Title
Structural Arrangement of the DmSNAP190 Myb repeats on U1 and U6 snRNA Gene Promoter Sequences

Permalink
https://escholarship.org/uc/item/58r246gm

Author
Kang, Yoon Soon

Publication Date
2014

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

SAN DIEGO STATE UNIVERISTY

Structural Arrangement of the DmSNAP190 Myb repeats on U1 and U6 snRNA

Gene Promoter Sequences

A Dissertation submitted in partial satisfaction of the

Requirements for the degree Doctor of Philosophy

in

Chemistry

by

Yoon Soon Kang

Committee in Charge:

University of California, San Diego

      Professor Navtej Toor
      Professor Rommie Amaro

San Diego State University

      Professor William E. Stumph, Chair
      Professor Peter van der Geer
      Professor Anca Segall

2014

The dissertation of Yoon Soon Kang is approved, and it is

acceptable in quality and form for publication on microfilm:

————————————————————————

————————————————————————

————————————————————————

————————————————————————

————————————————————————

————————————————————————
Chair

University of California, San Diego

San Diego State University

2014

DEDICATION

I dedicate this dissertation to the following people:

This dissertation is dedicated to my parents, and the rest of my family for their incredible support, continuous encouragement, and endless love.

.

TABLE OF CONTENTS

LIST OF FIGURES

General Introduction

Chapter 1

Concluding remarks

ACKNOWLEDGEMENTS

I would like to thank my thesis advisor Dr. Stumph for giving me the opportunity to work in his lab and continuous guidance throughout this study. I would also like to thank Dr. van der Geer, Dr. Segall, Dr. Toor, and Dr. Amaro for serving on my dissertation committee and for their valuable time.

I am grateful to all of my past and current colleagues of the Stumph lab for their advice and encouragement. I especially thank to Mun kyoung Kim for his great help in teaching me how to do cross-linking experiments. I thank Mun Kyoung Kim, Hsien-Tsung Lai, Nermeen H. Barakat and Deodato Magante for their contributions to the publication that constitutes Chapter 1 of this thesis. I am also grateful to Matthew T. Doherty, Cheryn Lee and Michelle Kurano for their contributions to the publications that constitute Chapter 2 and Chapter 3.

I would like to thank to my parents for their support. I am also deeply thankful to Jin-joo Kang for her encouragement and support.

Chapter one, in full, is a reprint of the material as it appears in Molecular and Cellular Biology, 2010. Identification of SNAPc subunit domains that interact with specific nucleotide positions in the U1 and U6 gene promoters. Kim, M. K.; Kang, Y. S.; Lai, H. T.; Barakat, N. H.; Magante, D.; Stumph, W. E. The dissertation author was a co-first author of the paper.

Chapter two, in full, is a reprint of the material as it appears in The Journal of Biological Chemistry, 2012. Architectural Arrangement of the Small Nuclear RNA (snRNA)-activating Protein Complex 190 Subunit (SNAP190) on U1 snRNA Gene

Promoter DNA. Doherty, M. T.; Kang Y. S.; Lee, C.; Stumph, W. E. The dissertation author was a co-first author of the paper.

The material in chapter three, in full, is being prepared to submit for publication. The dissertation author was the primary researcher and first author of this paper.

VITA

1997-2004 B.S., Biochemistry & Molecular Biology

Hanyang University

Ansan, Korea

1999-2001 Military Service

Seoul, Korea

2006-2014 Teaching Associate

Chemistry and Biochemistry Department, San Diego State University

2011-2011 Teaching Associate, Department of Chemistry and Biochemistry

University of California, San Diego

2013-2014 Pre-doctoral Candidate, Chemistry and Biochemistry

University of California, San Diego

San Diego State University

2014 Ph.D., Chemistry

University of California, San Diego and San Diego State University

 (Joint Doctoral Program)

PUBLICATIONS AND ABSTRACTS

<u>**Publications**</u>

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.

Kim MK*, **Kang YS***, Lai HT, Barakat NH, Magante D, and Stumph WE (2010) Identification of SNAPc subunit domains that in- teract with specific nucleotide positions in the U1 and U6 gene promoters. Mol. Cell Biol. 30**,** 2411–2423; Correction (2010) Mol. Cell Biol. 30, 5257. *Co-first authors.

Doherty MT*, **Kang YS***, Lee C and Stumph WE (2012) Architectural Arrangement of the Small Nuclear RNA (snRNA)-activating Protein Complex 190 Subunit (SNAP190) on U1 snRNA Gene Promoter DNA. J. Biol. Chem 287:39369-39379. *Co-first authors.

Kang YS, Kurano M, and Stumph WE (2014) The largest subunit of SNAPc adopts different architectural arrangements on U1 and U6 snRNA gene promoters in response to sequence differences in U1 and U6 promoter elements in preparation.

Abstracts

Stumph WE, Kim MK, and **Kang YS**. Mapping SNAPc Subunit Domains that Interact with Specific Nucleotide Positions in the U1 and U6 snRNA Gene Promoters. Experimental Biology 2010, Anaheim, CA (April 2010).

Stumph WE, **Kang YS**, and Kim MK. Different domains of SNAPc subunits closely interact with U1 and U6 snRNA gene promoter sequences for differential recruitment of RNA polymerases II and III. Experimental Biology 2011, Washington, DC (April 2011).

Kang YS, Doherty M, and Stumph WE. Arrangement of the 4.5 Myb Repeats of SNAP190 on the PSE of the U1 snRNA gene promoter. Cold Spring Harbor Meeting on "Mechanisms of Eukaryotic Transcription" at Cold Spring Harbor, NY (August, 2011).

Stumph WE, Doherty M, and **Kang YS**. Arrangement of the 4.5 Myb Domain Repeats of SNAP190 on the U1 Gene Promoter. Experimental Biology 2012, San Diego, CA (April 2012).

Doherty M, **Kang YS**, and Stumph WE. Arrangement of the 4.5 Myb repeats of SNAP190 on the U1 snRNA gene promoter. Keystone Symposium on "Eukaryotic Transcription" at Snowbird, UT (April 2012).

Kang YS, Doherty M, and Stumph WE. Arrangement of the 4.5 Myb repeats of SNAP190 on snRNA gene promoter sequences. 8th International Biennial Conference on RNA Polymerases I and III at Warrenton, VA (June 2012).

Stumph WE, **Kang YS**, and Kurano M. Differential Contacts of the 4.5 Myb Repeats of SNAP190 with the DNA of U1 and U6 snRNA Gene Promoters. Experimental Biology 2013, Boston, MA (April 2013).

Kang YS, Kurano M, and Stumph WE. The 4.5 Myb Repeats of SNAP190 Interact Differently with the DNA of U1 and U6 snRNA Gene Promoters. Cold Spring Harbor

Laboratory Meeting on "Mechanisms of Eukaryotic Transcription" at Cold Spring Harbor, NY (August, 2013).

Kang YS, Kurano M, and Stumph WE. The 4.5 Myb repeats of SNAP190 interact differently with the DNA of U1 and U6 snRNA gene promoters. International Conference on Transcription by RNA Polymerases I, III, IV, and V: OddPols 2014, Ann Arbor, MI (June 2014).

FIELDS OF STUDY

Major Field: Biochemistry and Molecular Biology

Studies in Transcriptional Regulation of Eukaryotic Gene Expression

Professor William E. Stumph.

ABSTRACT OF THE DISSERTATION

Structural Arrangement of the DmSNAP190 Myb repeats on U1 and U6 snRNA Gene

Promoter Sequences

by

Yoon Soon Kang

Doctor of Philosophy in Chemistry

University of California, San Diego

San Diego State University

2014

Professor William E. Stumph, Chair

The small nuclear RNA activating protein complex (SNAPc) is necessary for transcription of genes coding for the spliceosomal snRNAs (U1, U2, U4, U5, and U6). The *Drosophila melanogaster* SNAP complex (DmSNAPc) consists of three subunits (DmSNAP190, DmSNAP50, and DmSNAP43) that recognize a proximal sequence element A (PSEA) located approximately 40-60 base pairs upstream of the transcription start site. DmSNAPc recruits RNA polymerase II to the U1-U5 gene promoters, but RNA polymerase III to U6 gene promoters. Interestingly, the precise nucleotide sequence of the PSEA plays a dominant role in determining RNA polymerase specificity of

Drosophila snRNA genes. Indeed, there are a few key nucleotide positions that are "conserved-to-be-different" between the PSEAs of fly snRNA genes transcribed by RNA polymerase II and those transcribed by RNA polymerase III. Furthermore, site-specific protein-DNA photo-cross-linking assays have indicated that DmSNAPc adopts different conformations on U1 and U6 promoters. Thus, the investigation of the structural arrangement of DmSNAPc on U1 or U6 PSEAs is important for understanding RNA polymerase specificity.

Chapter 1 describes the mapping of protein domains of each of the three DmSNAPc subunits that cross-link to specific phosphate positions in the U1 or U6 PSEAs. To do this, novel methodology was developed that combined the site-specific protein-DNA photo-cross-linking technique with subsequent site-specific chemical cleavage of the protein. The results with DmSNAP190 oriented the subunit with its N-terminus facing the transcription start site of the U1 gene and its C-terminus facing the upstream direction.

Chapter 2 describes localization of each of the 4.5 Myb repeats of DmSNAP190 on the U1 PSEA by determining where each cross-links to the DNA. This determined the structural arrangement of the DmSNAP190 Myb repeats on the U1 PSEA.

Chapter 3 reports work that determines the structural arrangement of DmSNAP190 Myb repeat domains on the U6 PSEA. The cross-linking data suggest that there is significant movement of at least two of the Myb repeats on U6 promoter sequences relative to U1. These data allow us for the first time to put forth a comprehensive model of the different conformations adopted by DmSNAPc on U1 and U6 promoters.

GENERAL INTRODUCTION

The small nuclear RNAs (snRNAs) are metabolically stable, non-coding RNA molecules that are found in the nuclei of eukaryotic cells (Busch et al., 1982). The snRNAs play an important role in RNA processing events such as the splicing of pre-mRNAs, 3' end formation of histone mRNA, and ribosomal RNA processing (Steitz et al., 1988; Kass et al., 1990; Bond et al., 1991; Peculis and Steitz, 1993; Sharp, 1994; Hung and Stumph, 2011). In addition, some snRNAs are also involved in other facets of gene regulation, such as transcription initiation and elongation events (Nguyen et al., 2001; Yang et al., 2001; Kwek et al., 2002). Therefore, the regulation of snRNA genes is essential for maintaining proper gene expression, and the accurate and efficient control of the expression of snRNA genes is important for cellular survival (Reddy and Busch, 1988).

The structure of *Drosophila melanogaster* snRNA gene promoters

In animals, most of the genes coding for snRNAs (e.g. U1, U2, U3, U4, U5 and U7 genes) are transcribed by RNA polymerase II (Pol II), but other snRNA genes (e.g. U6 and 7SK) are transcribed by RNA polymerase III (Pol III) (Dahlberg and Lund, 1988; Hernandez, 1992; Lobo and Hernandez, 1994; Parry et al., 1989; Su et al., 1997; Hernandez et al., 2007). Although these genes are transcribed by different RNA polymerases, the promoter structure of both classes of genes share notable similarities (Fig. 1).

In *Drosophila melanogaster*, U1, U2, U4, and U5 snRNA genes contain two

conserved promoter sequences, a well conserved Proximal Sequence Element A (21 base pairs) and a less conserved Proximal Sequence Element B (8 base pairs) (PSEA and PSEB, respectively) that recruit Pol II to the DNA (Zamrod et al., 1993; Hernandez et al., 2007). Unlike the Pol II snRNA promoters, the U6 snRNA gene promoters contain an 8 bp TATA box, instead of a PSEB, located 12 bp downstream of the U6 PSEA (Hernandez et al., 2007). Comparison between a large number of *D. melanogaster* Pol II and Pol III PSEA sequences revealed that the 5' halves of the PSEAs are very similar to each other, whereas the 3' halves are more divergent (Hernandez et al., 2007).
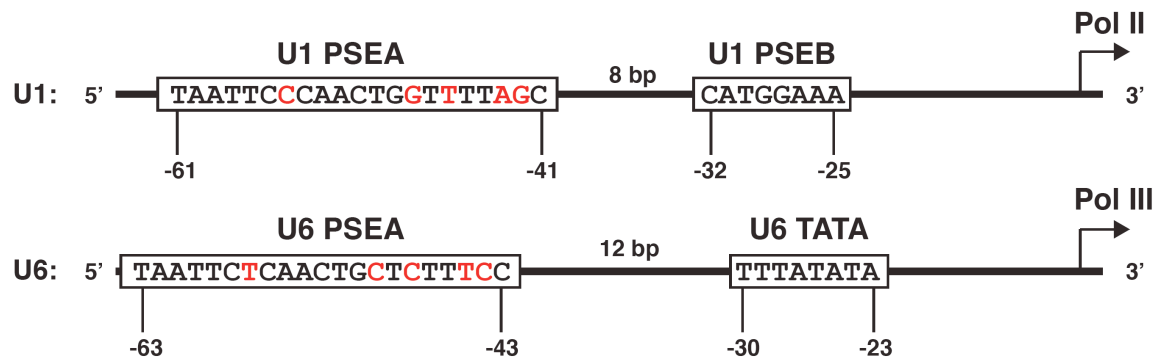
Figure 1. **Comparison of the promoter structure of *D. melanogaster* snRNA genes transcribed by Pol II or Pol III.** The PSEA, PSEB, and TATA sequences of the U1:95Ca and U6:96Ab genes are shown. Only 5 of 21 nucleotide positions (in red color) are different between the U1 and U6 PSEAs.

RNA polymerase specificity of snRNA gene promoters

Interestingly, *Drosophila* RNA polymerase specificity is dependent on a few nucleotide differences between the PSEAs of the two classes of snRNA genes, and the PSEA is the dominant element for determining the RNA polymerase specificity of the *Drosophila* snRNA genes (Jensen et al., 1998). Neither the PSEB vs. TATA box

sequence nor the spacing difference between promoter elements determined RNA polymerase specificity (although they do affect transcription efficiency) (Jensen et al., 1998; Lai et al., 2005; McNamara-Schroeder et al., 2001). Previously, our lab showed that the *Drosophila* U1 and U6 PSEAs are not interchangeable *in vitro* and *in vivo*: the U1 PSEA cannot function for Pol III recruitment, and the U6 PSEA cannot function for Pol II recruitment (McNamara-Schroeder et al., 2001; Barakat and Stumph, 2008), even though both PSEAs bind the same multi-subunit transcription factor, known as the small nuclear RNA activating protein complex (SNAPc). Therefore, in fruit flies, the divergent 3' half of the PSEA seems to play a dominant role in determining RNA polymerase specificity (Hernandez et al., 2007; Hung and Stumph. 2011).

Characterization of the *Drosophila* SNAPc

SNAPc, also known as PSE-binding protein (PBP) or PSE-binding transcription factor (PTF), was first identified in the human system in HeLa cell extract (Waldschmidt et al., 1991). The small nuclear RNA activating protein complex (SNAPc) recognizes the PSEs of both pol II and Pol III snRNA genes, and SNAPc is capable of activating appropriately both pol II and Pol III transcription of snRNA genes (Waldschmidt et al., 1991; Sadowski et al., 1993; Goomer et al., 1994; Henry et al., 1995; Yoon et al., 1995). The human SNAPc contains five integral polypeptide subunits (SNAP190 or PTFα, SNAP50 or PTFβ, SNAP45 or PTFδ, SNAP43 or PTFγ, and SNAP19) that were designated with apparent molecular weights of approximately 190, 50, 45, 43 and 19 kDa (Henry et al., 1995; Yoon et al., 1995). Significantly, a human SNAP complex containing only the three subunits SNAP190, SNAP50, and SNAP43 was sufficient to reconstitute

sequence-specific DNA binding as well as basal transcription activity of snRNA genes (Mittal et al., 1999; Ma and Hernandez, 2001; Hinkley et al., 2003; Jawdekar et al., 2006). Thus, these three subunits represent the "core subunits" of human SNAPc required for pre-initiation complex assembly on snRNA genes. The other two subunits, SNAP45 and SNAP19, may play roles in regulation of SNAPc activity and complex stability (Henry et al., 1998; Mittal et al., 1999; Ma and Hernandez, 2001).

Our lab has been characterizing *D. melanogaster* SNAPc (DmSNAPc) (Su et al., 1997; Wang and Stumph, 1998; Li et al., 2004; Lai et al., 2005). DmSNAPc is able to recognize and bind to U1 and U6 PSEAs and can activate transcription of the *Drosophila* U1 and U6 snRNA genes (Su et al., 1997). DmSNAPc consists of three subunits: DmSNAP190, DmSNAP50, and DmSNAP43 that are orthologous to the SNAP190, SNAP50, and SNAP43 subunits of human SNAPc (Li et al., 2004). These three subunits were originally identified by site-specific protein-DNA photo-cross-linking (Wang and Stumph, 1998).

The photo-cross-linking studies furthermore indicated that there were significant differences in the way that the three subunits contacted the DNA of the U1 and U6 PSEAs (Wang and Stumph, 1998; Li et al., 2004). Those studies are summarized in Fig. 2.
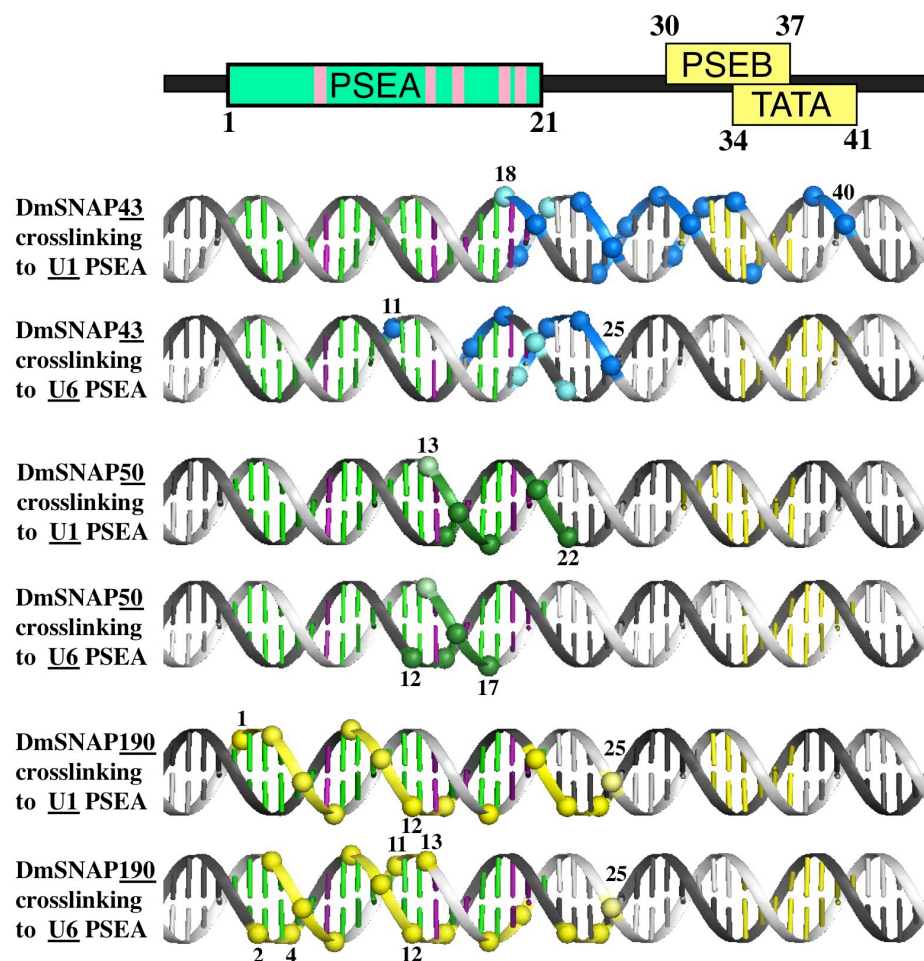
Figure 2. **Crosslinking of DmSNAPc to DNA probes that contain the U1 or U6 PSEAs.** At the top of the figure, the positions of the PSEA (U1 or U6) and the PSEB (U1) or TATA box (U6) are shown relative to the DNA helices shown below. The magenta bars indicate the five positions where the U1 and U6 PSEAs differ. The crosslinking patterns for each of the three subunits are shown separately. The colored spheres (blue for DmSNAP43, green for DmSNAP50, or yellow for DmSNAP190) indicate phosphate positions that crosslinked strongly to the respective subunits. Lighter colored spheres indicate positions of weaker cross-linking. Even-numbered phosphates are on the template strand, and odd are on the non-template strand.

These studies from the photo-cross-linking assays suggest that the structural conformations of SNAPc-DNA-complexes are different when SNAPc binds on either the U1 or U6 PSEA. A schematic diagram depicting these differences in conformation is shown in Fig. 3.
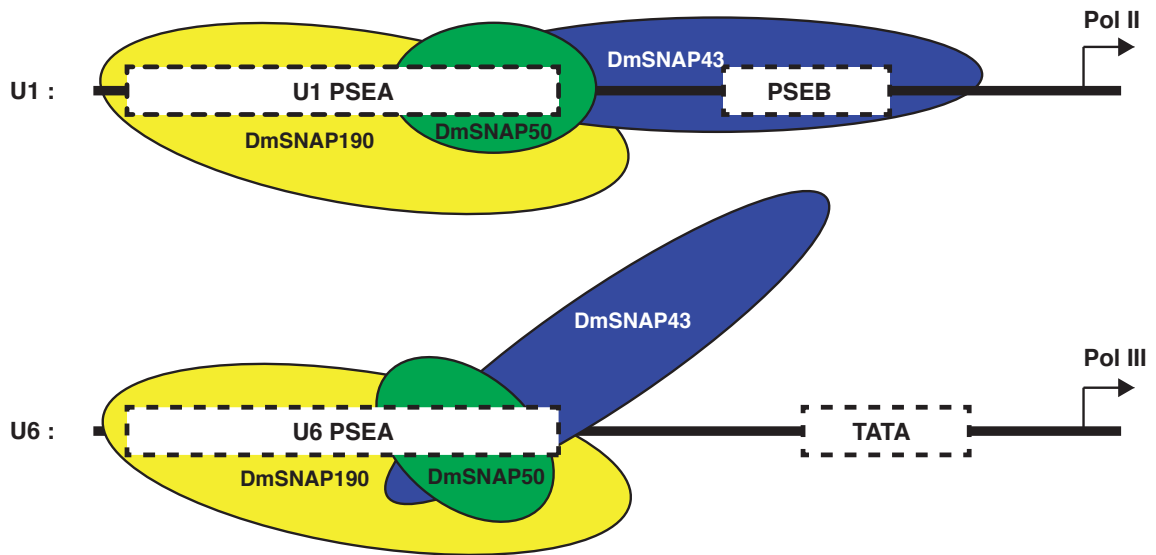
Figure 3. **Schematic model illustrating different conformations of DmSNAPc contacting the U1 or U6 snRNA gene promoters**. The yellow ovals represent DmSNAP190, which contacts the entire PSEA of both the U1 and U6 snRNA genes. DmSNAP50 (green) and DmSNAP43 (blue) each bind to the 3' half of the U1 and U6 PSEAs. Most notably, DmSNAP43 contacts DNA far downstream of the U1 PSEA but not far downstream of the U6 PSEA.

DmSNAPc subunit domains required for complex assembly and DNA binding

The subunits of DmSNAPc stably interact with each other in solution prior to binding to DNA (Su et al., 1997), and all three subunits are required for sequence-specific binding by DmSNAPc (Stumph lab, unpublished data). Our lab has used mutational analysis to identify domains within each DmSNAP subunit that are required for interaction with the other two subunits (Hung et al., 2009). One of the main

conclusions from those findings was that, with one exception, the most evolutionarily conserved region of each DmSNAPc subunit was required for its association with the other two subunits. Regarding the DmSNAP190 interaction with the two other DmSNAP subunits, the conserved Myb domain of DmSNAP190 (see section below) was sufficient for its interaction with the N-terminal region of DmSNAP43, but a non-conserved region of DmSNAP190 (residues 63-175) was required for interaction with DmSNAP50 in order to form the DmSNAP complex (Hung et al., 2009). However, DNA binding by DmSNAPc is dependent not only upon the conserved regions but also upon regions outside of the conserved regions, including regions of DmSNAP190 both N-terminal and C-terminal of the Myb domain (Hung et al., 2009; Hung et al., 2012).

DmSNAP190 contains a Myb domain

My own work, reported in this thesis, has concentrated on DmSNAP190, the largest subunit of DmSNAPc. SNAP190 has a unique domain that consists of 4.5 tandem Myb repeats, termed respectively Rh, Ra, Rb, Rc, and Rd (Fig. 4) (Wong et al., 1998; Li et al., 2004). Originally, Myb repeats were first characterized as three imperfect repeats, termed R1, R2, and R3 (each ~52 amino acids residues in length) that form the DNA-binding domain of the Myb oncoprotein (Biedenkapp et al., 1988; Sakura et al., 1989; Tanikawa et al., 1993). Each Myb repeat contains three conserved tryptophan amino acid residues (sometimes replaced by phenylalanine or tyrosine residues containing aromatic side chains) that form a hydrophobic core important for the structural integrity of the DNA-binding domain as well as the structure of the individual repeats (Ogata et al.,

1994; Wong et al., 1998; Tahirov et al., 2002). The structural analysis of the Myb-DNA interaction by using NMR spectrometry and X-ray diffraction analysis revealed that each Myb repeat contains three helices with the last two forming a variant of the helix-turn-helix motif. The R2 and R3 Myb repeats are closely packed in the major groove of the DNA (Ogata et al., 1994; Tahirov et al., 2002).

```
                                           ___α1_____      __α2___      ___α3____
Mouse c-Myb R1 repeat    LGKTRWTREEDEKLKKLVEQN---GTDDW-KVIANYLPNRTDVQCQHRWQKVLNPE
Fly   Rb Myb repeat      KNSHRWSEEDNDKLRAIVDRNTANSVINW-KKVVEYFPDKSKSTLIGRYYYVLHPS
Human Rb Myb repeat      LKRKEWTEEEDRMLTQLVQEMRVGSHIPY-RRIVYYMEGRDSMQLIYRWTKSLDPG

Mouse c-Myb R2 repeat    LIKGPWTKEEDQRVIELVQKY---GPKRW-SVIAKHLKGRIGKQCRERWHNHLNPE
Fly   Rc Myb repeat      ISHEPFTTKEDMMLFAAVEEY---NGK-FHCFPRSLFPNRSLTQLRTRYHNVLAQR
Human Rc Myb repeat      LKKGYWAPEEDAKLLQAVAKY---GEQDW-FKIREEVPGRSDAQCRDRYLRRLHFS

Mouse c-Myb R3 repeat    VKKTSWTEEEDRIIYQAHKRLG----NRW-AEIAKLLPGRTDNAIKNHWNSTMRRKV
Fly   Rd Myb repeat      NKTDSWSVQDDTRLMSFVTQYG---ASQW-LNCATFLGNHTRTSCRTRFLVIKKFLE
Human Rd Myb repeat      LKKGRWNLKEEEQLIELIEKYG---VGHW-AKIASELPHRSGSQCLSKWKIMMGKKQ
```

Figure 4. **Schematic structure of the DmSNAP190 protein and alignment with mouse c-Myb repeats and human SNAP190 Rb, Rc, Rd Myb repeats**. (A) DmSNAP190 consists of 721 amino acids and includes a Myb domain comprised of 4.5 tandem Myb repeats (Ra, Rb, Rc, and Rd are full length repeats while Rh is a half Myb repeat). (B) Amino acid alignment of mouse c-Myb repeats with *Homo sapiens and Drosophila melanogaster* Rb, Rc, and Rd Myb repeats. A multiple sequence alignment of Myb repeats by using ClustalW (Thompson et al., 1994).

Subject matter of this dissertation

Although the atomic structure of the SNAPc–DNA complex has not been determined, the site-specific protein–DNA photo-cross-linking assay has allowed us to learn much about the architecture of the protein-DNA complex on U1 and U6 promoter sequences. The photo-cross-linking technique previously allowed our group to identify the specific nucleotide positions contacted by each of the three DmSNAPc subunits on U1 and U6 PSEAs (Fig. 2) (Wang and Stumph, 1998; Li et al., 2004). However, we were interested in mapping or localizing regions or sub-regions within each DmSNAP subunit that cross-link to specific phosphate positions of the U1 or U6 PSEAs. To do this, our lab developed a novel protocol that combines the site-specific protein-DNA photo-cross-linking technique with subsequent site-specific chemical cleavage of the polypeptides (Kim et al., 2010).

Chapter 1 describes work in which my fellow graduate student, Mun Kyoung Kim, applied this technique to map several domains in DmSNAP43 and DmSNAP50 that cross-linked to a number of discrete phosphate positions in the U1 and U6 PSEAs. This chapter also describes my work that determined the orientation of DmSNAP190 on the U1 PSEA. Results from this first chapter have been published in the journal Molecular and Cellular Biology (Kim et al., 2010) and the dissertation author is a co-first author of this research.

Chapter 2 describes our further mapping and localization of DmSNAP190 subunit domains that cross-link to all 13 nucleotide positions within the U1 PSEA that can be cross-linked to DmSNAP190 (Fig. 2). This work provided detailed information about the arrangement of all 4.5 Myb repeats of DmSNAP190 on the U1 PSEA, and further revealed surprisingly that the region of DmSNAP190 N-terminal of the Myb domain

repeats makes extensive contacts with the 3' half of the U1 PSEA (Doherty et al. 2012).

Results from this second chapter have been published in the Journal of Biological

Chemistry (2012), and the dissertation author is a co-first author of this research.

Chapter 3 describes the localization of DmSNAP190 domains that cross-link with

each of the 16 nucleotide positions of the U6 PSEA that can cross-link to DmSNAP190

when DmSNAPc binds to a U6 PSEA (Fig. 2). This work demonstrated that there are

significant differences in the positioning of at least two of the DmSNAP190 Myb repeats

when DmSNAPc binds to a U6 PSEA as opposed to a U1 PSEA. By combining all of

these results, we are able for the first time to present comprehensive models of the

conformational differences adopted by DmSNAPc on U1 and U6 promoters.

REFERENCES

Barakat N. H., and Stumph W. E. (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.

Biedenkapp, H., Borgmeyer, U., Sippel, A. E., and Klempnauer, K. H. (1988). Viral myb oncogene encodes a sequence-specific DNA binding activity. Nature *335*, 835-837.

Bond, U. M., Yario, T. A., Steitz, J. A. (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*, 1709-22.

Busch, H., Reddy, R., Rothblum L., Choi Y. C. (1982). SnRNAs, SnRNPs, and RNA processing. Annu Rev Biochem, *51*, 617-54.

Dahlberg, J. E., and Lund, E. (1988). The genes and transcription of the major small nuclear RNAs, In Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles, M. L. Birnstiel, ed. (Heidelberg, Federal Republic of Germany: Springer Verlag KG), pp. 38-70.

Doherty, M. T., Kang Y. S., Lee, C. and Stumph, W. E. (2012) Architectural Arrangement of the Small Nuclear RNA (snRNA)-activating Protein Complex 190 Subunit (SNAP190) on U1 snRNA Gene Promoter DNA. J. Biol. Chem *287*:39369-39379.

Goomer, R. S., Urso, O., and Kunkel, G. R. (1994). A complex that contains proteins binding to the PSE and TATA sites in a human U6 small nuclear RNA promoter. Gene *148*, 269-275.

Henry, R. W., Sadowski, C. L., 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, R. W., Mittal, V., Ma, B. C., 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, 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.

Hernandez, G., Valafar, F., and Stumph, W. E. (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. (2001). Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription. J Biol Chem *276*, 26733-6.

Hinkley, C. S., Hirsch, H. A., Gu, L. P., LaMere, B., Henry, R. W. (2003). The small nuclear RNA-activating protein 190 Myb DNA binding domain stimulates TATA box-binding protein-TATA box recognition. J Biol Chem, *278*, 18649-57.

Hung, K. H., Titus, M., Chiang, S. C. and Stumph, W. E. (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.

Hung, K. H., Stumph, W. E. (2011). Regulation of snRNA gene expression by the *Drosophila melanogaster* small nuclear RNA-activating protein complex (DmSNAPc). Crit. Rev. Biochem. Mol. Biol. *46*, 11–26.

Hung, K. H., and Stumph, W. E. (2012) Localization of residues in a novel DNA-binding domain of DmSNAP43 required for DmSNAPc DNAbinding activity. FEBS Lett. *586*, 841–846.

Jawdekar, G. W., Hanzlowsky, A., Hovde, S. L., Jelencic, B., Feig, M., Geiger, J. H. and Henry, R. W. (2006). The unorthodox SNAP50 zinc finger domain contributes to cooperative promoter recognition by human SNAPc. J Biol Chem *281*:31050–31060.

Jensen, R. C., Wang, Y., Hardin, S. B., and Stumph, W. E. (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, J. A., Sollner-Webb, B. (1990). The U3 small nucleolar ribonucleoprotein functions in the first step of preribosomal RNA processing. Cell, *60*, 897-908.

Kim, M. K., Kang, Y. S., Lai, H. T., Barakat, N. H., Magante, D., and Stumph, W. E. (2010) Identification of SNAPc subunit domains that in- teract with specific nucleotide positions in the U1 and U6 gene promoters. Mol. Cell Biol. *30*, 2411–2423; Correction (2010) Mol. Cell Biol. *30*, 5257.

Kwek, K. Y., Murphy, S., Furger, A., Thomas, B., O'Gorman, W., Kimura, H., Proudfoot, N. J., Akoulitchev, A. (2002). U1 snRNA associates with TFIIH and regulates transcriptional initiation. Nat Struct Biol, *9*, 800-5.

Lai, H. T., Chen, H., Li, C., McNamara-Schroeder, K. J., and Stumph, W. E. (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.

Li, C., Harding, G. A., Parise, J., McNamara-Schroeder, K. J., and Stumph, W. E. (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.

Lobo, S. M., and Hernandez, N. T. (1994). Transcription of snRNA genes by RNA polymerases II and III, In Transcription: Mechanisms and Regulation, R. C. Conaway, and J. W. Conaway, eds. (New York, NY: Raven Press), pp. 127-159.

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.

McNamara-Schroeder, K. J., Hennessey, R. F., Harding, G. A., Jensen, R. C., and Stumph, W. E. (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 B. C. 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.

Nguyen, V. T., Kiss, T., Michels, A. A., Bensaude, O. (2001). 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. Nature, *414*, 322-5

Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, A., Inoue, T., Kanai, H., Sarai, A., Ishii, S., and Nishimura, Y. (1994) Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recog- nition helices. Cell *79*, 639 – 648

Parry, H. D., Scherly, D., and Mattaj, I. W. (1989). Snurpogenesis: the transcription and assembly of U snRNP components. Trends BiochemSci *14*, 15-19.

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

Reddy, R., and Busch, H. (1988). Small nuclear RNAs: RNA sequences, structure, and modifications, In Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles, M. L. Birnstiel, ed. (Heidelberg, Germany: Springer Verlag KG), pp. 1-37.

Sadowski, C. L., Henry, R. W., Lobo, S. M., 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.

Sakura, H., Kanei-Ishii, C., Nagase, T., Nakagoshi, H., Gonda, T. J., and Ishii, S. (1989) Delineation of three functional domains of the transcrip- tional activator encoded by the c-myb protooncogene. Proc. Natl. Acad. Sci. U.S.A. *86***,** 5758 –5762

Sharp, P. A. (1994). Split genes and RNA splicing. Cell *77*, 805-816.

Steitz, J. A., Black, D. L., Gerke, V., Parker, K. A., Kramer, A., Frendewey, D., Keller, W. (1988). Functions of the abundant U-snRNPs. In: Birnsteil ML, ed. Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles. Heidelberg, Federal Republic of Germany: Springer Verlag KG, 115-54.

Su, Y., Song, Y., Wang, Y., Jessop, L., Zhan, L. C., and Stumph, W. E. (1997). Characterization of a *Drosophila* proximal-sequence-element-binding protein involved in transcription of small nuclear RNA genes. Eur J Biochem *248*, 231-237.

Tahirov,T. H., Sato,K., Ichikawa-Iwata, E., Sasaki,M., Inoue-Bungo, T., Shiina, M., Kimura, K., Takata, S., Fujikawa, A., Morii, H., Kumasaka, T., Yamamoto, M., Ishii, S., and Ogata, K. (2002) Mechanism of c-Myb-C/EBPβ cooperation from separated sites on a promoter. Cell *108***,** 57–70

Tanikawa, J., Yasukawa, T., Enari, M., Ogata, K., Nishimura, Y., Ishii, S., and Sarai, A. (1993) Recognition of specific DNA sequences by the c-myb protooncogene product. Role of three repeat units in the DNA-binding domain. Proc. Natl. Acad. Sci. *U.S.A. 90*, 9320 –9324

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Research, *22*(22):4673-4680

Waldschmidt, R., Wanandi, I., and Seifart, K. H. (1991). Identification of transcription factors required for the expression of mammalian U6 genes *in vitro*. EMBO J *10*, 2595-2603.

Wang, Y., and Stumph, W. E. (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, M. W., Henry, R. W., Ma, B. C., Kobayashi, R., Klages, N., Matthias, P., Strubin, M., and Hernandez, N. (1998).. Mol Cell Biol *18*, 368-377.

Yang, Z. Y., Zhu, Q. W., Lu,o K. X., Zhou, Q. (2001). The 7SK small nuclear RNA

inhibits the CDK9/cyclin T1 kinase to control transcription. Nature, *414*, 317-22.

Yoon, J. B., Murphy, S., Bai, L., Wang, Z., and Roeder, R. G. (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.

Zamrod, Z., Tyree, C. M., Song, Y., and Stumph, W. E. (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.

CHAPTER 1

Identification of SNAPc Subunit Domains That Interact with Specific Nucleotide Positions in the U1 and U6 Gene Promoters

Identification of SNAPc Subunit Domains That Interact with Specific Nucleotide Positions in the U1 and U6 Gene Promoters[▽]

Mun Kyoung Kim,[1,2][†][‡] Yoon Soon Kang,[3][†] Hsien-Tsung Lai,[1,2][¶] Nermeen H. Barakat,[3][§]
Deodato Magante,[3][#] and William E. Stumph[1,3][*]

*Molecular Biology Institute,[1] Department of Biology,[2] and Department of Chemistry and Biochemistry,[3]
San Diego State University, San Diego, California 92182-1030*

The small nuclear RNA (snRNA)-activating protein complex (SNAPc) is essential for transcription of genes coding for the snRNAs (U1, U2, etc.). In *Drosophila melanogaster*, the heterotrimeric DmSNAPc recognizes a 21-bp DNA sequence, the proximal sequence element A (PSEA), located approximately 40 to 60 bp upstream of the transcription start site. Upon binding the PSEA, DmSNAPc establishes RNA polymerase II preinitiation complexes on U1 to U5 promoters but RNA polymerase III preinitiation complexes on U6 promoters. Minor differences in nucleotide sequence of the U1 and U6 PSEAs determine RNA polymerase specificity; moreover, DmSNAPc adopts different conformations on these different PSEAs. We have proposed that such conformational differences in DmSNAPc play a key role in determining the different polymerase specificities of the U1 and U6 promoters. To better understand the structure of DmSNAPc-PSEA complexes, we have developed a novel protocol that combines site-specific protein-DNA photo-cross-linking with site-specific chemical cleavage of the protein. This protocol has allowed us to map regions within each of the three DmSNAPc subunits that contact specific nucleotide positions within the U1 and U6 PSEAs. These data help to establish the orientation of each DmSNAPc subunit on the DNA and have revealed cases in which different domains of the subunits differentially contact the U1 versus U6 PSEAs.

The *Drosophila melanogaster* small nuclear RNA (snRNA)-activating protein complex (DmSNAPc) is a heterotrimeric transcription factor (21) that is required for the synthesis of the U1, U2, U4, U5, and U6 spliceosomal snRNAs (2, 22, 31). Homologous protein complexes are required for snRNA gene expression in humans (1, 11, 12, 18, 33–35) and for spliced leader RNA synthesis in trypanosomes (6, 7, 30). This indicates that SNAPc appeared early in eukaryotic evolution and continues in contemporary times to be utilized for the transcription of important noncoding RNA molecules in diverse organisms. DmSNAPc binds sequence-specifically to an essential, conserved ~21-bp promoter element termed the proximal sequence element A (PSEA) that is located approximately 40 to 60 bp upstream of the transcription start site of fly snRNA genes (8, 13, 19, 23).

In animals, the U1 to U5 snRNA genes are transcribed by RNA polymerase II (Pol II), but U6 snRNA genes are transcribed by RNA Pol III (4, 5, 10, 14, 19, 24, 29, 31). Surprisingly, the primary determinant of the RNA polymerase specificity of the *D. melanogaster* snRNA genes is the precise sequence of the PSEA. A few conserved nucleotide differences in the 3′ half of the 21-bp PSEA are sufficient to determine the polymerase specificity of the fly snRNA genes *in vitro* and to restrict the polymerase specificity *in vivo* (2, 19, 20, 26).

The three subunits of DmSNAPc are termed DmSNAP190, DmSNAP50, and DmSNAP43 so that the names correspond to the most widely used nomenclature for the three orthologous human SNAPc subunits, SNAP190, SNAP50, and SNAP43. These three human subunits are also known as PTFα, PTFβ, and PTFγ, respectively (1, 11, 12, 33–35). Although all three of these subunits of human and fly SNAPc are required for sequence-specific binding of the complex to DNA, only one of the subunits (SNAP190) contains a canonical DNA-binding domain (33).

In earlier studies, we used site-specific protein-DNA photo-cross-linking to map the nucleotide positions within the U1 and U6 PSEAs that are contacted by each of the three subunits of DmSNAPc (32). Those results revealed that the DmSNAP190 subunit could be cross-linked to nucleotide positions extending over the entire length of the PSEA. In contrast, DmSNAP50 and DmSNAP43 cross-linked to nucleotides in the 3′ half of the 21-bp PSEA but not to nucleotides in the 5′ half of the PSEA.

Interestingly, the cross-linking patterns of the subunits, particularly of DmSNAP50 and DmSNAP43, were found to be significantly different depending upon whether DmSNAPc was bound to a U1 versus a U6 PSEA (32). Most strikingly, a later study (22) revealed that DmSNAP43 cross-linked to DNA extending 20 bp downstream of the U1 PSEA, but it cross-linked to DNA only 5 bp downstream of a U6 PSEA. Based

───────────────

* Corresponding author. Mailing address: Department of Chemistry and Biochemistry, San Diego State University, 5500 Campanile Dr., San Diego, CA 92182-1030. Phone: (619) 594-5575. Fax: (619) 594-4634. E-mail: wstumph@sciences.sdsu.edu.

‡ Present address: Program in Molecular Biology, 15 North 2030 East, EIHG Bldg. 533, Rm. 1400, University of Utah, Salt Lake City, UT 84112.

¶ Present address: 5323 Harry Hines Blvd. North Campus, ND2.114, Simmons Comprehensive Cancer Center, University of Texas Southwestern Medical Center at Dallas, Dallas, TX 75390.

§ Present address: Division of Hematology/Oncology, Department of Medicine, Case Western University, 10900 Euclid Ave., Cleveland, OH 44106.

Present address: Diazyme-GA, P.O. Box 85608, San Diego, CA 92186-5608.

† These two authors have contributed equally to this work.

▽ Published ahead of print on 8 March 2010.

upon the cross-linking data as well as functional data relating to transcription, we have proposed a model in which the DNA sequence of the PSEA acts as a differential allosteric effector of DmSNAPc conformation that in turn leads to RNA polymerase specificity (2, 13, 20, 22, 32).

At this time, there is little or no knowledge about the three-dimensional structure of SNAPc or its subunits and how they bind to DNA. Thus, to better understand the mode of SNAPc binding to DNA, we have developed a novel protocol that combines the photo-cross-linking technique with subsequent site-specific chemical cleavage of the protein. This has allowed us to localize domains or regions of each of the three DmSNAP subunits that interact with specific individual nucleotide positions in the U1 and U6 PSEAs. Furthermore, novel regions of DmSNAP50 and of DmSNAP43 that contact the DNA have been mapped, and our results reveal for the first time the orientation of each subunit on the DNA. Finally, the data reveal that distinct domains of DmSNAP50 and of DmSNAP43 interact closely with the DNA depending upon whether DmSNAPc is bound to a U1 versus a U6 PSEA.

<div align="center">

MATERIALS AND METHODS

</div>

DmSNAPc constructs, expression, and purification. The preparation of untagged and N- and C-terminally FLAG-tagged constructs encoding each of the three DmSNAPc subunits has been recently described (16). Constructs with point mutations to eliminate or introduce hydroxylamine cleavage sites (asparaginyl-glycyl peptide bonds) were prepared by using Stratagene's QuikChange II site-directed mutagenesis kit. To minimize the possibility of harmful changes, conservative mutations (e.g., N to Q) were chosen for eliminating NG peptide bonds. In addition, new NG cleavage sites were introduced at residues that were not conserved in the orthologous proteins of other *Drosophila* species based upon genomic sequencing information available from Flybase (http://flybase .org/; data not shown).

Various forms of the three subunits of DmSNAPc were coexpressed in stably transfected S2 cells by using Invitrogen's *Drosophila* expression system as previously described (16, 21, 22). For mapping domains within DmSNAP190, N- or C-terminally tagged DmSNAP190 was coexpressed with untagged DmSNAP50 and DmSNAP43, whereas untagged DmSNAP190 was coexpressed with untagged DmSNAP50 and N-terminally tagged DmSNAP43 (to allow FLAG purification). For mapping DmSNAP50, N- or C-terminally tagged DmSNAP50 was coexpressed with untagged DmSNAP190 and DmSNAP43, whereas untagged DmSNAP50 was coexpressed with N-terminally tagged DmSNAP190 and untagged DmSNAP43. To map DmSNAP43, N- or C-terminally tagged DmSNAP43 was coexpressed with untagged DmSNAP190 and untagged DmSNAP50, whereas untagged DmSNAP43 was coexpressed with N-terminally tagged DmSNAP190 and untagged DmSNAP50. The DmSNAP50 construct used in the DmSNAP43 mapping experiments contained three point mutations that eliminated all NG sites within the coexpressed DmSNAP50. This was done to prevent cross-linked DmSNAP50 cleavage products from interfering with identification of the DmSNAP43 digestion products.

Subunit expression was induced with copper sulfate, and DmSNAPc was purified by FLAG immunoaffinity chromatography as recently described (16). Electrophoretic mobility shift assays were carried out to confirm the DNA-binding activity of the FLAG-purified DmSNAPc variants (16). Although the results were not quantitative, no significant differences in the DNA-binding activity of the point mutants or of the tagged variants were apparent compared to that of wild-type DmSNAPc (data not shown). The samples were dialyzed against 25 mM HEPES K$^+$ (pH 7.6), 12.5 mM MgCl$_2$, 0.01 mM ZnCl$_2$, 0.1 mM EDTA, 100 mM KCl, 1 mM dithiothreitol (DTT), and 10% (by volume) glycerol to remove 3× FLAG peptide in preparation for the photo-cross-linking reactions.

Site-specific protein-DNA photo-cross-linking and hydroxylamine digestion. (i) Preparation of double-stranded site-specific DNA photo-cross-linking probes. Double-stranded DNA probes that contained a photo-cross-linking agent (azidophenacyl group) attached at specific individual phosphate positions within or downstream of the U1 or U6 PSEA sequence were prepared as previously described in detail (32), with one major exception. In previous works (20, 22, 32), two shorter oligonucleotides (usually 20 to 30 nucleotides long) were utilized. One contained an azidophenacyl group attached to a phosphorothioate located 5′ of the third nucleotide from the 5′ end. The other served as an upstream primer. These were annealed to a longer oligonucleotide (about 70 nucleotides in length) and extended with T4 DNA polymerase to form the double-stranded probe. In the current work, longer versions of the "short" oligonucleotides were synthesized so that the double-stranded probe could be formed without the DNA polymerase step. As an example, to make a DNA probe with an azidophenacyl group at phosphate position 12 of the template strand of the U1 PSEA, the 67-mer of the nontemplate strand, 5′-ACGAATT CATTCTTA<u>TAATTCCCAACTGGTTTT</u>AGCGGTACCGCCATGGAAAG GTATGGGATCCTCA-3′ (underlined nucleotides corresponding to the U1 PSEA), was annealed to the following oligonucleotides of the template strand: 3′-TGCTTAAGTAAGAAT<u>ATTAAGGGTTxGA*</u>-5′ (where x indicates the position of a phosphorothioate derivatized with azidophenacyl bromide and the asterisk indicates a ^{32}P) and 3′-<u>CCAAAATCG</u>CCATGGCGGTACCTTTCCA TACCCTAGGAGT-5′. After annealing, the nick was sealed by using T4 DNA ligase, and the double-stranded probe was purified by gel electrophoresis as previously reported (32).

Three additional long oligonucleotides were used for probe preparation depending upon the strand chosen to contain the cross-linker or the U6 versus the U1 PSEA. These were the U1 PSEA template strand 67-mer, 5′-TGAGGATC CCATACCTTTCCATGGCGGTACC<u>GCTAAAACCAGTTGGGAATTATA AGAATGAATTCGT</u>-3′; U6 PSEA nontemplate strand 67-mer, 5′-ACGAAT TCATTCTTA<u>TAATTCTCAACTGCTCTTT</u>CCGGTACCGCCATGGAAAG GTATGGGATCCTCA-3′; and U6 PSEA template strand 67-mer, 5′-TGAGG ATCCCATACCTTTCCATGGCGGTACC<u>GGAAAGAGCAGTTGAGAATT AT</u>AAGAATGAATTCGT-3′. The position of the cross-linking agent in the probe was determined by appropriately increasing or decreasing the lengths of the two shorter oligonucleotides used in the annealing reaction so that a fully double-stranded probe would be generated. In each case, the phosphate positions containing cross-linker and ^{32}P radiolabel were separated by a single unmodified phosphate. All oligonucleotides were purchased from Integrated DNA Technologies.

(ii) Photo-cross-linking of DmSNAPc to the DNA probes. Photo-cross-linking reactions were carried out as previously described (32) but were scaled up 10-fold. Briefly, 40 μl of FLAG-purified DmSNAPc was incubated with 1,000,000 cpm DNA probe in the dark for 30 min in a volume of 80 μl in a final buffer composition of 12.5 mM HEPES (pH 7.6), 50 mM KCl, 6.25 mM MgCl$_2$, 0.05 mM EDTA, and 5% (by volume) glycerol. The samples were then irradiated with 313 nm UV light for 333 s. Following irradiation, cross-linked samples were subjected to DNase I and S1 nuclease digestion to remove all but 2 or 3 nucleotides of the DNA probe. The success of the cross-linking reaction was verified by analyzing 1/10 of each sample on denaturing polyacrylamide gels as previously described (32).

(iii) Chemical digestion of cross-linked protein and gel electrophoresis. The remaining 9/10 of each reaction (volume, ~160 μl) was placed into dialysis tubing (Spectra/Por 1, 6,000 to 8,000 molecular weight cutoff; catalog no. 132 645) and dialyzed at 45°C against 400 ml of 1.8 M hydroxylamine (Sigma catalog no. 159417) solution (pH 9.0) freshly prepared as described in *Current protocols in protein science* (3). After 5 h of dialysis to allow chemical cleavage, the samples were transferred and dialyzed against 6 mM Tris-HCl (pH 6.8), 0.25 M urea, 0.2% sodium dodecyl sulfate, and 0.5% β-mercaptoethanol for 4 h at 25°C to remove the hydroxylamine salts in preparation for gel electrophoresis. The samples were then evaporated to dryness in a Savant SpeedVac concentrator and then redissolved in an amount of water that was 1/10 the volume prior to the drying step. Then a solution of 50% glycerol, 25% β-mercaptoethanol, and 0.125% bromophenol blue was added to produce final concentrations of 10% glycerol, 5% β-mercaptoethanol, and 0.025% bromophenol blue. Samples were then run on 11 to 15% denaturing polyacrylamide gels. Following electrophoresis, gels were dried and subjected to autoradiography.

Immunoblots of chemically digested DmSNAPc. FLAG affinity-purified DmSNAPc samples (200 μl of each) were treated exactly as described above for digestion with hydroxylamine, except that the incubation with DNA photo-cross-linking probe and UV irradiation was omitted. Immunoblotting was carried out as recently described (16). N-terminal FLAG-tagged fragments of the DmSNAP subunits were detected by using alkaline phosphatase-conjugated monoclonal antibody against the FLAG epitope (Sigma catalog no. A9469). C-terminal fragments of DmSNAP190 were detected by using a polyclonal rabbit antibody generated against a 14-amino-acid peptide from the C terminus of DmSNAP190 (22) as the primary antibody. Antibodies that were generated by injecting rabbits with full-length bacterially expressed DmSNAP50 or DmSNAP43 (2) were uti-

A

U1 PSEA Non-template Strand   **U6 PSEA Non-template Strand**

— DmSNAP190 —

— DmSNAP50 —
— DmSNAP43

```
        1  3  5  7  9 11 13 15 17 19 21 23 25            1  3  5  7  9 11 13 15 17 19 21 23 25
        *  *  *  *  *  *  *  *  *  *  *  *  *            *  *  *  *  *  *  *  *  *  *  *  *  *
-61 TAATTCCCAACTGGTTTTAGCGGTAC            -63 TAATTCTCAACTGCTCTTTCCGGTAC
    ATTAAGGGTTGACCAAAATCGCCATG                ATTAAGAGTTGACGAGAAAGGCCATG
        *  *  *  *  *  *  *  *  *  *  *  *            *  *  *  *  *  *  *  *  *  *  *  *
        2  4  6  8 10 12 14 16 18 20 22 24            2  4  6  8 10 12 14 16 18 20 22 24
```

— DmSNAP190 —

— DmSNAP50 —
— DmSNAP43 —

U1 PSEA Template Strand   **U6 PSEA Template Strand**

B

```
                          358
        1                  ↓                                    721
DmSNAP190:  N[        |Rh|Ra|Rb|Rc|Rd|          ]C

      179 234,237
        1    ↓ ↓         377              1          192        363
DmSNAP50:  N[   |Zn++|]C   DmSNAP43:  N[        ↓        ]C
```
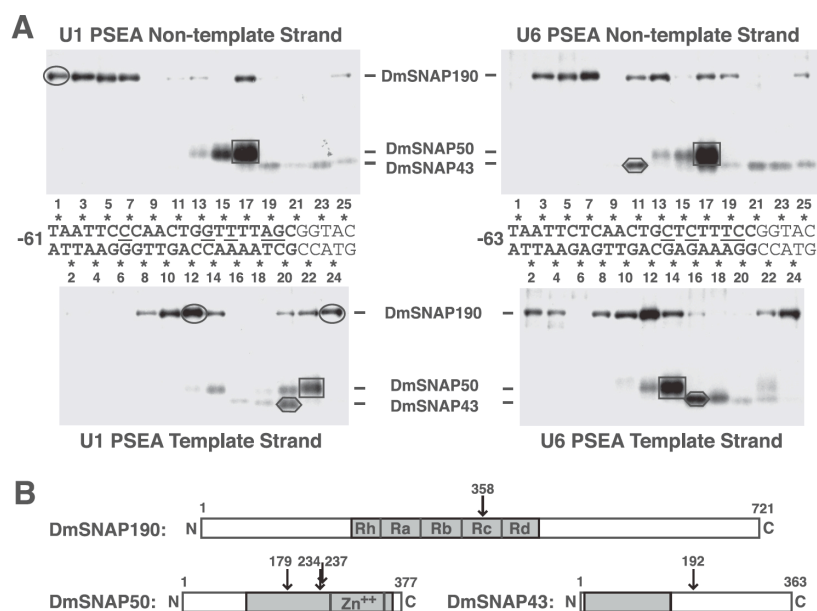
FIG. 1. (A) Site-specific protein-DNA photo-cross-linking data taken from Wang and Stumph (32). Cross-linking positions utilized in the current study are indicated by ovals, squares, and hexagons for DmSNAP190, DmSNAP50, and DmSNAP43, respectively. The sequences of the PSEAs are from the *D. melanogaster* U1:95Ca gene (left) and U6:96Ab gene (right) and differ at the five positions indicated by underlining (13). The 21-bp PSEA sequences (in bold type) extend from position −61 to −41 (left to right) relative to the U1 transcription start site and from position −63 to −43 (left to right) relative to the U6 start site. (B) Structural features of DmSNAP190, DmSNAP50, and DmSNAP43 which consist of 721, 377, and 363 amino acid residues, respectively. The shaded areas indicate the regions that are greater than 26% identical and 42% similar to the human SNAP orthologs. The illustration also shows that SNAP190 contains 4.5 Myb repeats designated Rh, Ra, Rb, Rc, and Rd. SNAP50 contains a unique zinc binding domain as indicated. Downward arrows indicate the locations of NG peptide bonds in the wild-type proteins.

lized to detect cleavage fragments from those two subunits. The population of anti-DmSNAP50 polyclonal antibodies readily detected both N- and C-terminal fragments when DmSNAP50 was cleaved at position 234 but detected the N-terminal fragment very weakly when cleavage was at position 179. The N-terminal cleavage fragment that contained only 102 amino acid residues could not be detected by these antibodies. The population of DmSNAP43 polyclonal antibodies strongly detected the C-terminal fragments when cleavage of DmSNAP43 occurred at either position 192 or 272. The N-terminal fragments in each case yielded very weak signals, but the corresponding bands were clearly present with longer development times. In all cases, the secondary antibody was alkaline phosphatase-conjugated goat anti-rabbit IgG(Fc) (Promega catalog no. S3731).

RESULTS

Figure 1A shows previously published photo-cross-linking data that initially defined the interaction of the three DmSNAPc subunits with specific nucleotide positions within the U1 and U6 PSEAs (32). For those experiments, photo-cross-linking reagent (together with an adjacent ^{32}P radiolabel) was incorporated into the phosphate backbone individually at every other nucleotide position (indicated by the asterisks above and below the DNA sequences in the middle of Fig. 1A) in the nontemplate (upper) strand or in the template (lower) strand of the DNA. This was done for both the U1 (left) and U6 (right) PSEAs (21-bp sequences shown in bold type). Following incubation of the DNA probe with DmSNAPc, subunits of DmSNAPc that cross-linked at each position were identified following UV irradiation, nuclease digestion, denaturing poly-

acrylamide gel electrophoresis, and autoradiography. Phosphate positions that were chosen for the protein-mapping studies in the current report are enclosed within ovals, rectangles, or hexagons for DmSNAP190, DmSNAP50, and DmSNAP43, respectively. Although not shown in Fig. 1A, for the current studies, we also employed positions 28 and 40 of the template strand downstream of the U1 PSEA, because positions 28 and 40 cross-linked exclusively to DmSNAP43 when DmSNAPc was bound to a U1 (but not U6) PSEA (22).

To localize domains within the individual DmSNAPc subunits that cross-linked to the chosen phosphate positions, we carried out site-specific protein-DNA photo-cross-linking and followed this with cleavage of the proteins with hydroxylamine reagent. This allowed the specific radiolabeled fragment(s) of the hydroxylamine-digested protein that cross-linked to the DNA to be identified by relative size on a denaturing polyacrylamide gel.

Hydroxylamine cleaves polypeptides preferentially at NG peptide bonds (3). The positions of NG peptide bonds in the wild-type DmSNAP subunits are shown in Fig. 1B. DmSNAP190 and DmSNAP43 each contain a single NG bond, whereas DmSNAP50 contains three NG bonds, two of which are located within three residues of each other. Site-directed mutagenesis was used to eliminate NG bonds and/or to introduce NG bonds at new positions in order to provide novel hydroxylamine cutting sites. This allowed us to increase the resolution of the mapping
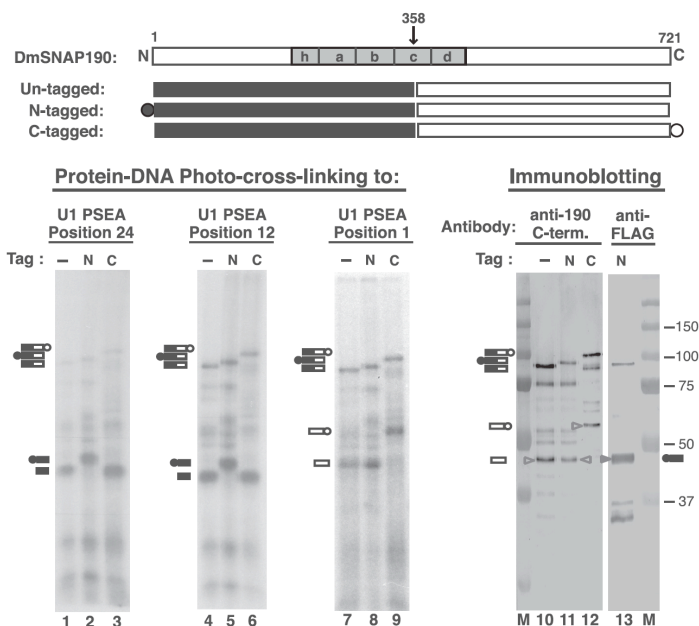
FIG. 2. U1 PSEA positions 24 and 12 cross-link to the N-terminal half of DmSNAP190, but position 1 cross-links to the C-terminal half of DmSNAP190. The upper line in the figure represents wild-type DmSNAP190 and the location of the 4.5 Myb repeats. Below this, the two halves of DmSNAP190 (darkly shaded and unshaded) that result from hydroxylamine cleavage following amino acid residue 358 in the untagged, N-tagged, and C-tagged DmSNAP190 constructs are shown. The circles at the N terminus or at the C terminus represent the tags on the DmSNAP190 constructs. The lower part of the figure shows autoradiograms of denaturing gels of the cross-linked protein fragments after hydroxylamine digestion when cross-linking was carried out to U1 PSEA position 24 (lanes 1 to 3), position 12 (lanes 4 to 6), or position 1 (lanes 7 to 9). The symbols alongside the gel indicate the origin of each band and correspond to the symbols in the upper part of the figure. Lanes 10 to 13 show immunoblots of cleaved protein fragments detected by using antibodies against a C-terminal peptide of DmSNAP190 or against the FLAG epitope. Lane M contained molecular mass markers (in kDa). Arrowheads (unshaded for C-terminal and shaded for N-terminal fragments) indicate the specific cleavage products of interest. Bands corresponding to residual undigested full-length DmSNAP190 (untagged, N-tagged, and C-tagged) are observed in the upper part of all panels.

experiments within the primary amino acid sequence of the proteins. For reasons indicated below, most DmSNAP constructs were expressed as untagged, N-terminally FLAG-tagged, and C-terminally FLAG-tagged variants (16). *In vivo*-assembled DmSNAPc was prepared by co-overexpressing tagged or untagged wild-type or mutant variants of all three subunits in stably transfected *D. melanogaster* S2 cells and purified by FLAG affinity chromatography (16, 22).

Mapping regions of DmSNAP190 that contact nucleotide positions 24, 12, and 1 of the U1 PSEA. As shown in Fig. 1A, DmSNAP190 could be cross-linked to nucleotides stretching over the entire length of the U1 PSEA. We wished to localize the regions of DmSNAP190 that cross-linked to the 3′ end, to the middle, and to the 5′ end of the U1 PSEA. For this, we selected positions 24, 12, and 1, respectively, because DmSNAP190 is the only subunit of DmSNAPc that cross-linked to each of these three positions (Fig. 1A). DmSNAP190 contains 4.5 Myb repeats termed Rh, Ra, Rb, Rc, and Rd (Fig. 1B), and this Myb domain is the region that is most evolutionarily conserved relative to human SNAP190 (22). The Myb domain of human SNAP190, in particular the Rc and Rd repeats, has been implicated in DNA binding by human

SNAPc (15, 25, 28, 33). The orientation of the 4.5 Myb repeats on the PSEA, however, is not known.

The NG site at position 358 in wild-type DmSNAP190 is located near the middle of the Rc Myb repeat; this is also very near the middle of the entire 721-amino-acid polypeptide chain (Fig. 1B). Thus, cleavage with hydroxylamine should yield two fragments of nearly equal size that would be indistinguishable on denaturing gels. To positively identify whether the N-terminal half or the C-terminal half of DmSNAP190 cross-linked to each of the PSEA positions 24, 12, or 1, we expressed three different versions of DmSNAP190: untagged, N-terminally tagged, and C-terminally tagged (Fig. 2, upper section). The N-terminal tag increased the size of the N-terminal fragment by 2.2 kDa, and the C-terminal tag increased the size of the C-terminal fragment by 5.7 kDa.

Figure 2 shows results from the cross-linking of DmSNAPc to positions 24, 12, and 1 of the U1 PSEA. When cross-linked to position 24, DmSNAP190 tagged at the N terminus (lane 2) produced a cross-linked radiolabeled fragment with a decreased mobility relative to the cross-linked fragments obtained from untagged or C-terminally tagged DmSNAP190 (lanes 1 and 3, respectively). From these results, we conclude

that position 24 of the U1 PSEA cross-linked to the N-terminal fragment of DmSNAP190 when the protein was cleaved at position 358 within the polypeptide chain. The same pattern was obtained when the cross-linking agent was placed at position 12 of the U1 PSEA (Fig. 2, lanes 4 to 6).

However, a substantially different result was obtained when photo-cross-linking was carried out to position 1 of the U1 PSEA. In this case (Fig. 2, lanes 7 to 9), the fragments from the untagged and N-terminally tagged DmSNAP190 migrated to the same position (lanes 7 and 8), but the fragment obtained from the C-terminally tagged DmSNAP190 migrated with a slower mobility. This indicated that nucleotide position 1, at the 5′ end of the U1 PSEA, cross-linked to the C-terminal half of DmSNAP190 cleaved at position 358.

To further confirm the identities of the bands on the auto-radiograms, immunoblots were carried out on the hydroxyl-amine-digested proteins. Lanes 10 to 12 of Fig. 2 show that antibodies directed against a peptide at the C terminus of DmSNAP190 produced the same pattern of bands observed in the cross-linking experiments in lanes 7 to 9, consistent with the evidence that position 1 cross-linked to the C-terminal fragment of DmSNAP190. Furthermore, the FLAG antibodies detected a band (lane 13) that migrated similarly to the cross-linked amino-terminally tagged bands observed in lanes 2 and 5. (This band also comigrated with the untagged C-terminal fragment.)

From the data in Fig. 2, we conclude that residues located within the N-terminal half of DmSNAP190 cross-linked to positions 24 and 12 of the U1 PSEA. On the other hand, residues located within the C-terminal half of DmSNAP190 cross-linked to U1 PSEA position 1.

To map the locations of the cross-linked residues of DmSNAP190 with greater precision (i.e., with a resolution of approximately one Myb repeat), additional DmSNAP190 constructs that each contained a single hydroxylamine cleavage site near the center of either the Ra, Rb, or Rd repeat, or just N-terminal or C-terminal to the Myb domain, were prepared (Fig. 3A). In each of the mutant constructs, the normal cleavage site in the wild-type sequence at position 358 was eliminated by changing asparagine 358 to glutamine. Schematic diagrams of these constructs are shown in Fig. 3B. Each of these constructs was prepared with the FLAG tag present at the N terminus. Cleavage at the single NG site in each construct H through E was expected to yield a pattern of C-terminal fragments of decreasing size and N-terminal fragments of increasing size. Following hydroxylamine cleavage of the protein, the C- and N-terminal fragments of DmSNAP190 were detected on immunoblots by using antibodies specific for the C terminus or for the FLAG-tagged N terminus (Fig. 3C, lanes 1 to 12). Arrowheads point out the position of the relevant band in each lane.

Figure 3C (lanes 13 to 18) shows the photo-cross-linking pattern obtained when the cross-linking agent was at position 24 of the U1 PSEA. Although the N- and C-terminal fragments of wild-type DmSNAP190 (construct C) did not resolve in lane 16, results shown in Fig. 2 had already demonstrated that position 24 cross-linked to the N-terminal fragment of Dm-SNAP190. When the cleavage site was moved farther C-terminal (constructs D and E), the N-terminal fragment, as expected, continued to cross-link (Fig. 3C, lanes 17 and 18).

However, when the cleavage site was moved to the center of repeat Rb (construct B), it was the C-terminal fragment that cross-linked (lane 15). Together, these results indicated that position 24 of the U1 PSEA cross-linked to amino acids of DmSNAP190 located between residues 306 and 358. When cross-linking was carried out to position 12 of the U1 PSEA (Fig. 3C, lanes 19 to 24), an identical pattern was observed. Thus, position 12 also cross-linked to DmSNAP190 residues located between residues 306 and 358.

Lanes 25 to 30 of Fig. 3C show that a different pattern of results was obtained when DmSNAPc was cross-linked to position 1 of the U1 PSEA. The pattern was similar to that observed in lanes 13 to 24 of Fig. 3C but with two important exceptions. Data in Fig. 2 had indicated that the band observed in lane 28 (construct C) arose from cross-linking to the C terminus of wild-type DmSNAP190 (in contrast to the N-terminal cross-linking shown in lanes 16 and 22). When the cleavage site was moved to the center of the Rd repeat (construct D), the band of strongest intensity (lane 29) corresponded also to the C-terminal fragment. Thus, amino acids C-terminal of residue 409 cross-linked most strongly to position 1 of the U1 PSEA. However, a weaker band that corresponded to the N-terminal fragment of construct D was also visible in lane 29 (compare its position to the corresponding fragment in lanes 17 and 23), suggesting that amino acid residues between 359 and 409 were also able to cross-link to position 1 of the U1 PSEA, but with lower intensity. When the protein cleavage site was moved C-terminal to the Myb domain (construct E), cross-linking could only be detected to the N-terminal fragment (lane 30). Thus, when DmSNAPc bound the U1 PSEA, the region of DmSNAP190 between positions 359 and 483 mapped in close proximity to position 1 of the U1 PSEA.

The results of the DmSNAP190 mapping experiments are summarized schematically in Fig. 3D. Amino acids localized to the C-terminal half of Rb and/or the N-terminal half of Rc contact positions 24 and 12 of the U1 PSEA. In contrast, amino acids localized to Rd and/or the C-terminal half of Rc closely approach position 1 of the U1 PSEA. These data further suggest that the DNA-binding domain of DmSNAP190 on the U1 PSEA is oriented with its N terminus proximal to the U1 transcription start site and its C terminus more distal from the start site.

Mapping domains of DmSNAP50 that contact nucleotide positions 22 and 17 of the U1 PSEA and positions 14 and 17 of the U6 PSEA. When DmSNAPc binds to the U1 or U6 PSEAs, the DmSNAP50 subunit cross-links most strongly to phosphate positions 17 and 22 of the U1 PSEA but to positions 17 and 14 of the U6 PSEA (Fig. 1A). We therefore chose to map regions within DmSNAP50 that contact these four nucleotide positions. DmSNAP50 contains no canonical DNA-binding domain, but it does contain an unorthodox zinc-binding domain near its C terminus (Fig. 1B) that has been termed the SNAP finger (17). Although this unique region has no homology to other types of DNA-binding zinc fingers, it has nevertheless been implicated in DNA binding by human SNAPc (17).

DmSNAP50 contains three NG peptide bonds (Fig. 1B). However, for the purposes of domain-mapping experiments, the two nearly adjacent sites at residues 234 and 237 can effectively be treated as a single hydroxylamine cleavage site. Initial mapping experiments were attempted with the wild-type

A

```
              S168N
163 TNALPSGSLKRKPIERHLSTLVSLLATADSSF---SIDWNQISTLDLEYRHSPYSCEAMWRVYLTPD---
                                                  Rh
                      R248G
Ra:  227 LRRDDWSPEEDETLLAVATANR---MQNWELIAASLDR--RSDYQCFVRFHTALRFLLEP

                          S306G
Rb:  282 KNSHRWSEEDNDKLRAIVDRNTANSVINWKKVVEYFPD--KSKSTLIGRYYYVLHPS---

                       358
Rc:  337 ISHEPFTTKEDMMLFAAVEEY----NGKFHCFPRSLFPN-RSLTQLRTRYHNVLAQR---

                         Y409N
Rd:  389 NKTDSWSVQDDTRLMSFVTQYG---ASQWLNCATFLGN--HTRTSCRTRFLVIKKFLE--

                                                    D484G
     442 QNPNAKVEDLPRRRSKKVSLVNSDNWAQRLQEWQEDPESLVNDNPPKGTRVRGPKSK...
```

B

```
                    168   247 305 358 409   483
                    H     A   B   C   D     E
          1         v     v   v   v   v     v            721
DmSNAP190: N  [          |Rh|Ra|Rb|Rc|Rd|              ] C
```

H DmSNAP190 (NG168):
A DmSNAP190 (NG247):
B DmSNAP190 (NG305):
C DmSNAP190 (WT-NG358):
D DmSNAP190 (NG409):
E DmSNAP190 (NG483):

C

Immunoblotting

α-190 C-terminus   α-FLAG (N-term.)

M H A B C D E   M H A B C D E

Protein-DNA Photo-cross-linking

U1 PSEA Position 24   U1 PSEA Position 12   U1 PSEA Position 1

H A B C D E   H A B C D E   H A B C D E

150 —
100 —
75 —
FL
50 —
37 —
25 —

M 1 2 3 4 5 6   M 7 8 9 10 11 12   13 14 15 16 17 18   19 20 21 22 23 24   25 26 27 28 29 30

D

```
           1      168  247 305 358 409  483        721
DmSNAP190: N [         |Rh|Ra|Rb|Rc|Rd|         ] C
```

Cross-linking to U1 PSEA Position 24:
Cross-linking to U1 PSEA Position 12:
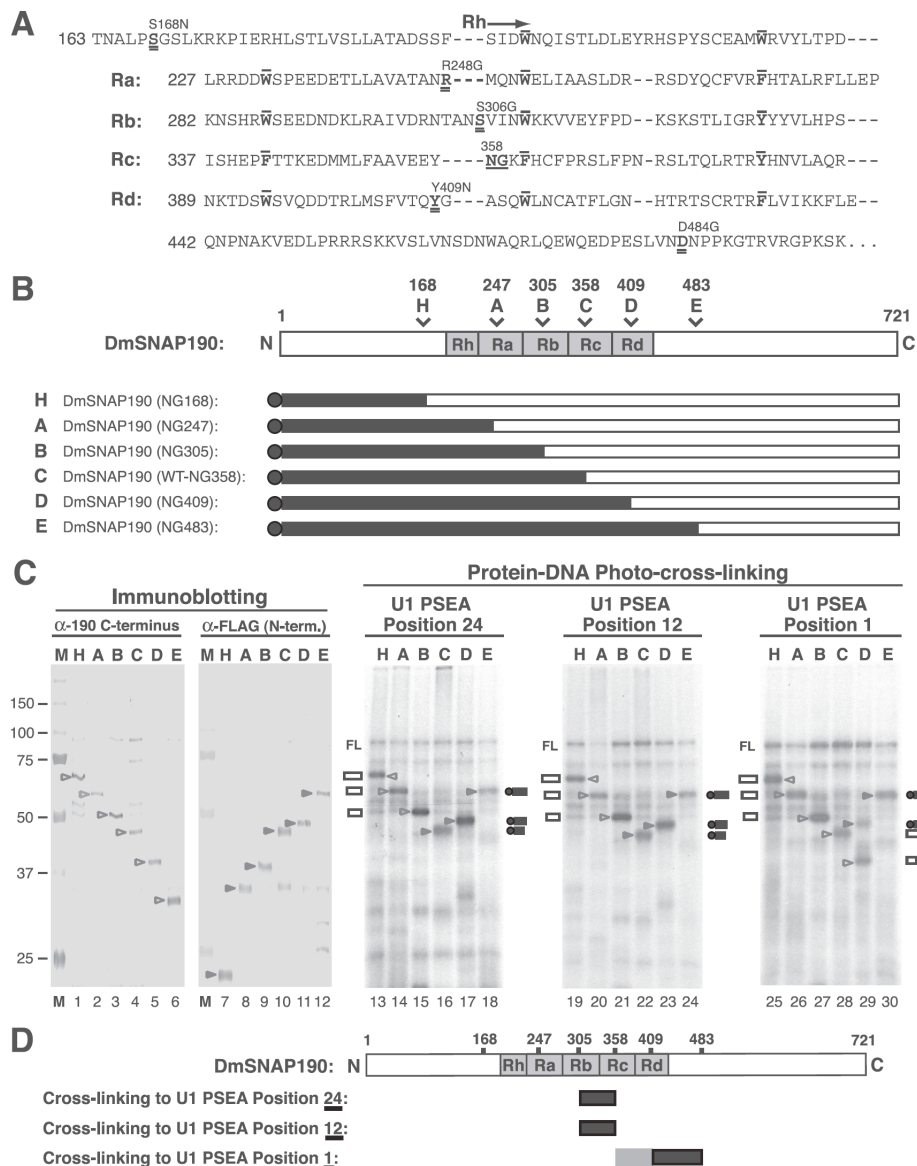Cross-linking to U1 PSEA Position 1:

FIG. 3. Localization of domains of DmSNAP190 that cross-link to U1 PSEA positions 24, 12, and 1. (A) Alignment of the Myb repeat sequences and flanking amino acids of DmSNAP190. The conserved aromatic amino acids characteristic of Myb repeats are shown in bold and are overlined. The NG hydroxylamine cleavage site at position 358 in the wild-type protein is bold and underlined. The positions where mutations were introduced to generate new NG cleavage sites are indicated above the sequence with the affected residues shown in bold and double underlined. (B) Schematic representation of the hydroxylamine cleavage patterns of six N-tagged DmSNAP190 constructs. The shaded and unshaded regions indicate the two cleavage fragments in each construct. (C) All gel results shown are subsequent to hydroxylamine digestion of DmSNAPc from cell lines separately expressing the six DmSNAP190 variants represented by constructs H to E shown in panel B. The first two panels show immunoblots using antibodies against a C-terminal peptide of DmSNAP190 (lanes 1 to 6) or antibodies against the FLAG epitope at the N terminus of each construct (lanes 7 to 12). Arrowheads (unshaded for C-terminal fragments and shaded for N-terminal fragments) point out the specific cleavage products of interest. The third, fourth, and fifth panels show results of protein-DNA photo-cross-linking of the six DmSNAP190 variants to U1 PSEA position 24 (lanes 13 to 18), position 12 (lanes 19 to 24), or position 1 (lanes 25 to 30). Symbols representing the fragments present in the first three lanes of each panel are shown to the left of each panel, and symbols representing the fragments in the last three lanes are shown to the right. In lane 29, the N- and C-terminal fragments both cross-linked to position 1. FL refers to the full-length undigested DmSNAP190. (D) Schematic drawing indicating the region of DmSNAP190 that cross-links to each of the positions 24, 12 and 1 of the U1 PSEA. The dark shading indicates the regions of strongest cross-linking, and the light shading indicates a region of weaker cross-linking. WT, wild type.

protein, but an excessive number of bands on the autoradiograms of the gels made it difficult to interpret the data.

Therefore, new constructs that lacked NG bonds at either position 179 or at positions 234 and 237 were made. We also made a third construct that lacked all three wild-type NG sites but contained a new NG bond at position 102. These constructs, with their expected hydroxylamine fragmentation patterns, are illustrated in Fig. 4A.

Figure 4B shows immunoblots of the various constructs after hydroxylamine digestion and detection with polyclonal antibodies produced against full-length recombinant DmSNAP50 (lanes 1 to 3, 5 to 7, and 9 to 11) or with monoclonal antibodies against the FLAG epitope (lanes 4, 8, and 12). In lanes 1 to 3, both the N- and C-terminal fragments were clearly detected by the anti-DmSNAP50 antibodies. In lanes 5 to 7, the C-terminal fragment gave a strong signal, but the signal from the N-terminal fragment, although still visible, was barely detectable. In lanes 9 to 11, only the large C-terminal fragments could be detected. In lanes 4, 8, and 12 of Fig. 3B, the anti-FLAG antibodies clearly revealed and confirmed the position of the FLAG-tagged N-terminal fragment for each of the three constructs.

Figure 4C shows the results of photo-cross-linking these constructs to either position 22 of the U1 PSEA (lanes 1 to 9) or position 14 of the U6 PSEA (lanes 10 to 18). When the construct that could be cleaved only at positions 234/237 was cross-linked to position 22 of the U1 PSEA, only N-terminal fragments cross-linked (Fig. 4C, lanes 1 to 3). (Compare with the N-terminal fragments in the immunoblots in Fig. 4B, lanes 1 to 3.) When cleavage of DmSNAP50 occurred at position 179, again only the N-terminal fragments cross-linked to the radiolabeled probe (Fig. 4C, lanes 4 to 6). (Compare to the weakly detected N-terminal bands in lanes 5 to 7 of Fig. 4B.)

When the DmSNAP50 cleavage site followed residue 102, the pattern changed (Fig. 4C, lanes 7 to 9). In this case, the cross-linked fragments from the untagged and the N-terminally tagged constructs migrated to identical positions, but the fragment from the C-terminally tagged construct migrated more slowly, matching the pattern of C-terminal fragments seen in Fig. 4B (lanes 9 to 11). This indicates that cross-linking occurred C-terminal to position 102.

By combining the data from lanes 1 to 9 of Fig. 4C, we conclude that position 22 of the U1 PSEA is contacted by amino acid residues located between positions 103 and 179 of DmSNAP50. Initially, this was an unexpected finding, because work on human SNAP50 had implicated the noncanonical zinc-binding domain (i.e., the SNAP finger) at the C terminus of the protein as being involved in DNA binding (17).

Besides the major bands that corresponded to the expected cleavage products and to the undigested full-length tagged and untagged DmSNAP50 polypeptides, a number of other bands were also visible in the autoradiograms. Many of these bands could be accounted for in a specific manner. The solid horizontal tick on the right side of each panel indicates the position of weak bands visible in lanes 2, 3, 5, 6, and 8 (but obscured in lane 9). These bands undoubtedly represent hydroxylamine digestion products of DmSNAP190, as this subunit could also cross-link to the U1 PSEA position 22 (Fig. 1). This band was not observed in lanes 1, 4, and 7, because DmSNAP190 in those lanes had an N-terminal tag to allow purification of

DmSNAPc that contained untagged DmSNAP50. If the N-terminal fragment of DmSNAP190 cross-linked to position 22, it should run with a slightly slower mobility and would thus be obscured by the full-length DmSNAP50 in lanes 1, 4, and 7. Thus, these data provide a "bonus" indication that the N-terminal half of DmSNAP190 cross-linked to position 22 of the U1 PSEA.

The asterisk to the right of lane 9 in Fig. 4C indicates a minor unexpected band. This band in lane 9 appears to run at a position that is just slightly above the C-terminal fragments that lacked tags (Fig. 4C, lanes 7 and 8). Although we do not understand the exact origin of this band, it appears to represent a C-terminal fragment from which most of the C-terminal tag has been removed. A weak band was also normally seen at this position in the immunoblots (migrating slightly slower than the untagged C-terminal fragments) as indicated by asterisks to the right of lanes 3, 7, and 11 in Fig. 4B. Perhaps the unnatural C-terminal tag is to some extent subject to *in vitro* proteolysis following FLAG purification or during hydroxylamine treatment. This would explain why this "starred" band very commonly appeared in the autoradiograms whenever cross-linking to a C-terminal fragment occurred (for further examples, see Fig. 4C, lanes 12, 15, and 18, as discussed below).

Next, we investigated the region of DmSNAP50 that contacted position 14 of the U6 PSEA. These data are shown in Fig. 4C (lanes 10 to 18). Lanes 10 to 12 demonstrate that the DmSNAP50 fragment C-terminal to residue 237 cross-linked to U6 PSEA position 14. Thus, amino acid residues located between positions 238 and 377, a region that includes the SNAP finger, are in close proximity to position 14 of the U6 PSEA. This cross-linking pattern and result were in sharp contrast to that obtained when the same construct was cross-linked to position 22 of the U1 PSEA (Fig. 4C, compare lanes 10 to 12 with lanes 1 to 3). Thus, position 22 of the U1 PSEA and position 14 of the U6 PSEA each cross-linked to a distinct domain of DmSNAP50.

When the other two DmSNAP50 constructs were used in the assay, it was again the C-terminal fragment that cross-linked to U6 PSEA position 14 (Fig. 4C, lanes 13 to 18). This was consistent with the conclusion drawn from the data in lanes 10 to 12 that DmSNAP50 residues C-terminal to residue 237 contacted position 14 of the U6 PSEA.

The results of cross-linking DmSNAP50 to position 22 of the U1 PSEA and to position 14 of the U6 PSEA are summarized diagrammatically in Fig. 4D. An important conclusion of these data is that these two nucleotide positions in the U1 and U6 PSEAs are contacted by two distinct regions of DmSNAP50.

There were again some unexpected bands in the gel. The bands marked with asterisks in lanes 12, 15, and 18 are similar to the band in lane 9, as previously discussed. Additional minor bands in lanes 13 to 18 are marked with dotted lines, and these bands migrate to the same positions as the major cleavage products in lanes 10 to 12. These unexpected bands therefore must arise from residual hydroxylamine cleavage at positions 234/237. Because the NG sites at these positions were changed to QG sites, the appearance of these bands suggests that hydroxylamine is capable of cleaving at the QG peptide bonds but with low efficiency. The appearance of this residual cleav-
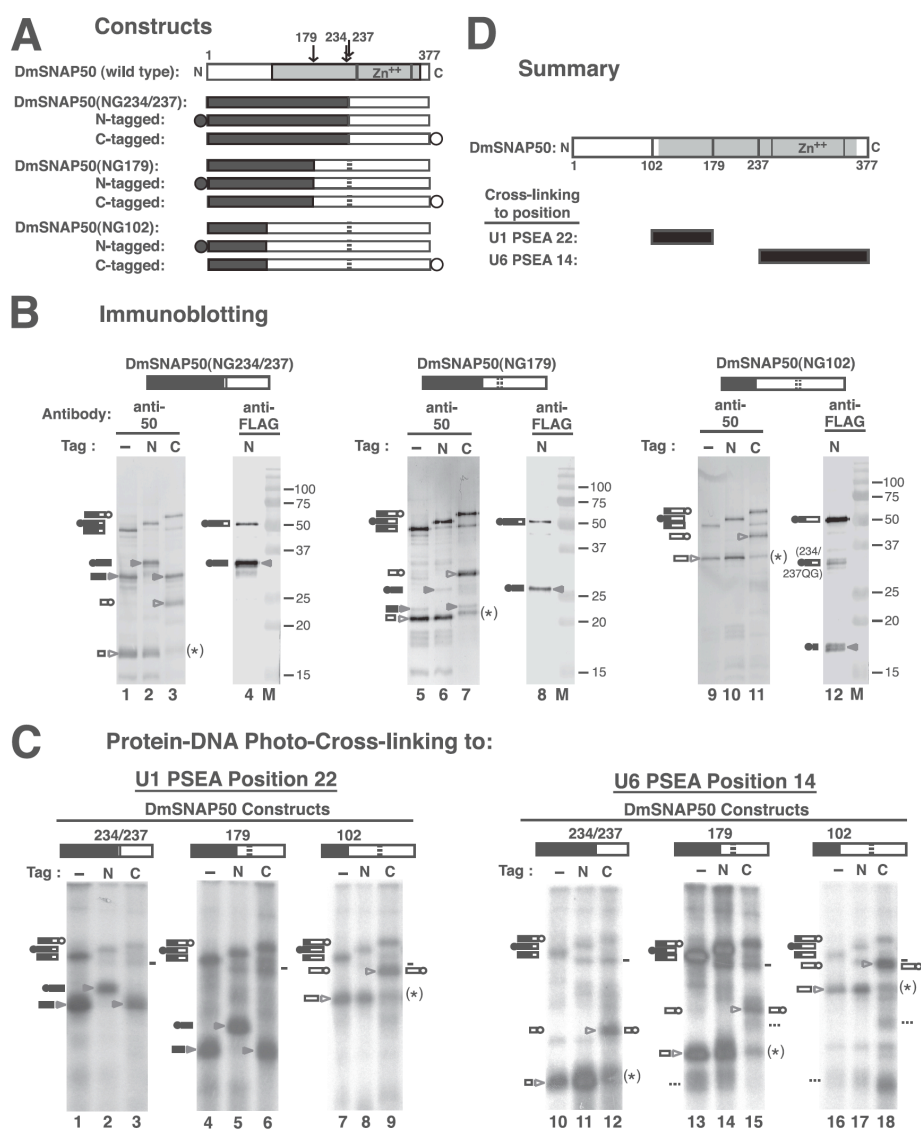
FIG. 4. Two distinct domains DmSNAP50 contact different nucleotides in the U1 and U6 PSEAs. (A) The top line shows the structure of wild-type DmSNAP50, including the evolutionarily conserved region (shaded), the hydroxylamine cutting sites (vertical arrows), and the zinc finger region. Below this are diagrammed nine DmSNAP50 constructs used for photo-cross-linking. (For one set of constructs, a new cutting site was introduced at position 102.) The vertical dashes at positions 234/237 indicate a point of weak residual cutting in the lower six constructs that arises from the two nearly adjacent sites that were mutated from NG to QG. (B) Immunoblots showing the locations of N-terminal and C-terminal fragments arising from hydroxylamine digestion of the constructs shown in panel A. Symbols are as described in the legends of Fig. 2 and 3. The asterisks in parentheses indicate the position of a band (in lanes 3, 7, and 11) that appears to correspond to a C-terminal fragment that has lost its C-terminal tag. See the text for further discussion. (C) Autoradiograms of chemical digestion products of the nine constructs cross-linked to U1 PSEA position 22 (lanes 1 to 9) or to U6 PSEA position 14 (lanes 10 to 18). The symbols alongside the gels designate the fragments responsible for the bands observed, and the arrowheads further clarify the relevant bands. The horizontal tick alongside the right of each panel indicates a digestion product of DmSNAP190. (D) The dark bars indicate the two different regions of DmSNAP50 that cross-linked to position 22 of the U1 PSEA or to position 14 of the U6 PSEA.
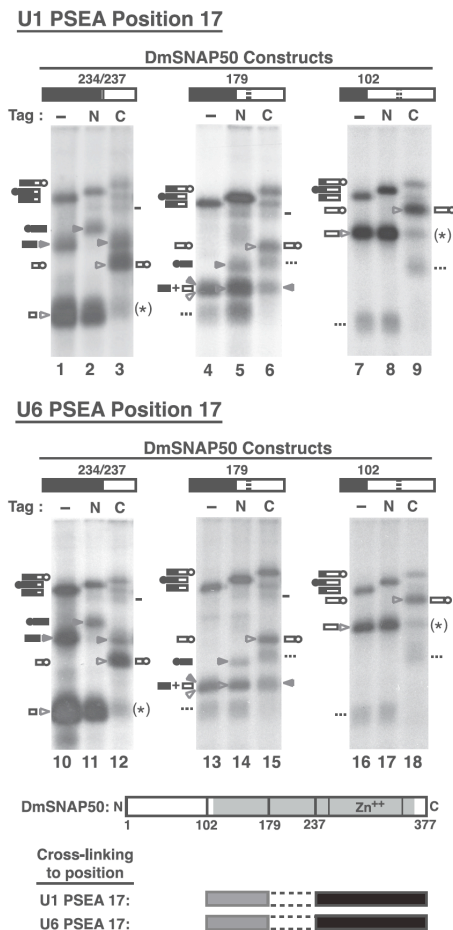
U1 PSEA Position 17

U6 PSEA Position 17

FIG. 5. Two domains of DmSNAP50 contact position 17 of both the U1 and U6 PSEAs. DmSNAPc separately containing nine different constructs of DmSNAP50 (shown in Fig. 4) was cross-linked to position 17 of the U1 PSEA or to position 17 of the U6 PSEA, digested with hydroxylamine, run on denaturing gels, and subjected to autoradiography. Symbols are the same as those described in previous figure legends. The bottom of the figure summarizes the regions of DmSNAP50 that cross-linked to position 17 of each PSEA. Dark shading indicates a region that cross-linked strongly, and lighter shading indicates a region of somewhat weaker cross-linking. It is undetermined whether residues 180 to 237 cross-link to position 17.

age is enhanced in this case because there are two QG sites nearly adjacent to each other.

Next, we investigated the cleavage patterns of the DmSNAP50 NG variants after cross-linking to position 17 of the U1 PSEA or to position 17 of the U6 PSEA (Fig. 5). When DmSNAP50 was cleaved at position 234/237, the N-terminal fragment and the C-terminal fragment both cross-linked to U1 position 17 (Fig. 5, lanes 1 to 3; compare to the immunoblot pattern in Fig. 4B, lanes 1 to 3). In fact, the autoradiography pattern obtained was a sum of the bands previously observed

when cross-linking was carried out to U1 position 22 and to U6 position 14 (Fig. 4C, lanes 1 to 3 and 10 to 12).

When the DmSNAP50 construct that had the hydroxylamine cleavage site at position 179 was used for cross-linking to U1 position 17 (Fig. 5, lanes 4 to 6), the pattern was initially difficult to interpret. However, upon further study, it became evident that this pattern also represented the sum of the bands observed when this construct was cross-linked to U1 position 22 and U6 position 14 (compare with Fig. 4C, lanes 4 to 6 and 13 to 15; also see the immunoblot in Fig. 4B, lanes 5 to 7). In Fig. 5, lane 4, the untagged N-terminal fragment and the untagged C-terminal fragment did not resolve on the autoradiogram. Besides these unresolved untagged N- and C-terminal fragments, a fragment with slower mobility due to its N-terminal tag was seen in Fig. 5, lane 5, and a fragment with still slower mobility due to its C-terminal tag was seen in lane 6. These correspond to the N-tagged and C-tagged fragments identified in the immunoblots of Fig. 4B, lanes 6 to 8.

When DmSNAP50 was fragmented at position 102, only the C-terminal fragment cross-linked (Fig. 5, lanes 7 to 9). This finding was consistent with the results shown in Fig. 4C, lanes 7 to 9 and 16 to 18. Additional bands marked in Fig. 5 with asterisks or dotted lines arose from the previously discussed artifactual phenomena.

Lanes 10 to 18 of Fig. 5 show the cross-linking patterns obtained when the DmSNAPc NG variants were cross-linked to position 17 of the U6 PSEA. In each case, the gel patterns were identical to those observed for position 17 of the U1 PSEA (compare each panel to the one directly above it). Thus, at the resolution obtained by these experiments, phosphate 17 is in a similar environment relative to the DmSNAP50 polypeptide when DmSNAPc binds either to a U1 PSEA or to a U6 PSEA. We can conclude from the data in Fig. 5 that position 17 of both PSEAs cross-linked to DmSNAP50 at residues located between 238 and 377 and with less intensity to residues between 103 and 179. From these data, however, we were unable to exclude the possibility that position 17 might additionally cross-link to residues located between 180 and 237. A schematic summary of these findings is shown at the bottom of Fig. 5.

Mapping regions of DmSNAP43 that contact nucleotide positions 40, 28, and 20 of the U1 PSEA and positions 16 and 11 of the U6 PSEA. DmSNAP43 has no known canonical DNA-binding domains; furthermore, it contacts the U1 and U6 PSEAs very differently. When DmSNAPc binds to a U6 PSEA, DmSNAP43 cross-links most strongly to positions 11 and 16 of the U6 PSEA (but not when it binds to a U1 PSEA) (Fig. 1A). On the other hand, when DmSNAPc binds to a U1 PSEA, DmSNAP43 cross-links much more strongly to position 20 and to positions 28 and 40 downstream of the U1 PSEA (Fig. 1A and reference 22). We therefore mapped domains of DmSNAP43 that cross-link to each of these five positions within the U1 or U6 PSEAs.

In initial experiments, the single NG peptide bond present at position 192 in wild-type DmSNAP43 was used for protein mapping. Figure 6, lanes 1 to 3, shows an immunoblot of hydroxylamine-cleaved untagged, N-tagged, and C-tagged DmSNAP43 detected using antibodies produced in rabbits against full-length recombinant DmSNAP43. These polyclonal antibodies strongly reacted with the C-terminal fragments but
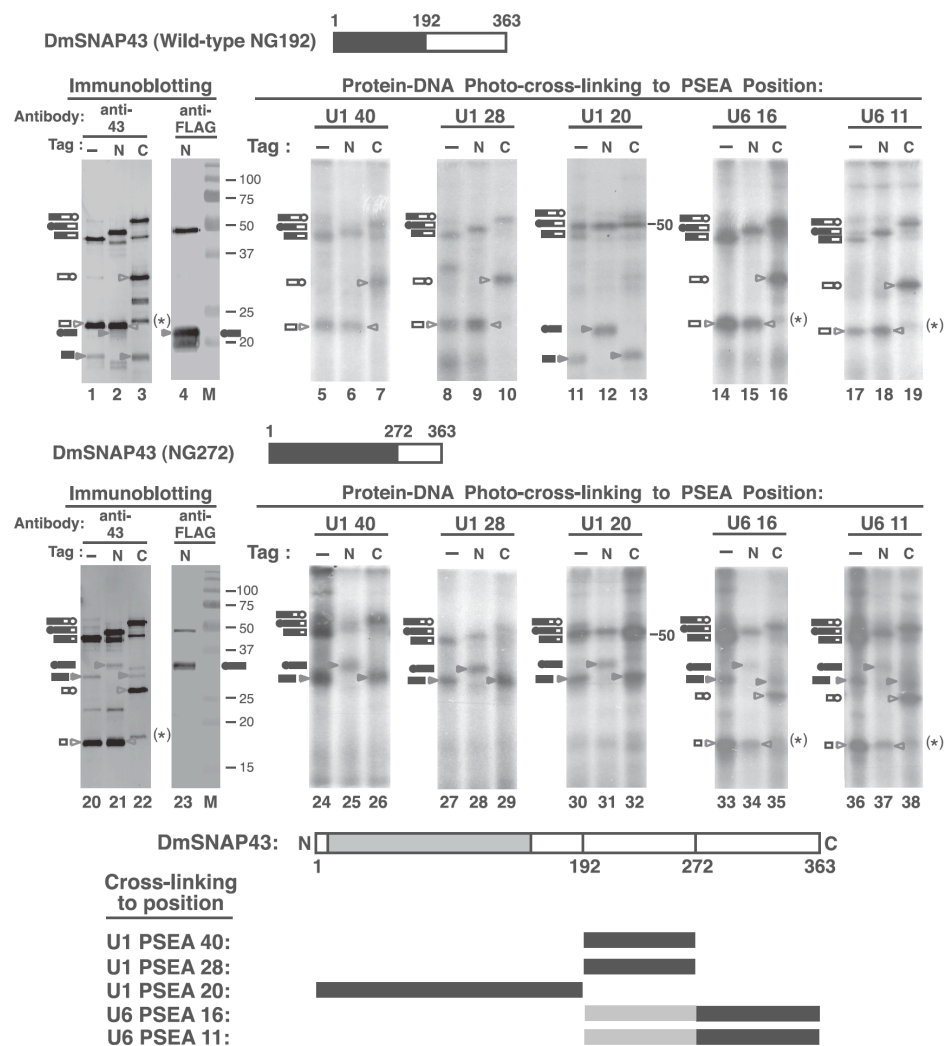
FIG. 6. Three distinct regions of DmSNAP43 differentially contact the U1 and U6 PSEAs. DmSNAPc was cross-linked to three phosphate positions in or downstream of the U1 PSEA and to two positions in the U6 PSEA. The DmSNAPc contained DmSNAP43 with a hydroxylamine cutting site either at the wild-type position (upper panels) or at position 272 (lower panels). Immunoblots shown in lanes 1 to 4 and 20 to 23 reveal the positions of the C-terminal and N-terminal fragments. The symbols that mark the band positions and identify the fragments in each panel are similar to those used in Fig. 2 to 5. The bottom of the figure summarizes the regions of DmSNAP43 that cross-link to the indicated positions in the U1 and U6 PSEAs. Dark shading indicates regions that cross-link strongly to the indicated nucleotide positions, and lighter shading indicates a region of weaker cross-linking.

detected the N-terminal fragments poorly. Nonetheless, the bands corresponding to the N-terminal fragments were visible when the blots were overdeveloped relative to the C-terminal fragment bands. Furthermore, an immunoblot with FLAG antibody (lane 4) confirmed the position of the tagged N-terminal fragment.

Figure 6, lanes 5 to 19, shows the results of photo-cross-linking to various nucleotide positions within the U1 and U6 PSEAs. When the cross-linking agent was at U1 PSEA position

40, only the C-terminal fragment cross-linked (lanes 5 to 7). The same result was obtained with the cross-linker positioned at nucleotide 28 of the U1 PSEA (lanes 8 to 10). However, when the cross-linker was at position 20, the pattern changed and revealed that the N-terminal fragment cross-linked to U1 PSEA position 20 (lanes 11 to 13).

Somewhat surprisingly, when cross-linking was carried out to positions 16 and 11 of the U6 PSEA, it was again the C-terminal half of DmSNAP43 that cross-linked to these posi-

tions (Fig. 6, lanes 14 to 19). Thus, the N-terminal half of DmSNAP43 cross-linked to position 20 of the U1 PSEA, but the C-terminal half cross-linked to positions 28 and 40 of the U1 PSEA and to positions 11 and 16 of the U6 PSEA.

To better localize the contact points within the C-terminal domain of DmSNAP43 for the U1 and U6 PSEAs, we next prepared a construct that contained an NG peptide bond only at position 272. The immunoblots in Fig. 6, lanes 20 to 23, show the positions of the N- and C-terminal fragments following hydroxylamine cleavage (the signals from the N-terminal fragments are weak in lanes 20 to 22). When this mutant construct was cross-linked to U1 PSEA positions 40 or 28 (lanes 24 to 29), only the N-terminal fragment (residues 1 to 272) cross-linked. In combination with the results shown in lanes 5 to 10, we can conclude that DmSNAP43 amino acid residues between 193 and 272 are in close proximity to nucleotide positions 40 and 28 downstream of the U1 PSEA. Figure 6 (lanes 30 to 32) shows that U1 position 20 cross-linked to the N-terminal fragment of the mutant construct; this was consistent with U1 position 20 cross-linking to the N-terminal half of DmSNAP43, as previously seen in lanes 11 to 13.

Importantly, and in contrast, positions 16 and 11 of the U6 PSEA cross-linked primarily to the most C-terminal fragment (residues 273 to 363) (Fig. 6, lanes 33 to 38). Less-intense bands of cross-linking to the N-terminal fragment were also visible in each of these lanes. From these results, we conclude that U6 PSEA positions 11 and 16 cross-linked most strongly to the last 91 amino acids of DmSNAP43, but residues between 193 and 272 could also be cross-linked with a lower efficiency.

A summary of the conclusions from the DmSNAP43 mapping experiments is shown at the bottom of Fig. 6. The N-terminal half of DmSNAP43 cross-linked to position 20 of the U1 PSEA, but positions 28 and 40 cross-linked to residues located between 193 and 272. Finally, positions 16 and 11 of the U6 PSEA cross-linked most strongly to residues 273 to 363 of DmSNAP43 but also cross-linked weakly to residues 193 to 272.

DISCUSSION

Previous studies had identified nucleotide positions contacted by each of the three subunits of DmSNAPc (22, 32). However, the domains of DmSNAPc that interact directly with the DNA had not been determined. In this study, we have been able to localize regions within the DmSNAPc subunits that are involved not only in contacting DNA but also in contacting specific nucleotide positions within the U1 and U6 PSEAs. From these data, we have mapped different domains of DmSNAP50 and of DmSNAP43 that contact different regions of the U1 and U6 PSEAs. In the absence of an atomic structure for SNAPc, these cross-linking results provide a basic understanding of significant structural differences of DmSNAPc on the U1 and U6 PSEAs. The results are summarized schematically in Fig. 7.

Domains of DmSNAP190 that interact with the U1 PSEA. Work with human SNAP190 revealed that the RcRd repeats alone were capable of binding to DNA but with relatively low sequence specificity (15, 25, 28, 33). In those studies, no DNA-binding activity was noted for the RhRaRb region (28, 33). Our results, within the resolution of the experiments shown in
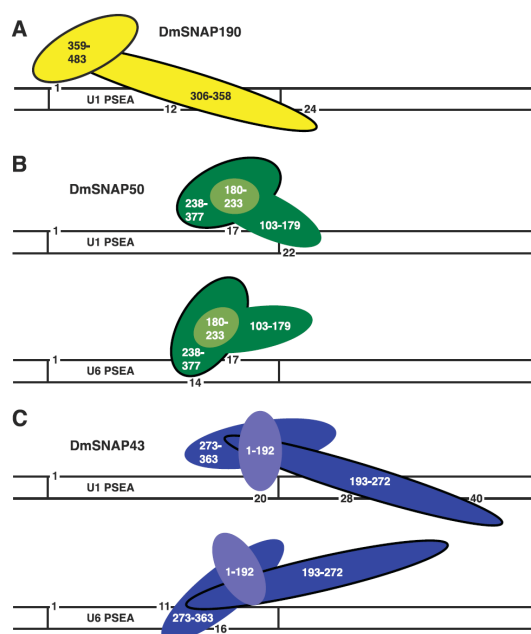

FIG. 7. Domains of DmSNAP190, DmSNAP50, and DmSNAP43 that interact with specific nucleotide positions within and downstream of the U1 and U6 PSEAs. Ovals represent amino acid domains within each of the subunits as labeled. (A) Domains of DmSNAP190 interacting with nucleotide positions 1, 12, and 24 of the U1 PSEA. (B) Domains of DmSNAP50 that interact with positions 17 and 22 of the U1 PSEA or with positions 14 and 17 of the U6 PSEA. (C) Domains of DmSNAP43 that interact with positions 20, 28, and 40 of the U1 PSEA or with positions 11 and 16 of the U6 PSEA.

Fig. 3, are consistent with those findings. We found that a region of the protein encompassed by residues 306 to 358 (located from the middle of the Rb repeat to the middle of the Rc repeat) interacted with positions 12 and 24 of the U1 PSEA (Fig. 3D and 7A). In contrast, a region encompassed by residues 359 to 483 (including the C-terminal half of the Rc repeat, the Rd repeat, and 42 amino acids C-terminal of the Rd repeat) was in proximity to position 1 of the U1 PSEA (Fig. 3D and 7A). This indicates that the C-terminal region of the Myb domain interacts with the 5′ end of the U1 PSEA and that the central region of the Myb domain interacts with the 3′ portion of the U1 PSEA. In a simplistic interpretation, this orientation of binding would place the N terminus of DmSNAP190 nearer the transcription start site and the C terminus farther from the start site. Whether this holds true as the protein folds in three dimensions is not yet clear. However, this orientation is consistent with the fact that the N-terminal domain of human SNAP190 interacts with the TATA-binding protein (TBP) (25), and TBP would most likely localize downstream of the PSEA toward the transcription start site. This orientation also makes sense from the perspective that the C-terminal domain of human SNAP190 interacts with Oct-1 protein bound to an upstream enhancer element (9).

Recent work from our laboratory indicated that a region of

DmSNAP190 C-terminal to the Myb domain was required for DmSNAPc to bind to DNA (16). However, in the studies reported here, we did not observe cross-linking of the C-terminal domain (residues 484 to 721) to the DNA. It is thus possible that the C-terminal domain of DmSNAP190 may contact nucleotides other than the three examined in the current study; alternatively, this region of DmSNAP190 may be required for DmSNAPc to adopt a proper DNA-binding conformation. Additional work will be required to answer this question.

Domains of DmSNAP50 interacting with the U1 and U6 PSEAs. DmSNAP50 and DmSNAP43 are the two subunits of DmSNAPc whose cross-linking patterns exhibit the most diversity when DmSNAPc binds to a U1 versus a U6 PSEA (22, 32). Moreover, the cross-linking of these two subunits is restricted to the 3′ half (and to sequences downstream) of the PSEAs. Therefore, we concentrated most of our effort on identifying the domains of these subunits that are closely associated with the DNA when DmSNAPc binds to either the U1 or U6 PSEA.

In the case of DmSNAP50 on the U6 PSEA, the SNAP finger at the C terminus of DmSNAP50 contacts phosphate position 14 (Fig. 4 and 7B), consistent with the finding that the SNAP finger is required for DNA binding by human SNAPc (17). However, when DmSNAPc binds to a U1 PSEA, a distinct region closer to the N terminus of DmSNAP50 (i.e., residues 103 to 179) associates closely with position 22 of the U1 PSEA (Fig. 4 and 7B).

Interestingly, both DmSNAP50 DNA-binding domains contact position 17 of the U1 and U6 PSEAs (Fig. 5 and 7B). This raises the possibility that position 17 may act as a fulcrum about which DmSNAP50 rotates so that residues 103 to 179 are in close proximity to position 22 of the U1 PSEA, but residues 237 to 377 are in close proximity to position 14 of the U6 PSEA (Fig. 7B).

Recent work from our laboratory found that deletion of the nonconserved N-terminal domain of DmSNAP50 (residues 1 to 91) eliminated the DNA-binding activity of DmSNAPc while having no effect on the ability of the three subunits to associate with each other. This N-terminal region of the protein did not cross-link to any of the four phosphate positions we examined (U1-22, U1-17, U6-14, and U6-17). As discussed above regarding DmSNAP190, it may be that this N-terminal domain of DmSNAP50 contacts nucleotides other than the four studied herein; alternatively, this region of DmSNAP50 may be required for DmSNAPc to adopt a proper DNA-binding conformation.

Domains of DmSNAP43 interacting with the U1 and U6 PSEAs. DmSNAP43 cross-links with significantly different patterns on U1 and U6 PSEAs (Fig. 1A and references 22 and 32) but contains no obvious sequence similarity to any characterized DNA-binding domains. Our earlier site-specific protein-DNA photo-cross-linking studies indicated that DmSNAP43 cross-linked strongly to the phosphate at position 20 in the U1 PSEA but only very weakly to that position when DmSNAPc was bound to the U6 PSEA (Fig. 1A). The new data indicate that this contact to the U1 PSEA at position 20 occurs through the N-terminal domain encompassed by residues 1 to 192 of DmSNAP43 (Fig. 6 and 7C).

On the other hand, phosphate positions 28 and 40, located

downstream of the U1 PSEA, cross-linked to residues located within an 80-residue domain stretching from residue 193 to 272 of DmSNAP43 (Fig. 6 and 7C). Interestingly, on a U6 PSEA, this same region of DmSNAP43 cross-linked weakly to phosphates 11 and 16 of the U6 PSEA (Fig. 6 and 7C). Thus, this domain of the protein (residues 193 to 272) may be in radically different positions when DmSNAPc binds to a U6 versus a U1 PSEA.

The strongest cross-linking contacts with phosphates 11 and 16 of the U6 PSEA, however, were made by the most C-terminal domain of DmSNAP43 (residues 273 to 363) (Fig. 6 and 7C). These contacts are nonexistent or extremely weak when DmSNAPc binds to a U1 PSEA (Fig. 1A). Thus, when the C-terminal domain of DmSNAP43 (residues 273 to 363) closely approaches positions 11 and 16 of the U6 PSEA, it probably also brings residues 193 to 272 into closer proximity to these nucleotides as well (Fig. 7C).

In work recently published by our laboratory (16), a C-terminal truncation of DmSNAP43 that deleted amino acids beyond residue 172 eliminated the DNA-binding activity of DmSNAPc even though it had no effect on the ability of DmSNAP43 to form a complex with DmSNAP50 and DmSNAP190. It is now clear from the photo-cross-linking data that the C-terminal domain of Dm-SNAP43, although not conserved in sequence during evolution, can make extensive contacts with the DNA that can extend over a 30-bp region. Furthermore, the pattern is different depending upon whether DmSNAPc is bound to a U1 or U6 PSEA.

The mutant constructs used in the structural investigations reported here were not tested for activity in a transcription assay. Despite this caveat, each of the mutant constructs continued to bind to DNA and to cross-link strongly to the same positions as the wild-type protein. Thus, it seems unlikely that the point mutations resulted in significant structural alterations to the DmSNAP complex.

In conclusion, we have localized domains within the DmSNAPc subunits that contact specific nucleotides in the U1 and U6 snRNA gene promoters. These data help to establish, to a first approximation, the orientation of the subunits on the DNA. Furthermore, the results reveal that different domains of the DmSNAP50 and DmSNAP43 subunits are in close proximity to different nucleotide positions when DmSNAPc binds to a U1 versus a U6 PSEA. These results support a model in which different conformations of DmSNAPc, induced by relatively minor DNA sequence differences, are important for determining the RNA polymerase specificity of snRNA gene promoters. In a similar vein, recent work established that the structure and the activity of the glucocorticoid receptor depend upon the specific DNA sequence that it recognizes and binds to as a homodimer (27).

REFERENCES

1. **Bai, L., Z. X. Wang, J. B. Yoon, and R. G. Roeder.** 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.

2. **Barakat, N. H., and W. E. Stumph.** 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 Lett. **582:**2413–2416.

3. **Crimmins, D. L., and S. M. Mische.** 1996. Chemical cleavage of proteins in solution, p. 11.4.1-11.4.8. *In* V. B. Chanda, vol. 2, suppl. 4. (ed.), Current protocols in protein science. John Wiley & Sons, Inc., New York, NY.

4. **Dahlberg, J. E., and E. Lund.** 1991. How does III x II make U6? Science **254:**1462–1463.

5. **Dahlberg, J. E., and E. Lund.** 1988. The genes and transcription of the major small nuclear RNAs, p. 38–70. *In* M. L. Birnstiel (ed.), Structure and function of major and minor small nuclear ribonucleoprotein particles. Springer Verlag KG, Heidelberg, Federal Republic of Germany.

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

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

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

9. **Ford, E., M. Strubin, and N. Hernandez.** 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.

10. **Henry, R. W., E. Ford, R. Mital, V. Mittal, and N. Hernandez.** 1998. Crossing the line between RNA polymerases: transcription of human snRNA genes by RNA polymerases II and III. Cold Spring Harbor Symp. Quant. Biol. **63:** 111–120.

11. **Henry, R. W., B. C. Ma, C. L. Sadowski, R. Kobayashi, and N. Hernandez.** 1996. Cloning and characterization of SNAP50, a subunit of the snRNA-activating protein complex SNAPc. EMBO J. **15:**7129–7136.

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

13. **Hernandez, G., F. Valafar, and W. E. Stumph.** 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.

14. **Hernandez, N.** 2001. snRNA genes: a model system to study fundamental mechanisms of transcription. J. Biol. Chem. **276:**26733–26736.

15. **Hinkley, C. S., H. A. Hirsch, L. P. Gu, B. LaMere, and R. W. Henry.** 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.

16. **Hung, K.-H., M. Titus, S. C. Chiang, and W. E. Stumph.** 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.

17. **Jawdekar, G. W., A. Hanzlowsky, S. L. Hovde, B. Jelencic, M. Feig, J. H. Geiger, and R. W. Henry.** 2006. The unorthodox SNAP50 zinc finger domain contributes to cooperative promoter recognition by human SNAPc. J. Biol. Chem. **281:**31050–31060.

18. **Jawdekar, G. W., and R. W. Henry.** 2008. Transcriptional regulation of human small nuclear RNA genes. Biochim. Biophys. Acta **1779:**295–305.

19. **Jensen, R. C., Y. Wang, S. B. Hardin, and W. E. Stumph.** 1998. The proximal sequence element (PSE) plays a major role in establishing the RNA poly-merase specificity of *Drosophila* U-snRNA genes. Nucleic Acids Res. **26:** 616–622.

20. **Lai, H.-T., H. Chen, C. Li, K. J. McNamara-Schroeder, and W. E. Stumph.** 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.

21. **Lai, H.-T., Y. S. Kang, and W. E. Stumph.** 2008. Subunit stoichiometry of the *Drosophila melanogaster* small nuclear RNA activating protein complex (SNAPc). FEBS Lett. **582:**3734–3738.

22. **Li, C., G. A. Harding, J. Parise, K. J. McNamara-Schroeder, and W. E. Stumph.** 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.

23. **Lo, P. C. H., and S. M. Mount.** 1990. *Drosophila melanogaster* genes for U1 snRNA variants and their expression during development. Nucleic Acids Res. **18:**6971–6979.

24. **Lobo, S. M., and N. T. Hernandez.** 1994. Transcription of snRNA genes by RNA polymerases II and III, p. 127–159. *In* R. C. Conaway and J. W. Conaway (ed.), Transcription: mechanisms and regulation. Raven Press, New York, NY.

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

26. **McNamara-Schroeder, K. J., R. F. Hennessey, G. A. Harding, R. C. Jensen, and W. E. Stumph.** 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.

27. **Meijsing, S. H., M. A. Pufall, A. Y. So, D. L. Bates, L. Chen, and K. R. Yamamoto.** 2009. DNA binding site sequence directs glucocorticoid receptor structure and activity. Science **324:**407–410.

28. **Mittal, V., B. C. Ma, and N. Hernandez.** 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.

29. **Parry, H. D., D. Scherly, and I. W. Mattaj.** 1989. 'Snurpogenesis': the transcription and assembly of U snRNP components. Trends Biochem. Sci. **14:**15–19.

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

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

32. **Wang, Y., and W. E. Stumph.** 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.

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

34. **Yoon, J. B., S. Murphy, L. Bai, Z. Wang, and R. G. Roeder.** 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.

35. **Yoon, J. B., and R. G. Roeder.** 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.

AUTHOR'S CORRECTION

Identification of SNAPc Subunit Domains That Interact with Specific Nucleotide Positions in the U1 and U6 Gene Promoters

Mun Kyoung Kim, Yoon Soon Kang, Hsien-Tsung Lai, Nermeen H. Barakat, Deodato Magante, and William E. Stumph

Molecular Biology Institute, Department of Biology, and Department of Chemistry and Biochemistry, San Diego State University, San Diego, California 92182-1030

Volume 30, no. 10, p. 2411–2423, 2010. Since publication of this paper, we have become aware of a mistake in the identity of the photo-cross-linking probe used to map the domain of DmSNAP190 that cross-linked to phosphate position 24 of the U1 proximal sequence element A (PSEA). The probe used to generate the data presented in Fig. 2 (lanes 1 to 3) and 3C (lanes 13 to 18) was in fact not a position 24 probe; instead, a U1 PSEA phosphate position 12 probe was used in error. Reperformance of these experiments with the correct probe revealed that phosphate position 24 of the U1 PSEA cross-linked to a region of DmSNAP190 located between amino acid residues 169 and 247 rather than between residues 306 and 358. None of the other conclusions of the publication were affected. We sincerely regret this error and any inconvenience that may have resulted to our colleagues.

Page 2414, Fig. 2, lanes 1–3: Although this panel was generated with the misidentified probe, the results of the experiment were the same when performed with the correct probe for U1 PSEA position 24 (data not shown). The conclusion that phosphate position 24 cross-linked to the N-terminal half of DmSNAP190 was not affected.

Page 2416: Figure 3C, lanes 13–18, should appear as shown below. These results indicate that position 24 of the U1 PSEA cross-linked to DmSNAP190 C terminal of amino acid residue 169 but N terminal of residue 247 inclusive.

Page 2416, Fig. 3D: The bar for cross-linking to U1 PSEA position 24 should be between residues 169 and 247.

Page 2421, column 2: Figure 7A should appear as shown below because DmSNAP190 residues between 169 and 247 contact position 24 of the U1 PSEA. (Although the domain of DmSNAP190 encompassed by residues 248 to 305 does not contact the U1 PSEA at nucleotide positions 1, 12, or 24, it may contact the U1 PSEA at other positions.)
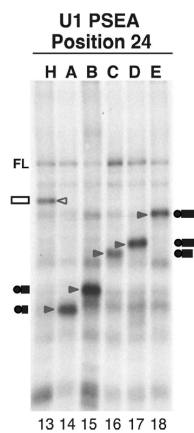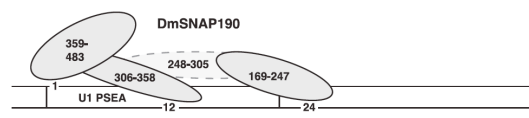

FIG. 3C.

FIG. 7A.

ACKNOWLEDGEMENTS:

This chapter, in full, is a reprint of the material as it appears in Molecular and Cellular Biology, 2010. Identification of SNAPc subunit domains that interact with specific nucleotide positions in the U1 and U6 gene promoters. Kim M.K.; Kang Y.S.;

Lai H.T.; Barakat N.H.; Magante D.; and Stumph W.E. The dissertation author was a primary investigator and a co-first author of this paper.

CHAPTER 2

Architectural Arrangement of the Small Nuclear RNA (snRNA)-activating Protein Complex 190 Subunit (SNAP190) on U1 snRNA Gene Promoter DNA

Architectural Arrangement of the Small Nuclear RNA (snRNA)-activating Protein Complex 190 Subunit (SNAP190) on U1 snRNA Gene Promoter DNA*[s]

Matthew T. Doherty[‡1]**, Yoon Soon Kang**[§1]**, Cheryn Lee**[§]**, and William E. Stumph**[§2]

From the Molecular Biology Institute and the Departments of [‡]*Biology and* [§]*Chemistry and Biochemistry, San Diego State University, San Diego, California 92182-1030*

Background: SNAP190, the largest subunit of the snRNA-activating protein complex (SNAPc), interacts with DNA via 4.5 Myb repeats.
Results: Each Myb repeat was mapped on U1 gene promoter DNA by site-specific protein-DNA photo-cross-linking.
Conclusion: The N-terminal repeats contacted DNA nearer the transcription start site, whereas the C-terminal repeats interacted farther upstream.
Significance: Structural insights were obtained into SNAPc bound to snRNA gene promoter sequences.

Myb repeats ~52 amino acid residues in length were first characterized in the oncogenic Myb transcription factor, which contains three tandem Myb repeats in its DNA-binding domain. Proteins of this family normally contain either one, two, or three tandem Myb repeats that are involved in protein-DNA interactions. The small nuclear RNA (snRNA)-activating protein complex (SNAPc) is a heterotrimeric transcription factor that is required for expression of small nuclear RNA genes. This complex binds to an essential promoter element, the proximal sequence element, centered ~50 base pairs upstream of the transcription start site of snRNA genes. SNAP190, the largest subunit of SNAPc, uncharacteristically contains 4.5 tandem Myb repeats. Little is known about the arrangement of the Myb repeats in the SNAPc-DNA complex, and it has not been clear whether all 4.5 Myb repeats contact the DNA. By using a site-specific protein-DNA photo-cross-linking assay, we have now mapped specific nucleotides where each of the Myb repeats of *Drosophila melanogaster* SNAP190 interacts with a U1 snRNA gene proximal sequence element. The results reveal the topological arrangement of the 4.5 SNAP190 Myb repeats relative to the DNA and to each other when SNAP190 is bound to a U1 promoter as a subunit of SNAPc.

The small nuclear RNA-activating protein complex (SNAPc),[3] also known as PSE-binding transcription factor (PTF), is required to activate transcription of genes encoding the snRNAs and certain other small stable RNA molecules (1–13). *Drosophila melanogaster* SNAPc (DmSNAPc) is a heterotrimer that consists of subunits known as DmSNAP190, DmSNAP50, and DmSNAP43 (14, 15). The three subunits of DmSNAPc form a stable complex in the absence of DNA and are all required together for sequence-specific interaction with a DNA promoter element called the proximal sequence element A (PSEA). The PSEA is conserved at a location about 40–60 base pairs upstream of the transcription start site of fly snRNA genes (7, 16).

The atomic structure of SNAPc is unknown. However, site-specific protein-DNA photo-cross-linking experiments have provided significant information about the arrangement of the three fruit fly subunits relative to the PSEA sequence and to each other on the DNA (14, 17). As summarized in Fig. 1, DmSNAP190 cross-linked to 13 specific phosphate positions extending between U1 PSEA nucleotide positions 1 and 25 (indicated by *yellow spheres* on the DNA double helix). Those studies furthermore indicated that DmSNAP50 cross-linked to phosphate positions located between 13 and 22, and DmSNAP43 cross-linked to phosphates between 18 and 40 when DmSNAPc was bound to a U1 snRNA gene PSEA (positions indicated by the *green* and *blue spheres*, respectively, in Fig. 1) (14, 17).

By combining the site-specific protein-DNA photo-cross-linking technique with subsequent site-specific chemical cleavage of the polypeptides, it was further possible to map yet smaller domains within DmSNAP50 and DmSNAP43 that interacted with specific nucleotide positions in the PSEA (18). A preliminary mapping was also accomplished of DmSNAP190 domains that interacted with the three positions 1, 12, and 24 of the U1 PSEA (18). In the current work, we have significantly extended those studies to localize the regions of DmSNAP190 that interact with each of the 10 remaining cross-linkable nucleotide positions of the U1 PSEA.

A distinctive feature of metazoan SNAP190 is that it contains 4.5 Myb repeats (14, 20). Myb repeats were first characterized as three imperfect repeats, termed R1, R2, and R3 (each ~52

[3] The abbreviations used are: SNAPc, small nuclear RNA-activating protein complex; DmSNAPc, *D. melanogaster* SNAPc; PTF, PSE-binding transcription factor; PSE, proximal sequence element; PSEA, proximal sequence element A; PSEB, proximal sequence element B.

Arrangement of SNAP190 Myb Repeats on U1 Gene Promoter DNA

amino acids residues in length), that constitute the DNA-binding domain of the c-Myb transcription factor (21, 22). Such Myb repeats have been found in a large family of fungal, plant, and animal DNA-binding proteins. Although c-Myb and many other Myb-related proteins contain three tandem repeats of this type, many other members of this family contain only two Myb repeats (*e.g.* the maize C1 and P proteins and budding yeast Rap1p) or a single repeat (*e.g.* human TRF1, rice RTBP1, and fission yeast Taz1p) (23–29). To our knowledge, metazoan SNAP190 is the only characterized polypeptide that contains more than three tandem Myb repeats.

The studies reported here reveal that all 4.5 Myb repeats of DmSNAP190 are located in close proximity to the DNA. Furthermore, based upon these data and the solved crystal structures of the repeats of the Myb protein itself, it became possible to accurately model the three-dimensional spatial arrangement of the 4.5 DmSNAP190 Myb repeats on the DNA when DmSNAPc binds to the U1 PSEA sequence. Somewhat surprisingly, we also discovered that several of the nucleotides in the 3′ half of the PSEA did not cross-link to any of the Myb repeats but rather cross-linked to the N-terminal domain of DmSNAP190 that precedes the Myb repeats.

EXPERIMENTAL PROCEDURES

DmSNAPc Constructs, Expression, and Purification—The preparation of untagged and N- and C-terminally FLAG-tagged constructs encoding DmSNAPc subunits has been described previously (30). The DmSNAP190 constructs with point mutations that eliminated and/or introduced hydroxylamine cleavage sites (asparaginyl-glycyl (NG) peptide bonds) have also been described previously (18). (One new N-terminally tagged DmSNAP190 construct was prepared for this work with a single NG peptide bond at residues 189–190.)

DmSNAP190, DmSNAP50, and DmSNAP43 were co-overexpressed in stably transfected *Drosophila* S2 cells each under the control of the metallothionein promoter (14, 15, 30). A FLAG tag was always present on the DmSNAP190 subunit except when we wished to express untagged DmSNAP190; in those cases, the DmSNAP43 subunit was expressed with an N-terminal FLAG tag. DmSNAP complexes were purified by FLAG immunoaffinity chromatography as described previously (30). Electrophoretic mobility shift assays were carried out to confirm the DNA-binding activity of each of the FLAG-purified DmSNAPc variants (30).

Site-specific Protein-DNA Photo-cross-linking, Hydroxylamine Digestion of Proteins, and Analysis of Cleaved Fragments—Double-stranded DNA probes that each contained a photo-cross-linking agent (azidophenacyl group) located at a specific individual phosphate position within or downstream of the U1 PSEA sequence were prepared exactly as described in detail previously (18). Briefly, two shorter synthetic oligonucleotides were annealed to a longer synthetic oligonucleotide and ligated to form a fully complementary double-stranded ^{32}P-labeled photo-cross-linking probe. To provide one example, a probe with cross-linker at position 7 of the non-template strand of the U1 PSEA was prepared by annealing each of the two oligonucleotides 5′-GCTATGACCATGATTACGAATTCATTCT-TA<u>TAATT</u>-3′ and 5′-*<u>CCxCAACTGGTTTTTAGC</u>GGTAC-

CGCCATGGAAAGGTATGGGATCC-3′ (where the PSEA is underlined, the asterisk indicates a ^{32}P, and x indicates the position of a phosphorothioate derivatized with azidophenacyl bromide) to the following 79-mer template strand oligonucleotide: 3′-CGATACTGGTACTAATGCTTAAGTAAGAAT<u>ATTA-AGGGTTGACCAAAAT</u>CGCCATGGCGGTACCTTTCCA-TACCCTAGG-5′. The annealed oligonucleotides were then ligated with T4 DNA ligase to generate the double-stranded probe. Probes with cross-linker at other specific phosphate positions on the non-template strand were generated by varying the position of the centrally located 3′ and 5′ ends of the two shorter synthetic oligonucleotides. Analogous strategies using complementary oligonucleotides were utilized to prepare probes with the cross-linking agent in the template strand of the DNA.

Photo-cross-linking reactions were carried out exactly as described previously (18). Briefly, FLAG-purified DmSNAPc was incubated with DNA probe in the dark for 30 min, followed by irradiation with UV light. Following irradiation, cross-linked samples were subjected to DNase I and S1 nuclease digestion to remove all but about 2 or 3 nucleotides of the DNA probe. Cross-linked proteins were cleaved specifically at NG peptide bonds by dialysis against freshly prepared 1.8 M hydroxylamine reagent at 45 °C for 5 h (18, 31). Samples were then prepared as described in detail recently (18) for electrophoresis through denaturing 11% polyacrylamide gels. Following electrophoresis, gels were dried and subjected to autoradiography.

Immunoblots of Hydroxylamine-digested DmSNAPc—Affinity-purified FLAG-tagged DmSNAPc samples were treated exactly as described above for digestion with hydroxylamine except that the incubation with the DNA photo-cross-linking probe and the UV irradiation were omitted. Immunoblotting was carried out as described previously (30). N-terminal FLAG-tagged fragments of DmSNAP190 were detected by using alkaline phosphatase-conjugated monoclonal antibody against the FLAG epitope (Sigma, catalog no. A9469). C-terminal fragments of DmSNAP190 were detected by using a primary antibody that was a rabbit polyclonal antibody generated against a 14-amino acid peptide from the C terminus of DmSNAP190 (14). The secondary antibody was alkaline phosphatase-conjugated goat anti-rabbit IgG(Fc) (Promega, catalog no. S3731).

RESULTS

Identification of DmSNAP190 Myb Domain Repeats That Cross-link to U1 PSEA Positions 3, 5, and 7 of the Non-template DNA Strand—In the 5′ portion of the U1 PSEA, DmSNAP190 can be cross-linked to phosphate positions 1, 3, 5, and 7 of the non-template strand (Fig. 1) (17). The photo-cross-linking protocol employed transfers a ^{32}P radiolabel in the DNA to nearby amino acid residues of any protein present in close proximity to the nucleotide position that contains the cross-linking agent. By carrying out hydroxylamine digestion of the protein at NG peptide bonds subsequent to the site-specific protein-DNA photo-cross-linking, it is possible to map or localize the region in the protein to which the cross-linking occurs. By employing such a strategy, we recently determined that phosphate position 1 of the U1 PSEA cross-linked to a region of DmSNAP190 located between amino acid residues 359 and 483 (18).
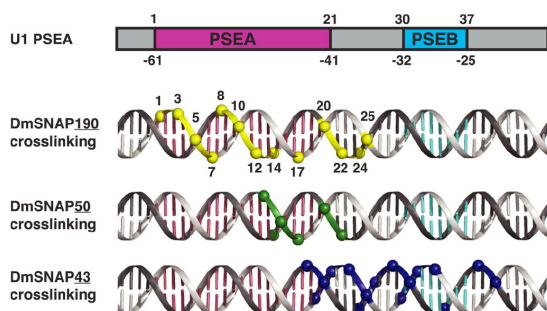
FIGURE 1. **Illustration summarizing the nucleotide positions where each of the three DmSNAPc subunits cross-link to U1 gene promoter sequences.** The *bar* at the *top* is a schematic representation of the U1 promoter region with the relative positions of the PSEA and PSEB indicated. The *negative numbers below* the *bar* indicate the positions of the PSEA and PSEB relative to the U1 gene transcription start site. The *top row* of *numbers* shows the numbering system used for identifying the PSEA positions (positions 1–21) and beyond the PSEA into the PSEB (positions 30–37). The phosphate positions to which DmSNAP190, DmSNAP50, and DmSNAP43 cross-link (as part of DNA-bound DmSNAPc) are indicated by the *yellow*, *green*, and *blue colored spheres*, respectively, on the double-helical DNA *below*. *Magenta bases* comprise the PSEA, and *cyan bases* comprise the PSEB. The template and non-template stands are rendered in *dark gray* and *light gray*, respectively. The phosphate positions to which DmSNAP190 cross-links are explicitly *numbered* on the *uppermost DNA helix*.

Fig. 2 shows results of experiments designed to map the regions of DmSNAP190 that cross-link to positions 3, 5, and 7 of the U1 PSEA. A schematic representation of DmSNAP190 is shown at the *top* of Fig. 2*A*. The 4.5 Myb repeats are labeled *Rh*, *Ra*, *Rb*, *Rc*, and *Rd*. The locations of hydroxylamine cutting sites (NG peptide bonds) engineered by site-directed mutagenesis into seven separate DmSNAP190 constructs are labeled *H*, *Ha*, *A*, *B*, *C*, *D*, and *E*. The N- and C-terminal fragments expected after hydroxylamine digestion of each construct are represented in the *lower part* of Fig. 2*A* by the *darkly shaded* and *unshaded* areas, respectively. Each construct contained a FLAG tag at its N terminus.

The *first two panels* in Fig. 2*B* (*lanes 1–14*) show immunoblots of these seven DmSNAP190 constructs following hydroxylamine digestion and detection with either antibodies specific for an epitope at the C terminus or specific for the tag at the N terminus. *Lanes 1–7* show the locations in the gel of the C-terminal fragments that decrease in size from *left* to *right*, and *lanes 8–14* reveal the positions of the N-terminal fragments that increase in size from *left* to *right*.

The *next panel* of Fig. 2*B* (*lanes 15–20*) shows an autoradiogram of ^{32}P-labeled fragments after photo-cross-linking of DmSNAPc site-specifically to position 3 of the U1 PSEA and subsequent hydroxylamine digestion of the protein. The pattern of decreasing fragment sizes in *lanes 15–19* is consistent with the cross-linking of phosphate position 3 to the C-terminal fragment in each of the constructs H through D. In contrast, it was the N-terminal fragment of construct E that cross-linked to position 3 (*lane 20*). Taken together, these results indicate that phosphate position 3 cross-linked to DmSNAP190 C-terminal of cut site D but N-terminal of cut site E. We thus can conclude that a domain of DmSNAP190 located between amino acid residues 410 and 483 cross-linked to phosphate position 3 of the U1 PSEA.

Lanes 21–26 of Fig. 2*B* show the results of photo-cross-linking the same six DmSNAPc constructs to position 5 of the U1 PSEA. Cross-linking of the C-terminal fragments of constructs H, A, and B was evident in *lanes 21–23*, whereas the N-terminal fragments from constructs D and E cross-linked (*lanes 25* and *26*). However, as previously noted (18), the N-terminal and C-terminal fragments from construct C migrated to very nearly the same position on the gel (*lanes 5* and *12*). Thus, it was not possible to unambiguously ascertain from these data alone whether the N-terminal or C-terminal fragment from construct C was responsible for the band in *lane 24*.

To distinguish between those two possibilities, photo-cross-linking was subsequently carried out with DmSNAPc that contained one of three different DmSNAP190 constructs: either untagged DmSNAP190, N-terminally tagged DmSNAP190, or C-terminally tagged DmSNAP190, each with the hydroxylamine cleavage site at position 358. These three constructs are shown at the *top* of Fig. 2*C*. (The tag at the N terminus increased the size of the N-terminal fragment by ~2.3 kDa compared with the untagged N-terminal fragment, and the tag at the C terminus increased the size of the C-terminal fragment by ~5.7 kDa.) These three cross-linking reactions to phosphate 5 were then run side-by-side on the same gel (Fig. 2*C*, *lanes 4–6*). In *lane 5*, the fragment from N-terminally tagged DmSNAP190 migrated more slowly than the cross-linked fragment from C-terminally tagged or untagged DmSNAP190. This result indicates that, after cleavage at position 358, the N-terminal fragment of DmSNAP190 cross-linked to phosphate position 5. The pattern from the converse result (if cross-linking occurred C-terminal of amino acid residue 358) is exemplified in *lanes 1–3* of Fig. 2*C*, which exhibits the C-terminal cross-linking pattern expected for position 3. Taken together, we can conclude that amino acids of DmSNAP190 between residues 306 and 358 cross-linked to phosphate position 5 of the U1 PSEA.

When experiments were carried out with the cross-linking agent at position 7 in the U1 PSEA, the pattern shown in Fig. 2*B*, *lanes 27–32*, was obtained. *Lanes 27* and *28* revealed that the C-terminal fragment of DmSNAP190 cross-linked when cleavage occurred following amino acid residue 168 or 247. However, when cleavage occurred following residue 305 (construct B, *lane 29*), the N-terminal fragment cross-linked most strongly, although the C-terminal fragment continued to cross-link but with less intensity. Thus, phosphate 7 of the U1 PSEA cross-linked to amino acid residues both N-terminal and C-terminal of residue 305 of DmSNAP190. When DmSNAP190 cleavage occurred at either position 409 or 483 (constructs D and E), only the N-terminal fragment cross-linked (Fig. 2*B*, *lanes 31* and *32*). This means that phosphate 7 of the U1 PSEA cross-linked to DmSNAP190 residues only to the N-terminal side of residue 409.

There remained a possibility that position 7 of the U1 PSEA could cross-link to residues C-terminal to residue 358 (although N-terminal to position 409). The experiments shown in Fig. 2*C*, *lanes 7–9*, discount that possibility because the banding pattern reveals that only the N-terminal fragment cross-linked. Taking all of these data into account, we conclude that phosphate 7 of the U1 PSEA cross-linked strongly to residues of DmSNAP190 localized between 248 and 305 and cross-linked

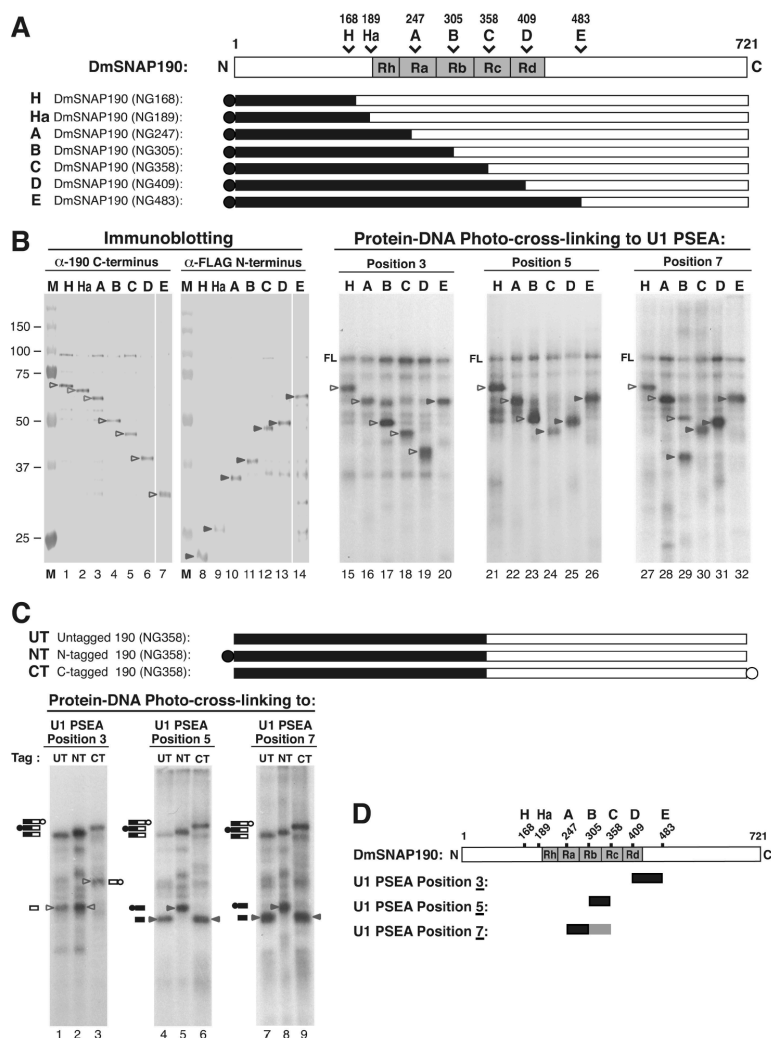*Arrangement of SNAP190 Myb Repeats on U1 Gene Promoter DNA*

FIGURE 2. **Localization of domains of DmSNAP190 that cross-link to U1 PSEA positions 3, 5, and 7 of the non-template strand.** *A,* schematic representation of the hydroxylamine cleavage patterns of DmSNAP190 constructs used in the photo-cross-linking studies. The *top bar* shows structural features of DmSNAP190, which consists of 721 amino acid residues and 4.5 Myb repeats designated Rh, Ra, Rb, Rc, and Rd. The position of each of the hydroxylamine cutting sites (NG peptide bonds) in each of the individual DmSNAP190 constructs is shown *above* the *bar* at the very *top.* The *shaded* and *unshaded regions below* represent the N-terminal and C-terminal fragments produced by chemical cleavage of each of the constructs. Constructs H through E each contain an N-terminal FLAG tag and a single NG peptide bond at a variable position within or flanking the Myb repeats. *B,* all gel results shown are subsequent to hydroxylamine digestion of DmSNAPc from cell lines separately expressing the DmSNAP190 variants represented by constructs H to E shown in *A.* The *first two panels* show immunoblots using antibodies against a C-terminal peptide of DmSNAP190 (*lanes 1–7*) or antibodies against the FLAG epitope at the N terminus of each construct (*lanes 8–14*). Lanes marked *M* contain molecular weight markers. *Arrowheads* (*unshaded* for C-terminal fragments and *shaded* for N-terminal fragments) point out the specific cleavage products of interest. *Lanes 7* and *14* are from the same gels as *lanes 1–6* and *8–13,* respectively. *Lanes 15–32* show autoradiography results of protein-DNA photo-cross-linking of the DmSNAP190 constructs H, A, B, C, D, and E to U1 PSEA position 3 (*lanes 15–20*), position 5 (*lanes 21–26*), or position 7 (*lanes 27–32*). In *lane 29,* both the N- and C-terminal fragments of construct B cross-linked to position 7. *FL,* position of bands corresponding to full-length undigested N-tagged DmSNAP190. *C,* constructs UT, NT, and CT are untagged, N-tagged, and C-tagged, respectively, and each has a single NG peptide bond at position 358. (Construct NT and construct C (from *A*) are actually identical but are labeled differently in different parts of the figure for clarity.) The *three panels below* show autoradiograms of protein-DNA photo-cross-linking of DmSNAP190 constructs UT, NT, and CT to U1 PSEA position 3 (*lanes 1–3*), position 5 (*lanes 4–6*), or position 7 (*lanes 7–9*). The *symbols alongside* the gel indicate the origin of each band and correspond to the schematic diagrams of the constructs *above. Unshaded arrowheads* point to bands representing C-terminal fragments, and *shaded arrowheads* point to N-terminal fragments. Bands corresponding to residual undigested full-length DmSNAP190 (untagged, N-tagged, and C-tagged) are observed in the *upper part* of each *panel. D,* schematic drawing indicating the region of DmSNAP190 that cross-linked to each of the positions 3, 5, and 7 of the U1 PSEA. *Dark shading,* stronger cross-linking; *lighter shading,* weaker cross-linking.
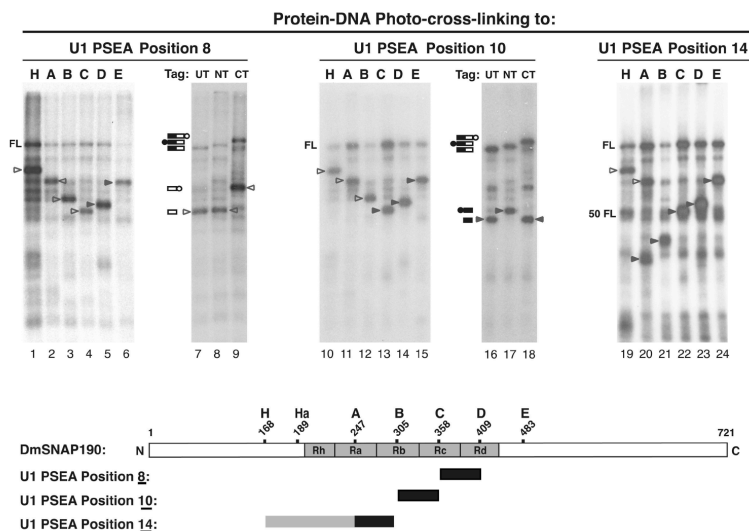
FIGURE 3. **Localization of domains of DmSNAP190 that cross-link to U1 PSEA positions 8, 10, and 14 of the template strand.** *Lanes 1–9*, *lanes 10–18*, and *lanes 19–24* show autoradiograms of the results of photo-cross-linking DmSNAP190 NG variants to U1 PSEA phosphate positions 8, 10, and 14, respectively. *Lanes* and *bands* are *labeled* as indicated in the legend of Fig. 2. *Unshaded arrowheads* point to bands corresponding to C-terminal fragments that cross-linked, and *shaded arrowheads* point to cross-linked N-terminal fragments. The term *50 FL* to the *left* of *lane 19* indicates a band in *lanes 19–24* that is due to the cross-linking of DmSNAP50 (full-length) to phosphate position 14 (see Fig. 1 and Ref. 17). The diagram at the *bottom* of the figure indicates the region of DmSNAP190 that cross-linked to each of the positions 8, 10, and 14 of the U1 PSEA. *Dark shading*, stronger cross-linking; *lighter shading*, somewhat weaker cross-linking that was more obviously weaker on the original autoradiograms.

with less intensity to residues between 306 and 358. The results of cross-linking phosphate positions 3, 5, and 7 to DmSNAP190 are summarized in Fig. 2*D*.

Identification of DmSNAP190 Myb Domain Repeats That Cross-link to U1 PSEA Positions 8, 10, and 14 of the Template Strand—Previous work revealed that DmSNAP190 cross-linked to phosphate positions 8, 10, 12, and 14 of the template strand in the central region of the U1 PSEA (Fig. 1) (17). More recent work determined that phosphate position 12 cross-linked to a region of DmSNAP190 between amino acid residues 306 and 358 (18). To further investigate the DmSNAP190 DNA-binding pattern to the U1 PSEA, we next examined the cross-linking of phosphate position 8 of the template strand to DmSNAP190. The pattern of cross-linking (Fig. 3, *lanes 1–9*) revealed that position 8 of the U1 PSEA cross-linked C-terminal to residue 306 (construct B, *lane 3*) but N-terminal to residue 409 (construct D, *lane 5*). Results shown in *lanes 7–9* indicated that the cross-linking occurred C-terminal to DmSNAP190 residue 358. We can therefore conclude that U1 PSEA position 8 cross-linked to amino acid residues of DmSNAP190 localized between positions 359 and 409.

Phosphate position 10 of the U1 PSEA exhibited a cross-linking pattern (Fig. 3 lanes 10–15) that was remarkably similar to that obtained with position 8 (*lanes 1–6*). Notably, however, the results in *lanes 16–18* are the converse of those observed in *lanes 7–9*. Thus, position 10 cross-linked to DmSNAP190 residues N-terminal of position 358. We therefore can conclude that U1 PSEA position 10 cross-linked to DmSNAP190 residues located between positions 306 and 358. Previous work

found that the nearby position 12 of the U1 PSEA likewise cross-linked to this same region of DmSNAP190 (18).

Consequently, we next examined the cross-linking of DmSNAP190 to position 14 of the U1 PSEA (Fig. 3, *lanes 19–24*). This position produced a significantly different pattern. (Background bands that result from the cross-linking of DmSNAP50 to phosphate position 14 are also observed on this gel.) Cross-linking to DmSNAP190 occurred C-terminal of position 168 (construct H, *lane 19*) but N-terminal of position 305 (construct B, *lane 21*). However, the results with construct A (*lane 20*) revealed that position 14 cross-linked to fragments both N-terminal and C-terminal of position 247. As a result, we can conclude that DmSNAP190 residues between 169 and 247 as well as residues between 248 and 305 are in close proximity to phosphate position 14 of the U1 PSEA. The results of cross-linking phosphate positions 8, 10, and 14 to DmSNAP190 are summarized at the *bottom* of Fig. 3.

Identification of DmSNAP190 Domains That Cross-link to U1 PSEA Positions 17 and 25 of the Non-template Strand—We next examined the domains in DmSNAP190 that cross-link to phosphate positions 17 and 25 in the non-template strand of the DNA. Each of these positions presented specific challenges in comparison with the previously described work. For example, phosphate position 17 cross-links with very much higher intensity to DmSNAP50 than it does to DmSNAP190 (17). As a result, very strong bands of cross-linked DmSNAP50 (full-length and hydroxylamine-cleaved products) appear on a gel after DmSNAPc cross-linking to position 17. At position 25, on the other hand, the DmSNAP190 cross-linking signal is very

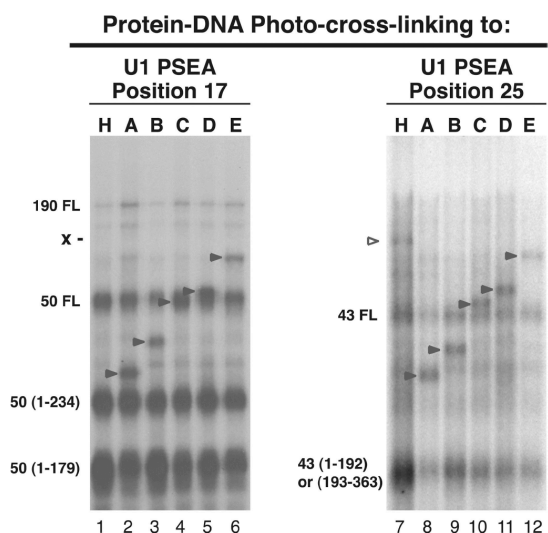*Arrangement of SNAP190 Myb Repeats on U1 Gene Promoter DNA*

Protein-DNA Photo-cross-linking to:

FIGURE 4. **Localization of domains of DmSNAP190 that cross-link to U1 PSEA positions 17 and 25 of the non-template strand.** *Lanes 1–6* and *lanes 7–12* show autoradiograms of the results of photo-cross-linking six different DmSNAP190 NG variants to U1 PSEA phosphate positions 17–25, respectively. *Labeling* of *lanes* and *bands* is as indicated in the legend of Fig. 2; *shaded arrowheads* point to cross-linked N-terminal fragments, and the *unshaded arrowhead* points to a faint cross-linked C-terminal fragment. ×, the absence of a band in *lane 1* that would be expected at the indicated position if there were cross-linking to the C-terminal fragment of construct H. Positions 17 and 25 cross-link not only to DmSNAP190 but also to DmSNAP50 and DmSNAP43, respectively (see Fig. 1 and Ref. 17). This necessarily results in additional bands present in all *lanes* of each *panel*. *50 FL* and *43 FL* indicate the positions of the cross-linked full-length (undigested) DmSNAP50 and DmSNAP43, respectively. The designations *50(1–234)*, *50(1–179)*, and *43(1–192)* or *(193–363)* point out the positions of DmSNAP50 and DmSNAP43 hydroxylamine digestion products with the *numbers* in *parentheses* indicating the amino acid positions at the beginning and end of those fragments. The bands near the bottom of the gel would be expected to obscure a possible cross-linked N-terminal fragment of DmSNAP190 produced by digestion of construct H (*lanes 1* and *7*).

Protein-DNA Photo-cross-linking to:

FIGURE 5. **Localization of domains of DmSNAP190 that cross-link to U1 PSEA positions 20 and 22 of the template strand.** *Lanes 1–6* and *lanes 7–12* show autoradiograms of the results of photo-cross-linking six different DmSNAP190 NG variants to U1 PSEA phosphate positions 20 and 22, respectively. *Labeling* of *lanes* and *bands* is as indicated in the legend of Fig. 2; *shaded arrowheads* point to cross-linked N-terminal fragments. ×, the absence of bands in *lanes 1* and *7* that would be expected at the indicated positions if there were cross-linking to the C-terminal fragment of construct H. Besides cross-linking to DmSNAP190, position 20 also cross-links to DmSNAP50 and DmSNAP43, and position 22 cross-links to DmSNAP50. Bands corresponding to the full-length DmSNAP50 and DmSNAP43 subunits and their digestion products are indicated *alongside* the *panels* as explained in the legend to Fig. 4.

weak compared with the positions already discussed (17), so long autoradiography exposures were required to detect the signal. Furthermore, position 25 cross-links to DmSNAP43 as well as to DmSNAP190 (17); therefore, DmSNAP43 bands appear in the gel in addition to the DmSNAP190 bands.

The results of these experiments with positions 17 and 25 are shown in Fig. 4. In the experiment with cross-linker at position 17 (*lanes 1–6*), N-terminal fragments from constructs A, B, and E could readily be identified (*lanes 2*, *3*, and *6*). Furthermore, upon careful observation, the N-terminal fragment from construct D could be observed just above the full-length DmSNAP50 band in *lane 5*, and the DmSNAP190 band in *lane 4* co-migrated with the DmSNAP50 band. It was clear from these data that position 17 cross-linked N-terminal of DmSNAP190 residue 247 (the cleavage site in construct A). It is further notable that no cross-linked C-terminal band was visible in *lane 1* (expected position marked by an ×). The absence of this C-terminal band implies that position 17 cross-linked to DmSNAP190 on the N-terminal side of residue 168, a region that precedes the Myb domain. Unfortunately, this smaller N-terminal fragment of DmSNAP190 (*i.e.* residues 1–168)
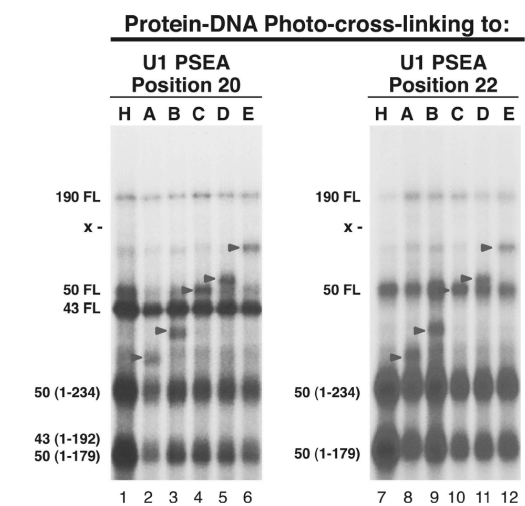
would be expected to co-migrate with and be obscured by the very strong DmSNAP50(1–179) band that lies near the bottom of the gel.

When the cross-linking agent was at position 25, the pattern obtained was very similar to that obtained with position 17. However, a weak yet clear band that corresponds to a C-terminal cross-linked fragment could be observed from construct H (Fig. 4, *lane 7*). This result indicates that position 25 cross-linked at least weakly to DmSNAP190 residues between 169 and 247. However, it does not exclude the possibility that position 25 might also cross-link to residues N-terminal of residue 168. The band corresponding to DmSNAP190 fragment 1–168 would be expected to overlap with and be obscured by the DmSNAP43 cleavage product band near the *bottom* of the *panel*.

Identification of DmSNAP190 Domains that Cross-link to U1 PSEA Positions 20 and 22 of the Template Strand—We next performed experiments to identify the regions of DmSNAP190 that are in close proximity to positions 20 and 22 of the template strand of the U1 PSEA. At position 20, both DmSNAP43 and DmSNAP50 cross-link more strongly than DmSNAP190 (17). At position 22, DmSNAP50 cross-links more strongly than DmSNAP190 (17). Despite the presence of intense bands that arise from DmSNAP43 and DmSNAP50 and their hydroxylamine digestion products in the autoradiograms of Fig. 5, most of the DmSNAP190 bands could also be readily discerned.

For phosphate positions 20 and 22 (Fig. 5, *lanes 1–12*), it was readily apparent that both positions cross-linked to a region of

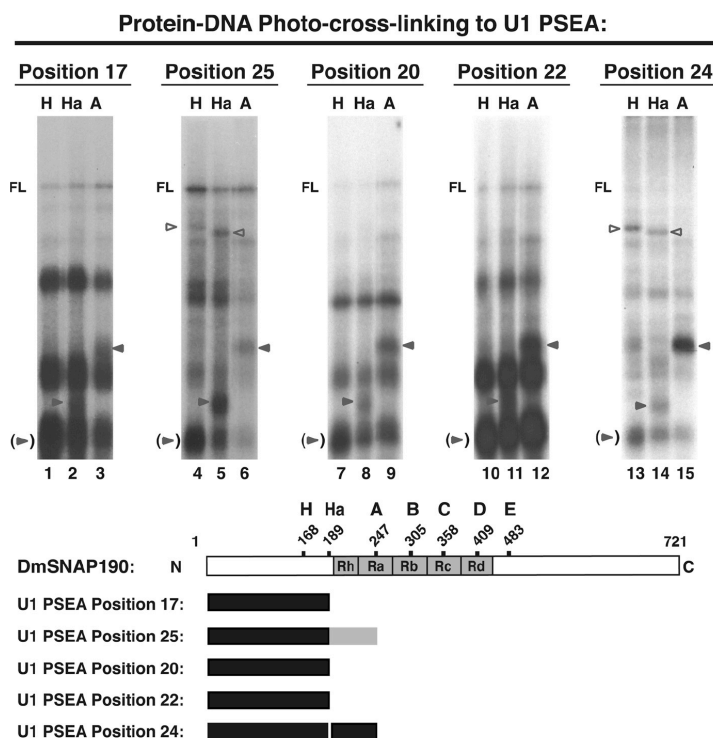*Arrangement of SNAP190 Myb Repeats on U1 Gene Promoter DNA*

FIGURE 6. **The N-terminal domain of DmSNAP190 cross-links to phosphate positions 17, 25, 20, 22, and 24 of the U1 PSEA.** Autoradiograms are shown of the results of photo-cross-linking DmSNAP190 constructs H, Ha, and A (Fig. 2) to U1 PSEA phosphate positions 17, 25, 20, 22, and 24. *Unshaded arrowheads* point to C-terminal fragments that were cross-linked in *lanes 4, 5, 13,* and *14,* whereas *shaded arrowheads* point to N-terminal fragments that cross-linked. The *shaded arrowheads* in *parentheses* point to the location of a presumed band in *lanes 1, 4, 7, 10,* and *13* that would correspond to an N-terminal cross-linked fragment following hydroxylamine digestion of DmSNAP190 construct H but is obscured by the strong bands arising from digestion products of DmSNAP50 or DmSNAP43. The *lower section* of the figure indicates the region of DmSNAP190 that cross-links to each of the positions 17, 25, 20, 22, and 24 of the U1 PSEA based upon the results shown in this figure as well as Figs. 4 and 5. *Dark shading*, regions of stronger cross-linking; *lighter shading*, region of weaker cross-linking. Position 24 cross-linked with approximately equal intensities both N-terminal and C-terminal of DmSNAP190 residue 189.

DmSNAP190 N-terminal of amino acid residue 247 (the cleavage site in construct A, *lanes 2* and *8*). Furthermore, and notably, there was no band visible in either *lane 1* or *lane 7* that would arise from cross-linking to the C-terminal fragment of construct H (expected position indicated by ✕). This result suggested that positions 20 and 22 cross-linked to the N-terminal fragment of construct H (residues 1–168) that would be obscured by the very dark DmSNAP50/43 bands at the bottom of the gel. To shed more light on this subject, the experiments shown in Fig. 6 were performed.

Phosphate Positions in the 3'-Half of the U1 PSEA Contact Residues within the N-terminal Domain of DmSNAP190—Results presented in Figs. 4 and 5 implied that the region N-terminal of the Myb domain of DmSNAP190 may cross-link to the U1 PSEA at phosphate positions 17, 25, 20, and 22. To investigate this subject further, we prepared a new DmSNAP190 construct with an NG peptide bond at position 189. This hydroxylamine cut site is just N-terminal of Rh of the Myb domain. We named this construct Ha (Fig. 2A). The relative positions migrated by the N- and C-terminal fragments following

hydroxylamine digestion of this construct are revealed in the immunoblot shown in Fig. 2B, *lanes 2* and *9*.

DmSNAPc that contained DmSNAP190 construct Ha was used in photo-cross-linking reactions with positions 17, 25, 20, and 22 of the U1 PSEA (Fig. 6, *lanes 2, 5, 8,* and *11*). In each case, this cross-linking reaction was flanked by lanes containing reactions with DmSNAP190 constructs H and A. Furthermore, because earlier work had indicated that a region of DmSNAP190 between residues 169 and 247 had cross-linked to phosphate position 24 (18), we decided to examine whether phosphate 24 might be capable of cross-linking in addition to the domain of DmSNAP190 between residues 1 and 168. Therefore, reactions (Fig. 6, *lanes 13–15*) were also carried out with the Ha construct and cross-linking agent at position 24 of the template strand to complement the previously published data.

Interestingly, each of the panels revealed a new band that migrated in the gel at the relative position corresponding to the N-terminal fragment of construct Ha (DmSNAP190 residues 1–189) cross-linked to the DNA (Fig. 2B). This band is quite

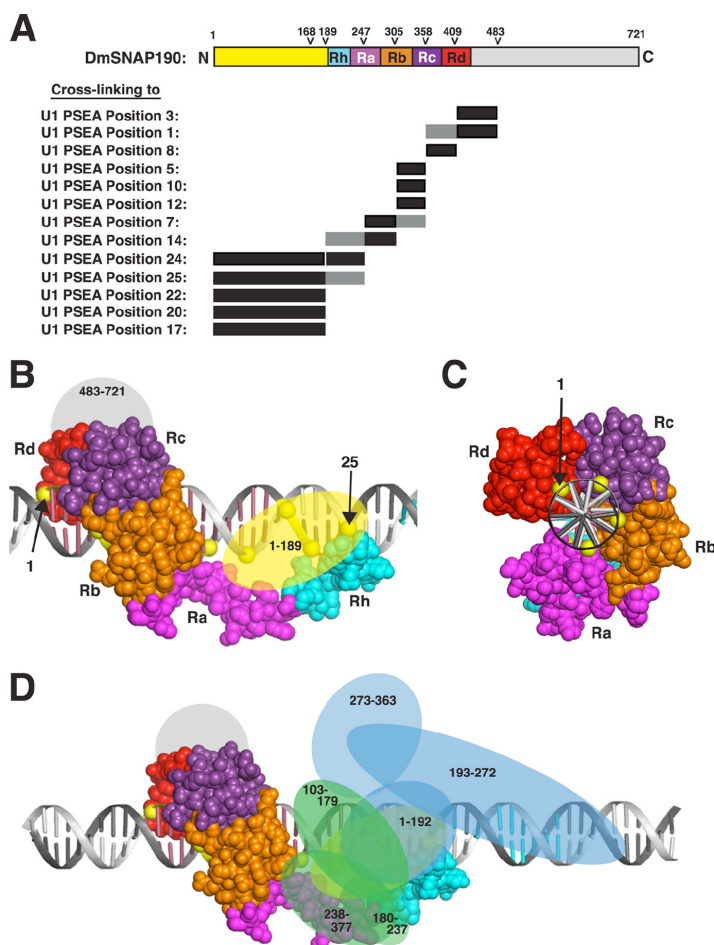*Arrangement of SNAP190 Myb Repeats on U1 Gene Promoter DNA*

FIGURE 7. **Localization and model of DmSNAP190 domains interacting with the U1 PSEA.** *A*, summary diagram indicating the regions of DmSNAP190 that cross-link to 13 phosphate positions in the U1 PSEA. In the *lower section* of *A*, stronger cross-linking is represented by the *dark shading*, and weaker cross-linking is shown by the *lighter shading*. *B*, model of the 4.5 Myb repeats of DmSNAP190 interacting with the U1 PSEA based upon the photo-cross-linking data summarized in *A*. The method and rationale of the modeling are explained in the text. Phosphate positions 1–25 of the U1 PSEA are explicitly labeled. The *yellow oval* represents the N-terminal domain of DmSNAP190 that precedes the Myb repeats and is shown interacting with phosphates 17, 20, 22, 24, and 25. The *gray circle* represents the C-terminal domain of DmSNAP190, which is not known to cross-link to the U1 DNA. *C*, the same model of the 4.5 Myb repeats shown in *B* but viewed from upstream of the U1 PSEA looking toward the transcription start site. Phosphate position 1 is explicitly *labeled*; the other visible phosphate positions are 3, 5, and 7. *D*, a more comprehensive model of the entire DmSNAP complex based upon photo-cross-linking studies, including DmSNAP50 (*green ovals*) and DmSNAP43 (*blue ovals*) (18). The *numbers within* the *ovals* indicate the extent of the amino acid residues in the domains of DmSNAP50 and of DmSNAP43 that have been mapped to specific nucleotide regions of the U1 PSEA (18).

distinct in *lanes 5*, *8*, and *14* of Fig. 6, corresponding to reactions with phosphate positions 25, 20, and 24. In *lanes 2* and *11*, the bands are less distinct due to the partial overlap with the very intense bands that result from cross-linking to DmSNAP50 at positions 17 and 22.

Further information can be gleaned from *lanes 4*, *5*, *13*, and *14*, which exhibit high molecular weight bands that correspond to the C-terminal fragment of DmSNAP190 in those lanes. The appearance of these bands indicates and confirms that amino acids C-terminal as well as N-terminal of residue 189 cross-linked to phosphate positions 25 and 24 of the U1 PSEA. Inter-

estingly, no evidence of a corresponding band was visible when the cross-linking agent was at position 17, 20, or 22 (*lanes 2*, *8*, and *11*), suggesting that these phosphate positions cross-linked only to the most N-terminal fragment (and furthermore consistent with the lack of C-terminal fragment bands in *lanes 1*, *7*, and *10*). The results of the photo-cross-linking reactions for positions 17, 25, 20, 22, and 24 are summarized in the *lower part* of Fig. 6.

Arrangement of DmSNAP190 Domains That Interact with the U1 PSEA—Fig. 7*A* provides a comprehensive summary that regionalizes the domains of DmSNAP190 that cross-link to

specific phosphate positions of the U1 PSEA. This diagram is based upon results presented in Figs. 2–6 and previously published for positions 1, 12, and 24 (18).

To better understand the molecular structure of DmSNAP190 and its interaction with the DNA, we took advantage of the known coordinates of the co-crystal structure of the c-Myb DNA-binding domain complexed with a specific DNA recognition sequence (Protein Data Bank entry 1H88) (32). To generate a theoretical model of the DmSNAP190 Myb repeats bound to DNA, we used a Web-based structural alignment tool (SWISS-MODEL Repository) (38). Repeats R2 and R3 of the c-Myb protein itself contain recognition helices closely packed in a head-to-tail fashion that together make sequence-specific contacts over a distance of about 6 base pairs in the major groove of the DNA. R1, on the other hand, lies farther from the DNA axis and does not make base-specific contacts (32, 33). Because DmSNAPc has 4.5 Myb repeats, whereas c-Myb itself has only three repeats, we modeled Rc and Rd of SNAP190 using the R2 and R3 repeats of c-Myb as templates. We then repeated the use of the alignment tool to model Rh, Ra, and Rb of DmSNAP190 using as templates R1 (latter half), R2, and R3, respectively, of c-Myb in its DNA-bound structure. All Protein Data Bank files were visualized and manipulated using PyMOL (Schroedinger LLC, New York).

After obtaining modeled structures for the DmSNAP190 Myb repeats bound to DNA as described above, we then manually aligned the phosphodiester backbone of the Myb DNA with the backbone of the U1:95Ca PSEA and flanking DNA sequence modeled as B-form DNA. The precise nucleotide positions used for the alignment were chosen to give the best agreement with the photo-cross-linking data summarized in Fig. 7*A*.

Indeed, we found that a good fit consistent with the cross-linking data could be obtained when the DmSNAP190 Myb repeats were placed on the DNA as shown in Fig. 7*B*. (Also see supplemental Protein Data Bank File 1.) The major criterion for placing the modeled protein domains was proper Myb repeat proximity to the corresponding cross-linked phosphates. In the model presented, the four "recognition helices" of Rd, Rc, Rb, and Ra all lie in the major groove of the DNA and potentially could contact base pairs 1–12 of the PSEA.

Rd, Rc, Rb, and the latter half of Ra (the helix-turn-helix portion) are all modeled as canonical Myb repeat structures in Fig. 7*B*. However, we found no way to model Rh and the first half of Ra as completely folded Myb repeats and still maintain consistency with the fact that the cross-linking data require Rh (or regions near Rh) to be in the vicinity of phosphates 24 and 25. In the Myb structure, each repeat is separated from its neighboring repeat by a relatively unstructured loop, and within each repeat, there are three α-helices that are separated by turns (32–34). By considering the loops and turns of Rh and Ra to be flexible pivot points, we could generate a feasible structure (as modeled in Fig. 7*B*) that fits the cross-linking data without disturbing the predicted α-helices of Rh and Ra.

Fig. 7*C* shows the same structure as Fig. 7*B* but viewed along the axis of the DNA from upstream of the PSEA toward the transcription start site. As the last 3.5 Myb repeats of DmSNAP190 (*i.e.* Rd, Rc, Rb, and the last half of Ra) track along

the major groove of the DNA, they nearly circumscribe the DNA. This is somewhat reminiscent of the structure of DNA-binding proteins that contain multiple zinc fingers of the C2-H2 type (35, 36).

To our knowledge, the N-terminal domain of DmSNAP190 (residues 1–189) has no sequence similarity to any other known protein, so we can make no predictions of its structure. However, we have found that it cross-links to U1 PSEA phosphate positions 17, 20, 22, and 25. Thus, this 189-residue domain is represented in Fig. 7*B* as a *yellow oval* in proximity to those phosphate positions in the PSEA. In contrast, the C-terminal domain of DmSNAP190 (residues 483–721, represented by the *gray circle* in Fig. 7*B*) has not been observed to contact the DNA of the U1 PSEA.

DISCUSSION

DmSNAP190 Myb Repeat Interaction with the U1 PSEA— We have used a site-specific protein-DNA photo-cross-linking assay combined with site-specific protein digestion to map domains in DmSNAP190 that contact each of the 13 phosphate positions that can be cross-linked to DmSNAP190 when DmSNAPc binds to a U1 PSEA. To our knowledge, SNAP190 is unique in that it is the only protein that possesses more than three tandem Myb repeats.

The data that we have obtained from our photo-cross-linking experiments have allowed us to model the DmSNAP190 Myb repeats on the DNA of the U1 PSEA. This work suggests a unique framework in which the four recognition helices of Ra, Rb, Rc, and Rd can fit into the major groove of the DNA, where they can potentially make base-specific contacts with a 10–12-base pair segment of DNA that constitutes the most highly conserved region of the insect PSEA (16).

On the other hand, the cross-linking data appear to be incompatible with the putative two helices of Rh and the first helix of Ra adopting a canonically folded Myb repeat structure. This region of the protein must stretch between phosphate 14 and phosphates 24 and 25 on the DNA. This is entirely feasible if the loops and turns of these repeats are considered to be flexible, allowing the three putative helices in this region of the protein to "unfold" while still retaining their local helical secondary structure. Indeed, we have made this assumption when carrying out the modeling of Rh and Ra illustrated in Fig. 7*B*. An alternative possibility for relieving this constraint imposed by the cross-linking data would be to introduce a downward bend into the DNA when viewed from the orientation shown in Fig. 7*B*. However, this would require a quite radical bend. Furthermore, earlier studies in our laboratory indicated that the DNA is only modestly bent upon the binding of DmSNAPc, and this modest bend is toward the upper surface of the DNA as oriented in Fig. 7*B* (37).

Although we have modeled the complex on B-form DNA, it is not our intention to suggest that the DNA is completely in B-form over this entire region. Indeed, the DNA-bending investigations from our laboratory cited above indicate that to be unlikely. Also, in a co-crystal with c-Myb and C/EBPβ, the DNA structure is bent but only modestly (32). Thus, in the absence of additional information, it seems reasonable to model

Arrangement of SNAP190 Myb Repeats on U1 Gene Promoter DNA

DmSNAP190 bound to B-form DNA, taking into account that some deviation from the B-form is likely.

Compatibility of the Model with DmSNAPc Subunit-Subunit Interaction Studies—In a previous study, we mapped domains of DmSNAP50 and of DmSNAP43 that cross-link to a number of specific nucleotide positions in the U1 PSEA (10, 18). An overlay of the results of those studies with the current findings is shown in Fig. 7D. It is likely from those photo-cross-linking results that the interfaces among the three proteins and the DNA are likely to be quite convoluted in the 3′-half of the PSEA, but nevertheless, more clarity is developing regarding the arrangement of the DmSNAPc subunits relative to the DNA and to each other.

In still another study, we mapped domains of DmSNAP190, of DmSNAP50, and of DmSNAP43 that were required for subunit-subunit interactions (30). Those findings indicated that a region of DmSNAP190 between amino acid residues 63 and 176 was required for interaction with DmSNAP50. This region is included in the N-terminal domain of DmSNAP190 that cross-linked to phosphate positions 17, 20, 22, 24, and 25. Interestingly, the DNA photo-cross-linking studies placed this N-terminal domain of DmSNAP190 into close proximity to DmSNAP50 residues 110–377 (Fig. 7, B and D) that are involved in interacting with DmSNAP190 (10, 30). The same protein-protein interaction studies also revealed that interaction between DmSNAP190 and DmSNAP43 was dependent upon residues 1–172 of DmSNAP43 interacting with Myb domain residues of DmSNAP190. The DNA cross-linking studies place this N-terminal domain of DmSNAP43 into the proximity of the Rh and Ra repeats of DmSNAP190 but probably not in proximity to the Rb, Rc, and Rd repeats (Fig. 7D). This suggests that DmSNAP43 interacts primarily with the Rh and Ra repeats of DmSNAP190. The interaction of DmSNAP43 with DmSNAP190 may prevent the Rh and Ra repeats of DmSNAP190 from adopting a canonical Myb repeat structure on the DNA.

In conclusion, these studies indicate that each of the 4.5 Myb repeats of DmSNAP190 is in close proximity to the DNA when DmSNAPc binds to the U1 PSEA, and the repeats are ordered with the C-terminal repeat (Rd) farthest from the transcription start site and the most N-terminal repeat (Rh) closest. Furthermore, and unexpectedly, the N-terminal domain of DmSNAP190 that precedes the Myb repeats can contact multiple phosphate positions in the 3′-half of the U1 PSEA. This finding places the N-terminal domain of DmSNAP190 in an ideal location, where it could possibly interact with promoter-bound TBP, an interaction that has been observed on the promoters of U6 genes in the human system (19).

REFERENCES

1. Waldschmidt, R., Wanandi, I., and Seifart, K. H. (1991) Identification of transcription factors required for the expression of mammalian U6 genes *in vitro*. *EMBO J.* **10**, 2595–2603
2. Murphy, S., Yoon, J. B., Gerster, T., and Roeder, R. G. (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
3. Wanandi, I., Waldschmidt, R., and Seifart, K. H. (1993) Mammalian transcription factor PBP. Characterization of its binding properties to the proximal sequence element of U6 genes. *J. Biol. Chem.* **268**, 6629–6640
4. Sadowski, C. L., Henry, R. W., Lobo, S. M., 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
5. Yoon, J. B., Murphy, S., Bai, L., Wang, Z., and Roeder, R. G. (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
6. Henry, R. W., Sadowski, C. L., 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
7. Su, Y., Song, Y., Wang, Y., Jessop, L., Zhan, L., and Stumph, W. E. (1997) Characterization of a *Drosophila* proximal-sequence-element-binding protein involved in transcription of small nuclear RNA genes. *Eur. J. Biochem.* **248**, 231–237
8. Hernandez, N. (2001) snRNA genes. A model system to study fundamental mechanisms of transcription. *J. Biol. Chem.* **276**, 26733–26736
9. Egloff, S., O'Reilly, D., and Murphy, S. (2008) Expression of human snRNA genes from beginning to end. *Biochem. Soc. Trans.* **36**, 590–594
10. Hung, K. H., and Stumph, W. E. (2011) Regulation of snRNA gene expression by the *Drosophila melanogaster* small nuclear RNA-activating protein complex (DmSNAPc). *Crit. Rev. Biochem. Mol. Biol.* **46**, 11–26
11. 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. U.S.A.* **100**, 80–85
12. Das, A., Zhang, Q., Palenchar, J. B., Chatterjee, B., Cross, G. A., 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
13. Schimanski, B., Nguyen, T. N., and Günzl, 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
14. Li, C., Harding, G. A., Parise, J., McNamara-Schroeder, K. J., and Stumph, W. E. (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
15. Lai, H. T., Kang, Y. S., and Stumph, W. E. (2008) Subunit stoichiometry of the Drosophila melanogaster small nuclear RNA activating protein complex (SNAPc). *FEBS Lett.* **582**, 3734–3738
16. Hernandez, G., Jr., Valafar, F., and Stumph, W. E. (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
17. Wang, Y., and Stumph, W. E. (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
18. Kim, M. K., Kang, Y. S., Lai, H. T., Barakat, N. H., Magante, D., and Stumph, W. E. (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; Correction (2010) *Mol. Cell Biol.* **30**, 5257
19. Ma, B., and Hernandez, N. (2002) Redundant cooperative interactions for assembly of a human U6 transcription initiation complex. *Mol. Cell Biol.* **22**, 8067–8078
20. Wong, M. W., Henry, R. W., Ma, B., 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
21. Sakura, H., Kanei-Ishii, C., Nagase, T., Nakagoshi, H., Gonda, T. J., and Ishii, S. (1989) Delineation of three functional domains of the transcriptional activator encoded by the c-myb protooncogene. *Proc. Natl. Acad.*

Sci. U.S.A. **86,** 5758–5762

22. Tanikawa, J., Yasukawa, T., Enari, M., Ogata, K., Nishimura, Y., Ishii, S., and Sarai, A. (1993) Recognition of specific DNA sequences by the c-myb protooncogene product. Role of three repeat units in the DNA-binding domain. *Proc. Natl. Acad. Sci. U.S.A.* **90,** 9320–9324

23. Grotewold, E., Athma, P., and Peterson, T. (1991) Alternatively spliced products of the maize P gene encode proteins with homology to the DNA-binding domain of myb-like transcription factors. *Proc. Natl. Acad. Sci. U.S.A.* **88,** 4587–4591

24. Bilaud, T., Koering, C. E., Binet-Brasselet, E., Ancelin, K., Pollice, A., Gasser, S. M., and Gilson, E. (1996) The telobox, a Myb-related telomeric DNA binding motif found in proteins from yeast, plants and human. *Nucleic Acids Res.* **24,** 1294–1303

25. Konig, P., Giraldo, R., Chapman, L., and Rhodes, D. (1996) The crystal structure of the DNA-binding domain of yeast RAP1 in complex with telomeric DNA. *Cell* **85,** 125–136

26. König, P., Fairall, L., and Rhodes, D. (1998) Sequence-specific DNA recognition by the myb-like domain of the human telomere binding protein TRF1. A model for the protein-DNA complex. *Nucleic Acids Res.* **26,** 1731–1740

27. Bianchi, A., Smith, S., Chong, L., Elias, P., and de Lange, T. (1997) TRF1 is a dimer and bends telomeric DNA. *EMBO J.* **16,** 1785–1794

28. Yu, E. Y., Kim, S. E., Kim, J. H., Ko, J. H., Cho, M. H., and Chung, I. K. (2000) Sequence-specific DNA recognition by the Myb-like domain of plant telomeric protein RTBP1. *J. Biol. Chem.* **275,** 24208–24214

29. Cooper, J. P., Nimmo, E. R., Allshire, R. C., and Cech, T. R. (1997) Regulation of telomere length and function by a Myb-domain protein in fission yeast. *Nature* **385,** 744–747

30. Hung, K. H., Titus, M., Chiang, S. C., and Stumph, W. E. (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

31. Crimmins, D. L., and Mische, S. M. (1996) Chemical cleavage of proteins in solution. *Curr. Protoc. Protein Sci.* **4,** 11.14.11–11.14.18

32. Tahirov, T. H., Sato, K., Ichikawa-Iwata, E., Sasaki, M., Inoue-Bungo, T., Shiina, M., Kimura, K., Takata, S., Fujikawa, A., Morii, H., Kumasaka, T., Yamamoto, M., Ishii, S., and Ogata, K. (2002) Mechanism of c-Myb-C/EBPβ cooperation from separated sites on a promoter. *Cell* **108,** 57–70

33. Ogata, K., Morikawa, S., Nakamura, H., Hojo, H., Yoshimura, S., Zhang, R., Aimoto, S., Ametani, Y., Hirata, Z., and Sarai, A. (1995) Comparison of the free and DNA-complexed forms of the DNA-binding domain from c-Myb. *Nat. Struct. Biol.* **2,** 309–320

34. Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, A., Inoue, T., Kanai, H., Sarai, A., Ishii, S., and Nishimura, Y. (1994) Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices. *Cell* **79,** 639–648

35. Pavletich, N. P., and Pabo, C. O. (1991) Zinc finger-DNA recognition. Crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* **252,** 809–817

36. Pavletich, N. P., and Pabo, C. O. (1993) Crystal structure of a five-finger GLI-DNA complex. New perspectives on zinc fingers. *Science* **261,** 1701–1707

37. Hardin, S. B., Ortler, C. J., McNamara-Schroeder, K. J., and Stumph, W. E. (2000) Similarities and differences in the conformation of protein-DNA complexes at the U1 and U6 snRNA gene promoters. *Nucleic Acids Res.* **28,** 2771–2778

38. Kiefer, F., Arnold, K., Künzli, M., Bordoli, L., and Schwede, T. (2009) The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res.* **37,** D387–D392

CHAPTER 3

The Largest Subunit of SNAPc Adopts Different Architectural Arrangements on U1 and U6 snRNA Gene Promoters in Response to DNA Sequence Differences in U1 and U6 Promoter Elements

ABSTRACT

The small nuclear RNA activating protein complex (SNAPc) is a multi-subunit transcription factor that is essential for transcription of snRNAs. In *Drosophila melanogaster*, three subunits (DmSNAP190, DmSNAP50, and DmSNAP43) form a stable complex that are required for sequence specific interaction with an snRNA gene promoter element called the PSEA. Only DmSNAP190 possesses a canonical DNA-binding domain that consists of 4.5 Myb repeats (Rh, Ra, Rb, Rc and Rd) that are homologous to the three tandem Myb repeats found in the DNA binding domain of the Myb oncoprotein. Previous work localized domains of DmSNAP190 on the U1 PSEA via site-specific protein-DNA photo-cross-linking followed by subsequent site-specific protein cleavage. In this study, we localize domains of DmSNAP190 on the U6 PSEA. The results indicate that the structural conformation of the DmSNAP190 Myb repeats on the DNA are significantly different when DmSNAPc binds to U1 and U6 snRNA gene promoters.

INTRODUCTION

Most of the spliceosomal small nuclear RNAs (snRNAs) are synthesized by RNA polymerase II (Pol II), but U6 snRNA is synthesized by RNA polymerase III (Pol III) (Dahlberg and Lund, 1988; Egloff et al., 2008; Henry et al., 1998; Hernandez, 2001; Hung and Stumph, 2011; Jawdekar and Henry, 2008; Jensen et al., 1998; Parry et al., 1989a; Su et al., 1997a). Despite this difference in RNA polymerase specificity, transcription of both classes of genes relies upon a unique DNA binding transcription factor called the small nuclear RNA activating protein complex (SNAPc) (Henry et al., 1995; Sadowski et al., 1993) or (PSE)-binding transcription factor (PTF) (Murphy et al., 1992; Yoon et al., 1995). In the fruit fly *Drosophila melanogaster*, DmSNAPc consists of three subunits known as DmSNAP190, DmSNAP50, and DmSNAP43 (Hung and Stumph, 2011; Li et al., 2004; Wang and Stumph, 1998). These three subunits are conserved throughout metazoan evolution, and homologous proteins are found even in the early-diverging trypanosomes (Das and Bellofatto, 2003; Das et al., 2005; Schimanski et al., 2005a, b). DmSNAPc recognizes a ~21 base pair (bp) promoter element termed the proximal sequence element A (PSEA) located approximately 40 to 60 bp upstream of the transcription start site of both the Pol II- and Pol III-transcribed genes (Hernandez et al., 2007; Hung and Stumph, 2011; Jensen et al., 1998).

The PSEAs of the two best-studied fly snRNA genes (U1:95Ca and U6:96Ab) differ at only five nucleotide positions (out of 21), yet these two PSEAs are not interchangeable. The sequence of the U1 PSEA is unable to promote transcription by Pol III, whereas the U6 PSEA cannot promote transcription by Pol II, even though both

sequences are bound by DmSNAPc in vitro and in vivo and function effectively in the context of their own native genes. Transcription assays in vitro indicated that the specific nucleotide sequence of the PSEA was the dominant element for determining polymerase specificity in flies, as a switch in RNA polymerase specificity occurred when the U1 and U6 PSEAs were exchanged (Jensen et al., 1998; McNamara-Schroeder et al., 2001). Specifically, the bases at nucleotide positions 19 and 20 within the PSEA appeared to be the most important for determining RNA polymerase specificity in an in vitro transcription assay (Jensen et al., 1998). In cellular transfection assays, switching the PSEAs resulted in a complete loss of gene expression (Lai et al., 2005; McNamara-Schroeder et al., 2001).

Evolutionary evidence also supports the concept that specific nucleotide differences within the PSEA are important for determining RNA polymerase specificity. When the *D. melanogaster* PSEAs of twenty-three Pol II snRNA genes and seven Pol III snRNA genes were compared, all twenty-three Pol II genes had a "G" or an "A" at position 19, whereas all seven Pol III genes had a "T" at this position. Similarly, a "G" was conserved at position 20 in the Pol II PSEAs, but a "C" was conserved at that position in the Pol III PSEAs (Hung and Stumph, 2011). Furthermore, in a bioinformatic study of Pol II and Pol III PSEAs from five other insect species, certain nucleotide positions in the 3' half of each species' PSEAs were "conserved-to-be-different" between the Pol II- and Pol III-transcribed genes, although the specific positions where these differences occurred varied among the different species (Hernandez et al., 2007).

Site-specific protein-DNA photo-cross-linking assays furthermore have revealed that all three subunits of DmSNAPc contact the DNA. In general terms, DmSNAP190

(the largest subunit) contacts the entire length of the PSEA whereas the two smaller subunits contact primarily the downstream half of the PSEA (Li et al., 2004; Wang and Stumph, 1998). Interestingly, the protein-DNA cross-linking studies also revealed that each subunit contacted the DNA differently depending upon whether DmSNAPc was bound to a U1 PSEA or to a U6 PSEA (Li et al., 2004; Wang and Stumph, 1998). This led us to propose a working model in which DmSNAPc adopts different conformations on U1 and U6 PSEAs (Figure 1A). We hypothesize these conformational differences subsequently lead to recruitment of different general transcription factors and distinct RNA polymerases.

Figure 1B shows the actual patterns of DmSNAP190 cross-linking to U1 or U6 PSEA-containing DNAs. The yellow spheres indicate phosphate positions that cross-linked to DmSNAP190 when DmSNAPc was bound to a U1 PSEA (left duplex) or to a U6 PSEA (right duplex). In both cases, DmSNAP190 cross-linked over the entire length of the PSEA. Although the cross-linking patterns exhibited many similarities, in each case there were several specific phosphate positions where DmSNAP190 cross-linked to the U1 PSEA only or to the U6 PSEA only. The differences in the cross-linking patterns of the two smaller subunits of DmSNAPc on the U1 and the U6 PSEAs were even more pronounced than observed with DmSNAP190 (not shown) (Li et al., 2004; Wang and Stumph, 1998).

In recent work we developed novel methodology in which we combined the site-specific protein-DNA photo-cross-linking assay with subsequent site-specific digestion of the protein (Kim et al., 2010a). This allowed us to map sub-regions or domains of DmSNAP50 and DmSNAP43 that cross-linked to specific phosphate positions in U1 or

U6 promoter DNA (Kim et al., 2010a). A similar analysis was subsequently carried out that identified domains of DmSNAP190 that cross-linked to each of the thirteen phosphate positions in the U1 PSEA (Figure 1) (Doherty et al., 2012; Kim et al., 2010a, b). However, similar work has not yet been reported for DmSNAP190 on a U6 PSEA.

The domain structure of DmSNAP190 is shown in Figure 1C. The most obvious feature of metazoan SNAP190 is that it contains 4.5 tandem Myb repeats, termed Rh, Ra, Rb, Rc, and Rd (Li et al., 2004; Wong et al., 1998). Myb repeats, each approximately 50 amino acids residues in length, were first characterized in the Myb oncoprotein. The Myb protein itself contains three tandem Myb repeats that form its DNA binding domain (Ogata et al., 2004; Tahirov et al., 2002; Tanikawa et al., 1993). To our knowledge, SNAP190 is the only protein that contains more than three Myb repeats. The N-terminal and C-terminal domains of DmSNAP190 that flank the Myb domain are not known to share significant homology with any other proteins.
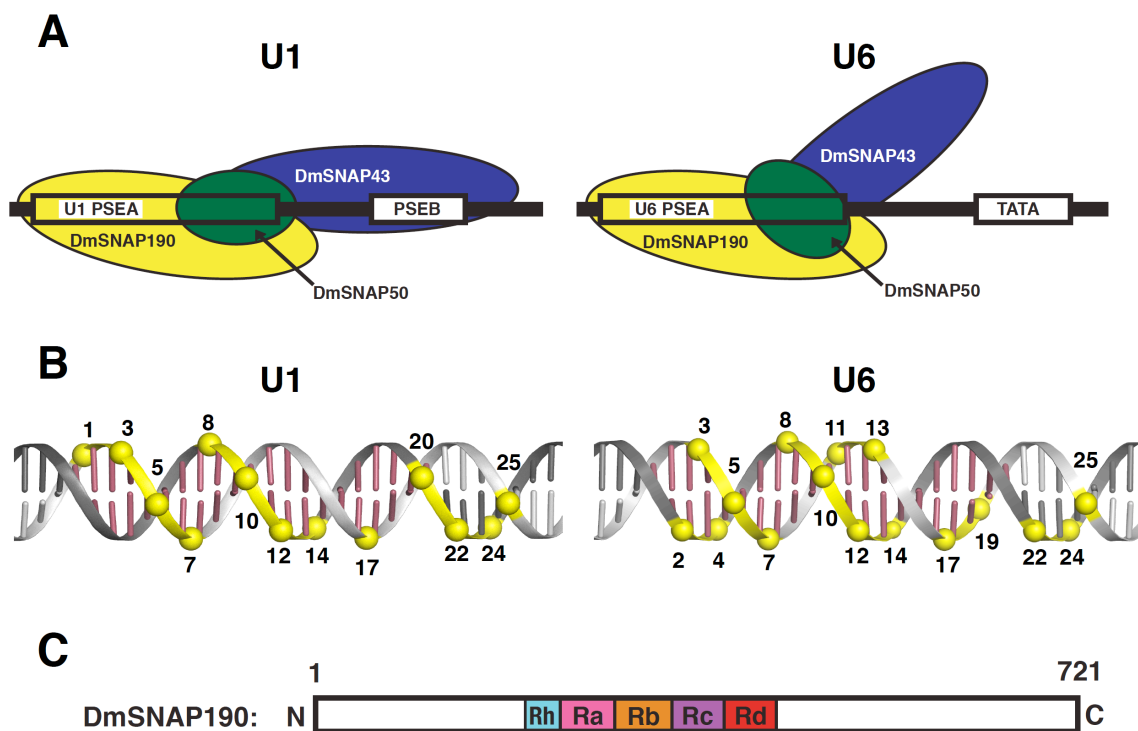
FIGURE 1. **Summary of previous results regarding DmSNAPc binding to U1 and U6 promoters** (A) Schematic drawing of DmSNAPc bound in different conformations to U1 and U6 snRNA gene promoters. DmSNAPc subunits are represented by yellow, green, and blue ovals (DmSNAP190, DmSNAP50, and DmSNAP43 respectively). (B) Contact points between the DmSNAP190 subunit and the U1 PSEA or U6 PSEA based upon site-specific protein-DNA photo-cross-linking (Wang and Stumph, 1998). Colored bases indicate the location of the 21 base pair U1 or U6 PSEA. The yellow spheres indicate phosphate positions that specifically cross-linked to DmSNAP190 when DmSNAPc was bound to DNA. (Only every second phosphate position was assayed on each strand. Odd numbered and even numbered phosphates are on the non-template and template strands respectively.) (C) A schematic diagram of DmSNAP190 protein, which consists of 721 amino acid residues and 4.5 Myb repeats designated Rh, Ra, Rb, Rc, and Rd.

Here we report the results of site-specific protein-DNA photo-cross-linking coupled with site-specific protein digestion to localize domains of DmSNAP190 on the U6 PSEA. Overall, we found that the general orientation of DmSNAP190 on U1 and U6 promoter sequences is similar. However, we also found that there is a significant shift in

the positions of at least two of the Myb repeats such that minor groove-spanning interactions take place on the U6 PSEA that are absent on the U1 PSEA. With these data, we are now able for the first time to draw a comprehensive picture of the conformational differences of the DmSNAP complex on U1 and U6 promoter sequences. Indeed, DmSNAPc appears to form a more "closed" structure on the U6 PSEA compared to the U1 PSEA, suggesting that distinct faces of the DmSNAPc subunits may be exposed (or occluded) on U1 and U6 promoters.

MATERIAL AND METHODS

DmSNAPc constructs, expression, and purification

Untagged and N- or C-terminally FLAG-tagged constructs encoding DmSNAPc subunits have been previously described (Doherty et al., 2012; Hung et al., 2009; Kim et al., 2010a). Various DmSNAP190 constructs with single hydroxylamine cleavage sites (asparaginyl-glycyl (NG) peptide bonds) have also been previously described (Doherty et al., 2012; Kim et al., 2010a). Using the same methods previously described, two new DmSNAP190 constructs were prepared for this work each with a single NG peptide bond at residues 445-446 and 463-464 respectively.

DmSNAPc variants were over-expressed in Drosophila S2 cells stably co-transfected with DmSNAP190, DmSNAP50, and DmSNAP43 constructs each driven by the metallothionein promoter (Doherty et al., 2012; Hung et al., 2009; Kim et al., 2010a). DmSNAP190 was FLAG-tagged at its N-terminus except when untagged DmSNAP190 was desired. In the latter case, N-terminally FLAG-tagged DmSNAP43 was co-

transfected. FLAG immunoaffinity chromatography was used as previously described as a purification step for the DmSNAP complexes (Doherty et al., 2012; Hung et al., 2009; Kim et al., 2010a). The DNA binding activity of each variant DmSNAPc was confirmed by electrophoretic mobility shift assays.

Site-specific protein-DNA photo-cross-linking, protein hydroxylamine digestion, and detection of cleaved fragments

Double-stranded DNA probes that each contained a photo-cross-linking agent (azidophenacyl group) located at a specific individual phosphate position within or downstream of the U6 PSEA were prepared exactly as described in detail previously for U1 probes (Doherty et al., 2012; Kim et al., 2010a). The azidophenacyl group was attached through a phosphorothioate, and each probe contained a ^{32}P radiolabel at the second phosphate position 5' of the cross-linker on the same DNA strand. Each double-stranded probe consisted of the following non-template strand DNA sequence, and its complement, with the cross-linker at one specific designated position in either the non-template or template DNA strand: 5'-GCTATGACCATGATTACGAATTCATTCTTA **TAATTC**T**CAACTG**C**TC**TTT**CC**CGGTACCGCCATGGAAAGGTATGGGATC-3'. The bold nucleotides indicate the sequence of the 21 bp U6 PSEA from the U6:96Ab gene. The underlined nucleotides indicate the five positions where the U6 PSEA bases differ from the PSEA bases of the U1:95Ca gene used in previous photo-cross-linking studies (Doherty et al., 2012). Outside of those five positions, the sequence of the U6 and U1 photo-cross-linking probes were identical to ensure that the U1/U6 differences in

cross-linking were due solely to the five nucleotide differences within the PSEA. Protein-DNA binding, photo-cross-linking, and hydroxylamine digestion of the cross-linked proteins were carried out exactly as previously described (Doherty et al., 2012; Kim et al., 2010a). Digested, radiolabeled protein fragments were identified by denaturing gel electrophoresis and autoradiography.

Immunoblots of hydroxylamine-digested DmSNAPc

Detection of N-terminal and C-terminal fragments of DmSNAP190 variants by immunoblotting was carried out exactly as previously described (Doherty et al., 2012; Kim et al., 2010a). Anti-FLAG monoclonal antibody was used to detect N-terminal fragments specifically, and antibodies directed against a 14 amino acid peptide from the C terminus of DmSNAP190 were used to detect C-terminal fragments.

RESULTS

Identification of DmSNAP190 Myb Domain Repeats that cross-link to U6 PSEA positions 3, 5, and 7 of the non-template DNA strand

As illustrated in Figure 1B, DmSNAP190 can be cross-linked to phosphate positions 3, 5, and 7 of the non-template stand of the U6 PSEA. To localize the region of DmSNAP190 that cross-linked to each of these three positions, we performed site-specific protein-DNA photo-cross-linking by using FLAG-immunoaffinity-purified DmSNAPc and U6 DNA probes that contained a cross-linker at U6 PSEA positions 3, 5,
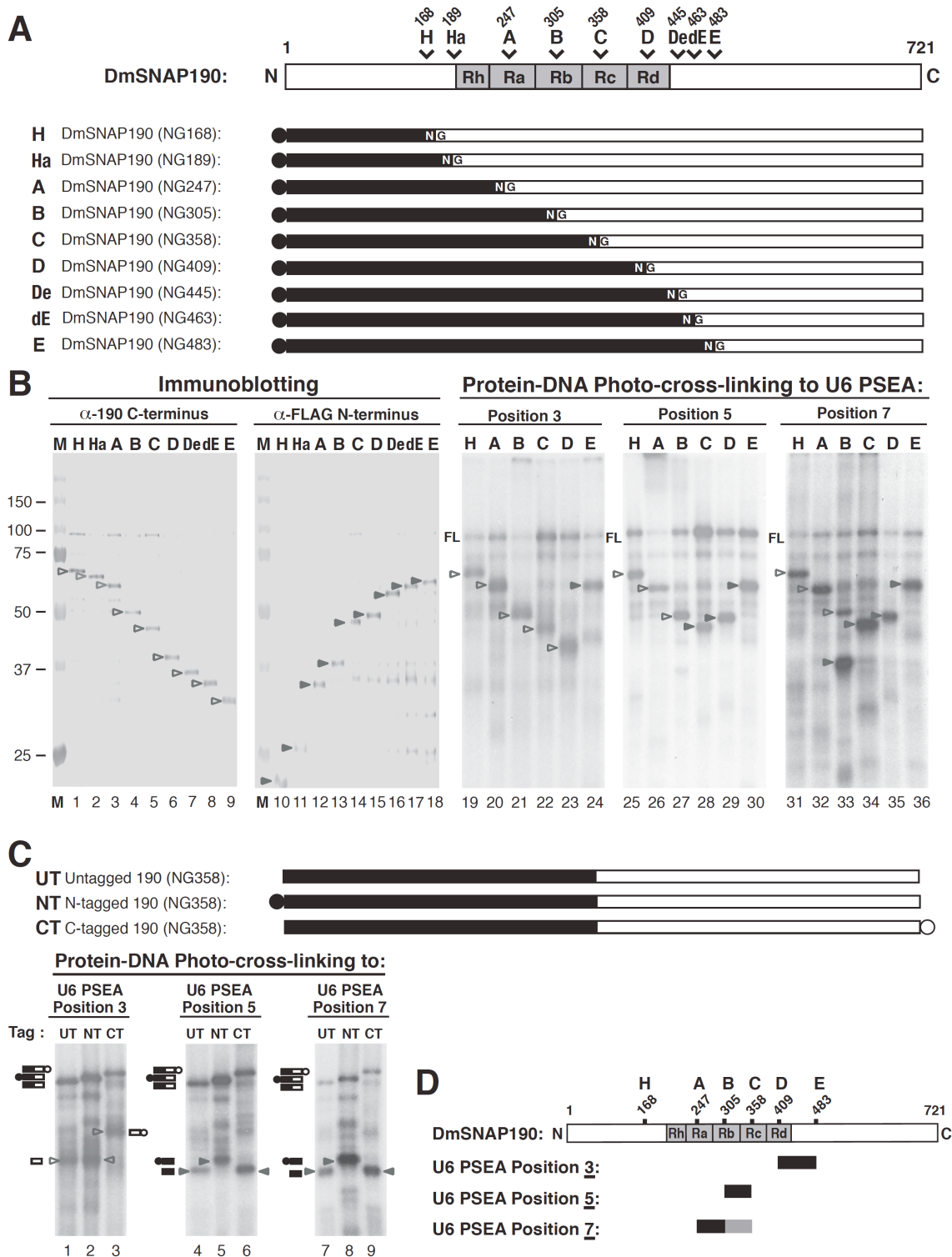
or 7. We then carried out hydroxylamine cleavage of the protein and then detected and identified cross-linked protein fragments by gel electrophoresis and autoradiography.

For different portions of this work, nine different DmSNAP190 constructs were utilized that each contained a single NG peptide bond that could be cleaved with hydroxylamine. The schematic diagram at the top of Figure 2A shows the position of the NG site within each protein construct. The locations of the 4.5 Myb repeats Rh, Ra, Rb, Rc, and Rd, are also indicated. Each of the nine constructs contained a FLAG tag at its N terminus. The N- and C-terminal fragments expected after hydroxylamine digestion of each construct are represented in the lower part of Figure 2A by the darkly shaded and the unshaded areas respectively.

Western blots of these protein constructs following hydroxylamine digestion are shown in the first two panels of Figure 2B (lanes 1-18). Lanes 1-9 show the results of detection with antibodies specific to the C terminus. Fragments of decreasing size were observed that corresponded to the unshaded fragments shown in Figure 2A. Lanes 10-18 show the N-terminal fragments of increasing size (detected using anti-FLAG antibodies) that corresponded to the shaded fragments in Figure 2A.

The third panel of Figure 2B (lanes 19-24) shows an autoradiogram of ^{32}P-labeled fragments after photo-cross-linking and hydroxylamine digestion of DmSNAP190 after site-specific cross-linking to position 3 of the U6 PSEA. The pattern of decreasing-size fragments in lanes 19-23 is consistent with the cross-linking of phosphate position 3 to the C-terminal DmSNAP190 fragment from each of the constructs H through D. In contrast, phosphate position 3 cross-linked to the N-terminal fragment of construct E

FIGURE 2. **Mapping DmSNAP190 Myb repeat domains that cross-link to U6 PSEA positions 3, 5, and 7 of the non-template strand.** (**A**) The top bar represents a schematic diagram of DmSNAP190 and shows the position of nine different hydroxylamine cleavage sites present individually in the nine different constructs represented below (constructs H through E). The darkly shaded and unshaded regions represent the N-terminal and C-terminal fragments expected to be produced from each construct by hydroxylamine digestion. (**B**) All gel results shown are subsequent to hydroxylamine digestion of DmSNAPc from cell lines separately expressing the DmSNAP190 variants represented by constructs H to E shown in *A* above. The first two panels show immunoblots using antibodies against a synthetic C-terminal peptide of DmSNAP190 (lanes 1–9) or antibodies against the FLAG epitope at the N terminus of each construct (lanes 10–18). Lanes marked M contain molecular weight markers. Arrowheads (unshaded for C-terminal fragments and shaded for N-terminal fragments) point out the specific cleavage products of interest. The three panels to the right show autoradiography results of site-specific protein-DNA photo-cross-linking of the DmSNAP190 constructs H, A, B, C, D, and E to U6 PSEA position 3 (lanes 19–24), position 5 (lanes 25–30), or position 7 (lanes 31–36). The label *FL* indicates the position of bands corresponding to full-length undigested N-tagged DmSNAP190. Arrowheads point out the bands corresponding to cross-linked C-terminal (unshaded) or N-terminal (shaded) fragments. In lane 33, both the N- and C-terminal fragments of construct B cross-linked to position 7. (**C**) The bars in the upper part of the figure represent three DmSNAP190 constructs (UT, NT, and CT) that are untagged, N-tagged, and C-tagged, respectively, and each contains a single NG cleavage site at position 358. Construct NT and construct C (from panel A) are actually identical but are labeled differently in different parts of the figure for clarity. The three panels below show the results of protein-DNA photo-cross-linking of the UT, NT, and CT constructs to U6 PSEA position 3 (lanes 1–3), position 5 (lanes 4–6), or position 7 (lanes 7–9). Unshaded arrowheads indicate bands representing C-terminal fragments after hydroxylamine digestion, and shaded arrowheads point out N-terminal fragments. The symbols alongside the gel schematically indicate the origin of each band and correspond to the diagrams of the constructs above. (**D**) Schematic diagrams representing the region of DmSNAP190 that cross-linked to U6 PSEA positions 3, 5, or 7. Dark shading represents stronger cross-linking and lighter shading represents weaker cross-linking.

A DmSNAP190 domain structure with truncation constructs H, Ha, A, B, C, D, De, dE, E

B Immunoblotting (α-190 C-terminus, α-FLAG N-terminus) and Protein-DNA Photo-cross-linking to U6 PSEA at Position 3, Position 5, and Position 7

C UT Untagged 190 (NG358), NT N-tagged 190 (NG358), CT C-tagged 190 (NG358); Protein-DNA Photo-cross-linking to U6 PSEA Position 3, Position 5, Position 7

D DmSNAP190 with U6 PSEA Position 3, Position 5, Position 7 cross-linking regions

(lane 24). These results, taken together, indicate that phosphate position 3 cross-linked to DmSNAP190 C-terminal of cut site D but N-terminal of cut site E. We therefore can conclude that a domain of DmSNAP190 located between amino acid residues 410 and 483 cross-linked to phosphate position 3 of the U6 PSEA.

Photo-cross-linking of the same six DmSNAPc constructs to position 5 of the U6 PSEA is shown in lanes 25-30 of Figure 2B. DmSNAP190 C-terminal fragment cross-linking is evident in lanes 25-27, and N-terminal fragment cross-linking is clearly seen in lanes 29 and 30. As previously observed (Doherty et al., 2012; Kim et al., 2010a), however, the band in lane 28 is ambiguous because the N- and C-terminal fragments of construct C migrate to the same position on the gel (lanes 5 and 14). We therefore performed an experiment like ones described previously in earlier work to resolve this ambiguity (Doherty et al., 2012; Kim et al., 2010a). We carried out photo-cross-linking with DmSNAPc that contained one of three different DmSNAP190 constructs: either untagged DmSNAP190, N-terminally tagged DmSNAP190, or C-terminally tagged DmSNAP190, each with the hydroxylamine cleavage site at position 358. The tags specifically increased the lengths of the N-terminal tagged fragment and the C-terminal tagged fragment by ~2.3 kDa and ~5.7 kDa respectively relative to the untagged fragments. Photo-cross-linking reactions to phosphate position 5 with these three constructs were then run side-by-side on the same gel (Fig. 2C, lanes 4-6). The cross-linked radiolabelled fragment ran more slowly when the construct was N-terminally tagged than when the DmSNAP190 construct was untagged or C-terminally tagged. This result allowed us to conclude that, when protein cleavage was at residue 358, the N-terminal fragment of DmSNAP190 cross-linked to phosphate position 5. The pattern of

the converse result (e.g., if cross-linking had occurred to the C-terminal fragment) is represented in Figure 2C, lanes 1-3, which exemplifies the C-terminal cross-linking pattern expected for phosphate position 3. Taken together, the results indicated that amino acids of DmSNAP190 between residues 306 and 358 cross-linked to phosphate position 5 of the U6 PSEA.

Results of experiments with the cross-linker at phosphate position 7 of the U6 PSEA are shown in Figure 2B, lanes 31-36. Only the C-terminal fragment cross-linked when DmSNAP190 was cleaved following either residue 168 or 247 (lanes 31 and 32). However, when cleavage was at residue 305 (construct B), both the N-terminal and C-terminal fragments were found to cross-link, although the N-terminal fragment cross-linked more intensely (Figure 2B, lane 33). This means that phosphate position 7 of the U6 PSEA was able to cross-link to amino acid residues of DmSNAP190 both N-terminal and C-terminal of residue 305. When cleavage occurred at positions 409 or 483 (lanes 35 and 36), only the N-terminal fragments cross-linked to phosphate position 7.

Due to the ambiguity associated with cleavage of the protein at position 358, the data shown in Figure 2B, lanes 31-36, could not formally exclude the possibility that phosphate position 7 cross-linked C-terminal of position 358 (although N-terminal of position 409). However, the results shown in Figure 2C, lanes 7-9, indicate that cross-linking of phosphate position 7 occurred only to the N-terminal fragment when DmSNAP190 was cleaved at position 358.

By taking all of these data into account, we conclude that phosphate position 7 of the U6 PSEA cross-linked strongly to residues of DmSNAP190 localized between 248 and 305 and cross-linked with less intensity to residues between 306 and 358. These

results, together with those from phosphate positions 3 and 5, are summarized in Figure 2D. These results in fact are identical to those obtained in earlier work with phosphate positions 3, 5, and 7 of the U1 PSEA (Doherty et al., 2012). This reveals that the same domains of DmSNAP190 are in close proximity to phosphate positions 3, 5, and 7 when DmSNAPc binds to either the U1 or U6 PSEA.

Identification of DmSNAP190 Myb Domain Repeats that cross-link to U6 PSEA positions 8, 10, and 12 of the template DNA strand

To further examine the DmSNAPc DNA-binding pattern to the U6 PSEA, we next investigated the cross-linking of DmSNAP190 to phosphate position 8 of the template strand of the U6 DNA. The pattern of cross-linking (Figure 3, lanes 1-9) revealed that position 8 cross-linked C-terminal to residue 306 (construct B, lane 3) but N-terminal to residue 409 (construct D, lane 5). Results shown in lanes 7-9 furthermore indicated that the cross-linking was localized to a region of DmSNAP190 C-terminal of residue 358. We therefore conclude that U6 PSEA position 8 cross-linked to amino acid residues of DmSNAP190 localized between positions 359 and 409.

We next examined the cross-linking pattern of phosphate position 10 of the U6 PSEA to DmSNAP190. The results shown in Figure 3 lanes 10-15 were remarkably similar to those seen for position 8 in lanes 1-6. However, the results in lanes 16-18 were the converse of those observed in lanes 7-9. That is, position 10 cross-linked to the N-terminal fragment when protein cleavage was at position 358. Thus, phosphate position

10 cross-linked to amino acid residues of DmSNAP190 located between positions 306 and 358.

The cross-linking pattern of U6 PSEA phosphate position 12 to DmSNAP190 is shown in Figure 3 lanes 19-27. These results were identical to those obtained for position
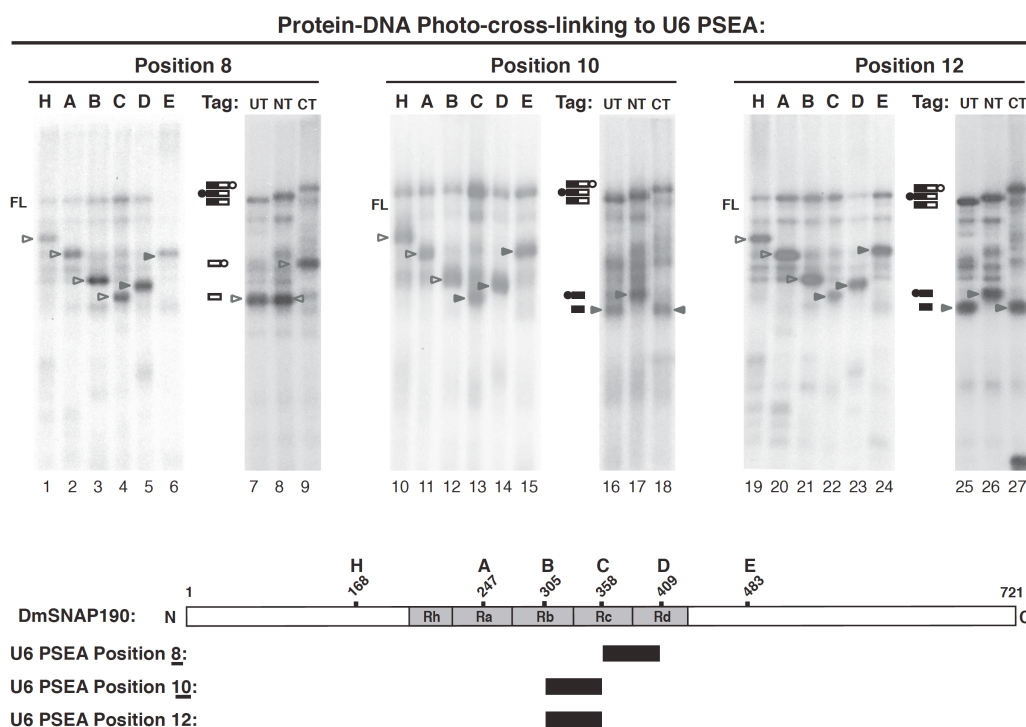

FIGURE 3. **Mapping DmSNAP190 Myb repeat domains that cross-link to U6 PSEA positions 8, 10, and 12 of the template strand.** Lanes 1-9, lanes 10-18, and lanes 19-24 show autoradiography results of photo-cross-linking DmSNAP190 NG variants to U6 PSEA phosphate positions 8, 10, and 12, respectively. Lanes and bands are labeled as indicated in the legend of Figure 2. Unshaded arrowheads point at bands corresponding to cross-linked C-terminal fragments, and shaded arrowheads point out cross-linked N-terminal fragments. The diagram at the bottom of the figure indicates the region of DmSNAP190 that cross-linked to each of positions 8, 10, and 12 of the U6 PSEA.

10 for all of the DmSNAP190 constructs. Thus, position 12 of the U6 PSEA, like position 10, cross-linked to a region of DmSNAP190 located between amino acid residues 306 and 358. The cross-linking results for U6 PSEA positions 8, 10, and 12 are summarized in the schematic diagram at the bottom of Figure 3. Notably, these are the same regions

of DmSNAP190 that cross-linked to phosphate positions 8, 10, and 12 of the U1 PSEA (Doherty et al., 2012; Kim et al., 2010a). Together with the results discussed in the preceding section, this suggests that phosphate positions 3, 5, 7, 8, 10, and 12 contact similar domains of DmSNAP190 on both the U1 and U6 PSEAs.

Identification of DmSNAP190 domains that cross-link to U6 PSEA positions 2 and 4 of the template DNA strand

When DmSNAPc binds to a U6 PSEA, DmSNAP190 can be cross-linked to phosphate positions 2 and 4; those cross-links do not occur when DmSNAPc binds to a U1 PSEA (Wang and Stumph, 1998). Experiments to localize domains of DmSNAP190 that cross-link to these positions in a U6 PSEA are shown in Figure 4. The cross-linking pattern shown in lanes 1- 6 revealed that cross-linking occurred C-terminal of amino acid residue 247 (construct A, lane 2). When DmSNAP190 was cleaved at position 305 (construct B, lane 3), the strongest cross-linking was to the N-terminal fragment. Consistent with this, strong bands corresponding to N-terminal fragments were also observed in lanes 4-6, indicating that strong cross-linking occurred between residues 248 and 305.

Interestingly, a less intense band consistent with cross-linking to the fragment C-terminal of residue 305 was also observed in lane 3. Although a C-terminal cross-linked fragment would be obscured in lane 4, bands consistent with C-terminal fragment cross-linking were observed in lanes 5 and 6 (clear arrowheads). These results suggested that the domain of DmSNAP190 C-terminal of residue 483 was cross-linking to position 2 of the U6 PSEA. Consistent with the above data, Figure 4 (lanes 7-9) revealed that both N-
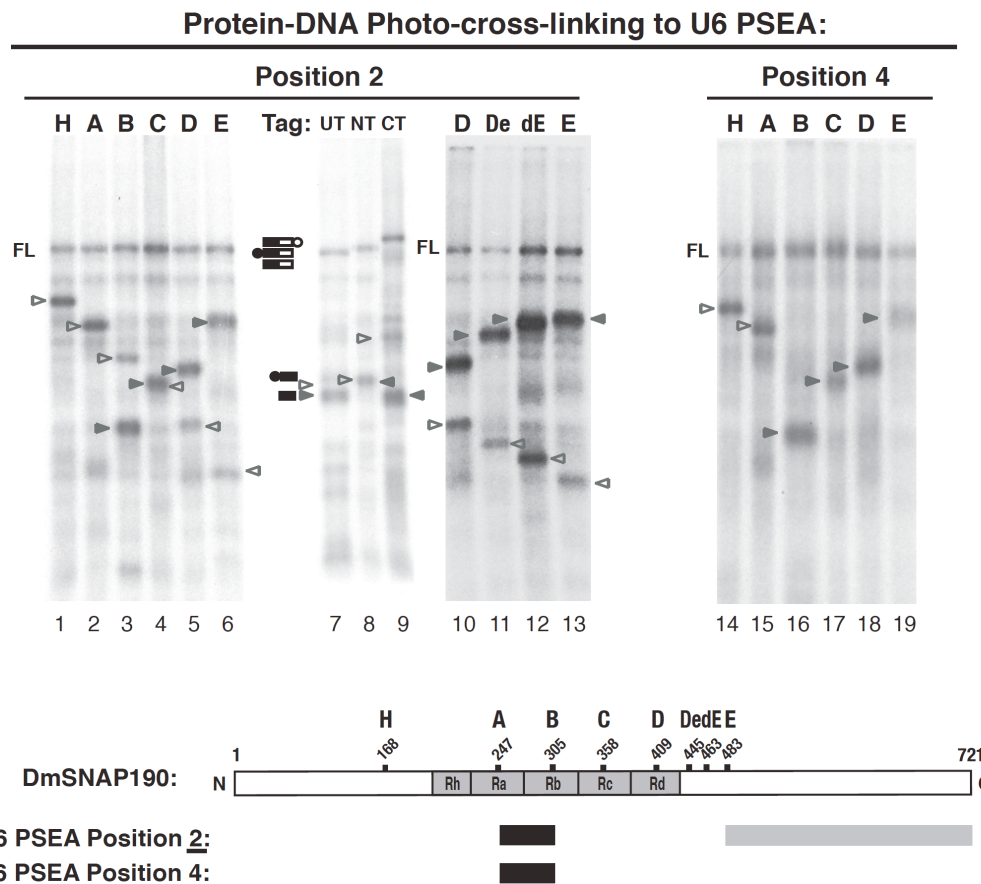
FIGURE 4. **Mapping DmSNAP190 domains that cross-link to U6 PSEA positions 2 and 4 of the template strand.** Lanes 1-13 and lanes 14-19 show autoradiograms of the results of photo-cross-linking DmSNAP190 NG variants to U6 PSEA phosphate positions 2 and 4, respectively. Lanes and bands are labeled as indicated in the legend of Figure 2. In lanes 3-13, both the N- and C-terminal fragments cross-linked to position 2. The diagram at the bottom of the figure indicates the region of DmSNAP190 that cross-linked to each of positions 2 and 4 of the U6 PSEA. Dark shading, stronger cross-linking; lighter shading, weaker cross-linking.

terminal and C-terminal fragments cross-linked to position 2 when the protein was

cleaved at residue 358. To further investigate the authenticity of the C-terminal bands

observed in lanes 1-9, we employed two new DmSNAP190 constructs (De and dE) that

could be cleaved by hydroxylamine following residues 445 or 463. Figure 4 (lanes 10-13)

shows results obtained with these constructs run alongside constructs D and E. In each

case, both the C-terminal fragment as well as the N-terminal fragment were observed to cross-link. These results provide convincing evidence that phosphate position 2 of the U6 PSEA cross-linked not only to a domain of DmSNAPc between residues 248 and 305 but cross-linked also to a domain of DmSNAP190 C-terminal of residue 483.

The pattern of cross-linking of DmSNAP190 to position 4 is shown in Figure 4 lanes 14-19. The pattern is similar to position 2 except there is no evidence of cross-linking to any region C-terminal of residue 305. A summary of the cross-linking results for U6 PSEA phosphate positions 2 and 4 is shown at the bottom of Figure 4. Both of these positions cross-linked to a domain of DmSNAP190 between amino acid residues 247 and 305. In addition, the C-terminal domain of DmSNAP190 (bounded by residues 484-721) was capable of cross-linking to position 2.

Identification of DmSNAP190 Myb domain repeats that cross-link to U6 PSEA positions 11 and 13 of the non-template DNA strand

When DmSNAPc binds to DNA, DmSNAP190 cross-links to phosphate positions 11 and 13 of the U6 PSEA but not to those same positions in a U1 PSEA. To map out the domains of DmSNAP190 that cross-link to those positions of the U6 PSEA, the experiments shown in Figure 5 were carried out. Because positions 11 and 13 cross-linked not only to DmSNAP190 but also to DmSNAP43 and DmSNAP50 respectively (Wang and Stumph, 1998), additional bands corresponding to those two smaller subunits appear on the gel. From previous work and knowledge of the primary structure of those subunits (Doherty et al., 2012; Kim et al., 2010a), it is possible to identify the

DmSNAP43 and DmSNAP50 bands on the gel and, except when overlapping bands occur, avoid confusion with the DmSNAP190 data.
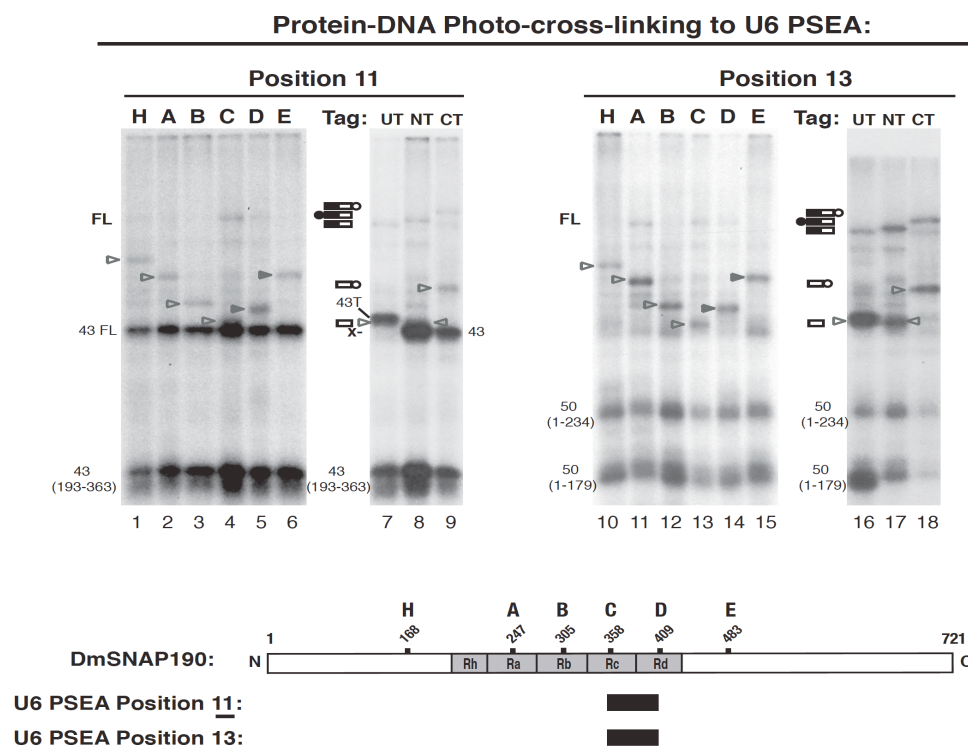

FIGURE 5. **Mapping DmSNAP190 Myb repeats domains that cross-link to U6 PSEA positions 11 and 13 of the non-template strand.** Lanes 1-9 and lanes 10-18 show autoradiograms of the results of photo-cross-linking DmSNAP190 NG variants to U6 PSEA phosphate positions 11 and 13, respectively. Lanes and bands are labeled as indicated in the legend of Figure 2. Unshaded arrowheads point out the bands corresponding to cross-linked C-terminal fragments and shaded arrowheads point to the bands corresponding to N-terminal fragments. Positions 11 and 13 cross-link not only to DmSNAP190 but also DmSNAP50 and DmSNAP43, respectively (Wang and Stumph 1998; Kim et al, 2010). 43FL indicates the band that arises from cross-linking to full-length untagged DmSNAP43. In lane 7, 43T points out a band that arises from cross-linking to N-terminally tagged full-length DmSNAP43 that was incorporated into the DmSNAPc used in this lane as a means of purification when DmSNAP190 was not tagged. The designations 43(193-363), 50(1-234), and 50(1-179) indicate the positions of cross-linked DmSNAP43 and DmSNAP50 hydroxylamine digestion products with the numbers in parentheses indicating the amino acid positions at the beginning and end of those fragments. The diagram at the lower part of the figure indicates the region of DmSNAP190 that cross-linked to each of positions 11 and 13 of the U6 PSEA.

When cross-linking was carried out with phosphate position 11 of the U6 PSEA (Figure 5, lanes 1-9), the strongest bands corresponded to cross-links with full length (undigested) DmSNAP43 and to the DmSNAP43 fragment encompassed by residues 193 -363 in accord with previous observations (Doherty et al., 2012; Kim et al., 2010a). Nevertheless, bands were also visible that corresponded to the fragments derived from hydroxylamine digestion of the DmSNAP190 constructs (designated by arrowheads). By the same reasoning previously applied, the results shown in lanes 1-6 of Figure 5 localized the cross-linking of position 11 to the region of DmSNAP190 bounded by residues 306 and 409.

To ascertain whether this cross-linking was to the region of the protein N-terminal or C-terminal of residue 358, the cross-linking experiments shown in lanes 7-9 were performed. Although some of the DmSNAP190 bands were partially obscured by the strong DmSNAP43 bands, the evidence was consistent with cross-linking that occurred C-terminal of position 358. First, in lane 9 there was a band that corresponded to the tagged C-terminal fragment.

Second, there were faint "shoulder" bands where the untagged C-terminal fragment of DmSNAP190 was expected to run just below the tagged DmSNAP43 band in lane 7 and just above the untagged DmSNAP43 band in lane 8. Finally, and importantly, there was no band in lane 7 that would correspond to the mobility of an untagged N-terminal DmSNAP190 fragment (position denoted by "x"). Therefore, we conclude that phosphate position 11 of the U6 PSEA cross-linked to the domain of DmSNAP190 that contains amino acid residues 359-409.

Experiments in which cross-linker was placed at position 13 in the U6 PSEA are shown in Figure 5 lanes 10-18. Despite the presence of DmSNAP50-derived bands in the lower part of the gel, bands corresponding to DmSNAP190 fragments were easily discerned in the upper part of the gel. In each lane (10-18) the DmSNAP190 cross-linking pattern was exactly the same as in lanes 1-9. We therefore conclude that both phosphate positions 11 and 13 of the U6 PSEA cross-linked to the domain of DmSNAP190 bounded by amino acid residues 359-409. These results are summarized schematically at the bottom of Figure 5.

Identification of DmSNAP190 domains that cross-link to positions 17, 19, and 25 of the non-template DNA strand and to positions 14, 22, and 24 of the template DNA strand of the U6 PSEA

Cross-linking data for the remaining six phosphate positions of the U6 PSEA that cross-link to DmSNAP190 are shown in Figure 6.

Position 17. At position 17, DmSNAP50 cross-links much more intensely than DmSNAP190 (Wang and Stumph, 1998). Nevertheless, the bands derived from the DmSNAP190 constructs can be discerned in most lanes and can be surmised when occluded by DmSNAP50 bands. Figure 6 lanes 2-6 established that phosphate position 17 of the U6 PSEA cross-linked to the N-terminal side of residue 247 (construct A). The lack of a high molecular weight band in lane 1 at the position marked by an "x" furthermore suggests that phosphate position 17 cross-linked N-terminal of amino acid residue 168. However, the N-terminal fragment derived from construct H would be

occluded near the bottom of the gel by the very strong band derived from DmSNAP50. To shed further light on the region of DmSNAP190 that cross-linked to position17, an experiment was carried out with the construct Ha (lanes 7-9). In this case, the N-terminal fragment of Ha resolved from the DmSNAP50 bands and revealed that phosphate 17 cross-linked, after hydroxylamine digestion of the protein, to the N-terminal fragment of DmSNAP190 localized between residues 1 and 189 (and most likely between 1 and 168). No cross-linking C-terminal of position 189 was observed.

Position 19. The cross-linking pattern of U6 PSEA phosphate position 19 to DmSNAPc is shown in Figure 6 lanes 10-18. Besides the bands derived from the DmSNAP190 constructs, at least three bands corresponding to full-length and digested fragments of DmSNAP43 were observed that had minimal effect on the DmSNAP190 analysis. The DmSNAP190 fragments formed a pattern revealing that position 19 of the U6 PSEA cross-linked exclusively to the N-terminal side of the cut site in construct A at residue 247 (lane 11). In lane 10, a very weak high molecular weight band corresponding to a cross-link to the C-terminal fragment could be observed. This band indicated that weak cross-linking occurred C-terminal to the hydroxylamine cut site at residue 168 in construct H. However, it was still quite possible that cross-linking could occur also N-terminal of residue 168. To further investigate this possibility, the experiments shown in Figure 6 lanes 16-18 were performed. In this case, stronger cross-linking to the N-terminal fragment and weaker cross-linking to the C-terminal fragment were observed when the hydroxylamine cut site was at position 189 (lane 17). From these results, we conclude that position 19 of the U6 PSEA cross-linked most intensely to the region of

DmSNAP190 located between amino acid residues 1 and 189 and with less intensity to the domain of DmSNAP190 between residues 190 and 247.

Position 25. When the cross-linking reagent was placed at position 25 of the U6 PSEA, the results of the DmSNAP190 cross-linking were essentially identical to those obtained with position 19 (compare the DmSNAP190 bands in lanes 19-27 with those in lanes 10-18). Thus, we likewise conclude that phosphate position 25 cross-linked more strongly to the region of DmSNAP190 encompassed by amino acid residues 1-189 and more weakly to the DmSNAP190 domain between residues 190 and 247.

Position 24. Placement of the cross-linker at position 24 of the U6 PSEA produced a relatively clear cross-linking pattern (Figure 6, lanes 28-36) because the two smaller subunits did not cross-link at this position (Wang and Stumph, 1998). Cross-linking at position 24 occurred to the N-terminal fragment of constructs A through E (lanes 29-33). A high molecular weight band was clearly observed that corresponded to the cross-linked C-terminal fragment of construct H (lane 28), but the band near the bottom of the gel was darker in lane 28 compared to the non-specific band of similar mobility in the remainder of the lanes, suggesting that cross-linking might also have occurred to the N-terminal fragment derived from construct H. To further assess whether the N-terminal domain of DmSNAP190 cross-linked to position 24, the reactions shown in Figure 6, lanes 34-36, that included the Ha construct were carried out. Lane 35 shows clear evidence for the cross-linking of position 24 to the N-terminal fragment of the Ha construct as well as to the C-terminal fragment. From these results, we conclude that position 24 of the U6 PSEA cross-linked with approximately equal intensities to the N-

terminal region of DmSNAP190 between amino acid residues 1 to 189 and to the adjacent region of DmSNAP190 located between residues 190 and 247.

Position 22. Results with the cross-linking reagent at phosphate position 22 of the U6 PSEA were similar to those obtained at position 24, except that the cross-linking intensity C-terminal of residue 189 was very, very weak (lanes 43 and 44). Thus, position 22 of the U6 PSEA cross-linked to the N-terminal domain of DmSNAP190 (residues 1-189) and very weakly to the domain of DmSNAP190 encompassed by residues 190 to 247.

Position 14. Cross-linker placed at position 14 of the U6 PSEA resulted in perhaps the most complicated pattern observed. The results indicated that when DmSNAP190 was cleaved at sites B, C, D, or E, cross-linking was exclusively to the N-terminal fragment, placing the region of DmSNAP190 cross-linking entirely N-terminal of amino acid residue 305 (Figure 6, lanes 48-51). When the hydroxylamine cleavage site was at residue 247, both the N-terminal fragment (more strongly) and the C-terminal fragment (more weakly) cross-linked (lane 47). This indicated that U6 PSEA position 14 cross-linked primarily N-terminal of residue 247, but also cross-linked to the domain of DmSNAP190 located between amino acid residues 248 and 305.

FIGURE 6. **Localization of DmSNAP190 domains that cross-link to positions 17, 19, and 25 of the non-template strand and to positions 14, 22, and 24 of the template strand of the U6 PSEA.** Autoradiograms are shown representing the results of photo-cross-linking various DmSNAP190 constructs to U6 PSEA phosphate positions 17, 19, 25, 24, 22, and 14. Lanes and bands are labeled as indicated in the legend of Fig. 2. Unshaded arrowheads point out the bands corresponding to cross-linked C-terminal fragments and shaded arrowheads points to the cross-linked N-terminal fragments. 190FL indicates the position of the full-length undigested N-tagged DmSNAP190. The labels 43FL and 50FL point out bands that correspond to cross-linked full-length DmSNAP43 and DmSNAP50, respectively. The numbers in parentheses following 50 or 43 [e.g., 50(1-234)] indicate the positions of hydroxylamine digestion fragments of DmSNAP43 or DmSNAP50 bounded by those amino acid residues. The shaded arrowheads in parentheses point to the location of a surmised DmSNAP190 N-terminal fragment derived from construct H (in lanes 1, 7, 10, 16, 19, 25, 37, 43, 46, and 52) that would be cross-linked but is obscured by the strong bands arising from digestion products of DmSNAP50 or DmSNAP43. The diagram at the bottom of the figure indicates the region of DmSNAP190 that cross-linked to each of positions 17, 19, 25, 24, and 14 of the U6 PSEA. Dark shading, stronger cross-linking; lighter shading, weaker cross-linking.

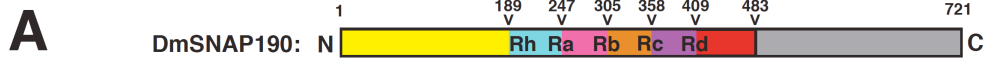## Protein-DNA Photo-cross-linking to U6 PSEA:
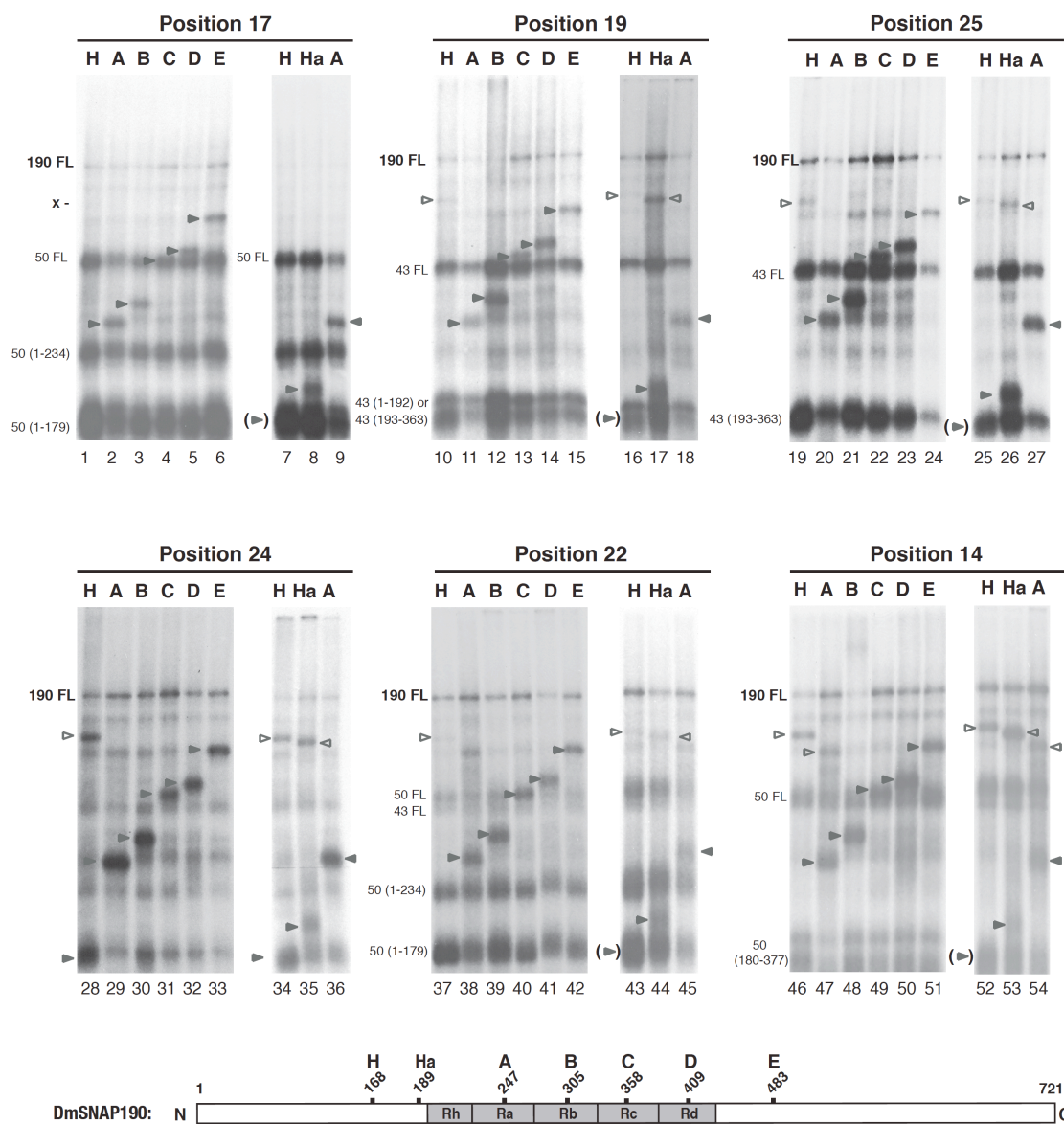
As expected with construct H, cross-linking to the C-terminal fragment was observed (lane 46), but there remained the possibility that cross-linking also occurred to the N-terminal fragment (which would be occluded by the DmSNAP50 fragment near the bottom of the gel). The cross-linking reactions shown in Figure 6 lanes 52-54 shed further light on that situation. When DmSNAPc was cleaved at position 189 (construct Rh, lane 53), stronger cross-linking was observed to the C-terminal fragment, but cross-linking was also observed to the N-terminal fragment. Taking all of the data into consideration, the most consistent interpretation is that position 14 of the U6 PSEA cross-linked most strongly to the domain of DmSNAP190 located between amino acid residues 190 and 247, but that cross-linking also occurred to DmSNAP190 between residues 1 to 189 as well as to residues between 248 and 305. A schematic summary of the results for U6 PSEA positions 17, 19, 25, 24, 22, and 14 is shown at the bottom of Figure 6.

Comparison of the pattern of DmSNAP190 cross-linking to U1 and U6 PSEAs

The results described in Figures 2 through 6 localized and identified domains within DmSNAP190 that cross-linked to all sixteen phosphate positions within and downstream of the U6 PSEA to which DmSNAP190 can be cross-linked when DmSNAPc binds to the U6 PSEA. In previous work, we carried out a similar study on all thirteen positions of a U1 PSEA that cross-linked to DmSNAP190 (Doherty et al., 2012; Kim et al., 2010a, b). Figure 7A is a summary that schematically illustrates the

FIGURE 7. **A summary of the regions of DmSNAP190 that interact with specific nucleotide positions of the U1 or U6 PSEAs.** (**A**) The top bar schematicall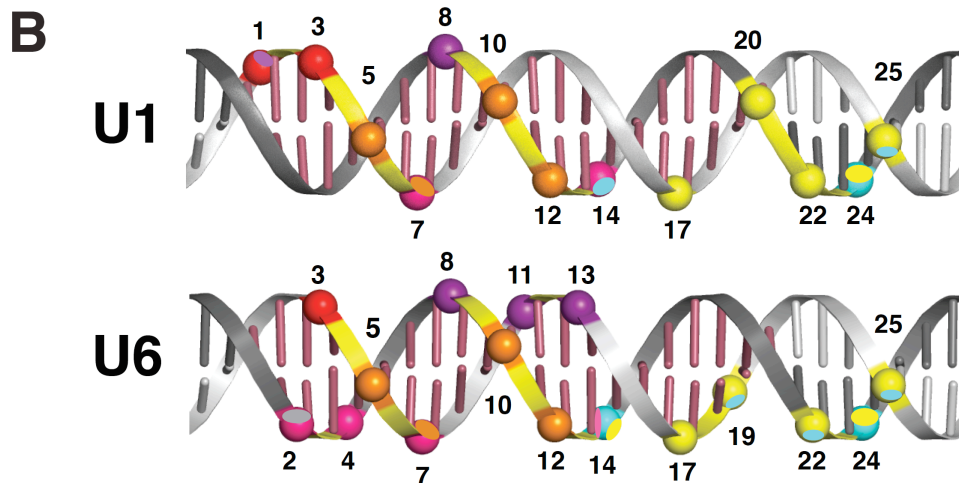y represents DmSNAP190 shown in different colors based upon the hydroxylamine cleavage sites. The central region summarizes the regions of DmSNAP190 that cross-linked to each of the cross-linkable phosphates in either the U1 PSEA (Kim et al, 2010; Doherty et al, 2012) or the U6 PSEA (Figures 2 through 6 of this chapter). Stronger cross-linking is represented by dark shading, and weaker cross-linking is indicated by lighter shading. (**B**) Projection of the cross-linking data on B-form DNA. Each phosphate position is color-coded to match the color of the domain of DmSNAP190 (as indicated in the schematic at the top of the figure) that cross-links to the individual phosphate position. Oval overlays are color-coded to the match the color of the protein regions of weaker cross-linking. Brownish-colored bases represent the bases of the U1 or U6 PSEAs.

A

DmSNAP190 Cross-linking to U1 PSEA

U1 PSEA Position 3:
U1 PSEA Position 1:
U1 PSEA Position 8:
U1 PSEA Position 5:
U1 PSEA Position 10:
U1 PSEA Position 12:
U1 PSEA Position 7:
U1 PSEA Position 14:
U1 PSEA Position 24:
U1 PSEA Position 25:
U1 PSEA Position 22:
U1 PSEA Position 20:
U1 PSEA Position 17:

DmSNAP190 Cross-linking to U6 PSEA

U6 PSEA Position 3:
U6 PSEA Position 8:
U6 PSEA Position 11:
U6 PSEA Position 13:
U6 PSEA Position 5:
U6 PSEA Position 10:
U6 PSEA Position 12:
U6 PSEA Position 2:
U6 PSEA Position 7:
U6 PSEA Position 4:
U6 PSEA Position 14:
U6 PSEA Position 24:
U6 PSEA Position 25:
U6 PSEA Position 22:
U6 PSEA Position 19:
U6 PSEA Position 17:

B

U1

U6

DmSNAP190 domains that cross-link to specific phosphate positions in the U1 or U6 PSEAs. Figure 7B shows these same data projected upon B-form DNA double helices that contain either a U1 or a U6 PSEA. The protein domains and the phosphates on the DNA are color-coded to indicate the mapped region of the DmSNAP190 protein that cross-linked to the similarly-colored phosphate spheres on the DNA. In cases where more than one domain of the protein cross-linked to the DNA, an oval of the color of the more weakly cross-linking domain was overlaid over a portion of the phosphate backbone sphere that was initially colored to match the protein region that most strongly cross-linked.

The general orientation of DmSNAP190 on the U1 and U6 promoters is similar but there are also significant differences. In both cases, the Myb Rh half-repeat and the DmSNAP190 N-terminal domain (shown in yellow) interact with the 3' half of the PSEA (Figure 7). The 3' half of the fly PSEA is less conserved than the 5' half and differs at specific nucleotide positions between the snRNA genes that are transcribed by RNA polymerase II and those transcribed by RNA polymerase III. The 3' half of the PSEA is also that portion of the PSEA that is contacted by the DmSNAP50 and DmSNAP43 subunits (Kim et al., 2010a; Wang and Stumph, 1998).

Perhaps the most striking difference between the U1 and U6 binding patterns is that unique minor groove-spanning contacts occur on the U6 PSEA that are absent when DmSNAPc binds to the U1 PSEA (Figure 7). On a U6 PSEA, a domain of DmSNAP190 (in violetpurple, amino acid residues 359-409) spans the minor groove between phosphate 8 on the template strand and phosphates 11 and 13 on the non-template strand; this suggests that there is a shift of this domain in the downstream direction (toward the

transcription start site) compared to its position on a U1 PSEA. Similarly, a second domain of DmSNAP190 (magenta, residues 248-305) spans the U6 PSEA minor groove between phosphate 7 of the template strand and phosphates 2 and 4 on the non-template strand, suggesting a movement of the corresponding DmSNAP190 domain in the upstream direction compared to its location on the U1 PSEA.

Model of the differential interaction of DmSNAP190 with the U1 and U6 PSEAs

There are no molecular structures available for the components of DmSNAPc or of its orthologs from other species. However, there is a co-crystal structure available for the DNA-binding domain of the c-Myb protein bound to DNA (Protein Data Bank entry 1H88) (Tahirov et al., 2002). Each Myb repeat consists of three alpha helices separated by turns, and each Myb repeat is separated from adjacent Myb repeats by a flexible loop. We previously generated a three-dimensional theoretical model of the DmSNAP190 Myb repeats bound to a U1 PSEA (Doherty et al., 2012) by using a web-based structural alignment tool (SWISS-MODEL repository) (Kiefer et al., 2009) to thread the DmSNAP190 Myb repeat sequences onto the structure of the Myb repeats of the c-Myb protein. We have now performed the same type of structure-building to generate a model of DmSNAP190 on the U6 PSEA. We have also revised the original U1 model to bring it better into accord with the constraints of the U6 model. All protein data files were visualized and manipulated using PyMOL (Schroedinger LLC, New York).

While developing both the U1 and U6 models, we kept all the predicted folds and turns of each of the four Myb repeats Ra, Rb, Rc, and Rd intact; we manipulated only the flexible loops that connect the Myb repeats (Ogata et al., 1995) to generate models consistent with the U1 and U6 cross-linking data. We took an extra liberty only with the Rh half-repeat that contains just the last two of the three conserved helices present in a full Myb repeat. In order permit the Rh region of the protein to stretch over the relatively long length of DNA to which this half-repeat is able to cross-link on both the U1 and U6 PSEAs, it was necessary to "straighten out" the turn that links the two helices. This seems reasonable because the stability of a single Myb repeat structure is dependent upon a hydrophobic core whose formation depends upon contributions by all three helices within a repeat (Ogata et al., 1995).

Figure 8 shows models of the arrangement of the domains of DmSNAP190 on the DNA when DmSNAPc binds either to a U1 PSEA or to a U6 PSEA. Because the primary data were obtained using hydroxylamine cut sites near the middle of each Myb repeat (in a non-conserved area), the models were developed by using PyMol and manually placing the Myb structures (modelled using the SWISS-MODEL repository software) into the proximity of the cross-linked phosphates in accordance with the color code used in Figure 7. For purposes of presentation and visualization in Figure 8, however, the regions of the protein were re-colored to represent each individual Myb repeat as a distinct color.

Work from other investigators demonstrated that the R2 and R3 Myb repeats of c-Myb interact with DNA by inserting a recognition alpha helix into the major groove of the DNA (Ogata et al., 1995; Ogata et al., 2004; Tahirov et al., 2002). We have followed that scheme precisely for the Rb repeat of DmSNAP190 and placed its "recognition

helix" into the major groove at almost precisely the same location both in the U1 PSEA and in the U6 PSEA (Figure 8). Interestingly, this placed the recognition helix of Rb in proximity to base pairs 5, 6, 7, and 8 of the U1 and U6 PSEAs, which are the most stringently conserved base pairs among the PSEAs of all the Pol II- and Pol III-transcribed snRNA genes of both *D. melanogaster* and other insect species that have been analyzed (Hernandez et al., 2007). It is also worthwhile noting that the cross-linking patterns that are most similar between the U1 and U6 PSEAs are the cross-links to the Rb repeat as opposed to the other domains of DmSNAP190 (Figure 7). Thus, interactions between the Rb recognition helix and base pairs in the vicinity of PSEA positions 5 through 8 likely represent the most conserved core of DmSNAP190 interactions with both the U1 and U6 PSEAs.

The model on a U1 PSEA (upper part of Figure 8) illustrates that repeats Ra, Rb, Rc, and Rd follow the contour of the major groove of the DNA between approximately positions 1 and 12 of the U1 PSEA. In contrast, on a U6 PSEA, repeats Rc and Ra adopt a more "vertical" orientation appearing directly "above" and "below" Rb when the complex is viewed from the direction shown in the figure. On the U6 PSEA, Rc spans the minor groove between phosphate 8 and phosphates 11 and 13. Likewise, specific regions of repeats Ra and Rb (primarily the unusually long flexible loop that joins these two repeats) span the minor groove to contact phosphates 2 and 4. Within the constraints of the conformation of the Myb repeats and molecular bond distances, it is not possible to place the Rc or Ra "recognition helices" deep into the major groove and at the same time have these repeats span the minor groove as required by the cross-linking data on the U6
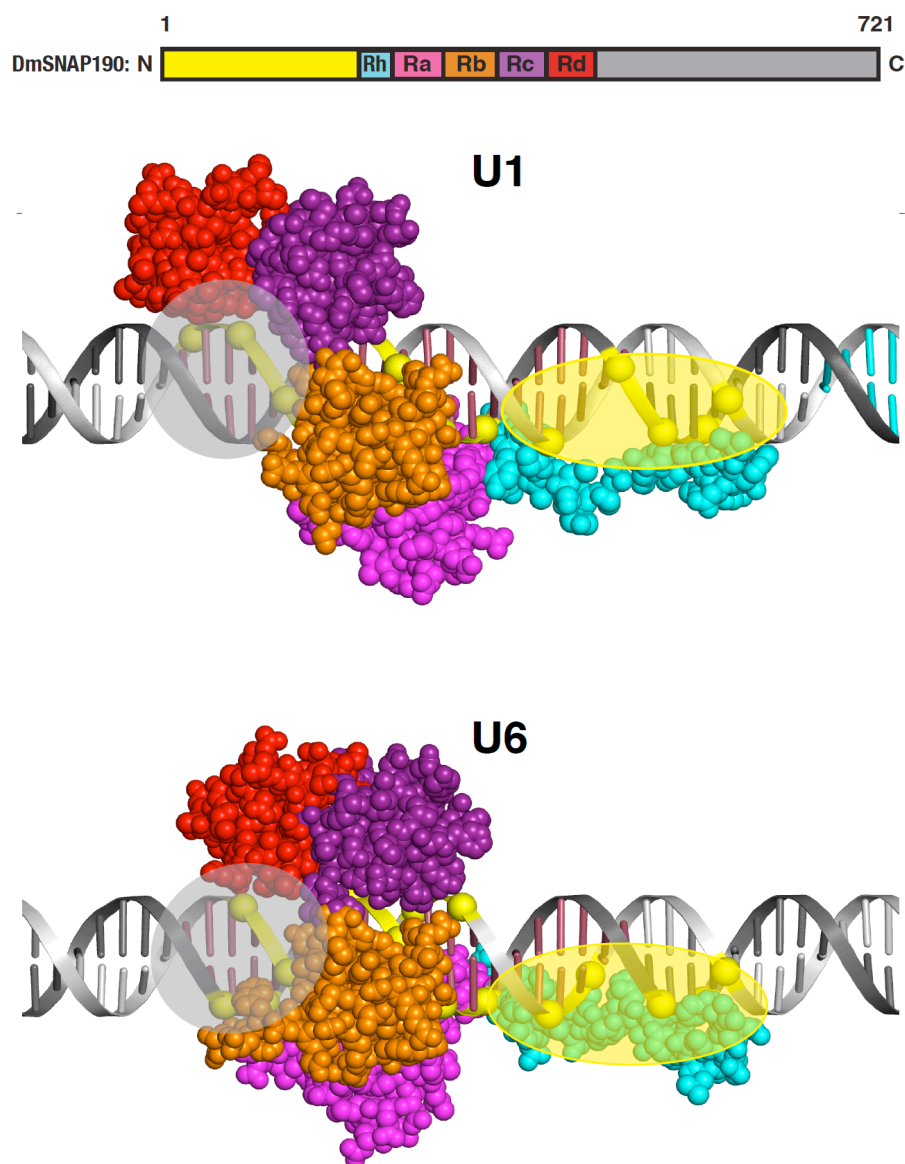
FIGURE 8. **Model of DmSNAP190 domains interacting differentially with the U1 and U6 PSEAs.** The top bar represents a schematic diagram of DmSNAP190. The N-terminal domain, Rh, Ra, Rb, Rc, Rd, and the C-terminal domain are shown in yellow, cyan, magenta, orange, deep purple, red, and gray colors, respectively. In the lower part of the figure, models are shown DmSNAP190 differentially bound to U1 and U6 PSEAs. The domains were situated on the B-form DNA backbone in accord with the cross-linking data summarized in Fig. 7. The domains of DmSNAP190 are color-coded to match the schematic diagram at the top of the figure. See the text for further discussion.

PSEA. In fact, the R1 repeat of c-Myb in the NMR and co-crystal structures of c-Myb does not directly interact with the major groove of the DNA (Ogata et al., 1995; Tahirov et al., 2002), so there is clear precedent for the lack of major groove to "recognition helix" contacts involving Myb repeats.

We considered the possibility that the DNA of the U1 and U6 PSEAs could be differentially bent, which in principle could partially account for the unique Myb repeat minor groove-spanning interactions observed for U6. However, extensive experiments to examine the bending of the U1 and U6 promoter DNA while bound to DmSNAPc indicated that there was only a minor bend in the DNA, and the magnitude and direction of the bend were very similar for DmSNAPc-bound U1 and U6 promoter DNA (Hardin et al., 2000). Furthermore, base positions 1 through 12 in the PSEAs of both the Pol II-transcribed and Pol III-transcribed snRNA genes of *D. melanogaster* are very strongly conserved in nucleotide sequence (Hung and Stumph, 2011), so it seems very unlikely that the nearly identical sequences of the 5' half of the PSEA (the region contacted by the Ra, Rb, Rc, and Rd repeats) would be differentially bent on U1 and U6 promoters. Although it is unlikely that the DNA bound by DmSNAPc exists as simple B-form DNA over its entire length as represented in Figure 8, in the absence of any additional information it seems reasonable to utilize B-form DNA for the molecular modeling.

The Rh repeat of DmSNAP190 contacts phosphates 14, 24, and 25 on both the U1 and U6 PSEAs, but on the U6 PSEA there are additional contacts of Rh with positions 19 and 22, perhaps indicating that Rh is positioned more toward the rear face of the U6 PSEA in Figure 8 compared to its position on the U1 PSEA. The N-terminal domain of DmSNAP190 (residues 1-189, shown in yellow in Figure 8) contacts phosphate 20

(template strand) only on the U1 PSEA but contacts position 19 (non-template strand) only on the U6 PSEA, probably indicating that this domain is rotated slightly clockwise around the DNA on a U6 PSEA relative to a U1 PSEA when viewed from upstream toward the transcription start site.

A surprising result from our U6 data was that the domain of DmSNAP190 C-terminal to the Myb repeats cross-linked to phosphate position 2 of the U6 PSEA. No cross-linking of this C-terminal domain (represented in grey in Figure 8) was observed to DNA of the U1 PSEA. However, we cannot rule out the possibility that this domain of DmSNAP190 might cross-link to phosphate positions in the U1 PSEA that were not examined.

DISCUSSION

Model of the differential interaction of the DmSNAP complex with the U1 and U6 PSEAs.

In previous work, we mapped domains of DmSNAP50 and DmSNAP43 that cross-linked to specific nucleotide positions of the U1 and U6 PSEAs (Kim et al., 2010a). As a result, we know the specific positions in the DNA where individual domains of these two smaller subunits interact with the U1 and U6 PSEAs. To obtain a better understanding of the binding of the entire DmSNAP complex to U1 and U6 promoter sequences, we superimposed the earlier data from the two smaller subunits upon the results obtained with DmSNAP190. The result is shown in Figure 9. On a U1 PSEA, a domain of DmSNAP43 bounded by amino acid residues 193-272 contacted the DNA

quite far downstream of the U1 PSEA in the region surrounding the PSEB (Hung and Stumph, 2011; Kim et al., 2010a; Li et al., 2004). Those contacts were absent when DmSNAPc was bound to a U6 PSEA. Conversely, the domain of DmSNAP43 bounded by amino acid residues 273-363 cross-linked to phosphates 11 and 16 on the U6 PSEA but not when bound to a U1 PSEA. With regard to DmSNAP50, there were also significant but more subtle differences that existed in its conformation on the U1 and U6 PSEAs (Hung and Stumph, 2011; Kim et al., 2010a; Wang and Stumph, 1998).

Based upon all available site-specific protein-DNA photo-cross-linking data, the illustrations in Figure 9 show comprehensive schematic models of DmSNAPc bound differentially to U1 or U6 PSEAs. On the U1 PSEA, DmSNAPc appears to adopt a more open conformation that allows residues 193-272 of DmSNAP43 to contact the DNA in the region of the PSEB. In contrast, on the U6 PSEA new protein-DNA contacts occur at phosphates 11 and 13 (with the Rc repeat of DmSNAP190) and at phosphates 11 and 16 (with the C-terminal domain of DmSNAP43). This seems to bring DmSNAPc into a more "closed" conformation with potentially new interfacial interactions among the subunits, particularly between DmSNAP190 and DmSNAP43. (Note that DmSNAP190 and DmSNAP43 both cross-link to phosphate position 11 on the U6 PSEA, but neither cross-link to position 11 of the U1 PSEA.) As a consequence of DmSNAP190(Rc domain) and DmSNAP43(273-363) moving toward each other on the U6 PSEA, the interactions of DmSNAP43(193-272) with the phosphates that surround the PSEB downstream of the U1 PSEA no longer occur.

Compatibility of the model with other data

DmSNAPc subunit-subunit interaction studies in the absence of DNA have demonstrated the interaction of DmSNAP43 residues 1-172 with the Myb repeat domain of DmSNAP190 (Hung and Stumph, 2011; Hung et al., 2009). In agreement with that observation, the cross-linking data strongly suggest that the N-terminal domain of DmSNAP43 is in close proximity to the Rh Myb repeat on both the U1 and U6 PSEAs. Likewise, DmSNAP50 was found to interact with the N-terminal domain (residues 63-176) of DmSNAP190 (Hung and Stumph, 2011; Hung et al., 2009). The cross-linking studies place DmSNAP50 in close proximity to this domain on both the U1 and U6 PSEAs. Thus, the protein interaction studies and the photo-cross-linking data are highly consistent. Recent ChIP-seq studies combined with previous subunit interaction studies in the human system also provided an indication that DmSNAP190 is oriented with its N-terminal domain facing toward the transcription start site (James Faresse et al., 2012), in close accord with our model. Finally, other studies in the human system indicated that the N-terminal domain of SNAP190 was involved in recruiting TBP to the U6 promoter (Ma and Hernandez, 2002). Our studies place this N-terminal domain of DmSNAP190 at the 3' end of the U6 PSEA where it would be ideally positioned to interact with TBP.

Concluding remarks

By combining site-specific protein-DNA photo-cross-linking with site-specific digestion of the three subunits of DmSNAPc, we have obtained data that allow us to formulate a model of how DmSNAPc differentially interacts with U1 and U6 promoter
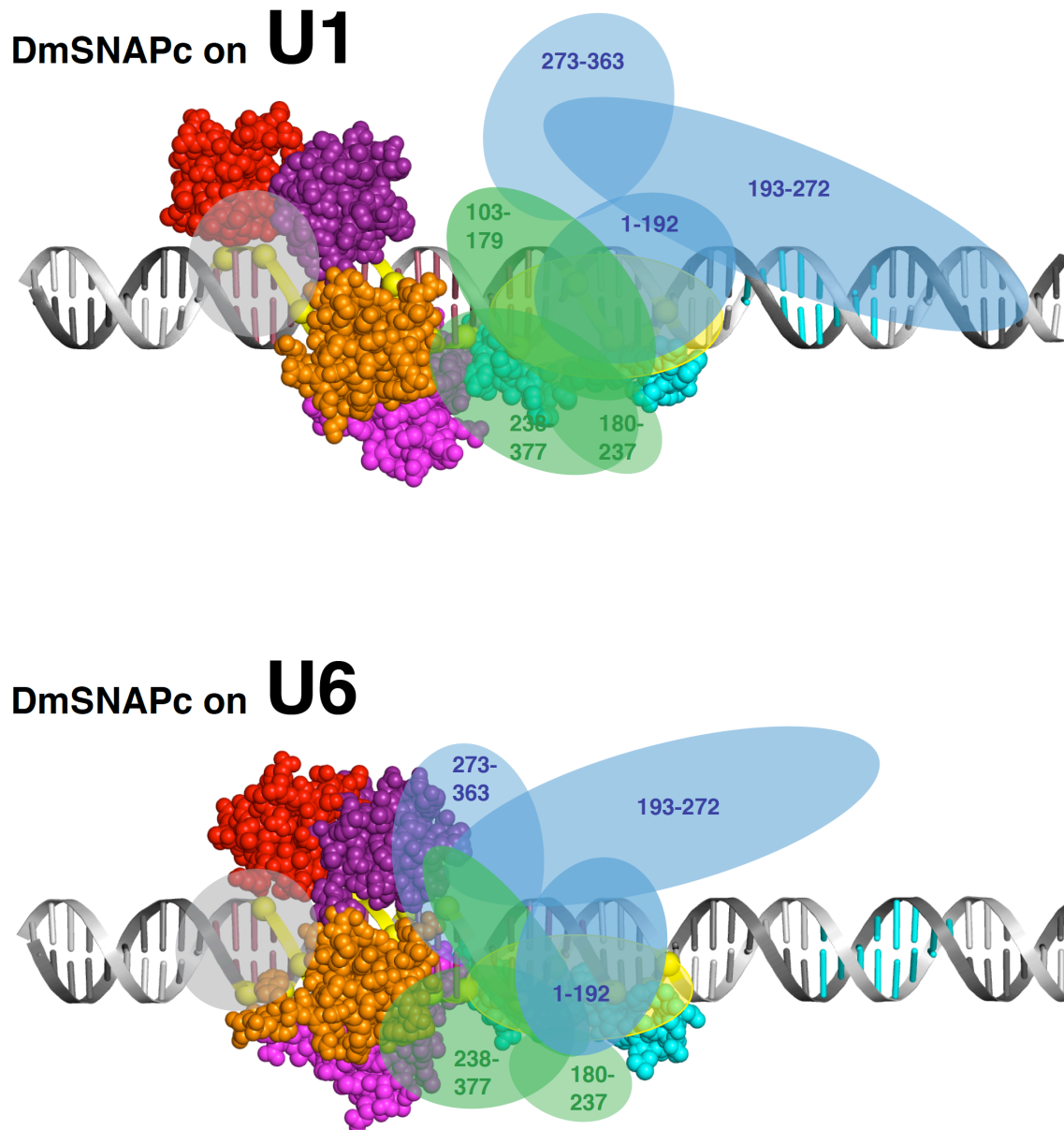
FIGURE 9. **Structural models of the entire DmSNAP complex interacting with the U1 and U6 PSEAs.** The models are based upon photo-cross-linking studies reported in this chapter and in Kim et al. (2010) and in Doherty et al. (2012). The green and blue ovals represent domains within DmSNAP50 and DmSNAP43, respectively. The numbers inside the ovals indicate the amino acid positions at the beginning and end of the fragments of each DmSNAPc subunit that have been mapped on U1 and U6 PSEAs. DmSNAP190 domains are shown as indicated in the legend of Fig. 8. The cyan colored bases represent the location of U1 PSEB or U6 TATA box in the upper and lower illustrations respectively.

sequences. From these data, we suspect that DmSNAPc is a quite flexible protein conformations that depend upon the PSEA sequence bound. Because the DNA sequence of the PSEA plays an important role in determining the RNA polymerase specificity of the U1 and U6 promoters, it seems likely that the specificity is mediated through the different conformations of DmSNAPc on the U1 and U6 promoter sequences. Determining the mechanisms by which the different DmSNAPc conformations recruit different RNA polymerases will be an important focus of future investigation.

REFERENCES

Dahlberg, J.E., and Lund, E. (1988). The genes and transcription of the major small nuclear RNAs. In Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles, M.L. Birnstiel, ed. (Heidelberg, Federal Republic of Germany: Springer Verlag KG), pp. 38-70.

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, J.B., Chatterjee, B., Cross, G.A., 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.

Doherty, M.T., Kang, Y.S., Lee, C., and Stumph, W.E. (2012). Architectural arrangement of the small nuclear RNA (snRNA)-activating protein complex 190 subunit (SNAP190) on U1 snRNA gene promoter DNA. J Biol Chem *287*, 39369-39379.

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

Hardin, S.B., Ortler, C.J., McNamara-Schroeder, K.J., and Stumph, W.E. (2000). Similarities and differences in the conformation of protein-DNA complexes at the U1 and U6 snRNA gene promoters. Nucleic Acids Res *28*, 2771-2778.

Henry, R.W., Ford, E., Mital, R., Mittal, V., and Hernandez, N. (1998). Crossing the line between RNA polymerases: Transcription of human snRNA genes by RNA polymerases II and III. Cold Spring Harbor SympQuantBiol *63*, 111-120.

Henry, R.W., Sadowski, C.L., 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.

Hernandez, G., Valafar, F., and Stumph, W.E. (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. (2001). snRNA genes: a model system to study fundamental mechanisms of transcription. J Biol Chem *276*, 26733-26736.

Hung, K.H., and Stumph, W.E. (2011). Regulation of snRNA gene expression by the Drosophila melanogaster small nuclear RNA activating protein complex (DmSNAPc). Crit Rev Biochem Mol Biol *46*, 11-26.

Hung, K.H., Titus, M., Chiang, S.C., and Stumph, W.E. (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.

James Faresse, N., Canella, D., Praz, V., Michaud, J., Romascano, D., and Hernandez, N. (2012). Genomic study of RNA polymerase II and III SNAPc-bound promoters reveals a gene transcribed by both enzymes and a broad use of common activators. PLoS genetics *8*, e1003028.

Jawdekar, G.W., and Henry, R.W. (2008). Transcriptional regulation of human small nuclear RNA genes. Biochim Biophys Acta 1779, 295-305.
Jensen, R.C., Wang, Y., Hardin, S.B., and Stumph, W.E. (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.

Kiefer, F., Arnold, K., Kunzli, M., Bordoli, L., and Schwede, T. (2009). The SWISS-MODEL Repository and associated resources. Nucleic Acids Res *37*, D387-392.

Kim, M.K., Kang, Y.S., Lai, H.T., Barakat, N.H., Magante, D., and Stumph, W.E. (2010a). Identification of SNAPc subunit domains that interact with specific nucleotide positions in the U1 and U6 gene promoters. Mol Cell Biol *30*, 2411-2423.

Kim, M.K., Kang, Y.S., Lai, H.T., Barakat, N.H., Magante, D., and Stumph, W.E. (2010b). Identification of SNAPc subunit domains that interact with specific nucleotide positions in the U1 and U6 gene promoters. Mol Cell Biol *30,* 5257.

Lai, H.T., Chen, H., Li, C., McNamara-Schroeder, K.J., and Stumph, W.E. (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.

Li, C., Harding, G.A., Parise, J., McNamara-Schroeder, K.J., and Stumph, W.E. (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.

Ma, B.C., and Hernandez, N. (2002). Redundant cooperative interactions for assembly of a human U6 transcription initiation complex. Mol Cell Biol *22*, 8067-8078.

McNamara-Schroeder, K.J., Hennessey, R.F., Harding, G.A., Jensen, R.C., and Stumph, W.E. (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.

Murphy, S., Yoon, J.B., Gerster, T., and Roeder, R.G. (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.

Ogata, K., Morikawa, S., Nakamura, H., Hojo, H., Yoshimura, S., Zhang, R., Aimoto, S., Ametani, Y., Hirata, Z., Sarai, A., et al. (1995). Comparison of the free and DNA-complexed forms of the DNA-binding domain from c-Myb. Nature Struct Biol *2*, 309-320.

Ogata, K., Tahirov, T.H., and Ishii, S. (2004). The c-Myb DNA binding domain. In Myb Transcription Factors: Their Role in Growth, Differentiation, and Disease, J. Framp;ton, ed. (Dordrecht: Kluwer), pp. 223-238.

Parry, H.D., Scherly, D., and Mattaj, I.W. (1989). Snurpogenesis: the transcription and assembly of U snRNP components. Trends Biochem Sci *14*, 15-19.

Sadowski, C.L., Henry, R.W., Lobo, S.M., 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., Nguyen, T.N., and Gunzl, A. (2005a). Characterization of a multisubunit transcription factor complex essential for spliced-leader RNA gene transcription in Trypanosoma brucei. Mol Cell Biol *25*, 7303-7313.

Schimanski, B., Nguyen, T.N., and Gunzl, A. (2005b). Highly efficient tandem affinity purification of trypanosome protein complexes based on a novel epitope combination. Eukaryot Cell *4*, 1942-1950.

Su, Y., Song, Y., Wang, Y., Jessop, L., Zhan, L.C., and Stumph, W.E. (1997). Characterization of a Drosophila proximal-sequence-element-binding protein involved in transcription of small nuclear RNA genes. Eur J Biochem *248*, 231-237.

Tahirov, T.H., Sato, K., Ichikawa-Iwata, E., Sasaki, M., Inoue-Bungo, T., Shiina, M., Kimura, K., Takata, S., Fujikawa, A., Morii, H., et al. (2002). Mechanism of c-Myb-C/EBP beta cooperation from separated sites on a promoter. Cell *108*, 57-70.

Tanikawa, J., Yasukawa, T., Enari, M., Ogata, K., Nishimura, Y., Ishii, S., and Sarai, A. (1993). Recognition of specific DNA sequences by the c-myb protooncogene product: Role of three repeat units in the DNA-binding domain. Proc Natl Acad Sci USA *90*, 9320-9324.

Wang, Y., and Stumph, W.E. (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, M. W., Henry, R. W., Ma, B. C., 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, J. B., Murphy, S., Bai, L., Wang, Z., and Roeder, R. G. (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.

CONCLUDING REMARKS

Mapping domains of DmSNAP190 that interact with specific nucleotide positions in the U1 promoter

In work described in chapters 1 and 2, I and my colleagues localized domains or sub-regions of DmSNAP190 that cross-linked to all 13 positions in a U1 PSEA that had previously been determined to cross-link to DmSNAP190. The work revealed that all 4.5 Myb repeats of DmSNAP190 reside close to the DNA of the U1 PSEA. Furthermore, we also determined the orientation of the Myb repeats on the promoter DNA. The Myb repeats are ordered with the C-terminal repeat (Rd) farthest from the transcription start site and the N-terminal half-repeat (Rh) closest to the start site. Although there are no crystal or NMR structures yet available for SNAPc from any organism, we were able, by using the known structure of the Myb repeats of the c-Myb protein bound to DNA (Ogata et al., 1995; Tahirov et al., 2002), to develop a model of the DmSNAP190 Myb repeats bound to the U1 promoter in a manner consistent with both the cross-linking data and the structures of Myb repeats (Doherty et al., 2012) (Chapter 3, Figure 8). In the model, the Rd, Rc, Rb, and Ra repeats follow the contour of the major groove of the DNA throughout the 5' half of the U1 PSEA.

Although the 5' half of the U1 PSEA interacts with repeats Rd, Rc, Rb, and Ra, our cross-linking results showed that the 3' half of the U1 PSEA interacts with the Rh half-repeat of DmSNAP190. To our initial surprise, this same region of the U1 PSEA (i.e., the 3' half) interacts even more strongly and extensively with the DmSNAP190 N-terminal domain that precedes the Myb repeat domain. This N-terminal domain of DmSNAP190 contains no known DNA-binding motifs. Furthermore, the 3' half of the

PSEA cross-links also to DmSNAP43 and DmSNAP50 (Wang and Stumph, 1998; Kim et al., 2010a). Thus, there seems to be an extremely complicated and intricate network of protein-DNA and protein-protein interactions in the region associated with the 3' half of the PSEA. The determination of an atomic-resolution model of these interactions would be very interesting and informative as part of future work on SNAPc.

The non-conserved C-terminal region of DmSNAP190 (residues 442 to 721) did not cross-link to any of the phosphate positions mapped in the U1 PSEA. Moreover, it was not found to interact with any other subunits of the DmSNAP complex, which would be appropriate for its location facing upstream of the PSEA. However, our lab has previously determined that this region contributes to the ability of DmSNAPc to bind to U1 DNA (Hung et al., 2009). Thus, the C-terminal domain of DmSNAP190 may be responsible for the stabilization of DmSNAPc on U1 DNA.

Mapping domains of DmSNAP190 that interact with specific nucleotide positions in the U6 promoter

Early site-specific protein-DNA photo-cross-linking studies carried out in our lab indicated that DmSNAP50 and DmSNAP43 cross-linked in quite different patterns to U1 and U6 PSEAs (Wang et al., 1998; Li et al., 2004). On the other hand, DmSNAP190 bound over the entire length of both PSEAs and showed relatively similar binding patterns on U1 and U6 PSEAs except for a few specific nucleotide positions. However, the work described in chapter 3 revealed that there are indeed significant differences in

the structural arrangement of DmSNAP190 when DmSNAPc binds to U1 versus U6 promoters.

Compared to the Rc and Ra positions on a U1 PSEA, the Rc repeat on a U6 PSEA moves slightly in the downstream direction and the Ra repeat shifts to a position in the upstream direction relative to Myb repeat Rb, which is situated similarly on U1 and U6 promoters. Both the movement of Rc and of Ra result in new minor groove-spanning protein-DNA interactions on U6 promoters that are absent on U1 promoters. Furthermore, as a result of these positional shifts, the Rc repeat of DmSNAP190 and a domain of DmSNAP43 (residues 273-363) appear to come into much closer proximity on a U6 promoter than on a U1 promoter (Chapter 3, Figure 9).

The N-terminal domain of DmSNAP190 that precedes the Myb repeats also contacts the 3' half of the U1 and U6 PSEAs differently. The N-terminal domain cross-links to position 20 (template strand) of the U1 PSEA, but it does not contact position 20 on the U6 PSEA; instead, it contacts position 19 (non-template strand) of the U6 PSEA. Because positions 19 and 20 in our numbering system are on nearly opposite sides of the DNA helix from each other, this suggests a partial rotation of the DmSNAP190 N-terminal domain around the DNA double helix on U6 compared to U1.

It also came as a surprise that the non-conserved C-terminal region of DmSNAP190 (residues 442 to 721) cross-linked to position 2 near the 5' end of the U6 PSEA. We previously observed no cross-linking of the DmSNAP190 C-terminal domain to the DNA of the U1 PSEA. However, previous work in our lab found that the DmSNAP190 C-terminal domain contributed to the DNA-binding activity of DmSNAP190 (Hung et al., 2009). Because we did not assay phosphate positions in our

photo-cross-linking assay upstream of the 21 base pair PSEA, it is possible that the C-terminal domain of DmSNAP190 might possibly contact the DNA upstream of the PSEA on both the U1 and U6 PSEAs.

Different conformations of DmSNAPc on U1 and U6 snRNA gene promoters and role in determining RNA polymerase specificity

DmSNAPc binds to the promoters of both U1 and U6 snRNA genes, but a few nucleotide sequence differences of the PSEAs results in the recruitment of different RNA polymerases (Jensen et al., 1998; McNamara-Schroeder et al., 2001). The results of the photo-cross-linking assays indicate that each of the DmSNAPc subunits interact differently with the U1 and U6 PSEAs. Therefore, we hypothesize that these different conformations of DmSNAPc act as foundations that lead to the differential recruitment of different classes (Pol II or Pol III) of general transcription factors.

In the human system, RNA polymerase II transcription on snRNA genes requires the general transcription factors TBP, TFIIA, TFIIB, TFIIE, and TFIIF (and probably TFIIH) (Sadowski et al., 1993; Kuhlman et al., 1999). Transcription of human U6 genes also requires TBP and an snRNA gene-specific form of the TFIIB-related factor (Brf) termed Brf2 (as well as the Pol III-specific factor Bdp1) (Chong et al., 2001). Human TBP was found to be capable of interacting with all three core subunits of SNAPc (SNAP43, SNAP50 and SNAP190 (Ma and Hernandez, 2002; Hinkley et al., 2003). In the fruit fly, TBP is required for transcription from both the U1 and U6 snRNA gene promoters, as shown by *in vitro* transcription assays and chromatin immunoprecipitation
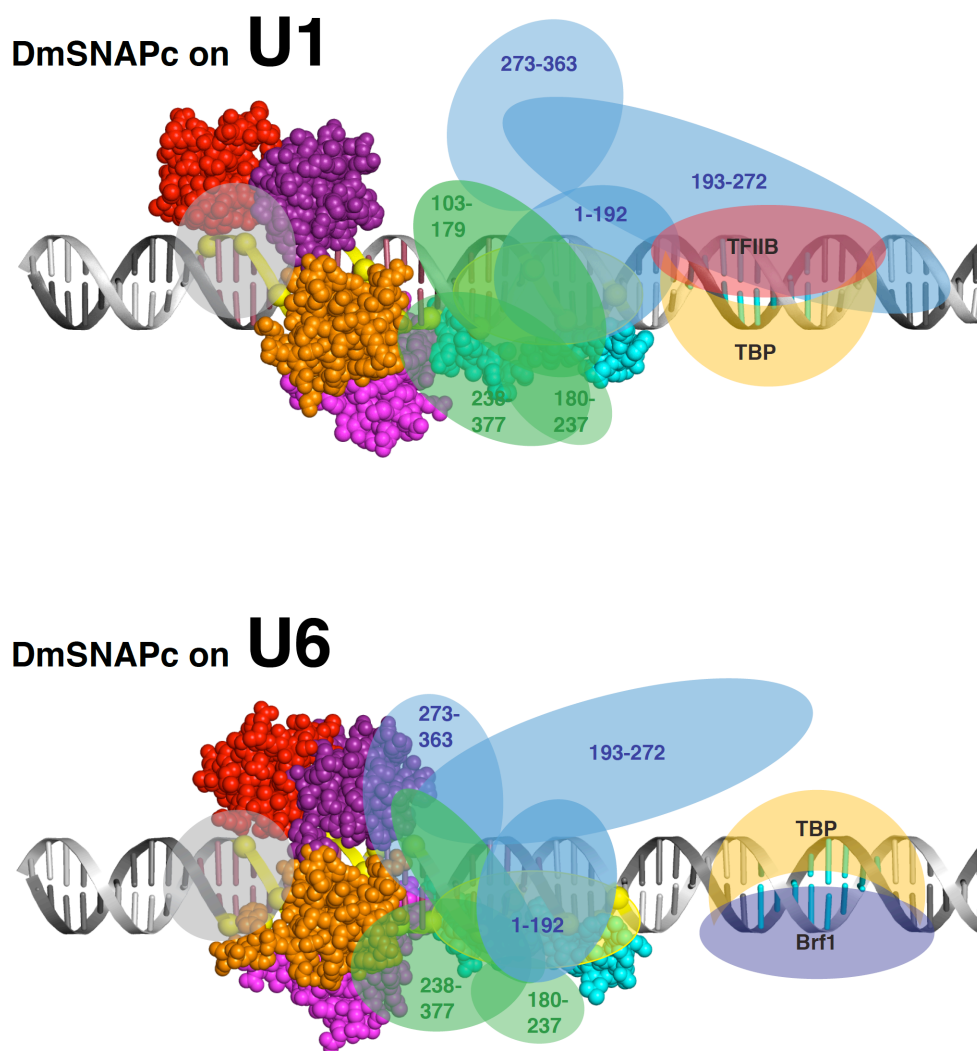
FIGURE 1. **Structural models illustrating how the differential interaction of the DmSNAP complex with the U1 and U6 PSEAs may recruit different general transcription factors.** We postulate that different conformations of DmSNAPc on U1 and U6 promoters lead to the recruitment of different general transcription factors (TBP and TFIIB on the U1 promoter and TBP and Brf1 on the U6 promoter). The binding of different general transcription factors would recruit Pol II to the U1 promoter but Pol III to the U6 snRNA gene promoter. Note that, due to a difference in spacing from the PSEA to the PSEB versus TATA box (Introduction Fig. 1), if TBP binds to the PSEB in the U1 promoter but to the TATA box on the U6 promoter, TBP will bind to nearly opposite sides of the DNA double helix relative to DmSNAPc on U1 and U6 promoters.

assays (Zamrod et al., 1993; Jensen et al., 1998; Barakat and Stumph, 2008; Verma et al., 2013). In the *D. melanogaster* genome, there is only one Brf gene, so Brf1 is presumably used for U6 transcription in fruit flies.

The upper part of Fig. 1 shows a representation of DmSNAPc on the U1 promoter together with schematic drawings of TBP and TFIIB interacting with the DNA. The lower part of Fig. 1 similarly shows a representation of DmSNAPc on a U6 promoter together with TBP and Brf1. It seems safe to assume that TBP binds to the TATA box on the U6 promoter. It is also logical that TBP interacts with the PSEB on the U1 promoter, based upon the fact that the PSEB is an 8 base pair conserved sequence located 25 to 32 base pairs upstream of the transcription start site.

It is worth keeping in mind that the TATA box is separated from the PSEA by 12 base pairs, but the PSEB is separated from the PSEA by only 8 base pairs. Therefore, if we assume that TBP interacts with the PSEB (on U1) and TATA box (on U6), TBP (as well as TFIIB/Brf1) must be binding on nearly opposite sides of the DNA helix on the U6 and U1 promoters relative to DmSNAPc (Fig. 1). It will be interesting to dissect how the interactions between DmSNAPc and the general transcription factors occur differentially on the snRNA Pol II and Pol III promoters.

Future directions

In future work, it will be important to dissect the interactions that occur between DmSNAPc and the general transcription factors on U1 and U6 promoters. For example, are different DmSNAPc subunits or different domains of the same subunits involved in recruiting the Pol II and Pol III general transcription factors to the U1 and U6 promoters

respectively? Which general transcription factors interact with DmSNAPc on the U1 and U6 promoters? These questions can potentially be answered by deletion or mutational analysis of DmSNAP subunits and of the general transcription factors. Assuming the diagram in Fig. 1 is more-or-less accurate, it will also be useful to precisely map the locations where TBP, TFIIB, and Brf1 interact with the DNA. This mapping can potentially be done by using the photo-cross-linking assay or by using a lambda exonuclease assay combined with immunoprecipitation.

Finally, and ideally, we would like to know the atomic structure of DmSNAPc bound to both U1 and U6 promoter DNA. Such knowledge could potentially be obtained by X-ray crystallographic studies. Indeed, it would be ideal to crystallize DmSNAPc together with the respective general transcription factors bound to U1 and U6 DNA. However, those desires represent long-term projects that have no guarantee of success. Perhaps the initial determination of smaller domains of the DmSNAP subunits would be more feasible on a shorter-term basis. For example, it would be interesting to know the atomic structure of the N-terminal domain of DmSNAP190 that precedes the 4.5 Myb repeats. This domain makes extensive contacts with the 3' half of the PSEA yet has no known DNA binding motifs. With the knowledge obtained from the cross-linking assays, it may be possible to piece together some of the DmSNAPc-DNA structure from the individual parts.

REFERENCES

Barakat, N. H. and Stumph W. E. (2008). TBP recruitment to the U1 snRNA gene promoter is disrupted by substituting a U6 proxi- mal sequence element A (PSEA) for the U1 PSEA. FEBS Letters *582*, 2413–2416.

Chong, S.S., Hu, P., Hernandez, N. (2001). Reconstitution of transcription from the human U6 small nuclear RNA promoter with eight recombinant polypeptides and a partially purified RNA polymerase III complex. J Biol Chem. *276*, 20727-34.

Doherty, M. T., Kang, Y. S., Lee, C., and Stumph, W. E. (2012). Architectural arrangement of the small nuclear RNA (snRNA)-activating protein complex 190 subunit (SNAP190) on U1 snRNA gene promoter DNA. J Biol Chem *287*, 39369-39379.

Hardin, S. B., Ortler, C. J., McNamara-Schroeder, K. J., and Stumph, W. E. (2000). Similarities and differences in the conformation of protein-DNA complexes at the U1 and U6 snRNA gene promoters. Nucleic Acids Res *28*, 2771-2778.

Henry, R.W., Sadowski, C.L., 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.

Hinkley, C. S., Hirsch, H. A., Gu, L. P., LaMere, B., Henry, R. W. (2003). The small nuclear RNA-activating protein 190 Myb DNA binding domain stimulates TATA box-binding protein-TATA box recognition. J Biol Chem, *278*, 18649-57.

Hung, K. H., Titus, M., Chiang, S. C., and Stumph, W.E. (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.

Jensen, R. C., Wang, Y., Hardin, S. B., and Stumph, W. E. (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.

Kim, M. K., Kang, Y. S., Lai, H. T., Barakat, N. H., Magante, D., and Stumph, W. E. (2 010a). Identification of SNAPc subunit domains that interact with specific nucleotide positions in the U1 and U6 gene promoters. Mol Cell Biol *30*, 2411-2423.

Kim, M. K., Kang, Y. S., Lai, H. T., Barakat, N. H., Magante, D., and Stumph, W. E. (2010b). Identification of SNAPc subunit domains that interact with specific nucleotide positions in the U1 and U6 gene promoters. Mol Cell Biol *30*, 5257.

Kuhlman, T. C., 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.

Li, C., Harding, G.A., Parise, J., McNamara-Schroeder, K.J., and Stumph, W.E. (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.

Ma, B.C., and Hernandez, N. (2002). Redundant cooperative interactions for assembly of a human U6 transcription initiation complex. Mol Cell Biol *22*, 8067-8078.

McNamara-Schroeder, K. J., Hennessey, R. F., Harding, G. A., Jensen, R. C., and Stumph, W. E. (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.

Ogata, K., Morikawa, S., Nakamura, H., Hojo, H., Yoshimura, S., Zhang, R., Aimoto, S., Ametani, Y., Hirata, Z., Sarai, A.*,* and Nishimura, Y. (1995). Comparison of the free and DNA-complexed forms of the DNA-binding domain from c-Myb. Nature Struct Biol *2*, 309-320.

Tahirov, T.H., Sato, K., Ichikawa-Iwata, E., Sasaki, M., Inoue-Bungo, T., Shiina, M., Kimura, K., Takata, S., Fujikawa, A., Morii, H.*,* Kumasaka, T., Yamamoto, M., Ishii, S., and Ogata, K. (2002). Mechanism of c-Myb-C/EBP beta cooperation from separated sites on a promoter. Cell *108*, 57-70.

Sadowski, C.L., Henry, R.W., Lobo, S.M., 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.

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

Wang, Y., and Stumph, W.E. (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.

Verma, N., Hung, K.H., Kang, J.J., Barakat, N.H., Stumph W.E. (2013). Differential utilization of TATA box-binding protein (TBP) and TBP-related factor 1 (TRF1) at different classes of RNA polymerase III promoters. J Biol Chem. *288*, 27564-70.

Zamrod, Z., Tyree, C.M., Song, Y., and Stumph, W.E. (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. MolCellBiol *13*, 5918-5927.