*Critical Reviews in Biochemistry and Molecular Biology*, 2011; 46(1): 11–26

informa
healthcare

REVIEW ARTICLE

# Regulation of snRNA gene expression by the *Drosophila melanogaster* small nuclear RNA activating protein complex (DmSNAPc)

Ko-Hsuan Hung[1], and William E. Stumph[2]

[1]*Department of Biology, and* [2]*Department of Chemistry and Biochemistry and Molecular Biology Institute, San Diego State University, San Diego, CA, USA*

**Abstract**

The small nuclear RNAs (snRNAs) are an essential class of non-coding RNAs first identified over 30 years ago. Many of the well-characterized snRNAs are involved in RNA processing events. However, it is now evident that other small RNAs, synthesized using similar mechanisms, play important roles at many stages of gene expression. The accurate and efficient control of the expression of snRNA (and related) genes is thus critical for cell survival. All snRNA genes share a very similar promoter structure, and their transcription is dependent upon the same multi-subunit transcription factor, termed the snRNA activating protein complex (SNAPc). Despite those similarities, some snRNA genes are transcribed by RNA polymerase II (Pol II), but others are transcribed by RNA polymerase III (Pol III). Thus snRNA genes provide a unique opportunity to understand how RNA polymerase specificity is determined and how distinct transcription machineries can interact with a common factor. This review will describe efforts taken toward solving those questions by using the fruit fly as a model organism. *Drosophila melanogaster* SNAPc (DmSNAPc) binds to a proximal sequence element (PSEA) present in both Pol II and Pol III snRNA promoters. Just a few differences in nucleotide sequence in the Pol II and Pol III PSEAs play a major role in determining RNA polymerase specificity. Furthermore, these same nucleotide differences result in alternative conformations of DmSNAPc on Pol II and Pol III snRNA gene promoters. It seems likely that these DNA-induced alternative DmSNAPc conformations are responsible for the differential recruitment of the distinct transcriptional machineries.

**Keywords:** *protein–DNA interaction; transcription initiation; pre-initiation complex; RNA polymerase specificity; promoter sequences; site-specific protein–DNA photo-cross-linking*





## Introduction

The small nuclear RNA activating protein complex (SNAPc) is a unique multi-subunit protein complex required for the synthesis of small nuclear RNAs (snRNAs) (Murphy *et al.*, 1992; Sadowski *et al.*, 1993; Yoon *et al.*, 1995; Henry *et al.*, 1995; Su *et al.*, 1997; Li *et al.*, 2004). The snRNAs are non-coding RNA molecules highly expressed in eukaryotic cells, and each is a product of its own independent transcription unit. The snRNAs are involved in many essential cellular functions such as pre-mRNA splicing, rRNA processing and histone mRNA 3′ end-formation (Steitz *et al.*, 1988; Kass *et al.*, 1990; Bond *et al.*, 1991; Guthrie, 1991; Peculis and Steitz, 1993; Sharp, 1994). In animals, most snRNAs (e.g. U1, U2, U3, U4, U5 and U7) are synthesized by RNA polymerase II (Pol II), but other small RNAs (e.g. U6 snRNA, 7SK RNA, $tRNA^{Sec}$, H1 RNA, and MRP RNA) are synthesized by RNA polymerase III (Pol III) (Zieve *et al.*, 1977; Dahlberg and Lund, 1988; Lee *et al.*, 1989; Parry *et al.*, 1989; Baer *et al.*, 1990; Yuan and Reddy, 1991; Hernandez, 1992; Lobo and Hernandez, 1994).

Interestingly, despite this differential requirement of RNA polymerase, the genes coding for all these small RNAs share a very similar promoter structure. In animals, snRNA promoters are characterized by a unique and essential upstream promoter element termed the proximal sequence element (PSE) (more specifically called

*Address for Correspondence*: William E. Stumph, Department of Chemistry and Biochemistry, San Diego State University, San Diego, CA 92182-1030, USA. Tel: +1 619 594 5575; E-mail: wstumph@sciences.sdsu.edu

*(Received 01 July 2010; revised 19 August 2010; accepted 20 August 2010)*

ISSN 1040-9238 print/ISSN 1549-7798 online 
DOI: 10.3109/10409238.2010.518136

the PSEA in *Drosophila melanogaster*) (Das *et al.*, 1987; Dahlberg and Lund, 1988; Zamrod *et al.*, 1993; Lobo and Hernandez, 1994). Despite this commonality, evidence indicates that the mechanisms involved in determining RNA polymerase specificity can be different in evolutionarily divergent organisms. In vertebrates, a TATA box present downstream of the PSE acts as a dominant determinant of Pol III specificity, whereas the absence of the TATA box results in snRNA gene transcription by Pol II (Mattaj *et al.*, 1988; Lobo and Hernandez, 1989). In flies, on the other hand, the exact sequence of the PSEA itself acts as the primary determinant of RNA polymerase specificity (Jensen *et al.*, 1998; McNamara-Schroeder *et al.*, 2001; Lai *et al.*, 2005). As described in greater detail in the following section, plant snRNA genes utilize a still different mechanism to determine polymerase specificity.

The PSEs of all snRNA genes (whether Pol II or Pol III-specific) are recognized and bound by the same evolutionarily conserved transcription factor, SNAPc. The interaction between the PSE and SNAPc initiates the recruitment of Pol II- or Pol III-specific factors for snRNA transcription (Sadowski *et al.*, 1993; Kuhlman *et al.*, 1999; Schramm *et al.*, 2000; Teichmann *et al.*, 2000; Cabart and Murphy, 2001; 2002; Das *et al.*, 2005; Schimanski *et al.*, 2005; Lee *et al.*, 2007; Barakat and Stumph, 2008). However, as mentioned above, the pathway for achieving RNA polymerase specificity can vary among distantly related organisms. This unusual scenario makes the snRNA genes an intriguing system for investigating how RNA polymerase specificity is determined and how a common factor (SNAPc in this case) is able to recruit different transcription machineries.

The transcription of animal snRNA genes has been studied most thoroughly in vertebrates (particularly the human system) and in the fruit fly *D. melanogaster*. Interestingly, fruit fly snRNA gene promoters exhibit a higher degree of conservation with regard to both sequence and location of promoter elements than generally observed in other animals, particularly vertebrates (Dahlberg and Lund, 1988; Lo and Mount, 1990; Hernandez *et al.*, 2007; Jawdekar and Henry, 2008). As a result, the specific molecular interactions that govern snRNA gene activation and RNA polymerase specificity may be more apparent and more accessible for study in fruit flies in comparison to other systems. Reviews on the transcriptional regulation of human snRNA genes have recently been published by Jawdekar and Henry (2008) and by Egloff *et al.* (2008). This article will therefore concentrate on mechanisms of snRNA transcription in the fruit fly while attempting to place this information into the context of the knowledge available from other systems.

First, the structure of snRNA (and snRNA-like) promoters and the contribution of these promoter sequences to determining RNA polymerase specificity will be reviewed with an emphasis on the fruit fly. This will be followed by a discussion of the evolutionarily conserved and non-conserved structural features of DmSNAPc in comparison to the homologous proteins from other organisms. Next the mapping and localization of functional domains within each of the DmSNAPc subunits will be described and compared to results published in the human system. We will then review evidence that DmSNAPc adopts different conformational states on Pol II and Pol III snRNA promoters as an allosteric effect of DNA sequence recognized. Finally, we will speculate on how these conformational differences of the DmSNAPc–DNA complex may lead to RNA polymerase specificity on Pol II and Pol III snRNA gene promoters.

## Structure and RNA polymerase specificity of Drosophila snRNA gene promoters

Early comparisons of DNA sequences identified conserved blocks of sequence upstream of cloned *D. melanogaster* snRNA genes as putative promoter elements (Beck *et al.*, 1984; Saba *et al.*, 1986; Das *et al.*, 1987; Lo and Mount, 1990). The functionality of these conserved elements was subsequently demonstrated by *in vitro* and *in vivo* transcription assays of mutated versus wild-type templates (Zamrod *et al.*, 1993; Jensen *et al.*, 1998; McNamara-Schroeder *et al.*, 2001; Lai *et al.*, 2005). Figure 1A schematically shows the promoter structure of fly snRNA genes transcribed either by Pol II or by Pol III. For purposes of comparison, the positions of the analogous Pol II and Pol III snRNA gene promoter elements in vertebrates and plants are illustrated in the lower parts of Figure 1A.

The PSEA is a more specific name for the insect PSE and is a unique ~21 base pair (bp) element located within a region 40–65 nucleotides upstream of the transcription start site. It was named the PSEA to distinguish it from the PSEB, a moderately conserved 8 bp promoter element located downstream of the PSEA in the insect Pol II-transcribed snRNA genes (Figure 1A). In contrast, the Pol III-transcribed snRNA genes possess a strongly conserved 8 bp TATA box instead of the PSEB downstream of the PSEA (Figure 1A). Interestingly, there is an 8 bp separation of the PSEB from the PSEA but a 12 bp separation of the TATA sequence from the PSEA, and these distinctive separations are strictly conserved among the Pol II and Pol III fly snRNA genes (Figure 1B). Like the vertebrate PSE (Dahlberg and Lund, 1988; Parry *et al.*, 1989; Lobo and Hernandez, 1994), the fly PSEA is the dominant element for specifying the transcription start site and is essential for snRNA promoter activity (Zamrod and Stumph, 1990; Jensen *et al.*, 1998; McNamara-Schroeder *et al.*, 2001; Lai *et al.*, 2005).

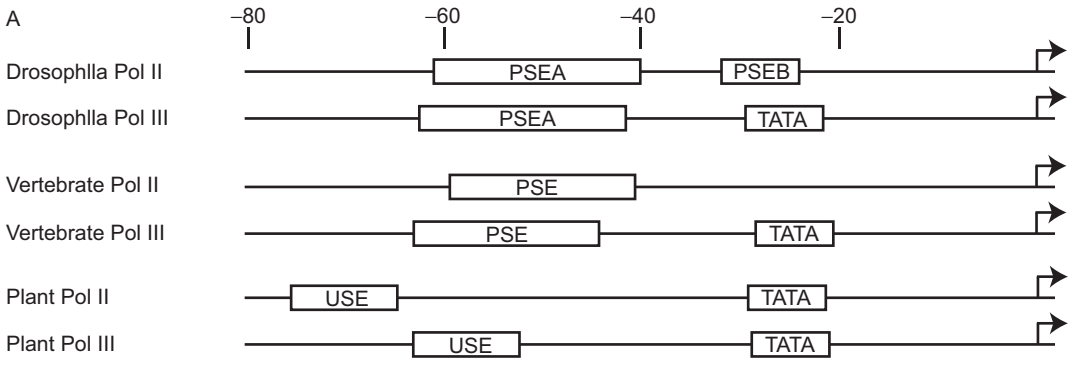



**Figure 1.** Promoter structure of snRNA genes. (A) Schematic representation of conserved promoter elements in the 5′ -flanking DNA of snRNA genes transcribed either by Pol II or by Pol III in *Drosophila*, vertebrates, and plants. The numbers at the top indicate the distance in base pairs upstream of the transcription start site. Abbreviations: PSE, proximal sequence element; PSEA, proximal sequence element A; PSEB, proximal sequence element B; USE, upstream sequence element; TATA, TATA box. (B) Alignment and analysis of snRNA and snRNA-like gene promoters of *D. melanogaster*. The sequences of the promoter regions of 23 Pol II-transcribed *D. melanogaster* snRNA genes are shown in the upper section. The schematic diagram above the sequences indicates the boundaries of the Pol II PSEA and of the PSEB that were originally identified as 21 bp and 8 bp elements respectively. The shaded areas represent possible extensions of the PSEA based upon the more comprehensive set of sequences shown in the figure. The numbers below the schematic diagram (–10, 1, 21, 30, 37) indicate the base positions relative to the traditional first nucleotide of the PSEA. The numbers above the schematic (–61, –41, –32, –25) indicate the approximate distances upstream of the transcription start site, but these can vary by up to 2 bp depending upon the particular gene. The lower section shows the sequences of the promoter regions of four snRNA and three snRNA-like genes transcribed by Pol III. The schematic drawing at the bottom indicates the boundaries of the Pol III PSEA and TATA box. The shading and numbering are similar to that described above for the Pol II schematic drawing. Nucleotides shown in boldface type at PSEA positions 19 and 20 indicate nucleotides never found at the corresponding position in the promoter sequences of genes transcribed by the opposite RNA polymerase (Pol II versus Pol III). Immediately below the Pol II gene sequences and above the Pol III gene sequences are shown consensus promoter sequences for the Pol II- and Pol III-transcribed snRNA (and snRNA-like genes), respectively. The nucleotides overlined in the Pol II consensus sequence and those underlined in the Pol III consensus sequence indicate nucleotides that are 100% conserved in all aligned Pol II and Pol III promoter sequences, respectively. Between the Pol II and Pol III consensus promoter sequences, a Pol II/III consensus PSEA is shown. A Pol II/III consensus nucleotide is shown only if it is present in both individual consensus promoter sequences. The bold Xs indicate positions where *different* nucleotides are always (or nearly always) preferred in the Pol III versus Pol II gene PSEAs.

Figure 1. continued on next page

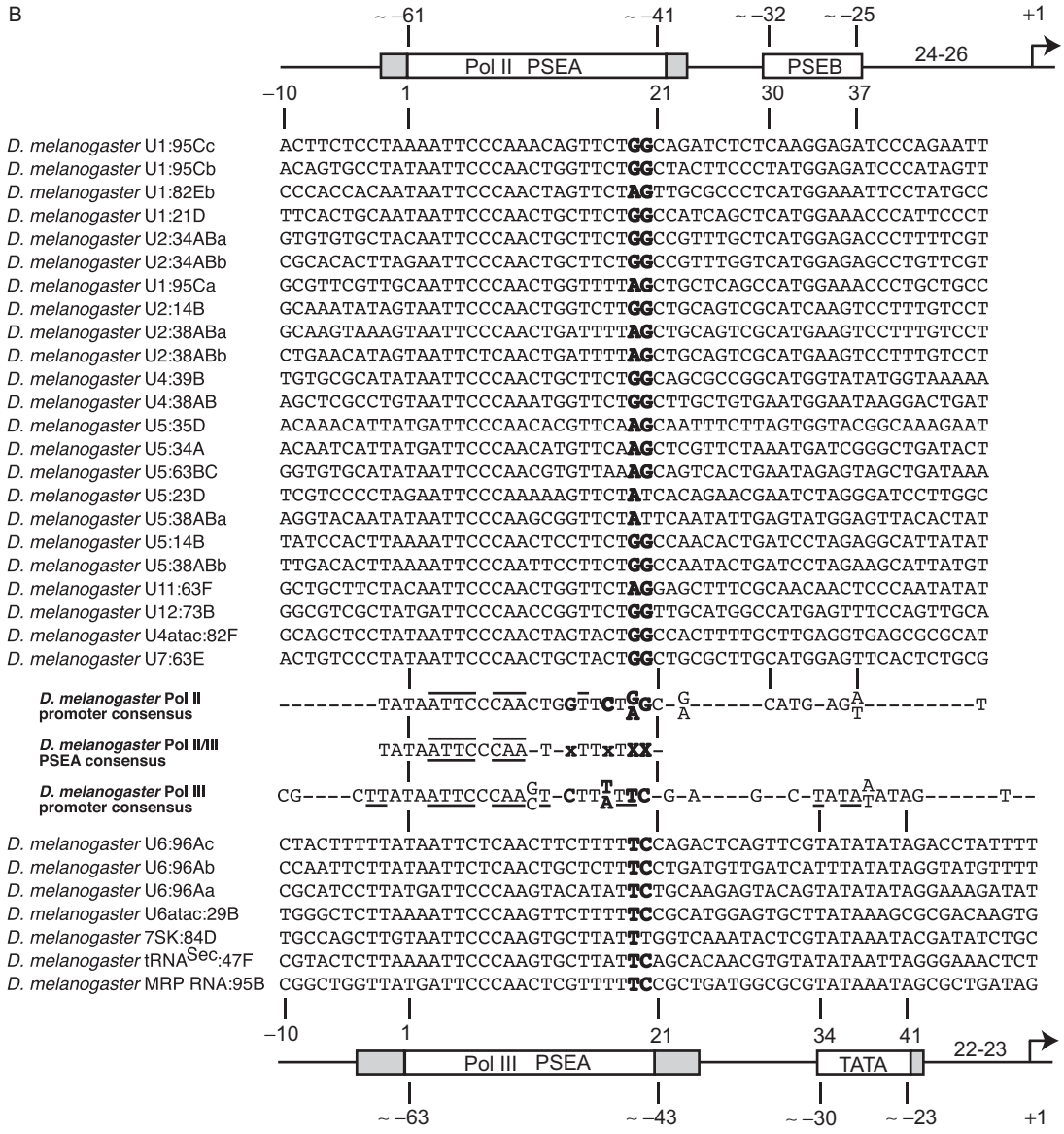


The extent of conservation of these elements is evident from the comparison of the sequences of a large set of snRNA and snRNA-like gene promoters from *D. melanogaster*, as shown in Figure 1B. A detailed sequence analysis of snRNA genes from six different insects revealed that this promoter structure is well conserved throughout the insect sub-phylum (Hernandez *et al.*, 2007). Generally, the insect PSEA is longer than the vertebrate PSE and exhibits a higher degree of sequence conservation.

Comparison between the sequences of the Pol II PSEAs and the Pol III PSEAs of *D. melanogaster* revealed significant differences as well as the obvious similarities. (Promoter consensus sequences, including the PSEA, are shown in the central area of Figure 1B.) It is notable that the 5′ half of the PSEA is very highly conserved between the Pol II and Pol III promoters. (Seven nucleotides in the 5′ half of the PSEAs are 100% conserved among the 30 genes listed.) In contrast, the Pol II and Pol III PSEAs diverge significantly from each other in their 3′ halves especially at certain specific nucleotide positions, most prominently at positions 19 and 20 denoted by Xs in the Pol II/III consensus PSEA (Figure 1B). For example, position 19 is always an A or G in the Pol II PSEAs versus always a T in the Pol III PSEAs, and position 20 is nearly always a G in the Pol II PSEAs but a C in the Pol III PSEAs (Figure 1B). These nucleotide positions have been said to be "conserved to be different" in the Pol II and Pol III PSEAs (Hernandez *et al.*, 2007). Other positions where there are often "conserved nucleotide differences" between the fly Pol II and the Pol III PSEAs are at positions 14 and 17 (denoted by lower case Xs in the Pol II/III consensus PSEA, Figure 1B).

Remarkably, the divergent 3′ half of the PSEA was found to play a key role in determining the RNA polymerase specificity of *Drosophila* snRNA genes. Altering as few as three nucleotide pairs in a U1 PSEA (including positions 19 and 20) to those found in a U6 PSEA completely switched the RNA polymerase specificity of the U1 promoter *in vitro* from Pol II to Pol III even in the absence of a TATA sequence (Jensen *et al.*, 1998). Moreover, the PSEAs from Pol II and Pol III promoters are not interchangeable *in vivo*. Substituting a U6 PSEA into a U1 promoter (a total of five nucleotide changes) resulted in the complete loss of U1 gene promoter activity *in vivo* (McNamara-Schroeder *et al.*, 2001; Lai *et al.*, 2005; Barakat and Stumph, 2008). Similarly, a reciprocal substitution of the U1 PSEA into the U6 promoter completely inactivated the U6 promoter in living cells (McNamara-Schroeder *et al.*, 2001; Lai *et al.*, 2005). Thus, the U6 PSEA cannot function

**Figure 1.** Continued.




for Pol II transcription, and the U1 PSEA cannot function for Pol III transcription, even though they both bind DmSNAPc. Conversely, other experiments showed that swapping the PSEB and TATA box of fly U1 and U6 snRNA promoters affected primarily transcription efficiency, but not the RNA polymerase selectivity of these promoters (Jensen *et al.*, 1998; Lai *et al.*, 2005).

The results described in the paragraph above were very surprising because, in vertebrates, snRNA gene Pol II and Pol III PSEs had been reported to be interchangeable (Mattaj *et al.*, 1988; Lobo and Hernandez, 1989). In vertebrates, the presence or absence of the TATA box (Figure 1A) was found to be the primary and dominant determinant of RNA polymerase specificity. For example, mutation of the vertebrate U6 TATA box to an unrelated sequence changed the promoter specificity to Pol II, while introduction of a TATA sequence into the vertebrate U1 or U2 promoters altered their specificity to Pol III (Mattaj *et al.*, 1988; Lobo and Hernandez, 1989).

In plants, the RNA polymerase specificity of snRNA genes is determined by a still different mechanism. In that case, both classes of promoters contain interchangeable USEs as well as interchangeable TATA boxes (Figure 1A). In this instance, RNA polymerase specificity is determined by the distance between the USE and the TATA box (Waibel and Filipowicz, 1990; Goodall *et al.*, 1991; Kiss *et al.*, 1991).

After the above differences in Pol II and Pol III snRNA gene promoters were discovered, the next question of general interest was the following: How are these differences in promoter structure read out by the transcription machinery to effect the recruitment of the

required polymerase? To seek answers to this question, it is necessary to have some knowledge of the structure and function of the small nuclear RNA activating protein complex, SNAPc.

## SNAPc: the key regulator of snRNA transcription

The PSEs of both Pol II- and Pol III-transcribed snRNA genes are recognized by the same multi-subunit transcription factor, SNAPc. SNAPc, also known as PSE-binding transcription factor (PTF) and PSE-binding protein (PBP), was first identified in the human system (Waldschmidt *et al.*, 1991; Murphy *et al.*, 1992; Sadowski *et al.*, 1993). Transcription of both Pol II and Pol III-transcribed snRNA genes was shown to be dependent upon SNAPc. Human SNAPc (HsSNAPc) contains five distinct polypeptide chains (HsSNAP190 or PTFα, HsSNAP50 or PTFβ, HsSNAP45 or PTFδ, HsSNAP43 or PTFγ, and SNAP19) for which the HsSNAP nomenclature reflects the apparent molecular weights of these subunits (Henry *et al.*, 1995; 1998; Yoon and Roeder, 1996). Significantly, a complex containing only the three subunits HsSNAP190, HsSNAP50, and HsSNAP43 was sufficient to reconstitute sequence-specific DNA binding as well as the basal transcription activity of human SNAPc (Mittal *et al.*, 1999; Ma and Hernandez, 2001; 2002; Hinkley *et al.*, 2003; Jawdekar *et al.*, 2006). Thus, these three subunits represent the "core subunits" of HsSNAPc required for pre-initiation complex assembly on snRNA genes. The other two subunits, HsSNAP45 and HsSNAP19, may play roles in regulation of SNAPc activity and complex stability (Henry *et al.*, 1998; Mittal *et al.*, 1999; Ma and Hernandez, 2001).

In the *Drosophila* system, DmSNAPc was first identified in a soluble nuclear fraction prepared from fly embryos. Partially purified DmSNAPc (originally termed DmPBP) exhibited sequence-specific PSEA-binding activity and was capable of stimulating PSEA-dependent activation of U1 and U6 snRNA gene transcription (Su *et al.*, 1997). The fly genome contains genes that code for proteins homologous to the HsSNAP190, HsSNAP50, and HsSNAP43 subunits, but no recognizable genes capable of encoding fly homologs of HsSNAP45 or HsSNAP19 have been detected (Li *et al.*, 2004). The evolutionary conservation of the SNAP190, SNAP50, and SNAP43 subunits further argues that these three subunits comprise the essential core of SNAPc.

In fact, orthologs of these three subunits have even been characterized in the anciently diverged trypanosomes, where tSNAPc is required for Pol II transcription of the spliced leader RNA, a small nuclear RNA that is trans-spliced onto the 5′ end of trypanosomal mRNAs (Huie *et al.*, 1997; Luo *et al.*, 1999; Das and Bellofatto, 2003; Schimanski *et al.*, 2004; 2005; Das *et al.*, 2005). The existence of tSNAPc indicates that the SNAP complex appeared very early in eukaryotic evolution and continues to be essential for snRNA transcription in diverse contemporary eukaryotes. A comparison of the structural features of the three orthologous fly, human, and trypanosomal (*T. brucei*) SNAPc subunits is presented in Figure 2. The shading indicates the regions most evolutionarily conserved among the three organisms shown.

In flies and humans, the most conserved region of the largest subunit, SNAP190, is a unique domain that consists of 4.5 tandem Myb repeats, termed respectively Rh, Ra, Rb, Rc, and Rd (Wong *et al.*, 1998; Li *et al.*, 2004). Myb repeats were first identified in the Myb oncoprotein and are involved in DNA binding (Klempnauer and Sippel, 1987; Biedenkapp *et al.*, 1988). In contrast to the 4.5 Myb repeats found in animal SNAP190, all other known Myb-domain proteins (to our knowledge) contain only one to three Myb repeats (reviewed in Rosinski and Atchley1998). Thus, the binding of the Myb repeats of SNAP190 to DNA is likely to be more complicated than that of proteins that contain fewer Myb repeats. The overall length of the fly protein is only about half the length of the human protein. This most likely reflects the presence of additional functional domains in the C-terminal region of the human protein such as those important for interaction with the HsSNAP45 subunit and with the enhancer-binding protein Oct-1 (Ford *et al.*, 1998; Mittal *et al.*, 1999). Interestingly, the trypanosomal SNAP190 protein is still shorter in overall length (about half the length of DmSNAP190 and quarter the length of HsSNAP190); furthermore, it contains only 2.5 identifiable Myb repeats that align best with the human and fly Ra (C-terminal half), Rb, and Rc repeats (Schimanski *et al.*, 2005).

The SNAP50 orthologs are the most evolutionarily conserved of the SNAPc subunits (Das and Bellofatto, 2003; Li *et al.*, 2004; Jawdekar *et al.*, 2006). Perhaps the most unusual feature of the SNAP50 protein is that the C-terminal domain consists of a unique non-canonical zinc finger (specifically named the "SNAP finger") (Bai *et al.*, 1996; Henry *et al.*, 1996; Das and Bellofatto, 2003; Li *et al.*, 2004; Jawdekar *et al.*, 2006). This domain contains seven conserved cysteine or histidine residues that when changed to alanine significantly reduced both zinc-binding and DNA-binding by human SNAPc (Jawdekar *et al.*, 2006). However, it should be emphasized that the SNAP finger is unique and has no detectable homology to other well-characterized zinc finger DNA-binding domains.

SNAP43 is probably the least characterized of the SNAPc subunits. The most evolutionarily conserved region lies toward the N terminus of SNAP43. Neither this region nor the non-conserved region of SNAP43 has any clear homology to other proteins in existing databases. However, as described further below, SNAP43 may play a critical role in the determination of RNA polymerase specificity at snRNA promoters.

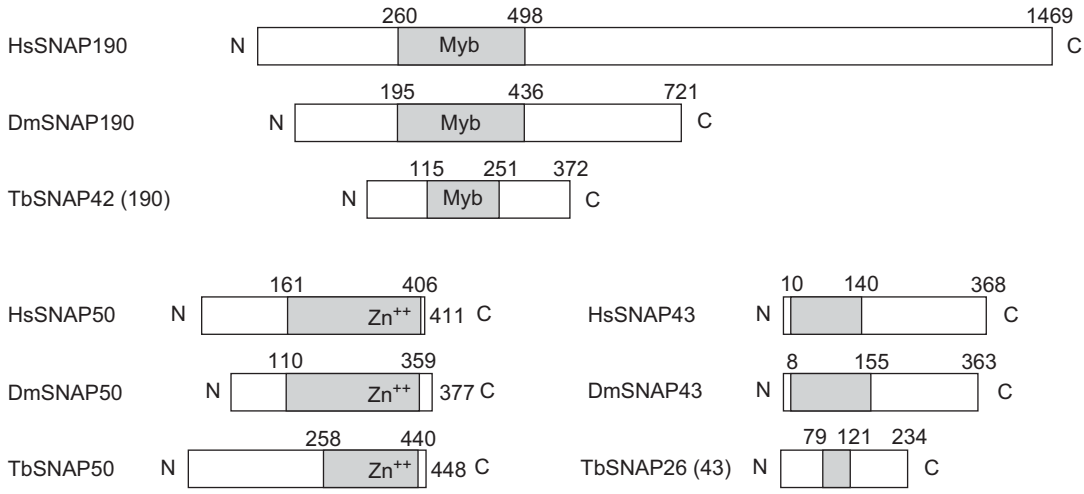



**Figure 2.** Comparison of *D. melanogaster* (Dm), human (Hs), and *Trypanosoma brucei* (Tb) SNAPc subunits. The rectangles indicate the relative lengths of the proteins, and the shaded areas indicate the evolutionarily conserved regions of the orthologous proteins. The numbers above and below the rectangles designate the amino acid positions in the proteins. The most conserved region in SNAP190 contains a Myb domain, which consists of 4.5 tandem Myb repeats in fly and human SNAP190 (Wong *et al.*, 1998; Li *et al.*, 2004), but just 2.5 Myb repeats in the trypanosome ortholog, TbSNAP42 (Schimanski *et al.*, 2005). This region of DmSNAP190 shares 27% identity and 44% similarity with HsSNAP190, but 15% identity and 57% similarity with TbSNAP42. For DmSNAP50, the most conserved region is located at the C terminus (residues 110 to 359), which includes an unorthodox zinc-binding domain (noted as $Zn^{++}$ in the figure) termed the "SNAP finger". This region shares 33% identity and 51% similarity with HsSNAP50, but 24% identity and 35% similarity with TbSNAP50. The most conserved region of DmSNAP43 resides in the N terminus (residues 8 to 155), which shares 31% identity and 48% similarity with HsSNAP43, but has 32% identity and 50% similarity to a shorter region (residues 79 to 121) of its trypanosome ortholog TbSNAP26.

Although it was known for quite some time that HsSNAPc and DmSNAPc were heteromeric complexes that contained a number of distinct polypeptides, the stoichiometry of the three core subunits in the complex was not absolutely clear (although often portrayed as 1:1:1). By using a mixture of tags on individual subunits and employing mobility supershift analysis with monoclonal antibodies against the tags, convincing evidence was recently obtained that the DmSNAP190, DmSNAP50, and DmSNAP43 subunits are each present in a single copy in native DmSNAPc bound to DNA (Lai *et al.*, 2008).

The SNAPc subunits co-purify with each other in solution, thus indicating a tight association with each other even when the complex is not associated with DNA (Murphy *et al.*, 1992; Sadowski *et al.*, 1993; Yoon *et al.*, 1995; Su *et al.*, 1997; Das and Bellofatto, 2003; Li *et al.*, 2004; Das *et al.*, 2005; Schimanski *et al.*, 2005). Although the isolated Rc and Rd repeats of HsSNAP190 can bind weakly but apparently without sequence specificity to DNA (Wong *et al.*, 1998; Ma and Hernandez, 2002; Hinkley *et al.*, 2003), all three core subunits of both human and fly SNAPc are essential for sequence-specific binding to the PSE(A). None of the three subunits can bind to the PSE(A) either individually or in any pair-wise combinations (Mittal *et al.*, 1999; Jawdekar *et al.*, 2006 and our unpublished observations). It is also clear from protein–DNA photo-cross-linking studies that each of the three core subunits, at least in flies, makes direct contact with the DNA and thereby contributes to the DNA-binding activity of the complex (Wang and Stumph, 1998; Li *et al.*, 2004; Kim *et al.*, 2010). The photo-cross-linking studies are described in a later section of this review.

## DmSNAPc subunit domains involved in DmSNAPc assembly

Mutational analyses have been used to identify domains within each of the core subunits that are required for complex formation with each of the other two subunits. Such studies have been carried out in both the fly (Hung *et al.*, 2009) and human systems (Mittal *et al.*, 1999; Ma and Hernandez, 2001; 2002; Hinkley *et al.*, 2003; Jawdekar *et al.*, 2006). Although there is a good deal of similarity in the findings from the two organisms, there also appear to be several significant differences.

Figure 3A schematically indicates the mapped domains within the fly proteins that are involved in subunit–subunit interactions. One of the main conclusions from those findings was that, with one exception, the evolutionarily most-conserved region of each DmSNAPc subunit was sufficient for its association with the other two subunits (Hung *et al.*, 2009). For example, the conserved Myb domain of DmSNAP190 was sufficient for its interaction with DmSNAP43. Furthermore, two adjacent regions within the conserved C-terminal region of DmSNAP50 (residues 110–291 and residues 292–377 respectively, which includes the SNAP finger) each interacted with both DmSNAP190 and DmSNAP43 (Hung *et al.*, 2009). That is, DmSNAP50 residues 292–377,

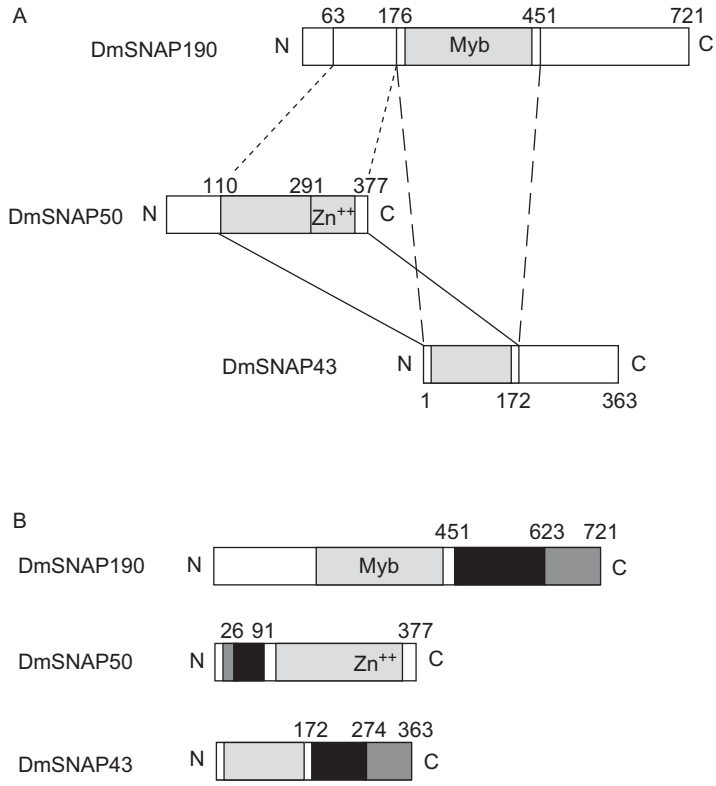



**Figure 3.** DmSNAPc subunit domains required for complex assembly and DNA binding. The lightly shaded area of each subunit represents the evolutionarily conserved region as shown in Figure 2. The numbers above and below the rectangles indicate the amino acid positions in the subunits where truncations were made to produce mutant proteins. (A) Map of DmSNAPc subunit-subunit interactions. The finely dashed lines indicate domains involved in interaction between DmSNAP190 and DmSNAP50. The longer dashed lines indicate the domains sufficient for interaction between DmSNAP190 and DmSNAP43, and the solid lines indicate regions of interaction between DmSNAP50 and DmSNAP43 (Hung *et al.*, 2009). (B) Map of subunit domains required for DmSNAPc DNA-binding activity. The lightly shaded areas represent the most conserved region in each subunit; these regions are required for subunit assembly and thus necessarily for DNA binding by DmSNAPc. The moderately shaded areas represent subunit domains not required for complex assembly but, when deleted, reduce (but do not eliminate) the DNA binding activity of DmSNAPc. The darkly shaded areas, when deleted, eliminate detectable DNA binding by DmSNAPc while not interfering with subunit assembly (Hung *et al.*, 2009).

which included the SNAP finger, interacted strongly with both DmSNAP43 and DmSNAP190. On the other hand, DmSNAP50 residues 110–291 interacted strongly with DmSNAP43 but more weakly with DmSNAP190. Finally, the conserved domain at the N terminus of DmSNAP43 (residues 1–172) was sufficient to associate with both DmSNAP190 and DmSNAP50. The only exception to the "conserved-domain rule" was that an evolutionarily non-conserved region of DmSNAP190 (residues 63–175) was required for interaction with DmSNAP50.

Comparison of these findings with the results of studies on human SNAPc revealed some important similarities and a number of surprising differences. As an example of similarity, in both organisms the conserved C-terminal region of SNAP50 interacts with the conserved N-terminal region of SNAP43 (Jawdekar *et al.*, 2006; Hung *et al.*, 2009). A significant difference, on the other hand, is that the strong interaction between DmSNAP190 and DmSNAP50 in flies has not been observed between the two homologous subunits of the human system. In fact available evidence suggests that there is no direct interaction between HsSNAP190 and HsSNAP50 (Wong *et al.*, 1998; Mittal *et al.*, 1999; Ma and Hernandez, 2001; Jawdekar *et al.*, 2006). The reason for this difference between the two systems is not clear.

Another significant difference is that Hung *et al.* (2009), working in the fly system, observed a strong interaction between the conserved N-terminal domain of DmSNAP43 and the Myb domain of DmSNAP190, but no interactions were observed between the analogous regions of the orthologous human subunits (Ma and Hernandez, 2001). Instead, an interaction was mapped between the non-conserved N-terminal region of HsSNAP190 and a non-conserved centrally-located region of HsSNAP43 (Ma and Hernandez, 2001). Interestingly, in the human system, stable association between HsSNAP190 and HsSNAP43 requires the additional subunit, HsSNAP19, which is not present in the fruit fly (Ma and Hernandez, 2001). Thus, it is quite possible that the lack of a strong direct interaction between the evolutionarily conserved regions

of HsSNAP190 and HsSNAP43 may be compensated in humans by the presence of HsSNAP19, which may act via non-analogous domains of the human subunits.

## DmSNAPc subunit domains required for DNA binding to the PSEA

Because the evolutionarily conserved regions of DmSNAPc are required for DmSNAPc assembly, the DNA-binding activity of DmSNAPc is also dependent upon those conserved regions of the subunits. But deletions within the non-conserved regions can also lead to reductions in DmSNAPc DNA binding activity while still allowing the assembly of all three subunits. In the case of fly SNAPc, even relatively short truncations of the DmSNAPc subunits often had a noticeable effect on DNA binding activity (Hung *et al.*, 2009).

Figure 3B indicates subunit domains required for effective DNA binding by DmSNAPc but not necessary for assembly of the three-subunit complex. As in previous figures, the lightest shading indicates the evolutionarily conserved region of each protein. The intermediate shading in Figure 3B indicates a region in each subunit that, when deleted, significantly decreased but did not completely eliminate DNA binding. However, the subsequent deletion of the darkly shaded region within any one of the subunits eliminated the DNA binding activity of the truncated DmSNAPc.

In the case of DmSNAP190, a C-terminal truncation following residue 623 greatly weakened but did not completely eliminate DNA-binding activity; however, a truncation following residue 451 of DmSNAP190 completely eliminated detectable DNA binding by DmSNAPc (Hung *et al.*, 2009). This was a very unexpected result when compared to findings in the human system because a human "mini-SNAPc", that completely lacked the non-conserved C-terminal region of HsSNAP190, bound very efficiently and with high specificity to PSE sequences (Mittal *et al.*, 1999; Ma and Hernandez, 2001; 2002; Hinkley *et al.*, 2003; Hanzlowsky *et al.*, 2006). In fact, the non-conserved region C-terminal to the Myb domain of HsSNAP190 was inhibitory to binding by HsSNAPc (Mittal *et al.*, 1999). It is not clear why the C-terminal domains of human and fly SNAP190 should have such differential effects on the DNA-binding activity of the complete SNAPc.

Figure 3B also indicates domains of DmSNAP50 and DmSNAP43 that are required for the DNA-binding activity of DmSNAPc (but not required for complex formation). Deletion of DmSNAP50 amino acids between residues 10 and 26 significantly reduced the DNA-binding activity of DmSNAPc, and deletion through residue 91 eliminated detectable DNA-binding activity (Hung *et al.*, 2009). We are not aware of any comparable studies that targeted the N-terminal function of human SNAP50.

In the case of DmSNAP43, deletion of residues following position 274 greatly weakened the DNA-binding activity of DmSNAPc, and deletion of nearly the entire non-conserved C terminus (that follows residue 172) resulted in a total loss of DmSNAPc DNA-binding activity (Figure 3B) (Hung *et al.*, 2009). These results are very similar to those obtained by Ma and Hernandez (2001) in the human system. In conclusion, it is clear from the truncation experiments described that domains of DmSNAP190, DmSNAP50, and DmSNAP43 that are not evolutionarily conserved contribute to the DNA-binding activity of DmSNAPc. Whether this contribution arises from direct contacts between the protein and the DNA is not addressed through the truncation experiments. It is possible that the non-conserved domains that are required for the DNA-binding activity may be necessary for DmSNAPc to adopt a conformation compatible with efficient DNA binding.

## The structure of the DmSNAPc–DNA complex

An atomic structure of SNAPc (or its subunits) or of the SNAPc–DNA complex is not yet available. However, site-specific protein–DNA photo-cross-linking studies carried out in the *Drosophila* system have provided a wealth of information regarding the architecture of the protein–DNA complex (Wang and Stumph, 1998; Li *et al.*, 2004; Lai *et al.*, 2005; Kim *et al.*, 2010). These studies have revealed the position of each DmSNAPc subunit along the length of the PSEA as well as their rotational positions relative to the DNA sequence and to each other. Most interestingly, these studies have provided considerable evidence that DmSNAPc assumes different conformations depending upon whether the protein is bound to a U1 or a U6 PSEA.

Figure 4 shows a summary of the results of these site-specific protein–DNA photo-cross-linking studies. The diagram at the top of the figure shows the position of the PSEA that is aligned in all the DNA diagrams below. The phosphate positions that cross-linked to each of the three subunits are indicated by colored spheres (yellow for DmSNAP190 cross-links; green and red for DmSNAP50, and blue and red for DmSNAP43). Odd-numbered phosphates indicate positions mapped on the non-template strand of the DNA, and even-numbered phosphates indicate positions mapped on the template strand. The colored areas thus represent the positions where each of the individual subunits closely approach the DNA when DmSNAPc binds either to a U1 PSEA or to a U6 PSEA (upper and lower diagrams respectively of each pair). DmSNAP190 cross-linked to phosphate positions that extended over the entire length of either a U1 or a U6 PSEA (Wang and Stumph, 1998). DmSNAP50 cross-linked to phosphate positions

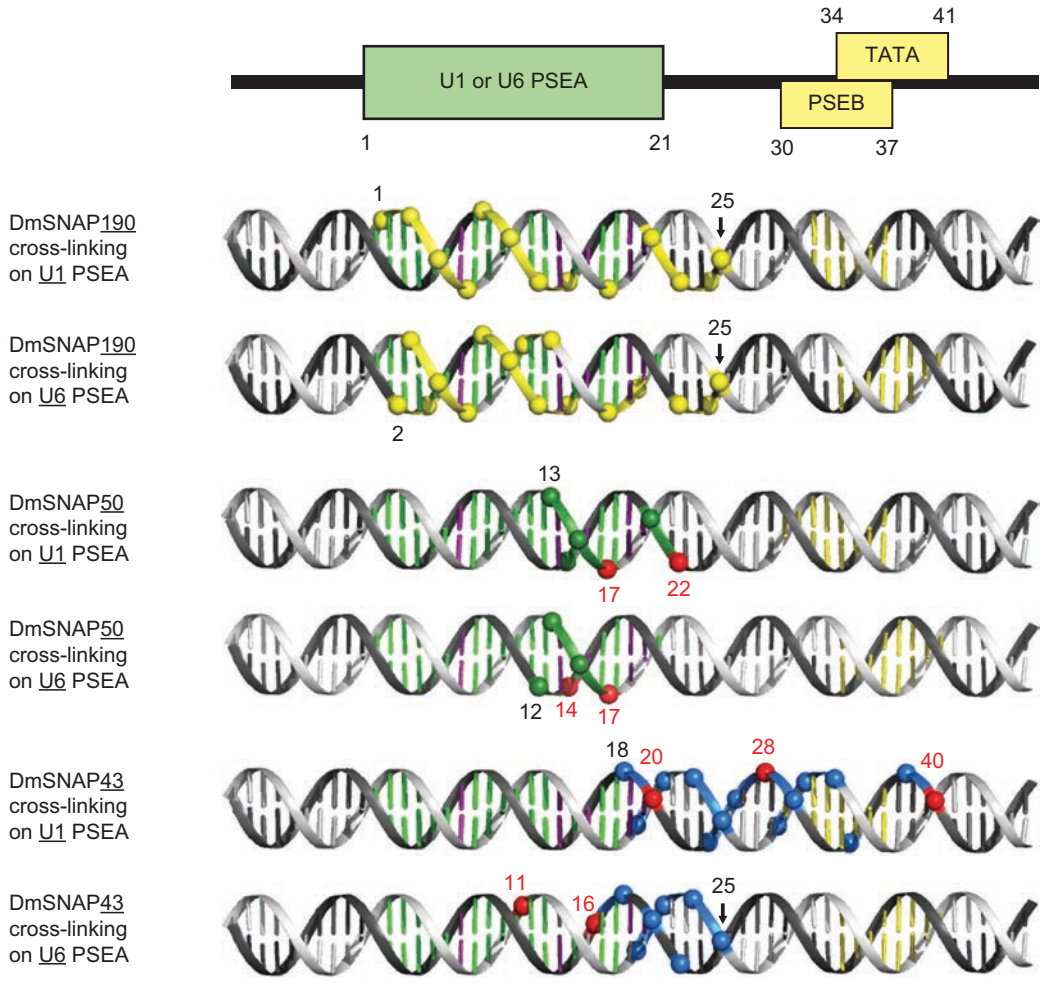



**Figure 4.** Contact points between snRNA gene promoter sequences and DmSNAPc. Site-specific protein–DNA photo-cross-linking data of DmSNAPc to DNA probes that contained either U1 or U6 PSEAs are shown projected onto B-form DNA (Wang and Stumph, 1998; Li *et al.*, 2004). (For experimental details, see below.) The colored spheres on the DNA structures indicate phosphate positions that specifically cross-linked to DmSNAP190 (yellow), DmSNAP50 (green and red), or DmSNAP43 (blue and red). (The red spheres indicate the phosphate positions specifically used for the DmSNAP50 and DmSNAP43 domain-mapping experiments summarized in Figure 5.) On each pair of DNA structures, the upper illustration shows the cross-linking pattern to DNA probes containing a U1 PSEA, and the lower illustration shows cross-linking to probes containing a U6 PSEA. The schematic drawing at the top of the figure shows the location of the PSEA (either U1 or U6) and the relative locations of the PSEB or TATA sequence in snRNA gene Pol II and Pol III promoters respectively. The locations of these promoter elements are also indicated by colored bases within the DNA structures below (green for PSEA and yellow for PSEB or TATA box). The five bases that differed in the DNA probes are indicated by magenta coloring (U1 versus U6 base pairs at PSEA positions 7, 14, 16, 19, and 20). The DNA sequence flanking the U1 or U6 PSEA was maintained identical in all probes (and contained a PSEB but not TATA sequence) in order to ensure that any differences in the cross-linking patterns were due entirely to the five different nucleotides within the U1 and U6 PSEAs. Experimental details: A series of 86 individual DNA probes were prepared each of which contained a photo-cross-linker at an individual phosphate position (with an adjacent $^{32}P$ radiolabel) within probes containing either a U1 or U6 PSEA. Cross-linker was placed at every second phosphate position on either the non-template strand (odd-numbered positions) or on the template strand (even-numbered positions). After binding and covalent cross-linking of DmSNAPc, all but two to three nucleotides of DNA were removed by nuclease digestion, and the radiolabeled proteins that cross-linked to each phosphate position were identified by gel electrophoresis and autoradiography. For further details see Wang *et al.* (1998) and Li *et al.* (2004).

extending from position 13 through position 22 of a U1 PSEA, but extending from phosphates 12 through 17 of a U6 PSEA (Wang and Stumph, 1998). DmSNAP43 exhibited the greatest differences in cross-linking patterns depending upon the source of the PSEA. This subunit cross-linked to phosphate positions extending from positions 18 to 40 of a U1 PSEA, but to a shorter and more upstream region of a U6 PSEA (phosphates extending from positions 11 to 25) (Wang and Stumph, 1998; Li *et al.*, 2004).

The experiments summarized in Figure 4 were carried out using DNA photo-cross-linking probes that were identical except at five nucleotide positions within the PSEA (U1 versus U6 bases at PSEA positions 7, 14, 16, 19, and 20). Therefore, it is certain that the observed differences in protein–DNA contacts arose from the five base

differences within the U1 and U6 PSEA sequences and not from the sequences flanking the PSEAs. It is particularly notable that DmSNAP43–DNA interactions occurred up to 20 bp (two turns of the DNA helix) downstream of the U1 PSEA. In contrast, DmSNAP43–DNA interactions were limited to only a distance of 4 bp downstream of a U6 PSEA (Figure 4) (Li *et al.*, 2004).

The cross-linking studies, besides providing information regarding the nucleotides contacted along the longitudinal axis of the PSEA, furthermore revealed the particular face of the DNA contacted by each of the three DmSNAPc subunits (Wang and Stumph, 1998). The cross-linking patterns indicated that, when the DNA is oriented as shown in Figure 4, DmSNAP190 interacted with the *front* face (and to some extent with the upper and lower faces) of the DNA in the 5′ half of the PSEA (either U1 or U6), but it contacted primarily the *lower* face of the DNA duplex in the 3′ half of the PSEA. On the other hand, DmSNAP50 occupied the *front* face of the DNA in the 3′ half of the PSEA. Finally, DmSNAP43 resided primarily on the *upper* face of the DNA in the 3′ portion of the PSEA (as well as far downstream of the U1 PSEA). It is worth noting that although this modeling is done on B-form DNA in Figure 4, it is certainly possible that the binding of DmSNAPc may distort the DNA. Indeed, a study that made use of circular permutation, mini-circle binding, and ligase-catalyzed circularization assays suggested that the DNA of both the U1 and U6 PSEAs was modestly but similarly bent by DmSNAPc toward the face of the DNA helix contacted by DmSNAP43 (Hardin *et al.*, 2000).

## Mapping protein domains within DmSNAP50 and DmSNAP43 that contact specific nucleotides of the U1 and the U6 PSEAs

Recently performed work has localized domains within the DmSNAPc subunits that cross-link strongly to certain individual phosphate positions (indicated by the positions of the red spheres in Figure 4) within the U1 and U6 PSEAs (Kim *et al.*, 2010). This was accomplished by combining the site-specific protein–DNA photo-cross-linking technique with site-specific chemical digestion of the protein. The cross-linked protein fragments were then identified by gel electrophoresis. For DmSNAP190, the comprehensive mapping of domains that cross-link to the U1 and U6 PSEAs is still in progress, but the domain-mapping experiments for DmSNAP50 and DmSNAP43 have been completed (Kim *et al.*, 2010) and the findings for these two smaller subunits are summarized in Figure 5.

There were four phosphate positions that cross-linked most strongly to DmSNAP50 (indicated by the red spheres in Figures 4 and 5). Those were phosphate positions 17 and 22 in the U1 PSEA, but phosphates 14 and 17 in the U6 PSEA (Wang and Stumph, 1998). Therefore, experiments were carried out to localize domains of DmSNAP50 associated with those four strongly-cross-linking phosphate positions (Kim *et al.*, 2010).

As indicated in Figure 5A, phosphate position 17 of both the U1 and U6 PSEAs cross-linked strongly to a polypeptide fragment of DmSNAP50 that spans amino acid residues 103–179. This same phosphate position (#17) of both PSEAs cross-linked even more strongly to the C-terminal fragment of DmSNAP50 comprised of residues 238–377 (which contains the zinc-binding SNAP finger domain). On the other hand, phosphate position 22 in the U1 PSEA cross-linked only to the DmSNAP50 fragment that contains residues 103–179. In a contrasting fashion, phosphate position 14 in the U6 PSEA cross-linked only to DmSNAP50 residues 238–377. Due to the limitations of the assay, it was not possible to know whether or not the region of DmSNAP50 between residues 180–237 interacts with phosphate position 17 of the U1 and U6 PSEAs. It is nonetheless apparent from the results that at least two distinct regions of DmSNAP50 (residues 103–179 and residues 238–377) are involved in contacting the DNA at different positions in the U1 and U6 PSEAs (phosphates 22 and 14 respectively).

An illustration of the mapped domains of DmSNAP50 interacting with the U1 and U6 PSEAs (represented on B-form DNA) is diagrammed in Figure 5B. When DmSNAPc binds to a U1 PSEA, a domain encompassing DmSNAP50 residues 103–179 closely approaches phosphates 17 and 22, and the SNAP finger domain at the C terminus (residues 238–377) is also close to phosphate 17. On the other hand, when DmSNAPc binds to a U6 PSEA, the SNAP finger domain (residues 238–377) closely approaches phosphate 14 as well as phosphate 17, and the region from residues 103–179 closely approaches phosphate 17 (but not 22) of the U6 PSEA.

An unanswered question from the results described above is whether the most N-terminal domain of DmSNAP50 (residues 1–102) is involved in contacting the DNA. The earlier truncation experiments indicated that amino acids in this region were required for DmSNAPc to bind to the DNA (Figure 3B). It is possible that the N-terminal domain of DmSNAP50 may contact nucleotides other than the four utilized in the photo-cross-linking/domain-mapping experiments described above, or it may be that this N-terminal domain is required for DmSNAPc to adopt a proper DNA-binding conformation.

Experiments were also carried out to map domains of the smallest subunit, DmSNAP43, that contact specific phosphates in the U1 and U6 PSEAs (Kim *et al.*, 2010). Because DmSNAP43 cross-linked strongly to phosphate positions 11 and 16 of the U6 PSEA,

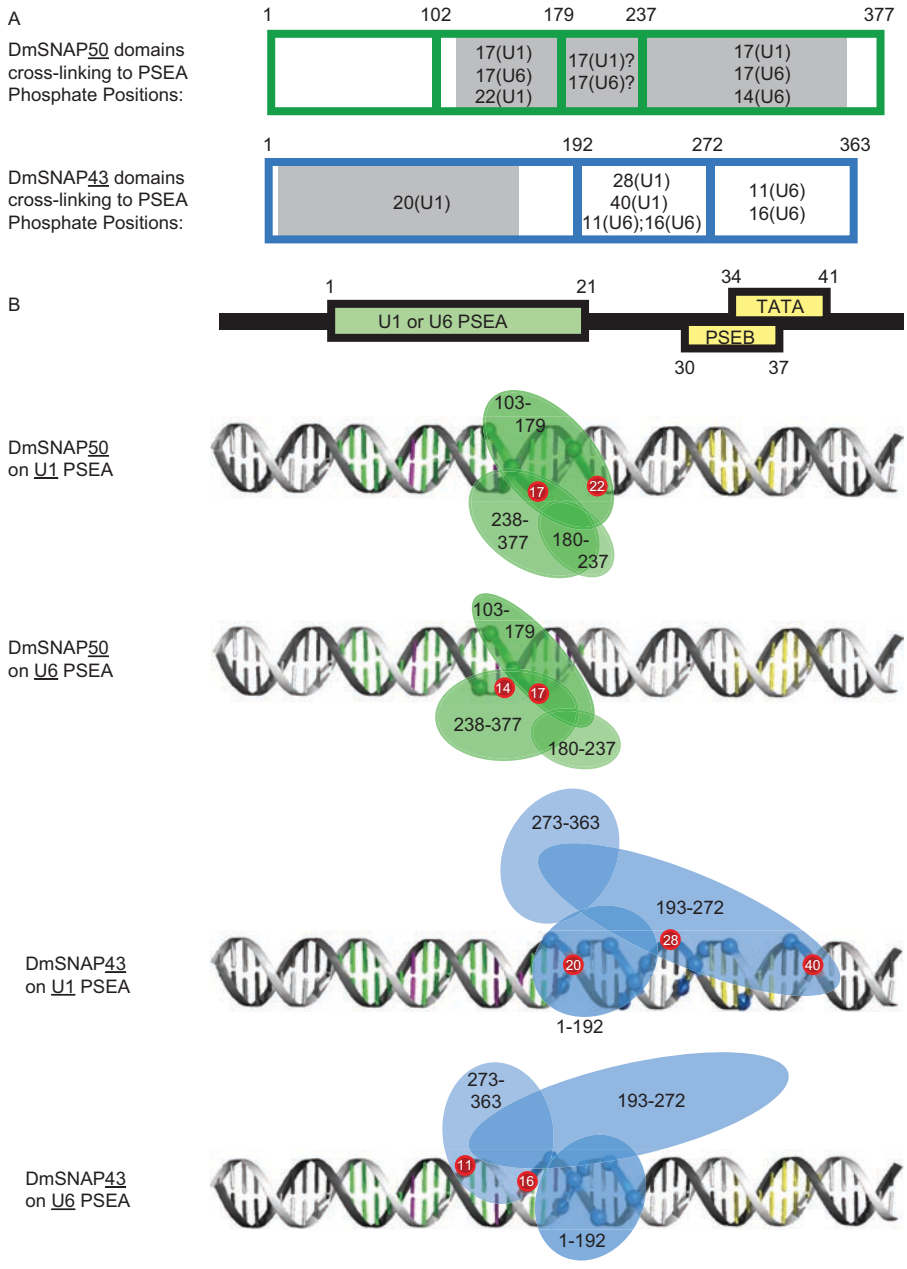



**Figure 5.** Domains of DmSNAP50 and DmSNAP43 that cross-link to specific phosphate positions on U1 or U6 DNA probes. (A) The rectangles at the top represent the linear amino acid sequences of the DmSNAP50 (green) and DmSNAP43 (blue) subunits. The shaded areas represent the evolutionarily conserved region of each subunit. The numbers above the rectangles indicate the amino acid positions at which the polypeptides were specifically cleaved with hydroxylamine to identify domains that cross-linked to specific nucleotide positions within the U1 and U6 PSEA probes (Kim *et al.*, 2010). Phosphate positions that cross-linked to the indicated domains of each subunit are indicated within the rectangles. (B) The results summarized in (A) are projected onto B-form DNA. The color-coding of the schematic drawing at the top, and of the DNA helices, is the same as described in the legend to Figure 4. The phosphate positions used in these domain-mapping experiments are specifically labeled in red with white lettering. The mapped domains of the DmSNAP50 and DmSNAP43 subunits that cross-linked to the red-colored phosphate positions are represented as individual ellipses. The cross-linking patterns for each subunit are shown separately. It should be noted that there are no data to reveal which domains contact the phosphate positions represented as the green or blue spheres. For further details, see Kim *et al.* (2010).

but strongly to positions 20 and 28 of the U1 PSEA (Figure 4), domains of DmSNAP43 were mapped that interact closely with those four positions. In addition, phosphate 40 of the U1 PSEA was employed in the mapping experiments since this is the phosphate furthest downstream of the PSEA that cross-linked to DmSNAP43. Those results are summarized in the lower part of Figure 5A.

First, phosphate 20 of the U1 PSEA cross-linked to the N-terminal fragment (residues 1–192) of DmSNAP43.

Phosphates 28 and 40, also of the U1 PSEA, cross-linked exclusively to the fragment of DmSNAP43 that spans residues 193–272. Finally, phosphates 11 and 16 of the U6 PSEA cross-linked strongly to the most C-terminal fragment of DmSNAP43 (residues 273–363) and much less strongly to a fragment encompassing residues 193–272.

A pictorial representation of these findings is presented in the lower part of Figure 5B. The N-terminal half of DmSNAP43 is in close proximity to position 20 within the U1 PSEA (and possibly relatively close to position 20 within the U6 PSEA, although no data was obtained for U6 position 20 due to the weakness of the cross-linking to that site; see Wang and Stumph, 1998). The DmSNAP43 domain encompassing residues 193–272 closely approaches phosphates 28 and 40 on the U1 PSEA, but not on a U6 PSEA. Residues within this same domain are within weak cross-linking distance to phosphates 11 and 16 when DmSNAPc binds to a U6 PSEA, but not when DmSNAPc binds to a U1 PSEA. Finally, the C-terminal domain of DmSNAP43 (residues 273–363) closely approaches phosphates 11 and 16 when DmSNAPc binds to a U6 PSEA, but not when DmSNAPc binds to a U1 PSEA. Thus, at least three separable regions of DmSNAP43 can be involved in interactions with the DNA, and these protein–DNA interactions differ depending upon the origin of the PSEA and its sequence (U1 or U6).

The locations of the mapped protein domains as drawn in Figure 5B are also consistent with the data that mapped the subunit domains involved in protein–protein interactions (illustrated in Figure 3A). Those studies indicated that the evolutionarily conserved region of DmSNAP50 is involved in protein–protein interactions with the evolutionarily conserved region of DmSNAP43 (Hung *et al.*, 2009). Interestingly, the protein–DNA photo-cross-linking studies described above place these same conserved domains of DmSNAP50 (DmSNAP50 residues 110 to 377) and DmSNAP43 (DmSNAP43 residues 1 to 155) into close proximity on the DNA (Figure 5B). This would seem to be a necessary consequence of the fact that the conserved domains of these two subunits participate in protein-protein interactions with each other.

## Role of DmSNAPc in the establishment of RNA polymerase specificity

Relatively minor sequence differences in the 3′ halves of the PSEAs of Pol II- and Pol III-transcribed snRNA genes are responsible for determining the differential RNA polymerase specificity of these promoters (Figure 1B) (Jensen *et al.*, 1998; McNamara-Schroeder *et al.*, 2001; Lai *et al.*, 2005). Interestingly, the 3′ half of the PSEA is the region contacted by all three subunits of DmSNAPc (Figure 4). How then are the signals encoded in the DNA sequence of the PSEAs transmitted to recruit distinct transcriptional machineries to the promoter? The protein–DNA photo-cross-linking experiments clearly indicate that the conformations of the DmSNAPc/DNA complexes are different on Pol II and Pol III PSEAs (Figures 4 and 5). We have hypothesized that the different PSEA sequences act as differential allosteric effectors of DmSNAPc conformation, and that the conformational differences of DmSNAPc on Pol II and Pol III snRNA promoters then lead to the differential recruitment of distinct sets of general transcription factors (GTFs) and subsequently different RNA polymerases to the different classes of snRNA genes.

A working model of how this might occur is shown in Figure 6. In this model, DmSNAPc adopts different conformations induced by the U1 and U6 PSEAs. The exposure of different surfaces of the DmSNAPc subunits, as well as differences in protein–DNA interactions, then results in the recruitment of Pol II GTFs to the U1 promoter but Pol III GTFs to the U6 promoter. This has the effect of recruiting Pol II to transcribe the U1 gene but Pol III to transcribe the U6 gene (Figure 6).

One of the Pol II GTFs required at the fly U1 promoter is the TATA binding protein (TBP). This was shown first by *in vitro* transcription assays and more recently by chromatin immunoprecipitation (ChIP) assays (Zamrod *et al.*, 1993; Barakat and Stumph, 2008). Importantly, when the U1 PSEA was changed to a U6 PSEA (i.e. five base pair changes in only the PSEA), TBP could not be detected on the U1 promoter *in vivo* as determined by ChIP assays, in spite of the fact that DmSNAPc was able to bind *in vivo* to the mutant U1 promoter that contained the U6 PSEA (Barakat and Stumph, 2008). When DmSNAPc was bound to a U6 PSEA in the context of the U1 promoter, it appears that DmSNAPc either failed to recruit TBP or alternatively inhibited the binding of TBP to the DNA (Barakat and Stumph, 2008). However, the negative ChIP results cannot definitively eliminate the possibility that TBP might bind the mutant U1 promoter but could not be cross-linked to DNA or recognized by the antibodies. In any case, when the U6 PSEA was substituted for the U1 PSEA in the U1 promoter, it is apparent that a structural change occurred that prevented the recruitment or detection of TBP and also prevented the formation of an active transcription complex (Barakat and Stumph, 2008). Any of these interpretations is consistent with the generalized model presented in Figure 6.

It is tempting to believe that the conserved PSEB, although having little resemblance to a TATA sequence, might be a site of interaction of TBP with the DNA of the U1 promoter. The PSEB, like a TATA sequence, is 8 bp in length, and it is located at precisely the expected distance

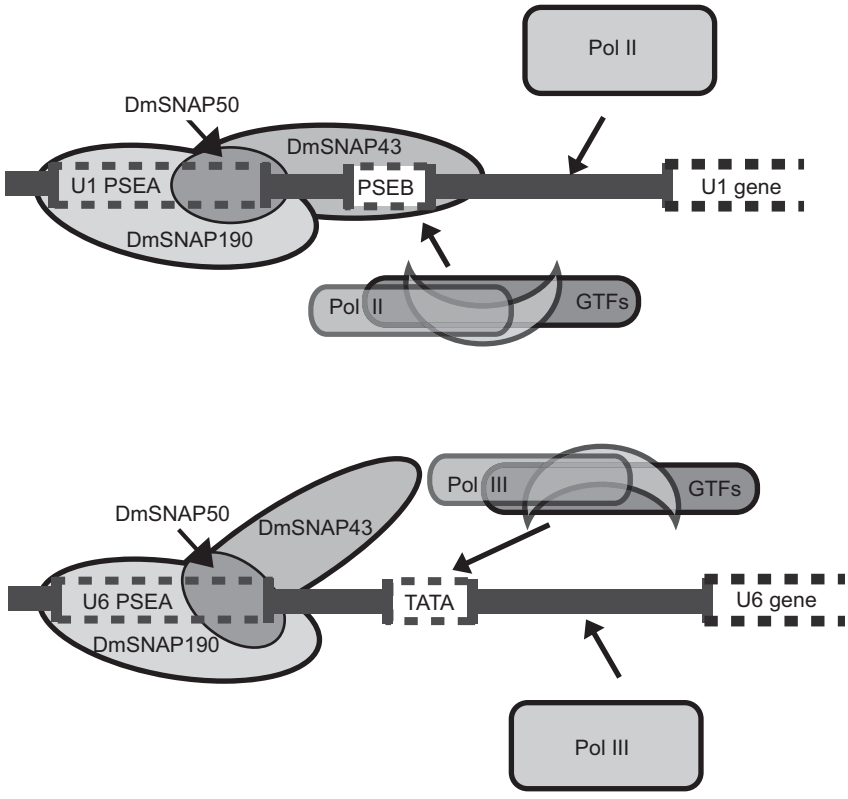



**Figure 6.** A working model illustrating how differential interactions of DmSNAPc with U1 or U6 PSEAs may lead to RNA polymerase specificity at *D. melanogaster* snRNA gene promoters. Minor sequence differences between U1 and U6 PSEAs are believed to act as allosteric effectors of DmSNAPc conformation, resulting in differential protein–DNA interactions as well as the potential exposure of different surfaces of the DmSNAPc subunits. The conformational differences of DmSNAPc are postulated to recruit Pol II GTFs to the U1 promoter (probably centered on the PSEB) but Pol III GTFs to the U6 promoter (probably centered on the TATA box). Due to the 12 bp versus 8 bp spacing difference between the PSEAs and either the TATA box or PSEB, a rotational as well as a longitudinal difference in the orientation and spacing of the Pol II and Pol III GTFs would be expected. The binding of the distinct GTFs would then recruit Pol II to the U1 promoter but Pol III to the U6 promoter to initiate transcription.

for a TATA box upstream of the Pol II transcription start site. However, there is not yet direct experimental evidence that TBP interacts with the PSEB. It is alternatively possible that TBP, rather than contacting the PSEB, could be tethered indirectly to the promoter by DmSNAPc or by other undefined components of the system. Additional work will be required to distinguish between these various possibilities. In any case, the fly U1 promoter may serve as an excellent system to study the role of TBP at TATA-less Pol II promoters.

The DmSNAP43 subunit in particular contacts the DNA very differently when DmSNAPc binds to the U1 and U6 PSEAs, and this subunit may be a particularly good candidate to play a role in recruitment of the Pol II GTFs. In the human system, transcription of snRNA genes by Pol II requires the GTFs TBP, TFIIA, TFIIB, TFIIE, and TFIIF (Sadowski *et al.*, 1993; Henry *et al.*, 1995; Kuhlman *et al.*, 1999). It is not clear whether TFIIH is involved, but a trypanosomal version of TFIIH was found to be essential for Pol II-dependent SL RNA transcription in those highly diverged organisms (Lee *et al.*, 2007). Because DmSNAP43 contacts the DNA at nucleotides within and flanking the PSEB on U1 promoters (Figures 4, 5, and 6), it seems possible that DmSNAP43 could be involved in recruiting TFIIA, TFIIB, and/or TBP by means of direct protein–protein contacts. Although human SNAP43 is capable of interacting directly with TBP in the absence of DNA (Hinkley *et al.*, 2003), it is not yet known if this is true in flies. It will be important in future experiments to examine the possible role of DmSNAP43 in the recruitment of the Pol II GTFs.

Pol III transcription of U6 (and 7SK) snRNA genes has been studied much more extensively in the human system than in flies (see Schramm and Hernandez, 2002; Jawdekar and Henry, 2008 and references therein). In terms of Pol III GTF recruitment, HsSNAPc was able to recruit both TBP and Brf2 to the human U6 promoter, and the recruitment of TBP occurred through direct interactions with HsSNAP190 (Ma and Hernandez, 2002; Hinkley *et al.*, 2003). Interestingly, Pol III transcription in fruit flies was found to employ the insect-specific TBP-related factor TRF1 rather than TBP itself (Takada *et al.*, 2000). It will be interesting to examine whether the DmSNAP190 subunit in flies plays an essential role in recruiting any of the Pol III GTFs to fly U6 snRNA gene promoters. Might it be possible that DmSNAP43

plays the most direct role in Pol II GTF recruitment, but that DmSNAP190 plays the most prominent role in Pol III GTF recruitment? Or are both of these subunits directly involved in GTF recruitment to both types of promoters? And what is the role of DmSNAP50? Further investigations will be required to identify all the players engaged in this process and to elucidate the molecular mechanisms by which distinct preinitiation complexes are assembled on either Pol II or Pol III snRNA gene promoters.

## Acknowledgements

We thank Cheryn Lee for substantial assistance in the preparation of figures involving graphical illustrations of DNA structure.

## Declaration of interest

This work was supported by a grant from the National Science Foundation (MCB-0842770) and in part by a grant from the California Metabolic Research Foundation. The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

## References

Baer M, Nilsen TW, Costigan C and Altman S. 1990. Structure and transcription of a human gene for H1 RNA, the RNA component of human RNase P. Nucleic Acids Res 18:97–103.

Bai L, Wang ZX, Yoon JB and Roeder RG. 1996. Cloning and characterization of the β subunit of human proximal sequence element-binding transcription factor and its involvement in transcription of small nuclear RNA genes by RNA polymerases II and III. Mol Cell Biol 16:5419–5426.

Barakat NH and Stumph WE. 2008. TBP recruitment to the U1 snRNA gene promoter is disrupted by substituting a U6 proximal sequence element A (PSEA) for the U1 PSEA. FEBS Letters 582:2413–2416.

Beck E, Jorcano JL and Alonso A. 1984. Drosophila melanogaster U1 and U2 small nuclear RNA genes contain common flanking sequences. J Mol Biol 173:539–542.

Biedenkapp H, Borgmeyer U, Sippel AE and Klempnauer KH. 1988. Viral myb oncogene encodes a sequence-specific DNA-binding activity. Nature 335:835–837.

Bond UM, Yario TA and Steitz JA. 1991. Multiple processing-defective mutations in a mammalian histone pre-mRNA are supressed by compensatory changes in U7 RNA both in vivo and in vitro. Genes Dev 5:1709–1722.

Cabart P and Murphy S. 2001. BRFU, a TFIIB-like factor, is directly recruited to the TATA-box of polymerase III small nuclear RNA gene promoters through its interaction with TATA-binding protein. J Biol Chem 276:43056–43064.

Cabart P and Murphy S. 2002. Assembly of human small nuclear RNA gene-specific transcription factor IIIB complex de novo on and off promoter. J Biol Chem 277:26831–26838.

Dahlberg JE and Lund E. 1988. The genes and transcription of the major small nuclear RNAs, pp. 38–70. In: Birnstiel ML, ed. Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles. Heidelberg, Federal Republic of Germany: Springer Verlag.

Das A and Bellofatto V. 2003. RNA polymerase II-dependent transcription in trypanosomes is associated with a SNAP complex-like transcription factor. Proc Natl Acad Sci USA 100:80–85.

Das A, Zhang Q, Palenchar JB, Chatterjee B, Cross GA and Bellofatto V. 2005. Trypanosomal TBP functions with the multisubunit transcription factor tSNAP to direct spliced-leader RNA gene expression. Mol Cell Biol 25:7314–7322.

Das G, Henning D and Reddy R. 1987. Structure, organization, and transcription of Drosophila U6 small nuclear RNA genes. J Biol Chem 262:1187–1193.

Egloff S, O'Reilly D and Murphy S. 2008. Expression of human snRNA genes from beginning to end. Biochem Soc Trans 36: 590–594.

Ford E, Strubin M and Hernandez N. 1998. The Oct-1 POU domain activates snRNA gene transcription by contacting a region in the SNAPc largest subunit that bears sequence similarities to the Oct-1 coactivator OBF-1. Genes Dev 12:3528–3540.

Goodall GJ, Kiss T and Filipowicz W. 1991. Nuclear RNA splicing and small nuclear RNAs and their genes in higher plants. Oxford Surv Plant Mol Cell Biol 7:255–296.

Guthrie C. 1991. Messenger RNA splicing in yeast: clues to why the spliceosome is a ribonucleoprotein. Science 253:157–163.

Hanzlowsky A, Jelencic B, Jawdekar G, Hinkley CS, Geiger JH and Henry RW. 2006. Co-expression of multiple subunits enables recombinant SNAPc assembly and function for transcription by human RNA polymerases II and III. Protein Expr Purif 48:215–223.

Hardin SB, Ortler CJ, McNamara-Schroeder KJ and Stumph WE. 2000. Similarities and differences in the conformation of protein-DNA complexes at the U1 and U6 snRNA gene promoters. Nucleic Acids Res 28:2771–2278.

Henry RW, Sadowski CL, Kobayashi R and Hernandez N. 1995. A TBP-TAF complex required for transcription of human snRNA genes by RNA polymerases II and III. Nature 374:653–656.

Henry RW, Ma BC, Sadowski CL, Kobayashi R and Hernandez N. 1996. Cloning and characterization of SNAP50, a subunit of the snRNA-activating protein complex SNAPc. EMBO J 15:7129–7136.

Henry RW, Mittal V, Ma BC, Kobayashi R and Hernandez N. 1998. SNAP19 mediates the assembly of a functional core promoter complex (SNAPc) shared by RNA polymerases II and III. Genes Dev 12:2664–2672.

Hernandez G, Valafar F and Stumph WE. 2007. Insect small nuclear RNA gene promoters evolve rapidly yet retain conserved features involved in determining promoter activity and RNA polymerase specificity. Nucleic Acids Res 35:21–34.

Hernandez N. 1992. Transcription of vertebrate snRNA genes and related genes, pp. 281–313. In: McKnight SL, Yamamoto KR, eds. Transcriptional Regulation. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Hinkley CS, Hirsch HA, Gu LP, LaMere B and Henry RW. 2003. The small nuclear RNA-activating protein 190 Myb DNA binding domain stimulates TATA box-binding protein-TATA box recognition. J Biol Chem 278:18649–18657.

Huie JL, He P and Bellofatto V. 1997. In vitro transcription of the Leptomonas seymouri SL RNA and U2 snRNA genes using homologous cell extracts. Mol Biochem Parasitol 90:183–192.

Hung KH, Titus M, Chiang SC and Stumph WE. 2009. A map of Drosophila melanogaster small nuclear RNA-activating protein complex (DmSNAPc) domains involved in subunit assembly and DNA binding. J Biol Chem 284:22568–22579.

Jawdekar GW and Henry RW. 2008. Transcriptional regulation of human small nuclear RNA genes. Biochim Biophys Acta 1779: 295–305.

Jawdekar GW, Hanzlowsky A, Hovde SL, Jelencic B, Feig M, Geiger JH and Henry RW. 2006. The unorthodox SNAP50 zinc finger domain contributes to cooperative promoter recognition by human SNAPc. J Biol Chem 281:31050–31060.

Jensen RC, Wang Y, Hardin SB and Stumph WE. 1998. The proximal sequence element (PSE) plays a major role in establishing the RNA polymerase specificity of Drosophila U-snRNA genes. Nucleic Acids Res 26:616–622.

Kass S, Tyc K, Steitz JA and Sollner-Webb B. 1990. The U3 small nucleolar ribonucleoprotein functions in the first step of preribosomal RNA processing. Cell 60:897–908.

Kim MK, Kang YS, Lai HT, Barakat NH, Magante D and Stumph WE. 2010. Identification of SNAPc subunit domains that interact with specific nucleotide positions in the U1 and U6 gene promoters. Mol Cell Biol 30:2411–2423.

Kiss T, Marshallsay C and Filipowicz W. 1991. Alteration of the RNA polymerase specificity of U3 snRNA genes during evolution and in vitro. Cell 65:517–526.

Klempnauer KH and Sippel AE. 1987. The highly conserved amino-terminal region of the protein encoded by the v-myb oncogene functions as a DNA-binding domain. EMBO J 6: 2719–2725.

Kuhlman TC, Cho H, Reinberg D and Hernandez N. 1999. The general transcription factors IIA, IIB, IIF, and IIE are required for RNA polymerase II transcription from the human U1 small nuclear RNA promoter. Mol Cell Biol 19:2130–2141.

Lai HT, Chen H, Li C, McNamara-Schroeder KJ and Stumph WE. 2005. The PSEA promoter element of the Drosophila U1 snRNA gene is sufficient to bring DmSNAPc into contact with 20 base pairs of downstream DNA. Nucleic Acids Res 33:6579–6586.

Lai HT, Kang YS and Stumph WE. 2008. Subunit stoichiometry of the Drosophila melanogaster small nuclear RNA activating protein complex (SNAPc). FEBS Lett 582:3734–3738.

Lee BJ, Kang SG and Hatfield D. 1989. Transcription of Xenopus selenocysteine tRNA Ser (formerly designated opal suppressor phosphoserine tRNA) gene is directed by multiple 5′-extragenic regulatory elements. J Biol Chem 264:9696–9702.

Lee JH, Nguyen TN, Schimanski B and Gunzl A. 2007. Spliced leader RNA gene transcription in Trypanosoma brucei requires transcription factor TFIIH. Eukaryot Cell 6:641–649.

Li C, Harding GA, Parise J, McNamara-Schroeder KJ and Stumph WE. 2004. Architectural arrangement of cloned proximal sequence element-binding protein subunits on Drosophila U1 and U6 snRNA gene promoters. Mol Cell Biol 24:1897–1906.

Lo PCH and Mount SM. 1990. Drosophila melanogaster genes for U1 snRNA variants and their expression during development. Nucleic Acids Res 18:6971–6979.

Lobo SM and Hernandez N. 1989. A 7 bp mutation converts a human RNA polymerase II snRNA promoter into an RNA polymerase III promoter. Cell 58:55–67.

Lobo SM and Hernandez NT. 1994. Transcription of snRNA genes by RNA polymerases II and III, pp. 127–159. In: Conaway RC, Conaway JW, eds. Transcription: Mechanisms and Regulation. New York, NY: Raven Press.

Luo H, Gilinger G, Mukherjee D and Bellofatto V. 1999. Transcription initiation at the TATA-less spliced leader RNA gene promoter requires at least two DNA-binding proteins and a tripartite architecture that includes an initiator element. J Biol Chem 274: 31947–31954.

Ma B and Hernandez N. 2001. A map of protein–protein contacts within the small nuclear RNA-activating protein complex SNAPc. J Biol Chem 276:5027–5035.

Ma BC and Hernandez N. 2002. Redundant cooperative interactions for assembly of a human U6 transcription initiation complex. Mol Cell Biol 22:8067–8078.

Mattaj IW, Dathan NA, Parry HD, Carbon P and Krol A. 1988. Changing the RNA polymerase specificity of U snRNA gene promoters. Cell 55:435–442.

McNamara-Schroeder KJ, Hennessey RF, Harding GA, Jensen RC and Stumph WE. 2001. The Drosophila U1 and U6 gene proximal sequence elements act as important determinants of the RNA polymerase specificity of snRNA gene promoters in vitro and in vivo. J Biol Chem 276:31786–31792.

Mittal V, Ma BC and Hernandez N. 1999. SNAPc: a core promoter factor with a built-in DNA-binding damper that is deactivated by the Oct-1 POU domain. Genes Dev 13:1807–1821.

Murphy S, Yoon JB, Gerster T and Roeder RG. 1992. Oct-1 and Oct-2 potentiate functional interactions of a transcription factor with the proximal sequence element of small nuclear RNA genes. Mol Cell Biol 12:3247–3261.

Parry HD, Scherly D and Mattaj IW. 1989. Snurpogenesis: the transcription and assembly of U snRNP components. Trends Biochem Sci 14:15–19.

Peculis BA and Steitz JA. 1993. Disruption of U8 nucleolar snRNA inhibits 5.8S and 28S rRNA processing in the Xenopus oocyte. Cell 73:1233–1245.

Rosinski JA and Atchley WR. 1998. Molecular evolution of the Myb family of transcription factors: evidence for polyphyletic origin. J Mol Evol 46:74–83.

Saba JA, Busch H, Wright D and Reddy R. 1986. Isolation and characterization of two putative full-length Drosophila U4 small nuclear RNA genes. J Biol Chem 261:8750–8753.

Sadowski CL, Henry RW, Lobo SM and Hernandez N. 1993. Targeting TBP to a non-TATA box cis-regulatory element: a TBP-containing complex activates transcription from snRNA promoters through the PSE. Genes Dev 7:1535–1548.

Schimanski B, Laufer G, Gontcharova L and Gunzl A. 2004. The Trypanosoma brucei spliced leader RNA and rRNA gene promoters have interchangeable TbSNAP50-binding elements. Nucleic Acids Res 32:700–709.

Schimanski B, Nguyen TN and Gunzl A. 2005. Characterization of a multisubunit transcription factor complex essential for spliced-leader RNA gene transcription in Trypanosoma brucei. Mol Cell Biol 25:7303–7313.

Schramm L and Hernandez N. 2002. Recruitment of RNA polymerase III to its target promoters. Genes Dev:16 2593–2620.

Schramm L, Pendergrast PS, Sun YL and Hernandez N. 2000. Different human TFIIIB activities direct RNA polymerase III transcription from TATA-containing and TATA-less promoters. Genes Dev 14:2650–2663.

Sharp PA. 1994. Split genes and RNA splicing. Cell 77:805–816.

Steitz JA, Black DL, Gerke V, Parker KA, Kramer A, Frendewey D and Keller W. 1988. Functions of the abundant U-snRNPs, pp. 115–154. In: Birnsteil ML, ed. Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles. Heidelberg, Federal Republic of Germany: Springer Verlag.

Su Y, Song Y, Wang Y, Jessop L, Zhan LC and Stumph WE. 1997. Characterization of a Drosophila proximal-sequence-element-binding protein involved in transcription of small nuclear RNA genes. Eur J Biochem 248:231–237.

Takada S, Lis JT, Zhou S and Tjian R. 2000. A TRF1:BRF complex directs Drosophila RNA polymerase III transcription. Cell 101:459–469.

Teichmann M, Wang ZX and Roeder RG. 2000. A stable complex of a novel transcription factor IIB-related factor, human TFIIIB50, and associated proteins mediate selective transcription by RNA polymerase III of genes with upstream promoter elements. Proc Natl Acad Sci USA 97:14200–14205.

Waibel F and Filipowicz W. 1990. RNA-polymerase specificity of transcription of Arabidopsis U snRNA genes determined by promoter element spacing. Nature 346:199–202.

Waldschmidt R, Wanandi I and Seifart KH. 1991. Identification of transcription factors required for the expression of mammalian U6 genes in vitro. EMBO J 10:2595–2603.

Wang Y and Stumph WE. 1998. Identification and topological arrangement of Drosophila proximal sequence element (PSE)-binding protein subunits that contact the PSEs of U1 and U6 snRNA genes. Mol Cell Biol 18:1570–1579.

Wong MW, Henry RW, Ma BC, Kobayashi R, Klages N, Matthias P, Strubin M and Hernandez N. 1998. The large subunit of basal transcription factor SNAPc is a Myb domain protein that interacts with Oct-1. Mol Cell Biol 18:368–377.

Yoon JB and Roeder RG. 1996. Cloning of two proximal sequence element-binding transcription factor subunits (γ and δ) that are required for transcription of small nuclear RNA genes by RNA polymerases II and III and interact with the TATA-binding protein. Mol Cell Biol 16:1–9.

Yoon JB, Murphy S, Bai L, Wang Z and Roeder RG. 1995. Proximal sequence element-binding transcription factor (PTF) is a multisubunit complex required for transcription of both RNA polymerase II- and RNA polymerase III-dependent small nuclear RNA genes. Mol Cell Biol 15:2019–2027.

Yuan Y and Reddy R. 1991. 5′ flanking sequences of human MRP/7-2 RNA gene are required and sufficient for the transcription by RNA polymerase III. Biochim Biophys Acta 1089:33–39.

Zamrod Z and Stumph WE. 1990. U4B snRNA gene enhancer activity requires functional octamer and SPH motifs. Nucleic Acids Res 18:7323–7330.

Zamrod Z, Tyree CM, Song Y and Stumph WE. 1993. In vitro transcription of a Drosophila U1 small nuclear RNA gene requires TATA box-binding protein and two proximal cis-acting elements with stringent spacing requirements. Mol Cell Biol 13: 5918–5927.

Zieve G, Benecke BJ and Penman S. 1977. Synthesis of two classes of small RNA species in vivo and in vitro. Biochemistry 16:4520–4525.

*Editor: Michael M. Cox*