defects. Proc Natl Acad Sci U S A *93*, 10333-7.

Ryan, K. J., and Wente, S. R. (2000). The nuclear pore complex: a protein machine bridging the nucleus and cytoplasm. Curr Opin Cell Biol *12*, 361-71.

Sachs, A. B., Sarnow, P., and Hentze, M. W. (1997). Starting at the beginning, middle, and end: translation initiation in eukaryotes. Cell *89*, 831-8.

Saha, N., Schwer, B., and Shuman, S. (1999). Characterization of human, Schizosaccharomyces pombe, and Candida albicans mRNA cap methyltransferases and complete replacement of the yeast capping apparatus by mammalian enzymes. J Biol Chem *274*, 16553-62.

Salditt-Georgieff, M., Harpold, M., Chen-Kiang, S., and Darnell, J. E., Jr. (1980). The addition of 5' cap structures occurs early in hnRNA synthesis and prematurely terminated molecules are capped. Cell *19*, 69-78.

Salinas, C. A., Sinclair, D. A., O'Hare, K., and Brock, H. W. (1998). Characterization of a Drosophila homologue of the 160-kDa subunit of the cleavage and polyadenylation specificity factor CPSF. Mol Gen Genet *257*, 672-80.

Salz, H. K., and Flickinger, T. W. (1996). Both loss-of-function and gain-of-function mutations in snf define a role for snRNP proteins in regulating Sex-lethal pre-mRNA splicing in Drosophila development. Genetics *144*, 95-108.

Salz, H. K., Flickinger, T. W., Mittendorf, E., Pellicena-Palle, A., Petschek, J. P., and Albrecht, E. B. (1994). The Drosophila maternal effect locus deadhead encodes a thioredoxin homolog required for female meiosis and early embryonic development. Genetics *136*, 1075-86.

Schroeder, S. C., Schwer, B., Shuman, S., and Bentley, D. (2000). Dynamic association of capping enzymes with transcribing RNA polymerase II. Genes Dev *14*, 2435-40.

Schwer, B., Mao, X., and Shuman, S. (1998). Accelerated mRNA decay in conditional mutants of yeast mRNA capping enzyme. Nucleic Acids Res *26*, 2050-7.

Schwer, B., and Shuman, S. (1994). Mutational analysis of yeast mRNA capping enzyme. Proc Natl Acad Sci U S A *91*, 4328-32.

Segalat, L., Perichon, R., Bouly, J. P., and Lepesant, J. A. (1992). The Drosophila pourquoi-pas?/wings-down zinc finger protein: oocyte nucleus localization and embryonic requirement. Genes Dev *6*, 1019-29.

Shatkin, A. J. (1976). Capping of eucaryotic mRNAs. Cell *9*, 645-53.

Shatkin, A. J. (1985). mRNA cap binding proteins: essential factors for initiating translation. Cell *40*, 223-4.

Shen, E. C., Stage-Zimmermann, T., Chui, P., and Silver, P. A. (2000). 7The yeast mRNA-binding protein Npl3p interacts with the cap-binding complex. J Biol Chem *275*, 23718-24.

Shibagaki, Y., Itoh, N., Yamada, H., Nagata, S., and Mizumoto, K. (1992). mRNA capping enzyme. Isolation and characterization of the gene encoding mRNA guanylytransferase subunit from Saccharomyces cerevisiae. J Biol Chem *267*, 9521-8.

Shimotohno, K., Kodama, Y., Hashimoto, J., and Miura, K. I. (1977). Importance of 5'-terminal blocking structure to stabilize mRNA in eukaryotic protein synthesis. Proc Natl Acad Sci U S A *74*, 2734-8.

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

Shuman, S., and Schwer, B. (1995). RNA capping enzyme and DNA ligase: a superfamily of covalent nucleotidyl transferases. Mol Microbiol *17*, 405-10.

Singleton, D. R., Chen, S., Hitomi, M., Kumagai, C., and Tartakoff, A. M. (1995). A yeast protein that bidirectionally affects nucleocytoplasmic transport. J Cell Sci *108*, 265-72.

Sonenberg, N., Morgan, M. A., Merrick, W. C., and Shatkin, A. J. (1978). A polypeptide in eukaryotic initiation factors that crosslinks specifically to the 5'-terminal cap in mRNA. Proc Natl Acad Sci U S A *75*, 4843-7.

Sonenberg, N., Rupprecht, K. M., Hecht, S. M., and Shatkin, A. J. (1979). Eukaryotic mRNA cap binding protein: purification by affinity chromatography on sepharose-coupled m7GDP. Proc Natl Acad Sci U S A *76*, 4345-9.

Speicher, S. A., Thomas, U., Hinz, U., and Knust, E. (1994). The Serrate locus of Drosophila and its role in morphogenesis of the wing imaginal discs: control of cell proliferation. Development *120*, 535-44.

Staehling-Hampton, K., Hoffmann, F. M., Baylies, M. K., Rushton, E., and Bate, M. (1994). dpp induces mesodermal gene expression in Drosophila. Nature *372*, 783-6.

Staley, J. P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. Cell *92*, 315-26.

Standart, N., and Jackson, R. (1994). Translational regulation. Y the message is masked? Curr Biol *4*, 939-41.

Stoffler, D., Fahrenkrog, B., and Aebi, U. (1999). The nuclear pore complex: from molecular architecture to functional dynamics. Curr Opin Cell Biol *11*, 391-401.

Svejstrup, J. Q., Vichi, P., and Egly, J. M. (1996). The multiple roles of transcription/repair factor TFIIH. Trends Biochem Sci *21*, 346-50.

Tanner, S., Stagljar, I., Georgiev, O., Schaffner, W., and Bourquin, J. P. (1997). A novel SR-related protein specifically interacts with the carboxy- terminal domain (CTD) of RNA polymerase II through a conserved interaction domain. Biol Chem *378*, 565-71.

Tarun, S. Z., Jr., Wells, S. E., Deardorff, J. A., and Sachs, A. B. (1997). Translation initiation factor eIF4G mediates in vitro poly(A) tail- dependent translation. Proc Natl Acad Sci U S A *94*, 9046-51.

Tautz, D., and Pfeifle, C. (1989). A non-radioactive in situ hybridization method for the localization of specific RNAs in Drosophila embryos reveals translational control of the segmentation gene hunchback. Chromosoma *98*, 81-5.

Torok, I., Strand, D., Schmitt, R., Tick, G., Torok, T., Kiss, I., and Mechler, B. M. (1995). The overgrown hematopoietic organs-31 tumor suppressor gene of Drosophila encodes an Importin-like protein accumulating in the nucleus at the onset of mitosis. J Cell Biol *129*, 1473-89.

Tower, J., Karpen, G. H., Craig, N., and Spradling, A. C. (1993). Preferential transposition of Drosophila P elements to nearby chromosomal sites. Genetics *133*, 347-59.

Uemura, H., and Jigami, Y. (1992). GCR3 encodes an acidic protein that is required for expression of glycolytic genes in Saccharomyces cerevisiae. J Bacteriol *174*, 5526-32.

Venables, J. P., and Eperon, I. (1999). The roles of RNA-binding proteins in spermatogenesis and male infertility. Curr Opin Genet Dev *9*, 346-54.

Vincent, M., Lauriault, P., Dubois, M. F., Lavoie, S., Bensaude, O., and Chabot, B. (1996). The nuclear matrix protein p255 is a highly phosphorylated form of RNA polymerase II largest subunit which associates with spliceosomes. Nucleic Acids Res *24*, 4649-52.

Visa, N., Izaurralde, E., Ferreira, J., Daneholt, B., and Mattaj, I. W. (1996). A nuclear cap-binding complex binds Balbiani ring pre-mRNA cotranscriptionally and accompanies the ribonucleoprotein particle during nuclear export. J Cell Biol *133*, 5-14.

Vorbruggen, G., Onel, S., and Jackle, H. (2000). Restricted expression and subnuclear localization of the Drosophila gene Dnop5, a member of the Nop/Sik family of the conserved rRNA processing factors. Mech Dev *90*, 305-8.

Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G. A., Winston, F., Buratowski, S., and Handa, H. (1998b). DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. Genes Dev *12*, 343-56.

Wada, T., Takagi, T., Yamaguchi, Y., Watanabe, D., and Handa, H. (1998a). Evidence that P-TEFb alleviates the negative effect of DSIF on RNA polymerase II-dependent transcription in vitro. Embo J *17*, 7395-403.

Wahle, E. (1991). Purification and characterization of a mammalian polyadenylate polymerase involved in the 3' end processing of messenger RNA precursors. J Biol Chem *266*, 3131-9.

Wahle, E., and Kuhn, U. (1997). The mechanism of 3' cleavage and polyadenylation of eukaryotic pre-mRNA. Prog Nucleic Acid Res Mol Biol *57*, 41-71.

Wahle, E., Lustig, A., Jeno, P., and Maurer, P. (1993). Mammalian poly(A)-binding protein II. Physical properties and binding to polynucleotides. J Biol Chem *268*, 2937-45.

Wahle, E., and Ruegsegger, U. (1999). 3'-End processing of pre-mRNA in eukaryotes. FEMS Microbiol Rev *23*, 277-95.

Wan, L., Dockendorff, T. C., Jongens, T. A., and Dreyfuss, G. (2000). Characterization of dFMR1, a drosophila melanogaster homolog of the fragile X mental retardation protein. Mol Cell Biol *20*, 8536-47.

Wang, S. P., and Shuman, S. (1997). Structure-function analysis of the mRNA cap methyltransferase of Saccharomyces cerevisiae. J Biol Chem *272*, 14683-9.

Weeks, J. R., Hardin, S. E., Shen, J., Lee, J. M., and Greenleaf, A. L. (1993). Locus-specific variation in phosphorylation state of RNA polymerase II in vivo: correlations with gene activity and transcript processing. Genes Dev *7*, 2329-44.

Wei, P., Garber, M. E., Fang, S. M., Fischer, W. H., and Jones, K. A. (1998). A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. Cell *92*, 451-62.

Wells, S. E., Hillner, P. E., Vale, R. D., and Sachs, A. B. (1998). Circularization of mRNA by eukaryotic translation initiation factors. Mol Cell *2*, 135-40.

Wen, Y., and Shatkin, A. J. (1999). Transcription elongation factor hSPT5 stimulates mRNA capping. Genes Dev *13*, 1774-9.

Wickens, M., Anderson, P., and Jackson, R. J. (1997). Life and death in the cytoplasm: messages from the 3' end. Curr Opin Genet Dev *7*, 220-32.

Wilkie, GS., Zimyanin, V., Kirby, R., Korey, C., Francis-Lang H., Van Vactor D., Davis I. (2001). Small bristles, the Drosophila ortholog of NXF-1, is essential for mRNA export throughout development. RNA *12*,1781-92.

Wilson, K. F., Fortes, P., Singh, U. S., Ohno, M., Mattaj, I. W., and Cerione, R. A. (1999). The nuclear cap-binding complex is a novel target of growth factor receptor-coupled signal transduction. J Biol Chem *274*, 4166-73.

Wilson, K. F., Wu, W. J., and Cerione, R. A. (2000). Cdc42 stimulates RNA splicing via the S6 kinase and a novel S6 kinase target, the nuclear cap-binding complex. J Biol Chem *275*, 37307-10.

Wu-Baer, F., Lane, W. S., and Gaynor, R. B. (1998). Role of the human homolog of the yeast transcription factor SPT5 in HIV- 1 Tat-activation. J Mol Biol *277*, 179-97.

Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., and Handa, H. (1999a). NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. Cell *97*, 41-51.

Yamaguchi, Y., Wada, T., Watanabe, D., Takagi, T., Hasegawa, J., and Handa, H. (1999b). Structure and function of the human transcription elongation factor DSIF. J Biol Chem *274*, 8085-92.

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

Yuryev, A., Patturajan, M., Litingtung, Y., Joshi, R. V., Gentile, C., Gebara, M., and Corden, J. L. (1996). The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins. Proc Natl Acad Sci U S A *93*, 6975-80.

Zeng, C., and Berget, S. M. (2000). Participation of the C-terminal domain of RNA polymerase II in exon definition during pre-mRNA splicing. Mol Cell Biol *20*, 8290-301.

Zhang, G., Taneja, K. L., Singer, R. H., and Green, M. R. (1994). Localization of pre-mRNA splicing in mammalian nuclei. Nature *372*, 809-12.

Zhang, P., and Spradling, A. C. (1993). Efficient and dispersed local P element transposition from Drosophila females. Genetics *133*, 361-73.

Zhao, J., Hyman, L., and Moore, C. (1999). Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. Microbiol Mol Biol Rev *63*, 405-45.

Zhou, M., Halanski, M. A., Radonovich, M. F., Kashanchi, F., Peng, J., Price, D. H., and Brady, J. N. (2000). Tat modifies the activity of CDK9 to phosphorylate serine 5 of the RNA polymerase II carboxyl-terminal domain during human immunodeficiency virus type 1 transcription. Mol Cell Biol *20*, 5077-86.

Zhu, Y., Pe'ery, T., Peng, J., Ramanathan, Y., Marshall, N., Marshall, T., Amendt, B., Mathews, M. B., and Price, D. H. (1997). Transcription elongation factor P-TEFb is required for HIV-1 tat transactivation in vitro. Genes Dev *11*, 2622-32.