Title

TBP recruitment to the U1 snRNA gene promoter is disrupted by substituting a U6 proximal sequence element A (PSEA) for the U1 PSEA

Author

Barakat, Nermeen H.

Publication Date

2008

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

SAN DIEGO STATE UNIVERISTY

TBP Recruitment to the U1 snRNA Gene Promoter Is Disrupted by Substituting a U6 Proximal Sequence Element A (PSEA) for the U1 PSEA

A Dissertation submitted in partial satisfaction of the

Requirements for the degree Doctor of Philosophy

in

Chemistry

by

Nermeen H. Barakat

Committee in Charge:

University of California, San Diego

Professor Partho Ghosh
Professor James T. Kadonaga
Professor Douglas Magde

San Diego State University

Professor William E. Stumph, Chair
Professor John J. Love
Professor Kathleen McGuire

2008

The dissertation of Nermeen H. Barakat is approved, and it is acceptable in quality and form for publication on microfilm:

_____

_____

_____

_____

_____

_____

Chair

University of California, San Diego

San Diego State University

2008

DEDICATION

I dedicate this dissertation to the following people:

The memory of my father, Hisham Barakat, who emphasized the importance of education.

My sweet mother, Souad Barakat, who has been my role-model of hard work.

My lovely sisters, Dr. Nora Barakat, Nesreen and Neveen Barakat who have been my anchors through my graduate school and my entire life.

and my dear husband, Dr. Mosen Istwani, who has been proud and supportive of my work.

TABLE OF CONTENTS

LIST OF FIGURES

General Introduction

Chapter 1

Chapter 2

ACKNOWLEDGEMENTS

I am extremely grateful to my thesis advisor and committee chair Dr. Stumph for giving me the opportunity to work in his lab, for his continuous guidance, unwavering patience to teach me through my scientific development process and his full support throughout the project. I would like to thank Dr. Love for his guidance and support. I thank Dr. Love, Dr. McGuire, Dr. Magde, Dr. Ghosh, and Dr. Kadonaga for their valuable time and for serving on my dissertation committee. I am grateful to Kathleen McNamara-Schroeder for excellent technical assistance. I would like to thank my colleagues Dr. Cheng Li, Dr. Mitch Titus, Gale A. Harding, Genaro Hernandez, Michelle Chen, Dr. Hsien-Tsung Lai, Anna Luy, Ko-Hsuan Alex Hung, Shu-Chi Kiki Chiang, Mun Kyung Kim, Yoon Soon Kang, and Jin Joo Kang and all the other members of the Stumph lab for making an exciting work environment. I would like to thank the California Metabolic Research Foundation for the Arne N. Wick Pre-doctoral Research fellowship in the last year.

I am deeply thankful to my best friend and my mother Souad Barakat, who taught me the importance of education and believed in me and for her continuous support and encouragement throughout the entire process. I would like to thank my colleagues and my sisters Dr. Nora Barakat, Nesreen, and Neveen for their endless support, help and contributions, and for making me proud to have such great sisters. I wish success and the best of luck to them in pursuing their education and career. Finally, I would like to thank my darling, my husband, Dr. Istwani who supported me and motivated me in my last year.

VITA

1989-1993	B.S., Biochemistry, University of Qatar
1994-1996	Diploma in Education, University of Qatar
1998-2000	Certificate of Achievement, Computer Science and Information Systems Grossmont Community College
2001-2003	Teaching Associate, Chemistry Department San Diego State University
2001-2003	M.A. Biochemistry, San Diego State University
2003-2004	Teaching Associate, Department of Chemistry and Biochemistry University of California, San Diego
2004-2007	Teaching Associate, Chemistry Department San Diego State University
2006-2008	Pre-doctoral Candidate, Chemistry and Biochemistry University of California, San Diego San Diego State University
2007-2008	Recipient of an Arne N. Wick Pre-doctoral Research Fellowship from the California Metabolic Research Foundation
2003-2008	Ph.D., Chemistry University of California, San Diego and San Diego State University (Joint Doctoral Program)

PUBLICATIONS AND ABSTRACTS

Peer-reviewed article

Barakat NH, Stumph WE. 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. 2008; 582(16):2413-6.

Manuscripts in preparation

Lai H-T, **Barakat NH**, Kim MK, Kang YS, Magante D, Stumph WE. Identification of subdomains within the DmSNAPc subunits required for interaction with specific nucleotides within the U1 and U6 gene promoters. *In preparation*.

Abstracts

Barakat NH, Stumph WE. Polymerase II transcription complex assembly on U1 snRNA gene promoter.

CSUPERB (California State University Program for Education and Research in Biotechnology, Los Angeles, CA, 2007).

Barakat NH, Stumph WE. TBP recruitment to the U1 gene promoter is disrupted by substituting a U6 proximal sequence element (PSEA) for the U1 PSEA.

(Mechanisms of Eukaryotic Transcription, Cold Spring Harbor Laboratory, NY, 2007).

Barakat NH, Stumph WE. TBP recruitment to the U1 gene promoter is disrupted by substituting a U6 proximal sequence element (PSEA) for the U1 PSEA.

(Experimental Biology, San Diego, CA, 2008).

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

TBP Recruitment to the U1 snRNA Gene Promoter Is Disrupted by Substituting a U6 Proximal Sequence Element A (PSEA) for the U1 PSEA

by

Nermeen H. Barakat

Doctor of Philosophy in Chemistry

University of California, San Diego, 2008

San Diego State University, 2008

Professor William E. Stumph, Chair

In eukaryotes, small nuclear RNAs (snRNAs) are required for pre-mRNA splicing. Most snRNAs, such as U1, U2 ,U4 and U5, are synthesized by RNA polymerase II, but U6 snRNA is synthesized by RNA polymerase III. Transcription of snRNA genes by either RNA polymerase is dependent upon a proximal sequence element (PSE) centered approximately 50-55 base pairs upstream of the start site. The PSE is recognized by the small nuclear RNA activating protein complex (SNAPc), a multi-subunit transcription factor. In *Drosphila melanogaster*, the PSE is more specifically termed the PSEA to distinguish it from a second conserved element termed the PSEB

present in the promoter of the Pol II transcribed fly snRNA genes. Interestingly, the fly U1 and U6 PSEAs are not functionally interchangeable, even though both are recognized by the same protein, DmSNAPc. A five-nucleotide substitution that changed the U1 PSEA to a U6 PSEA was shown to inactivate the U1 promoter.

In light of this knowledge I wished to investigate why the U6 PSEA cannot functionally substitute for the U1 PSEA. I sought to determine whether the U6 PSEA substitution disrupts a specific step in RNA polymerase II transcription pre-initiation complex assembly *in vivo*. To accomplish this, I used a chromatin immunoprecipitation (ChIP) assay. In chapter 1, I describe the preparation of reagents needed for the ChIP assays. I expressed TBP and two of the three subunits of DmSNAPc in bacteria. I then purified the proteins and used them for polyclonal antibody production.

In chapter 2, I demonstrate that the antibodies can be used in ChIPs to detect DmSNAP43, DmSNAP50 and TBP bound to the endogenous U1 promoter *in vivo*. I then generated cell lines stably transfected with reporter constructs for the U1 wild type promoter or the U1 promoter that contained a U6 PSEA. Interestingly, my ChIPs indicated that DmSNAPc assembled on both types of promoters. On the other hand, TBP assembled only on the wild type promoter. These results are consistent with a model in which DmSNAPc assumes a conformation on the U6 PSEA that prevents the assembly of a Pol II transcription pre-initiation complex.

GENERAL INTRODUCTION

GENERAL INTRODUCTION

Distinguishing features and functions of U-snRNAs

The U1, U2, U4, U5 and U6 small nuclear RNAs (snRNAs) are a metabolically stable class of RNA molecules that reside in eukaryotic nuclei. U-snRNAs were originally identified as being rich in uridylic acid, which differentiate them from messenger RNA (mRNA) or ribosomal RNA (rRNA) or transfer RNA (tRNA). However, upon discovery of new snRNAs, some of them were found not to contain an unusually high proportion of uridylic acid. The importance of the U-snRNAs is reflected by the fact that predicted secondary structures among orthologous U-snRNAs from distant species are virtually identical (Reddy and Busch, 1988).

Each of the snRNAs associates with about 6-10 polypeptides to form a small nuclear ribonucleoprotein particle (snRNP) (Berget and Robberson, 1986; Chabot et al., 1985; Krainer and Maniatis, 1985; Reddy and Busch, 1988). The U4, U5, and U6 snRNPs interact to form a single particle called the U4-U5-U6 tri-snRNP complex (Maniatis and Reed, 1987; Wassarman and Steitz, 1992). The snRNPs that contain the U1, U2, U4, U5 and U6 snRNAs are involved in splicing the precursors of mRNAs as components of the spliceosome (Guthrie, 1991; Sharp, 1994; Steitz, 1998).

Function of small nuclear ribonucleoprotein particles in pre-mRNA splicing

The first step of spliceosome formation is the U1 snRNP binding specifically to the 5′splice site in a mRNA precursor. This happens because the 5′ terminal region of U1 snRNA is complementary to the 5′splice site of intron-exon junctions. The second step in spliceosome formation is the specific binding of the U2 snRNP at the branch site, which also involves specific base pairing that "bulges out" the branch point adenosine (Wu and Manley, 1989; Zhuang and Weiner, 1989). After the binding of the U1 and U2 snRNPs to the pre-mRNA, the U4-U5-U6 tri-snRNP complex joins in. The U5 snRNA initially binds to a conserved exon sequence adjacent to the 5′ splice site (Wyatt et al., 1992). Following an ATP-dependent conformational change, the active site is created where the catalysis can take place.

The first step in catalysis involves the nucleophilic substitution at the 5′ splice junction by the 2′-hydroxyl group of the branch point adenosine. The resulting products are the lariat intermediate and the free 5′ exon. During this process, the U5 snRNA shifts and makes a firmer link to the 5′ exon, while the pairing between U1 snRNA and intron sequence is destabilized. The U6 snRNA take the place of U1 in binding to these sequences (Wassarman and Steitz, 1992). The U6 snRNA forms bonds to both the 5′ splice site and the branch point. The U1 and U4 snRNPs then leave the spliceosome.

The second step of the catalysis reactions involves the nucleophilic substitution at the 3′ splice junction by the 3′ OH group of the newly freed 5′ exon. Here, U5 retains its contact with the free 5′ exon and also establishes a new contact with a 3′ exon sequence

immediately downstream from the 3′ splice junction (Sontheimer and Steitz, 1993). It is probable that U5 anchors the 5′ exon after the first step (preventing it from dissociating), and then aligns it with the 3′ exon. Certain downstream elements ("exonic enhancers") such as 5′ splice sites of the next intron appear to be important in this step (Chiara and Reed, 1995). The mechanisms of the catalysis remain unknown and are the subject of extensive investigation.

Genes coding for the snRNAs

Genes that encode the snRNAs are present in multiple copies in eukaryotic genomes. As an example, in the fruit fly *Drosophila melanogaster*, there are several copies of each of the snRNA genes (Saluz et al., 1988). More specifically, there are 5 copies of true U1 snRNA genes, 5 copies of U2 snRNA genes, 4 copies of U4 snRNA genes, 7 copies of U5 snRNA genes and 3 copies of U6 snRNA genes in *D. melanogaster* (Hernandez et al., 2007). In humans, there are about 30 copies per haploid genome of U1 snRNA genes, and about 10 copies of the U2 genes (Hammarstrom et al., 1984; Lindgren et al., 1985) (Pavelitz et al., 1995). There are 5 copies of U5 genes (Wyatt et al., 1992). There are 9 full-length U6 loci in the human genome, five of them being true genes (Domitrovich and Kunkel, 2003) .

Expression of genes coding for the snRNAs

The genes coding for the U1 to U5 snRNAs are transcribed by RNA polymerase II (Pol II) but the U6 genes are transcribed by RNA polymerase III (Pol III) (Dahlberg

and Lund, 1988; Hernandez et al., 2007; Hernandez, 1992; Lobo and Hernandez, 1994; Parry et al., 1989; Su et al., 1997). According to dogma, DNA-dependent RNA polymerase I, II, and III direct transcription of various classes of genes with distinct promoter structures. Each RNA polymerase recognizes its target promoters via protein-protein interactions with a largely distinct set of basal transcription factors bound to core promoter elements. The genes that encode the snRNAs are interesting in that both the U1-U5 class of genes and the U6 genes are atypical transcription units, yet the promoters of both classes of genes have certain features in common with each other (Fig. I.1).

Most snRNA gene core promoters contain two cis-acting elements: a unique Proximal Sequence Element (PSE) about 55 bp upstream of the transcription start site and a TATA box or another conserved sequence at the usual position of the TATA box (Fig I.1). Vertebrate Pol II transcribed snRNA genes are an exception in that they lack any recognizably conserved sequence at the location the TATA box (Fig. I.1). Vertebrate U6 snRNA gene promoters in contrast contain TATA boxes. In the presence of the upstream PSE, the TATA box acts as a dominant element for determining the RNA polymerase III specificity of vertebrate U6 snRNA genes. Moreover, the vertebrate U1 and U6 gene PSEs are functionally interchangeable and do not appear to contribute to RNA polymerase specificity (Lobo and Hernandez, 1989; Mattaj et al., 1988).

In plants, both Pol II- and Pol III-transcribed snRNA genes contain a TATA box (Fig.I.1). In this case, RNA polymerase specificity is determined by a 10 bp difference in spacing that exists between the TATA box and the Upstream Sequence Element (USE)

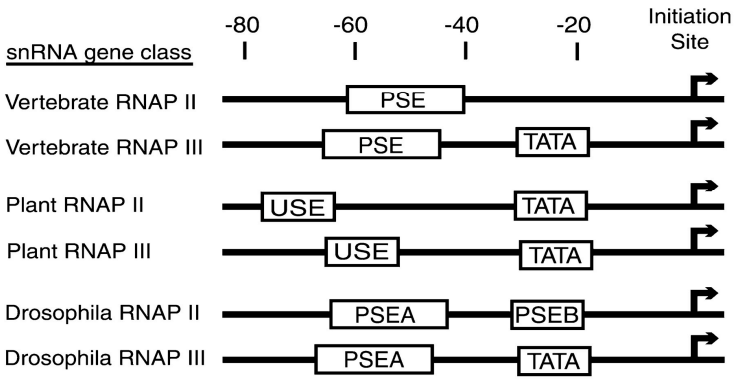


Figure I.1. **The approximate locations of the cis-acting elements conserved in the basal promoters of the snRNA genes of vertebrates, plants, and fruit flies.** Similarities and differences between the snRNA genes transcribed by RNA polymerase II and by RNA polymerase III are indicated.

(Fig. I.1). Inserting or deleting 10 bp of DNA between the USE and the TATA box of plant Pol III or Pol II snRNA promoters respectively switched the RNA polymerase specificity of the plant snRNA genes (Waibel and Filipowicz, 1990). Similarly to vertebrate PSEs, the plant USEs are interchangeable and do not contribute to RNA polymerase specificity (Kiss et al., 1991; Waibel and Filipowicz, 1990).

In *Drosophila*, the PSE is referred to as the PSEA to distinguish it from a second conserved core promoter element, called PSEB, present in the promoters of all *Drosophila* Pol II transcribed snRNA genes (Fig.I.1) (Hernandez et al., 2007; Zamrod et al., 1993). *Drosophila* U6 genes, similar to all other organisms studied, are transcribed by Pol III and contain canonical TATA boxes (Fig. I.1).

Fig I.2. shows a comparison of the promoter sequences of the *Drosophila* U1, U2, U4, U5 and U6 snRNA genes. Those transcribed by Pol II (upper part of figure) contain a well-conserved PSEA and a less well-conserved PSEB. The U6 genes (lower part of Fig. I.2) contain a similarly well-conserved PSEA and an 8 bp TATA box.

From earlier, less-comprehensive sequence comparisons, our group had defined the PSEA as a 21 bp sequence (region indicated by the un-shaded parts of the rectangles at the top and bottom of Fig. I.2). The comprehensive comparison shown in Fig. I.2 indicates that there is some sequence conservation extending a few nucleotides beyond this region in the 5′ direction. Similarly, the PSEB was originally defined as an 8 bp sequence, but it may extend a nucleotide further in the 5′ direction (Fig. I.2). However,

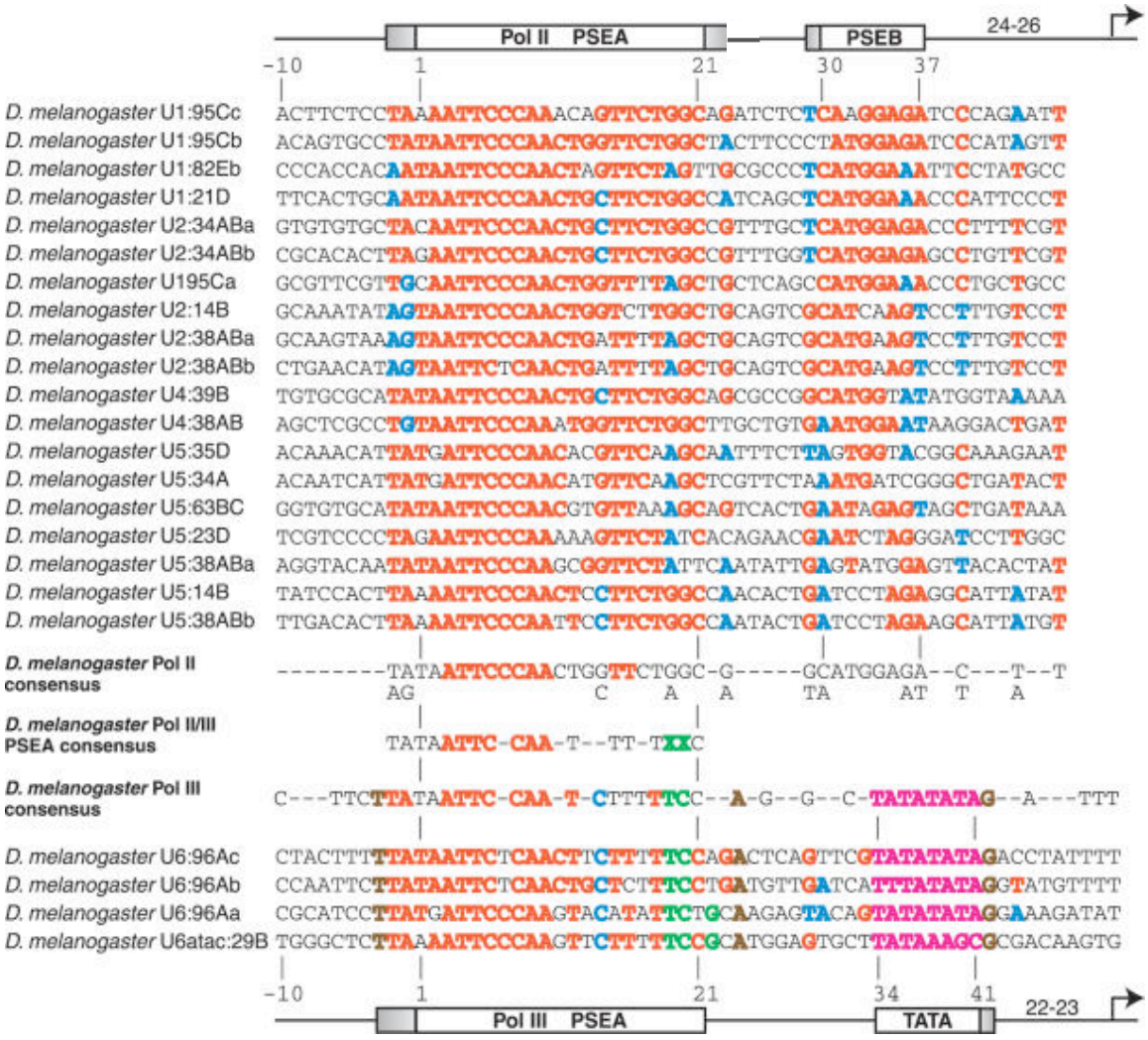
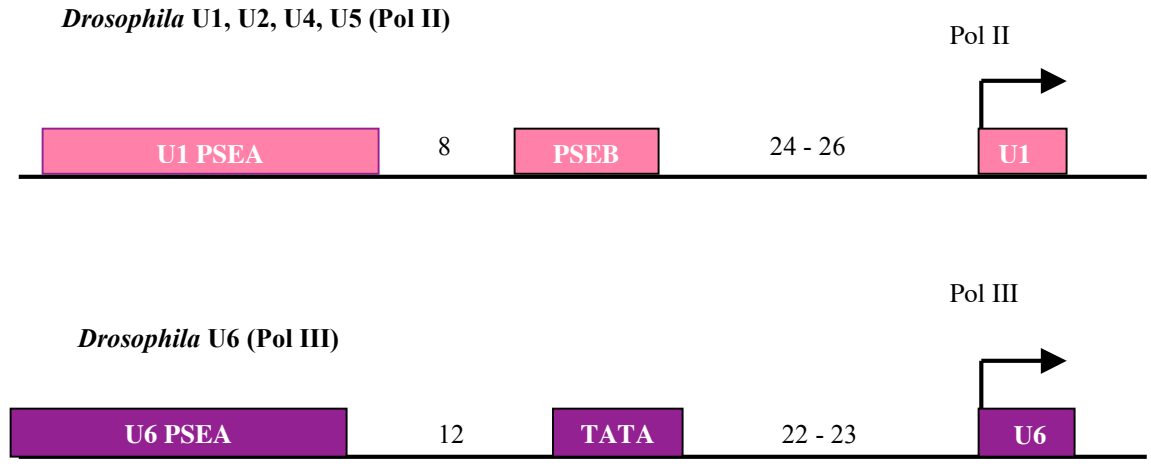


Figure I.2**. Alignment and analysis of snRNA gene promoters of *D. melanogaster*.** The sequences of the promoter regions of nineteen *D. melanogaster* snRNA genes transcribed by RNA polymerase II are shown in the upper section of the figure. In the lower section of the figure the promoter of three *D. melanogaster* U6 sequences and one U6-atac sequence are aligned and compared.

for historical and practical purposes, it remains simpler to still consider the *Drosophila* PSEA as a 21 bp sequence and the PSEB as an 8 bp sequence.

With these limitations taken into consideration, Fig. I.3A schematically diagrams the placement or location of the promoter elements in the *Drosophila* snRNA genes relative to the transcription start site. In the Pol II transcribed snRNA genes, a separation of 8 bp is strictly conserved between the PSEA and the PSEB. In contrast, in the U6 gene promoters, the PSEA and TATA box are always separated by 12 bp.

The sequence comparisons and diagrams shown in Figs.I.2 and I.3B reveal another important property of the PSEAs. There are "conserved nucleotide differences" at certain positions of the Pol II and Pol III PSEAs of *Drosophila* snRNA genes. At position 19, the Pol II PSEAs always contain an A or G, but the U6 PSEAs always have a T. At position 20, the Pol II PSEAs always contain a G, but the U6 PSEAs always have a C. Finally, at position 14 all the U6 PSEAs have a C, but C's are rare at this position in the Pol II PSEAs. Figure I.3B shows the nucleotide sequences of the specific U1 (U1:95Ca) and U6 (U6:96Ab) promoter elements used in the experiments described in the following section. The PSEA differed at only 5 of 21 nucleotide positions (7, 14, 16, 19 and 20). The PSEB and TATA sequences differed at 5 of 8 positions.




```
                                  7              14  16     19 20
Drosophila U1:95Ca PSEA :    TAATTCCCAACTGGTTTTAGC
                                   *              *  *     * *
Drosophila U6:96Ab PSEA :    TAATTCTCAACTGCTCTTTCC

Drosophila U1:95Ca PSEB :    CATGGAAA
                             ** **  *
Drosophila U6:96Ab TATA :    TTTATATA
```

Figure I.3. **(A) Schematic comparison of *Drosophila* U1 and U6 snRNA gene promoter structure.** There is a strictly conserved 8 bp spacing between the PSEA and the PSEB of the fly Pol II transcribed snRNA genes. In U6 promoters, in contrast, a 12 bp spacing between the U6 PSEA and the TATA box is strictly conserved. **(B) Sequence comparisons between the two PSEAs and between the PSEB and the TATA sequence from a *Drosophila* U1 gene (U1-95Ca) and a U6 gene (U6:96Ab).** The wild type U1 and U6 PSEA sequences used in our studies differed at only 5 of 21 positions. The five differences at positions 7,14,16,19, and 20 are indicated by asterisks. The PSEB and TATA box differ at 5 of 8 positions.

In *Drosophila*, the PSEA is the dominant element for determining the RNA polymerase specificity of snRNA genes

In *Drosophila*, both the sequence and the location of the promoter elements from various snRNA genes are unusually well conserved compared to other animals (Figs. I.2 and I.3) (Das et al., 1987; Hernandez et al., 2007; Lo and Mount, 1990). This makes *Drosophila* an unusually approachable system for studying the machinery of RNA polymerase specificity.

To investigate the determinatnts of RNA polymerase specificity, mix-and-match templates were constructed by our lab that contained all possible combinations of U1 or U6 PSEA, 8 or 12 bp spacing, and PSEB or TATA box. These were then transcribed *in vitro* (Jensen et al., 1998). For those experiments, the U1 and U6 PSEAs differed at only five nucleotide positions (Fig. I.3B). The constructs that contained the U1 PSEA were transcribed by Pol II, and those that contained the U6 PSEA were transcribed by Pol III. The PSEB and TATA elements, as well as the 8 vs. 12 bp spacing, affected transcriptional efficiency but did not directly affect the choice of RNA polymerase *in vitro* (Jensen et al., 1998). Thus, exchanging the U1 and U6 PSEAs resulted in switching the RNA polymerase specificity of the promoter as measured by *in vitro* transcription assays (Jensen et al., 1998; McNamara-Schroeder et al., 2001).

When expression of similar constructs was examined *in vivo* by transient transfection, the results were somewhat different but still revealed the dominance of the

PSEA (Jensen et al., 1998; McNamara-Schroeder et al., 2001). Transfection experiments were done with reporter constructs that contained U1 and U6 promoters with "swapped" PSEAs or heavily mutated PSEAs. Substitution of the U6 PSEA into the U1 promoter, or conversely of the U1 PSEA into the U6 promoter, inactivated both promoters *in vivo*. Transcription of reporter constructs that contained the "wrong" PSEA was reduced to levels not significantly different from complete mutation of the PSEA (Fig. I.4).

These results clearly indicated that the U1 PSEA cannot function for Pol III transcription and the U6 PSEA cannot function for Pol II transcription, even though they differ at only 5 of 21 nucleotide positions.

Characterization of the *Drosophila* PSEA-binding protein

The PSE-binding protein (PBP) was first identified in the human system in HeLa cell extract (Bungert et al., 1992; Waldschmidt et al., 1991). It was further characterized and variously termed proximal transcription factor (PTF) (Murphy et al., 1992; Yoon et al., 1995) or snRNA activator protein complex (SNAPc) (Henry et al., 1995; Sadowski et al., 1996). This factor was capable of activating both Pol II and Pol III transcription from snRNA promoters (Goomer et al., 1994; Henry et al., 1995; Sadowski et al., 1993; Waldschmidt et al., 1991; Yoon et al., 1995). The human protein contains integral polypeptide subunits with apparent molecular weights of approximately 19, 43, 45, 50 and 190 kDa. The gene for each has been cloned (Bai et al., 1996; Henry et al., 1998; Henry et al., 1995; Sadowski et al., 1996; Wong et al., 1998; Yoon and Roeder, 1996).

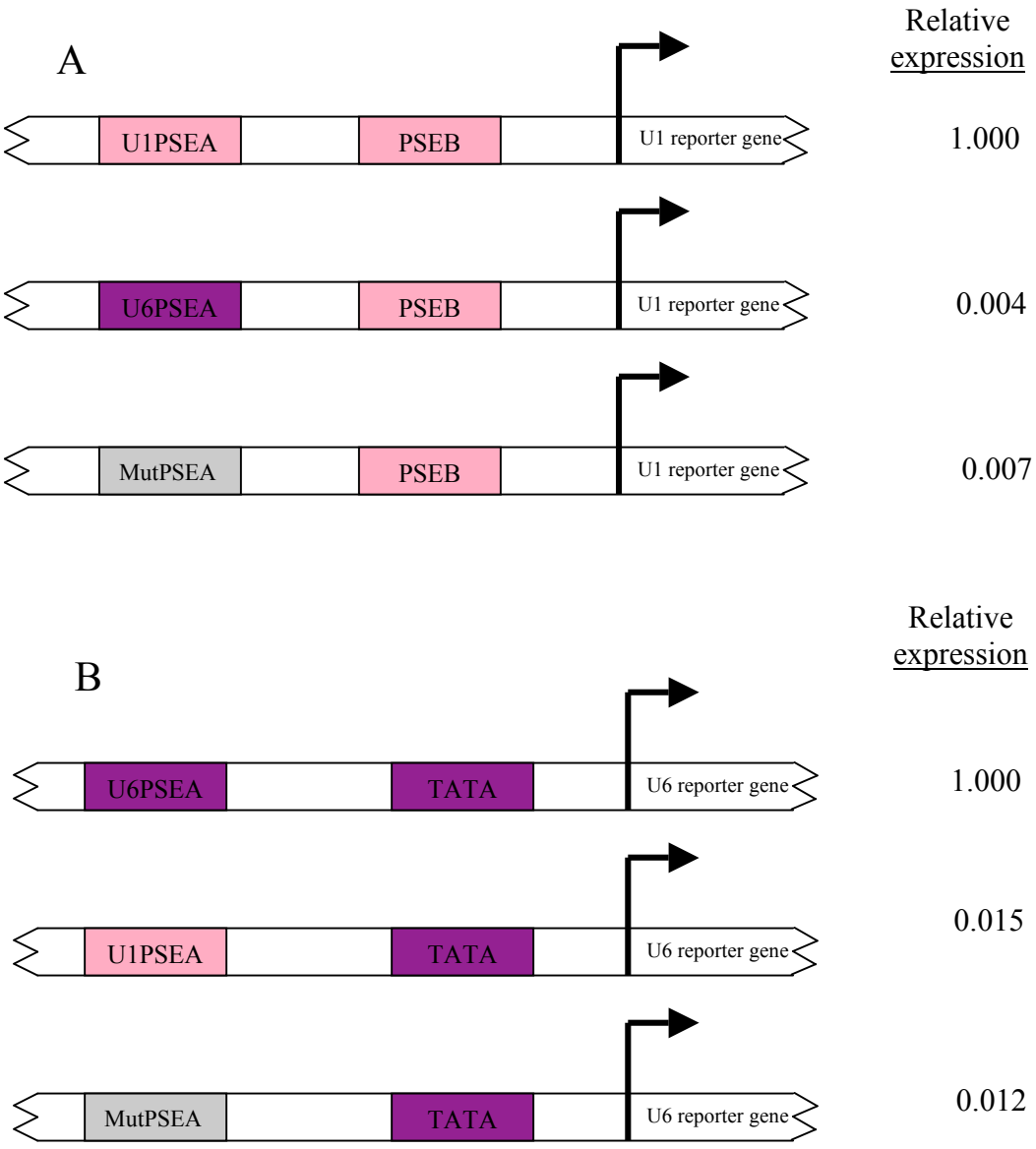


Figure I.4. **Relative Expression of U1 and U6 templates with wild type promoters or promoters with switched PSEAs.** S2 cells were transfected with the constructs shown and expression levels were determined by primer extension assays.
(A) Constructs contained the U1 5'-flanking DNA with either the wild type U1 PSEA, the wild type U6 PSEA, or mutant PSEA. The relative expression levels of the three constructs are shown in the column at the right.
(B) Constructs contained the U6 gene with wild type U6 PSEA, the wild type U1 PSEA, or mutant PSEA. The relative expression levels of the three constructs are shown in the column at the right.

These proteins, and the genes that encode them, are termed SNAP19, SNAP43, SNAP45, SNAP50 and SNAP190.

The *Drosophila melanogaster* PSEA-binding protein (DmPBP, re-termed DmSNAPc) has been characterized in our lab (Lai et al., 2005; Li et al., 2004; Su et al., 1997; Wang and Stumph, 1998). DmSNAPc binds to the U1 and U6 PSEAs and can activate transcription of the *Drosophila* U1 and U6 snRNA genes *in vitro* (Su et al., 1997). Our lab also showed that DmSNAPc contains three subunits: DmSNAP43, DmSNAP50, and DmSNAP190. These three subunits were originally identified by site-specific protein-DNA photo-cross-linking (Wang and Stumph, 1998). The cloning and characterization of the genes for the three subunits of DmSNAPc has also been carried out in our lab (Li et al., 2004). These three genes encode proteins with similarity to the SNAP43, SNAP50, and SNAP190 subunits of human SNAPc (Li et al., 2004).

The photo-cross-linking studies in our lab further indicated that the cross-linking pattern of DmSNAPc to the U1 PSEA and U6 PSEA was different (Li et al., 2004; Wang and Stumph, 1998). This suggests that the conformations of the U1 and U6 protein-DNA-complexes are different. Additional results from our lab indicated that the U1 and U6 PSEAs are both bent by a similar degree toward the face of the DNA contacted by the DmSNAP43 subunit (Hardin et al., 2000). Thus, we believe that the conformational differences in the U1 and U6 DNA-DmSNAP complexes exist primarily at the level of the protein rather than the level of the DNA (Fig.I. 5).

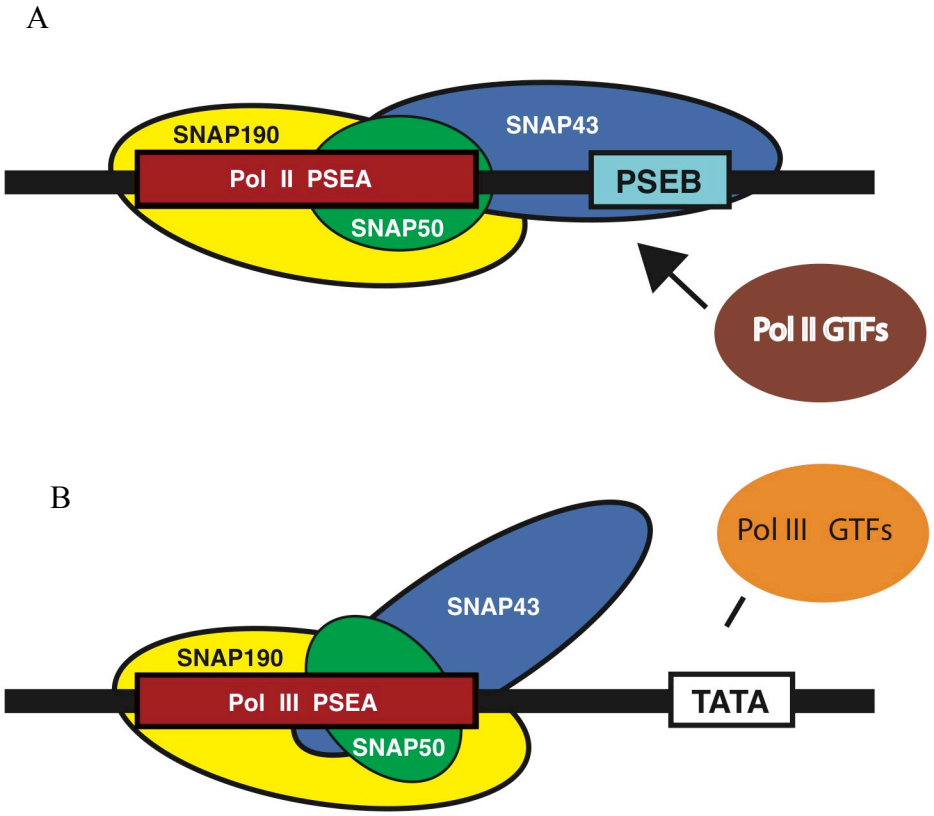


Figure I.5. **When DmSNAPc binds to U1 and U6 PSEAs, it adopts different conformations (as determined by site-specific protein-DNA photo-cross-linking).** We believe these conformational differences result in the recruitment of Pol II general transcription factors (GTFs) and RNA polymerase II to the U1 promoter (A) and of Pol III GTFs and RNA polymerase III to the U6 promoter (B).

We believe these conformational differences on U1 and U6 PSEAs result in the recruitment of RNA polymerase II basal transcription factors on a U1 promoter and RNA polymerase III basal transcription factors on a U6 promoter (Fig. I.5). The Hernandez and Reinberg groups have shown that all of the pol II GTFs (TBP, TFIIA, TFIIB, TFIIE, TFIIF), with the possible exception of TFIIH, are involved in U1 transcription in the human system (Kuhlman et al., 1999). Earlier work in our lab demonstrated a TBP requirement for U1 transcription also in fruit flies (Zamrod et al., 1993).

Human U6 transcription requires TBP, Bdp and a unique form of the TFIIB related factor termed Brf2. Surprisingly, TBP is not used for U6 transcription in fruit flies. Instead, the TBP-related factor TRF1 is used for U6 and other Pol III transcription in *Drosophila* (Isogai et al., 2007; Takada et al., 2000).

Subject matter of this dissertation

According to the working model (Fig.I.5), conformational differences in DmSNAPc, induced by binding to a U1 vs. U6 PSEA, lead to the differential recruitment of polymerase specific factors and to subsequent recruitment of either Pol II or Pol III. Previous transient transfection assays demonstrated that Pol II was unable to initiate transcription from a U1 promoter that contained a five-nucleotide substitution that converted the U1 PSEA to a U6 PSEA (Lai et al., 2005; McNamara-Schroeder et al., 2001). We were therefore interested in determining the answer to the following question: Which step of Pol II pre-initiation complex (PIC) assembly on the U1 promoter is disrupted *in vivo* when the U1 PSEA is changed to a U6 PSEA? Specifically, 1) does

DmSNAPc still assemble on the U1 promoter that contains the U6 PSEA substitution; and 2) does TBP still assemble on the U1 promoter that contains the U6 PSEA substitution? These questions are illustrated in Fig. I.6.

In chapter 1, I describe the preparation of reagents for a chromatin immunoprecipitation (ChIP) assay to answer the above questions. I expressed the DmSNAP43 and DmSNAP50 subunits of DmSNAPc in bacteria as 6xHis tagged fusion proteins and purified these by Ni^{+2} column chromatography. I then used the purified proteins for polyclonal antibody production in rabbits. I performed band shift and super-shift assays, which confirmed that the rabbits produced antibodies against each subunit. I also over-expressed the TATA-binding protein (TBP) in bacterial cells. I purified the TBP by Ni^{+2} column and S tag chromatography and used the purified TBP for polyclonal antibody production in rabbits. I also performed western blots to test antibody specificity against TBP.

In chapter 2, I verified that the antibodies could be used in ChIPs to detect DmSNAP43, DmSNAP50 and TBP bound to the endogenous U1 promoter *in vivo*. Subsequently, I generated *Drosophila* S2 cell lines stably transfected with reporter constructs driven from the wild type U1 promoter (as a positive control) or mutant U1 promoter that contained a U6 PSEA. I then used primer extension assays to determine the relative transcriptional activity of the wild type promoter and of the mutant promoter in the stably transfected cells. These assays confirmed that substitution of the U6 PSEA inactivated the U1 promoter.

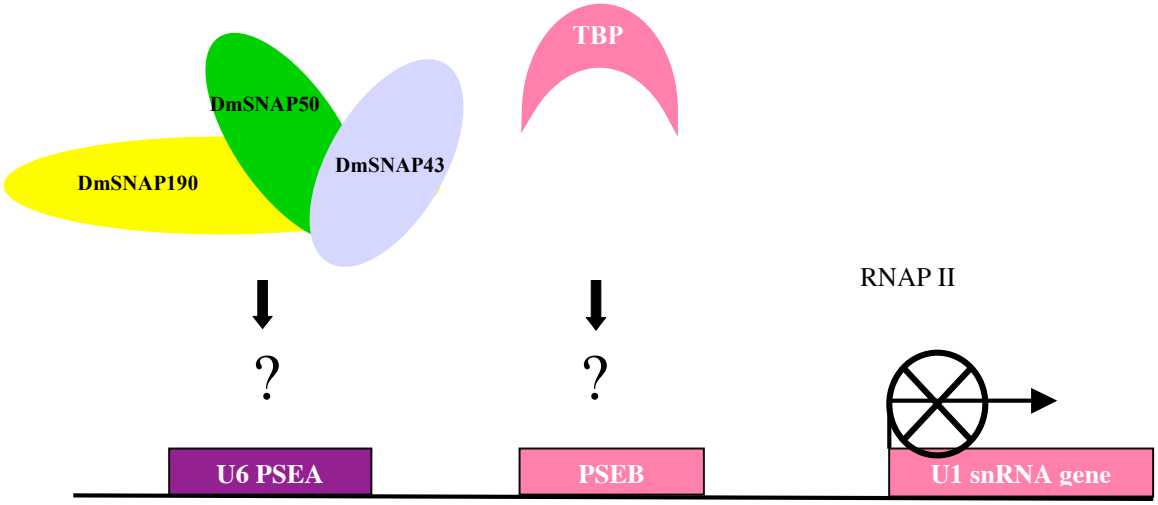


Figure I.6. **Subject matter of this dissertation.** Changing the U1 PSEA to a U6 PSEA inactivates the U1 promoter *in vivo*. This raised the following questions: If the U1 PSEA of the U1 promoter is changed to a U6 PSEA: (1) Does DmSNAPc still assemble on the promoter? and (2) does TBP still assemble on the promoter?

I then performed ChIP assays to determine if DmSNAPc and TBP assembled on the mutant as well as the wild type promoter *in vivo*. Interestingly, our results indicate that DmSNAPc assembles on the U1 promoter that contains a U6 PSEA, even though the gene was inactive by primer extension assay. In contrast TBP assembly failed to occur on the mutant promoter. This suggests that DmSNAPc, when bound to a U6 PSEA, is in the wrong conformation to recruit TBP to the U1 promoter. This is apparently one mechanism that ensures the RNA polymerase II specificity of the U1 promoter.

REFERENCES

Bai, L., Wang, Z. X., Yoon, J. B., and Roeder, R. G. (1996). Cloning and characterization of the ß 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.

Berget, S. M., and Robberson, B. L. (1986). U1, U2, and U4/U6 small nuclear ribonucleoproteins are required for *in vitro* splicing but not polyadenylation. Cell *46*, 691-696.

Bungert, J., Kober, I., Düring, F., and Seifart, K. H. (1992). Transcription factor eUSF is an essential component of isolated transcription complexes on the duck histone H5 gene and it mediates the interaction of TFIID with a TATA-deficient promoter. J Mol Biol *223*, 885-898.

Chabot, B., Black, D. L., LeMaster, D. M., and Steitz, J. A. (1985). The 3' splice site of pre-messenger RNA is recognized by a small nuclear ribonucleoprotein. Science *230*, 1344-1349.

Chiara, M. D., and Reed, R. (1995). A two-step mechanism for 5' and 3' splice-site pairing. Nature *375*, 510-513.

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, G., Henning, D., and Reddy, R. (1987). Structure, organization, and transcription of *Drosophila* U6 small nuclear RNA genes. J Biol Chem *262*, 1187-1193.

Domitrovich, A. M., and Kunkel, G. R. (2003). Multiple, dispersed human U6 small nuclear RNA genes with varied transcriptional efficiencies. Nucleic Acids Res *31*, 2344-2352.

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.

Guthrie, C. (1991). Messenger RNA splicing in yeast: clues to why the spliceosome is a ribonucleoprotein. Science *253*, 157-163.

Hammarstrom, K., Westin, G., Bark, C., Zabielski, J., and Petterson, U. (1984). Genes and pseudogenes for human U2 RNA. Implications for the mechanism of pseudogene formation. J Mol Biol *179*, 157-169.

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., 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.

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. (1992). Transcription of vertebrate snRNA genes and related genes, In Transcriptional Regulation, S. L. McKnight, and K. R. Yamamoto, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 281-313.

Isogai, Y., Takada, S., Tjian, R., and Keles, S. (2007). Novel TRF1/BRF target genes revealed by genome-wide analysis of *Drosophila* Pol III transcription. Embo J *26*, 79-89.

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.

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

Krainer, A. R., and Maniatis, T. (1985). Multiple factors including the small nuclear ribonucleoproteins U1 and U2 are necessary for pre-mRNA splicing *in vitro.* Cell *42*, 725-736.

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.

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.

Lindgren, V., Bernstein, L. B., Weiner, A. M., and Francke, U. (1985). Human U1 small nuclear RNA pseudogenes do not map to the site of the U1 genes in 1p36 but are clustered in 1q12-q22. Mol Cell Biol *5*, 2172-2180.

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

Lobo, S. M., and Hernandez, N. (1989). A 7 bp mutation converts a human RNA polymerase II snRNA promoter into an RNA polymerase III promoter. Cell *58*, 55-67.

Lobo, 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.

Maniatis, T., and Reed, R. (1987). The role of small nuclear ribonucleoprotein particles in pre-mRNA splicing. Nature *325*, 673-678.

Mattaj, I. W., Dathan, N. A., Parry, H. D., Carbon, P., and Krol, A. (1988). Changing the RNA polymerase specificity of U snRNA gene promoters. Cell *55*, 435-442.

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.

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

Pavelitz, T., Rusché, L., Matera, A. G., Scharf, J. M., and Weiner, A. M. (1995). Concerted evolution of the tandem array encoding primate U2 snRNA occurs in situ, without changing the cytological context of the RNU2 locus. EMBO J *14*, 169-177.

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., Kobayashi, R., and Hernandez, N. (1996). The SNAP45 subunit of the small nuclear RNA (snRNA) activating protein complex is required for RNA polymerase II and III snRNA gene transcription and interacts with the TATA box binding protein. Proc Natl Acad Sci USA *93*, 4289-4293.

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.

Saluz, H., Dudler, R., Schmidt, T., and Kubli, E. (1988). The localization and estimated copy number of *Drosophila melanogaster* U1, U4, U5 and U6 snRNA genes. Nucleic Acids Res *16*, 3582.

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

Sontheimer, E. J., and Steitz, J. A. (1993). The U5 and U6 small nuclear RNAs as active site components of the spliceosome. Science *262*, 1989-1996.

Steitz, T. A. (1998). A mechanism for all polymerases. Nature *391*, 231-232.

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.

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.

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

Waldschmidt, R., Wanandi, I., and Seifart, 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.

Wassarman, D. A., and Steitz, J. A. (1992). Interactions of small nuclear RNA's with precursor messenger RNA during *in vitro* splicing. Science *257*, 1918-1925.

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.

Wu, J., and Manley, J. L. (1989). Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing. Genes Dev *3*, 1553-1561.

Wyatt, J. R., Sontheimer, E. J., and Steitz, J. A. (1992). Site-specific cross-linking of mammalian U5 snRNP to the 5' splice site before the first step of pre-mRNA splicing. Genes Dev *6*, 2542-2553.

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.

Yoon, J. B., and Roeder, R. G. (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.

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.

Zhuang, Y., and Weiner, A. M. (1989). A compensatory base change in human U2 snRNA can suppress a branch site mutation. Genes Dev *3*, 1545-1552.

CHAPTER 1

Reagent Preparation For Chromatin Immunoprecipitation Assay of DmSNAPc and TBP

INTRODUCTION

The chromatin immunoprecipitation (ChIP) technique combines immunoprecipitation of chromatin fragments with the polymerase chain reaction to map sites of protein-DNA interaction *in vivo*. The method relies on the rapid cross-linking of protein/DNA complexes within the nucleus of living cells, followed by chromatin isolation, random shearing, and immunoprecipitation with antibodies directed towards proteins of interest. The amount of co-immunoprecipitated DNA is then analysed by PCR. The ChIP technique is schematically diagrammed in Fig.1.1.

The Chromatin Immunoprecipitation (ChIP) Assay

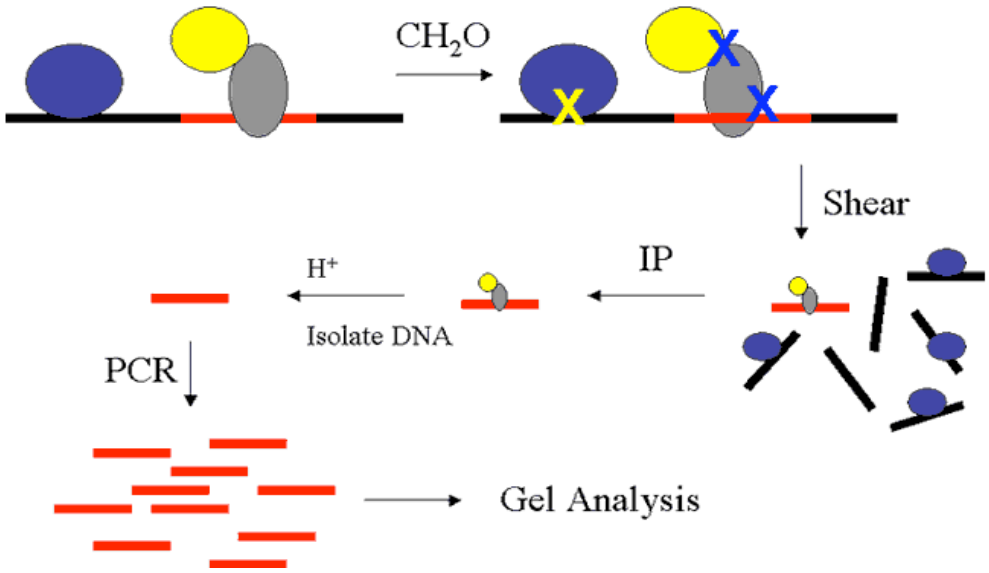


Figure 1.1. **Outline of chromatin immunoprecipitation (ChIP) assay**. Protein-DNA crosslinks are formed with formaldehyde followed by chromatin fragmentation by sonication. The DNA-protein complexes are immunoprecipitated with antibodies specific for the protein of interest and Protein A agarose beads. The crosslinks are reversed and the DNA is isolated for analysis. (This image was obtained from http://proteomics.swmed.edu/chiptochip.htm).

Although the ChIP method is powerful for analyzing *in vivo* protein-DNA interactions, appropriate experimental parameters must be empirically determined for each protein and antibody. ChIP requires a primary antibody for immunoprecipitation. In ChIP a polyclonal antibody is usually preferable to a monoclonal. Whereas monoclonal antibodies recognize only a single epitope, within a polyclonal antibody population there will be a number of antibodies that recognize different epitopes. A polyclonal population will reduce the probability that all specific epitopes will be masked within the chromatin or by the process of chemical cross-linking, so there is a better chance of a positive result by using polyclonal antibodies to recognize the protein of interest.

Previously published results in our lab demonstrated that Pol II is unable to initiate transcription from a U1 promoter that contains a 5 bp substitution that converts the U1 PSEA to a U6 PSEA. As a corollary of our model (General Introduction, Fig. I.6), we believe the presence of the U6 PSEA in the U1 promoter most likely disrupts a very early stage in pre-initiation complex (PIC) assembly. I therefore used ChIP assays to determine factors that fail to assemble at the U1 promoter *in vivo* when the promoter contains a U6 PSEA.

As preparation for ChIP, I expressed two of the three subunits of DmSNAPc in bacteria as V5-6xHis tagged fusion proteins. Also, I expressed TBP in bacteria as a 6xHis-S-tag fusion protein. I purified each of these essentially to homogeneity by metal-chelate chromatography, and in the case of TBP, by S-tag affinity chromatography.

Milligram quantities of these proteins were provided to a commercial service for polyclonal antibody production in rabbits.

Materials and Methods

Bacterial Expression and Purification of DmSNAP43 and DmSNAP50

Full-length cDNA clones that contained the DmSNAP50 and DmSNAP43 genes were purchased from Research Genetics. Each gene was amplified by PCR and re-cloned into the bacterial expression vector pCRT7-TOPO (Invitrogen), which provides a V5 and 6xHis tag at the C-terminus of the expressed protein. These DmSNAP43 and DmSNAP50 constructs are shown in Appendix A. These plasmids were used to transform competent E. coli BL21/dnaY for individual expression of DmSNAP43 and DmSNAP50. A detailed protocol for bacterial expression and purification is given in Appendix B. Briefly, cells were grown in two liters of Luria Broth (LB) to an optical density at 600 nm of approximately 0.5, and recombinant protein expression was induced by addition of IPTG to a concentration of 1 mM. Following three hours of induction cells were collected and lysed in a denaturing 6 M guanidine lysis buffer.

The 6xHis-tagged DmSNAP43 and DmSNAP50 were each purified by two applications to nickel chelating affinity resin. The first application was under denaturing conditions and elution was carried out at pH 4.0 in 8 M urea containing buffer. The pH for the pooled fractions was then adjusted to pH 7.8. The protein was then applied to fresh nickel chelating resin, bound to the resin for two hours at room temperature,

transferred to the cold room and washed with 500 mM NaCl, 25 mM sodium phosphate (pH 8.0), and eluted with 500 mM imidazole, 150 mM NaCl and 25 mM sodium phosphate (pH 8.0). Fractions containing significant amounts of protein were pooled and dialyzed against 150 mM NaCl and 25 mM sodium phosphate (pH 8.0) to remove the imidazole. Purified protein was electrophoresed through 10% SDS-polyacrylamide gels and stained with Coomassie blue or transferred to PVDF membranes. V5-tagged proteins were detected by using anti-V5/alkaline phosphatase conjugated antibody from Invitrogen.

Bacterial Expression and Purification of TBP

A full-length cDNA clone coding for *D. melanogaster* TBP was purchased from Research Genetics and re-cloned into the pET-30b (+) vector (Novagen). This construct expresses TBP with 6xHis and S-tags at the amino terminus of the encoded protein. The TBP construct is shown in Appendix A. The plasmid was used to transform competent Rosetta2(DE3) which improves expression of protein that contain codons rarely used in *E. coli* (AGG,AGA,AUA,GUA,CCC,GGA).

A detailed protocol for bacterial expression and purification of TBP is given in Appendix C. Briefly, cells were grown, induced and lysed as described above for DmSNAP50 and DmSNAP43. The 6xHis-S-tagged TBP was initially purified by two applications to a nickel chelating affinity resin as described above for the SNAP proteins. The TBP was then further purified by S-protein affinity chromatography.

Fractions containing significant amounts of TBP from the second nickel column were pooled and dialyzed against 20 mM Tris-HCl (pH 7.5), 0.15 M NaCl and 0.1% Triton X-100. The dialyzed TBP was applied to the S-protein resin, washed three times, and the tagged TBP was eluted with 0.2 M citrate (pH 2.0). The pH was then increased to pH 7.0 by using NaOH. The 6xHis and S-tags were then cleaved from the protein by overnight digestion with 20 units of recombinant Enterokinase (Novagen). EKapture Agarose was used to remove the Enterokinase. Untagged TBP was concentrated using centricon centrifugal devices (Amicon) prior to injection into rabbits for antibody production.

Polyacrylamide Gel Electrophorsis and Immunoblotting

Proteins were electrophoresed through SDS-10% polyacrylamide gels and either stained with Coomassie blue or prepared for immunoblotting. For immunoblotting, gels were equilibrated in transfer buffer (25 mM Tris, 190 mM glycine, and 20% methanol) for 20 minutes, and proteins were transferred to PVDF membranes for 2 hours at 100 V. Membranes were blocked for 30 minutes in TBS-Tween-20 containing 1% BSA.
Membranes were incubated for 2 hours with a 1:5000 dilution of anti-V5/alkaline phosphatase conjugated antibody (Invitrogen) in 20 ml of blocking solution for 2 hours. Following washing, membranes were developed using ~15 ml of alkaline phosphatase substrate (Promega Western Blue) until protein bands became distinct.

Electrophortic Mobility Shift Assay (EMSA)

EMSA reactions were carried out using 6 μl of native DmSNAPc partially purified from fly embryos by DEAE cellulose and heparin agarose chromatography (HA300 fraction) (Su et al., 1997). DNA probe was prepared by annealing the following two synthetic oligonucleotides that contain the PSEA sequences of the fly U1-95Ca gene: 5′ GTTCGTTGCAATTCCCAACTGGTTTTAGCTGCTCAGCCATGGAAACC 3′ and 3′ AAGCAACGTTAAGGGTTGACCAAAATCGACGAGTCGGTACCTTTGGG 5′. The double stranded DNA was then end-labeled with T4 polynucleotide kinase and γ-^{32}P-ATP. Each band shift reaction contained 50,000 cpm of radiolabeled DNA probe and 2 μg poly d(I-C). BCZ-100 buffer (200 mM Hepes pH 7.9, 50 mM $MgCl_2$, 0.1 mM $ZnCl_2$, 2 mM EDTA, 20% glycerol, 100 mM KCl, 3 mM dithiothrieitol and 0.5 mM PMSF) was added to give a final reaction volume of 20 μl.

The reactions were then incubated at 20°C for 15 minutes to allow DmSNAPc/DNA complexes to form. After 15 minutes of incubation, 2 μl of antisera produced against DmSNAP50 or DmSNAP43 were added and incubation was continued for another 15 minutes. Control reactions were carried out in which pre-immune serum was added (not shown) or in which no serum was added.

Results and Discussion

Purification of bacterially-expressed DmSNAP50 and DmSNAP43

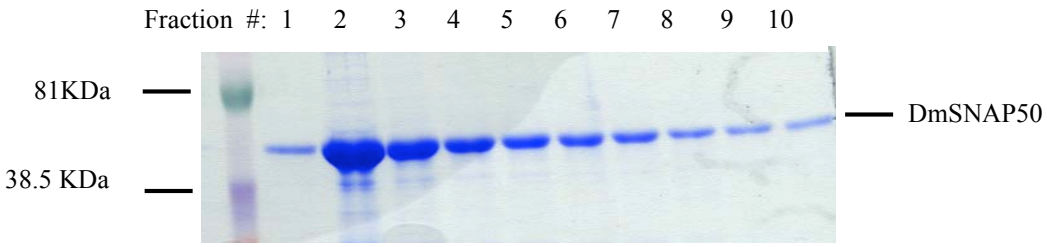
Bacterially-expressed 6xHis tagged DmSNAP50 was purified by two passes over a nickel-chelating column. Aliquots of elution fractions from the second column were

analyzed by SDS-PAGE and the gel was stained with Coomassie blue (Fig. 1.2A). A single strong band was detected that ran at approximately the expected molecular weight (Fig. 1.2A).

Since the recombinant DmSNAP50 should contain a V5 tag as well as a 6xHis tag, the identity of the purified protein was verified by probing immunoblots with anti-V5 antibodies. Fig. 1.2B shows that the purified protein was indeed recognized by the V5 antibodies. Together, these results indicate that bacterially expressed 6xHis and V5 tagged DmSNAP50 was purified essentially to homogeneity by the two passages over the nickel chelating column.

Similar experiments were carried out with bacterially expressed DmSNAP43 following a similar two rounds of purification on nickel chelating resin. In this case Coomassie blue stained gels revealed the presence of two major bands that co-purified in approximately equi-molar quantities (Fig. 1.3A). Immunoblots of the purified fraction probed with anti-V5 antibodies revealed that both bands contained the V5 tag (Fig. 3B). This strongly suggests that the lower band represents a DmSNAP43 degradation product from which an ~ 6 kDa fragment from the amino terminus has been removed. However, we reasoned that the presence of the degradation product should not be harmful in for polyclonal antibody production. Therefore, aliquots of the purified DmSNAP50 and DmSNAP43 (Figs. 1.2,1.3) were shipped to a commerical service for polyclonal antibody production in rabbits. Each protein was used to immunize two rabbits (DmSNAP43, rabbits 78 and 79; DmSNAP50, rabbits 80 and 81).

A



B

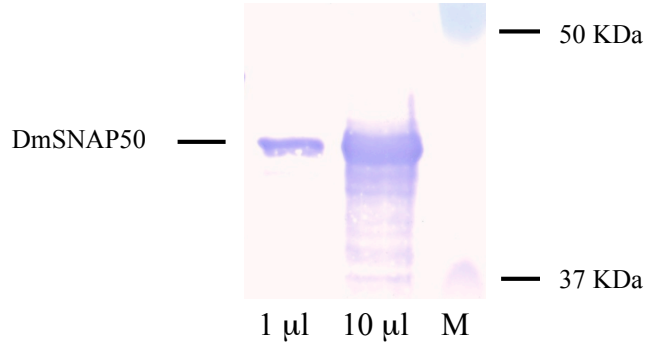


Figure 1.2. **Bacterial expression and purification of DmSNAP50 .**
(A) DmSNAP50 fractions eluting from the second of two nickel chelating columns were analyzed by SDS-PAGE and stained with Coomassie blue. (B) Either 1 µl or 10 µl aliquots of purified DmSNAP50 were analyzed by immunoblotting. The tagged DmSNAP50 was detected by using anti-V5 antibodies.

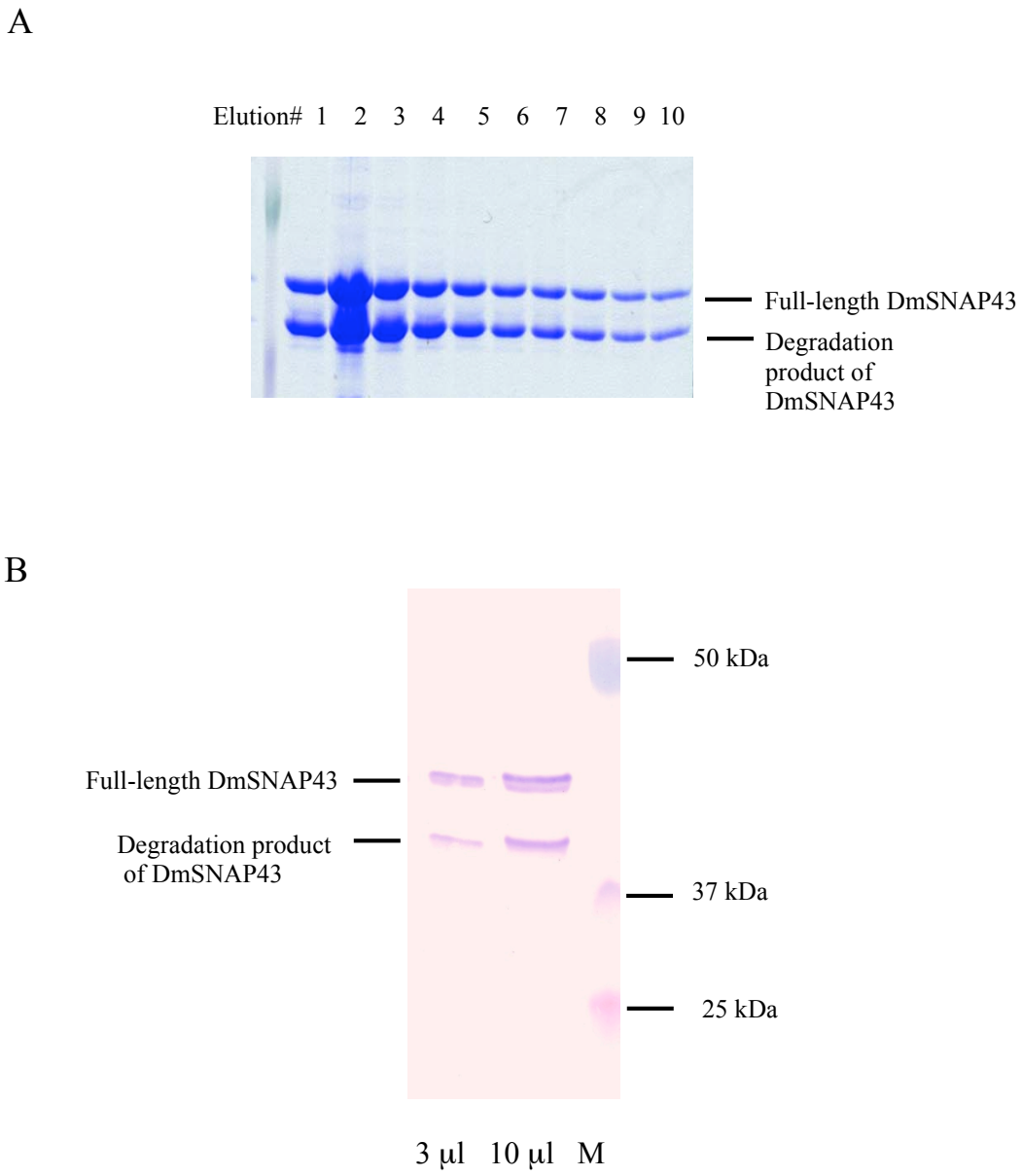


Figure 1.3. **Bacterial expression and purification of DmSNAP43.**
(A) DmSNAP43 fractions eluting from the second of two nickel chelating columns were analyzed by SDS-PAGE and stained with Coomassie blue. (B) Either 3 µl or 10 µl aliquots of purified DmSNAP43 were analyzed by immunoblotting. The tagged DmSNAP43 was detected by using anti-V5 antibodies.

The DmSNAP50 and DmSNAP43 antibodies can super-shift native DmSNAPc/DNA complexs.

Upon receiving the antisera, I wanted to determine whether the rabbits had indeed produced antibodies against DmSNAP43 and DmSNAP50 that could bind specifically to native DmSNAPc bound to DNA. To do this, I performed band shift assays and tested whether the antisera from the rabbits could super shift DmSNAPc/DNA complexes that were formed between native untagged DmSNAPc (purified from embryo nuclear extracts) and a radiolabeled DNA fragment that contained a U1 PSEA sequence. The results are shown in Fig. 1.4.

Lanes 1 and 12 of Fig. 1.4 demonstrate the position of the band-shift that results from the binding of DmSNAPc to the DNA probe. Addition of antisera from each of the rabbits immunized with either DmSNAP43 (lanes 2-5) or DmSNAP50 (lanes 6-11) resulted in a super-shift of the DmSNAPc-DNA complex. Super-shifts were not observed when pre-immune serum from any of the rabbits was used instead of the specific antibodies (data not shown). In summary, DmSNAPc-DNA complexes was super-shifted by the antisera but not by the pre-immune sera. These data suggest strongly that DmSNAP43 and DmSNAP50 polyclonal antibodies have been produced in the rabbits, and both antibodies have the ability to bind specifically to the native DmSNAPc bound to DNA.

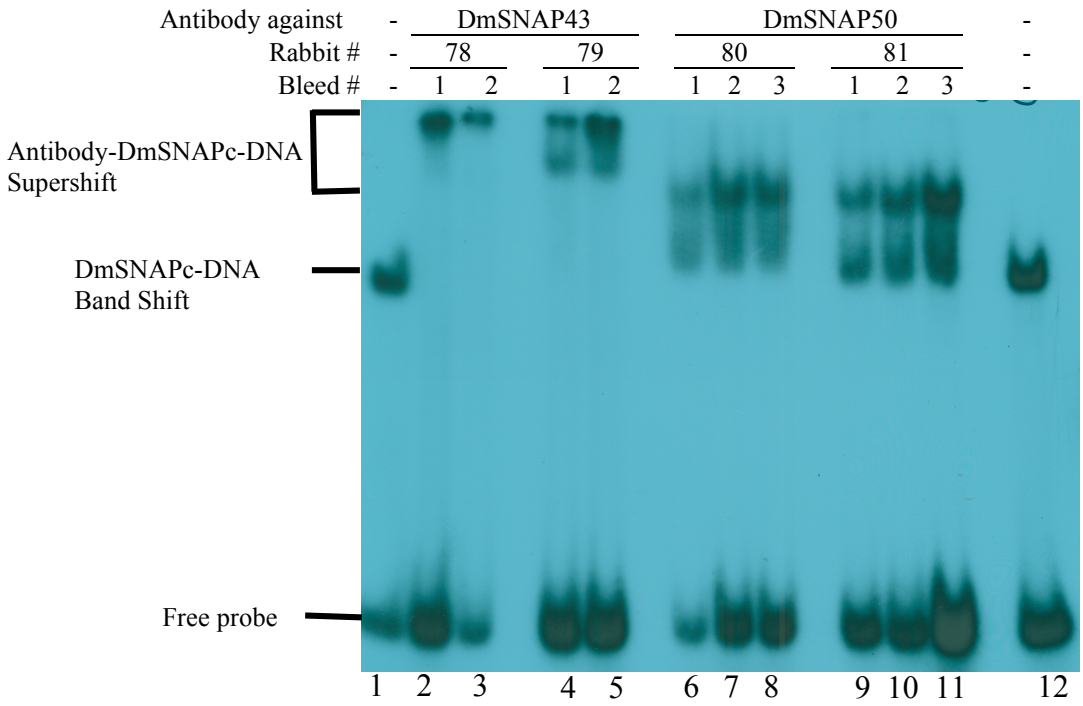


Figure 1.4. **Electrophoretic mobility shift and super-shift assay demonstrating that the antibodies produced in rabbits against DmSNAP43 and DmSNAP50 can react with native DmSNAPc bound to DNA**

Electrophoretic mobility shift assays were carried out using DmSNAPc prepared from fruit fly embryos and a probe that contains the U1 PSEA. Lanes 1 and 12 contained no added antibody to demonstrate the position of the DmSNAPc-DNA complex. In lanes 2-11, 2 μl of polyclonal antisera produced in rabbits against either DmSNAP43 (lanes 2-5) or DmSNAP50 (lanes 6-11) were added, which super-shifted the DmSNAPc-DNA complex. The specific rabbits and bleed numbers are indicated above the individual lanes.

Purification of bacterially-expressed TBP

Bacterially-expressed 6xHis and S-tag tagged TBP was purified by two passes over a nickel-chelating column and one purification using S-protein resin. The elution fraction from the S-protein purification was incubated with recombinant enterokinase to cleave off the 6xHis and S-tags. An aliquot of the elution fraction after cleaving the tag was analyzed by 10% SDS-PAGE and stained with Coomassie blue (Fig 1.5). A single major band was detected that ran at the expected molecular weight (Fig 1.5).

The TBP, from which the tags had been removed, was shipped to a commerical service for polyclonal antibody production in rabbits. After receiving the antisera, I used the antibodies for immunoblots to detect untagged TBP. A soluble nuclear fraction (SNF) isolated from fly embryos (Kadonaga, 1990; Kamakaka et al., 1991; Zamrod et al., 1993) was fractionated by SDS-PAGE and transferred to a PVDF membrane. The antibodies raised against the bacterially-expressed TBP detected a single band in the SNF that ran at the size expected for *Drosophila* TBP (Fig.1.6). These results provide reasonable evidence that the rabbits produced highly specific antibodies that can react with endogenous *Drosophila* TBP.

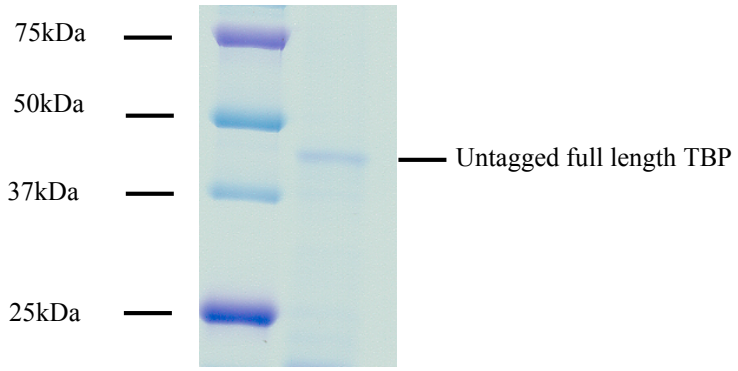


Figure 1.5. **Bacterial expression and purification of DmTBP.**
DmTBP was expressed in bacteria and purified by two passages over a nickel chelating resin followed by one passage over an S-protein affinity resin. After removal of the 6xHis and S-tags, an aliquot was analyzed by SDS-PAGE and stained with Coomassie blue.

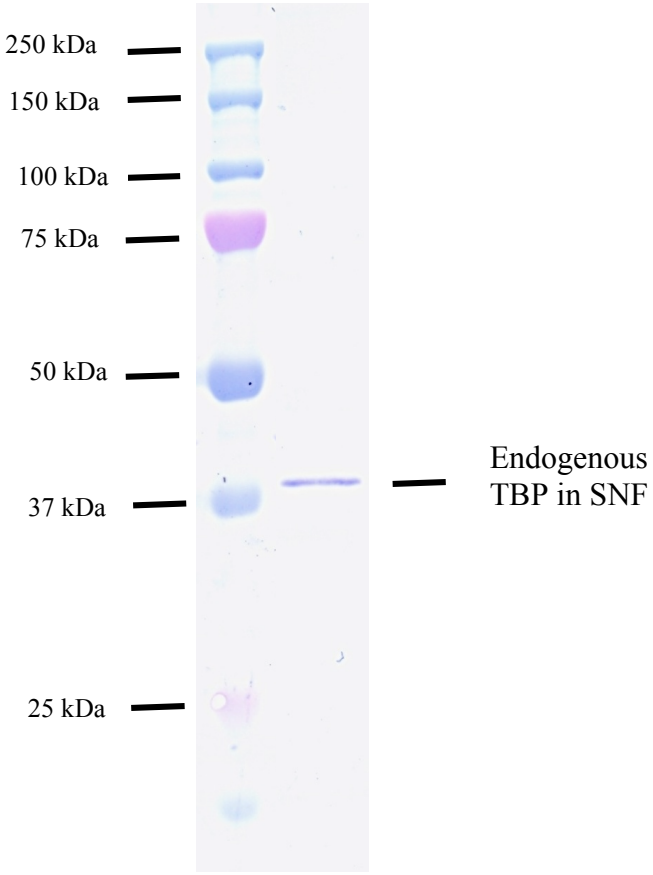


Figure 1.6. **Immunoblot demonstrating that antibodies raised against bacterially-expressed TBP can bind specifically to TBP in the *Drosophila* soluble nuclear fraction.**

REFERENCES

Kadonaga, J. T. (1990). Assembly and disassembly of the *Drosophila* RNA polymerase II complex during transcription. J Biol Chem *265*, 2624-2631.

Kamakaka, R. T., Tyree, C. M., and Kadonaga, J. T. (1991). Accurate and efficient RNA polymerase II transcription with a soluble nuclear fraction derived from *Drosophila* embryos. Proc Natl Acad Sci USA *88*, 1024-1028.

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.

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 2

TBP Recruitment to the U1 snRNA Gene Promoter Is Disrupted by Substituting a U6 Proximal Sequence Element A (PSEA) for the U1 PSEA

ABSTRACT

Most major spliceosomal snRNAs (U1, U2, U4, and U5) are synthesized by RNA polymerase II, but U6 snRNA is synthesized by RNA polymerase III. In *Drosophila melanogaster,* transcription of all snRNA genes requires a unique ~21 base pair promoter element termed the proximal sequence element A (PSEA) located ~40-60 base pairs upstream of the transcription start site. In both classes of genes, the PSEA is recognized by the multi-subunit small nuclear RNA activating protein complex (DmSNAPc). Previously we found that a 5 nucleotide substitution that changed a U1 PSEA to a U6 PSEA inactivated the U1 promoter *in vivo*. Thus, the U6 PSEA was unable to promote the formation of a functional RNA polymerase II transcription complex on the U1 promoter. We have now used chromatin immunoprecipitation assays to determine the step of RNA polymerase II pre-initiation complex assembly that is disrupted when the U1 promoter contains a U6 PSEA. Our results indicate that changing the U1 PSEA to a U6 PSEA had no effect on the ability of DmSNAPc to bind to the promoter but instead interfered with the assembly of TBP. These findings agree with a model in which DmSNAPc, bound to a U6 PSEA, assumes a conformation incapable of recruiting TBP to the U1 promoter.

INTRODUCTION

The U1, U2, U4, U5, and U6 snRNAs play essential roles in removing intron sequences from pre-mRNAs in eukaryotic cells (Guthrie, 1991; Sharp, 1994; Steitz et al., 1988). In animals, the genes that code for the U1, U2, U4, and U5 snRNAs are transcribed by RNA polymerase II (Pol II), but U6 genes are transcribed by RNA polymerase III (Pol III) (Dahlberg and Lund, 1988; Hernandez, 1992; Lobo and Hernandez, 1994; Parry et al., 1989a). Interestingly, transcription of both classes of genes is dependent upon a unique promoter element (termed the proximal sequence element, or PSE) centered approximately 50-55 base pairs (bp) upstream of the transcription start site (Dahlberg and Lund, 1988; Das et al., 1987; Das et al., 1988; Earley et al., 1984; Gruber et al., 1991; Hoffman et al., 1986; Kazmaier et al., 1987; Lescure et al., 1991; Li et al., 1996; Li et al., 1994; Lobo et al., 1990; McNamara and Stumph, 1990; McNamara et al., 1987; Morris et al., 1986; Parry et al., 1989a; Parry et al., 1989b; Southgate and Busslinger, 1989; Wendelburg and Marzluff, 1992; Zamrod et al., 1993). The PSE is recognized by the small nuclear RNA activating protein complex (SNAPc) (Sadowski et al., 1993), a multi-subunit transcription factor that has also been called PBP (Waldschmidt et al., 1991; Wanandi et al., 1993) or PTF (Murphy et al., 1992; Yoon et al., 1995).

Three subunits of SNAPc (SNAP190, SNAP50, and SNAP43) are strongly conserved throughout evolution and even have homologs in trypanosomes where tSNAPc is required for transcription of the spliced leader RNA (Das and Bellofatto, 2003; Das et

al., 2005; Schimanski et al., 2005). In higher eukaryotes, SNAPc seems to be uniquely required for transcription of the snRNAs and a number of other small stable RNAs whose genes have external Pol III promoters (Carbon and Krol, 1991; Murphy et al., 1992; Myslinski et al., 2001; Woodhams et al., 2007; Yoon et al., 1995).

In *Drosophila melanogaster*, the PSE is more specifically termed the PSEA to distinguish it from a second conserved element termed the PSEB present in the promoters of the Pol II-transcribed fly snRNA genes (Lo and Mount, 1990; Zamrod et al., 1993). The PSEB, although not essential, affects transcriptional efficiency and is located approximately 25-32 base pairs upstream of the Pol II transcription start site (Hernandez et al., 2007; Lai et al., 2005; Lo and Mount, 1990; Zamrod et al., 1993). This is the precise location that is occupied by a TATA box sequence in many Pol II promoters, but changing the PSEB to a TATA box actually decreased transcriptional efficiency of the U1 promoter *in vivo* approximately 2-fold (Lai et al., 2005).

Interestingly, the fruit fly U1 and U6 PSEAs are not functionally interchangeable even though they can both bind DmSNAPc and differ at only a few of the 21 nucleotide positions. *In vitro* transcription assays indicated that changing the U1 PSEA to a U6 PSEA (via 5 base-changes) switched the RNA polymerase specificity of the U1 promoter from Pol II to Pol III (Jensen et al., 1998; McNamara-Schroeder et al., 2001). *In vivo* (i.e., expression in transgenic flies or in transient transfection assays), however, transcription from the U1 promoter was suppressed by changing the U1 PSEA to a U6 PSEA (Lai et al., 2005; McNamara-Schroeder et al., 2001). Those experiments demonstrated that the

precise sequence of the PSEA plays a dominant role in determining the RNA polymerase specificity of snRNA promoters in fruit flies. Indeed, five other insect species that were examined have just a few conserved nucleotides in the PSEAs of the Pol II-transcribed snRNA genes that are different from those conserved in the PSEAs of the Pol III-transcribed snRNA genes (Hernandez et al., 2007). This indicates that mechanisms dependent upon the PSEA sequence are likely utilized to determine the RNA polymerase specificity of snRNA genes throughout at least the order insecta.

In agreement with the importance of sequence differences between the PSEAs of Pol II- and Pol III-transcribed insect snRNA genes, site-directed protein-DNA photo-cross-linking studies have indicated that *D. melanogaster* SNAPc (DmSNAPc) binds in different conformations to the U1 and U6 PSEAs (Li et al., 2004; Wang and Stumph, 1998). DmSNAP50 and DmSNAP43 in particular show distinct patterns of cross-linking to the U1 and U6 PSEAs. Most strikingly, DmSNAP43 cross-links to 20 bp of DNA downstream of the U1 PSEA but to only about 5 bp of DNA downstream of the U6 PSEA (Li et al., 2004). The region contacted by DmSNAP43 on the U1 promoter includes the PSEB which is located 9-16 bp downstream of the PSEA. Earlier *in vitro* studies have shown that the TATA-binding protein, TBP, is required for transcription of snRNA genes by Pol II in both vertebrates and insects (Sadowski et al., 1993; Zamrod et al., 1993). We suspect that the DmSNAP43 subunit of DmSNAPc, because it contacts the PSEB of the U1 promoter, may be involved in recruiting TBP to the unconventional PSEB sequence (Lai et al., 2005).

Because we were interested in examining the mechanism by which substitution of the U6 PSEA for the U1 PSEA inactivates the U1 promoter *in vivo*, we used chromatin immunoprecipitation (ChIP) assays to examine the step in Pol II pre-initiation complex assembly *in vivo* that is disrupted by such a PSEA substitution. Our data indicate that binding of DmSNAPc is unaffected when the U1 PSEA is changed to a U6 PSEA in the context of the U1 promoter; interestingly, however, TBP fails to assemble as a result of this 5 bp substitution within the PSEA.

MATERIALS AND METHODS

Bacterial expression of *D. melanogaster* SNAP43, SNAP50, and TBP and purification for antibody production

The genes that code for the subunits of DmSNAPc have been previously cloned and characterized (Li et al., 2004). The coding regions of the DmSNAP43 and DmSNAP50 genes were each amplified and inserted into the plasmid PCRT7/V5-His-TOPO (Invitrogen) for individual expression in *Escherichia coli* BL21(DE3). (The BL21[DE3] cells also harbored pUBS520 that contains the E. coli dnaY gene which codes for a minor arginine $tRNA_{AGA/AGG}$.) Bacterial cell lysates were prepared following a 3-hour induction with isopropyl-β-D-thiogalactopyranoside (IPTG) and milligram quantities of the DmSNAPs were purified by affinity chromatography using Invitrogen's ProBond nickel chelating resin as follows: Lysates were applied to the ProBond resin under denaturing conditions at pH 7.8, and the 6xHis tagged DmSNAPs were eluted at pH 4.0. The elutions fractions were pooled, re-adjusted to pH 7.8, and re-

applied to a fresh batch of ProBond resin under denaturing conditions. The resin was washed with buffer containing 500 mM NaCl, 25 mM sodium phosphate [pH 8], and the tagged protein was then eluted in buffer containing 150 mM NaCl, 500 mM imidazole, 25 mM sodium phosphate [pH 8]. These two steps of ProBond chromatography purified the overexpressed DmSNAP43 and DmSNAP50 proteins to near homogeneity as determined by Coomassie blue stained gels (not shown). Following dialysis to remove the imidazole, the recombinant DmSNAPs were provided to a commercial vendor (Chemicon) for polyclonal antibody production in rabbits.

The *D. melanogaster* gene for TBP was purchased from Research Genetics and re-cloned into the pET-30b(+) vector (Novagen) that provided 6xHis and S tags at the amino terminus of the encoded protein. The recombinant TBP was expressed in Rosetta2(DE3) cells and purified as described above for DmSNAP43 and DmSNAP50. The TBP was further purified by S-protein agarose affinity chromatography (Novagen) and the 6xHis and S tags were removed by digestion with enterokinase. The purified TBP was then used for antibody production as described above.

Reporter constructs and stably transfected cell lines

Constructs that contained the firefly luciferase gene driven by the promoter of the *D. melanogaster* U1:95Ca gene (formerly called the U1-95.1 gene) have been previously described (Lai et al., 2005; McNamara-Schroeder et al., 2001). Nineteen µg of reporter construct [that contained either the wild type U1 promoter or the U1 promoter with the U1 PSEA changed to a U6 PSEA (McNamara-Schroeder et al., 2001)] and 1 µg of

pCoBLAST were used to cotransfect S2 cells according to conditions recommended by Invitrogen. Stably transfected cell lines were selected and maintained on blasticidin-containing media.

Primer extensions

To carry out primer extensions to compare wild type and mutant U1 promoter activities in the stably transfected cells, total RNA was isolated from the cells as previously described (McNamara-Schroeder et al., 2001). Then 25 μg RNA aliquots were each co-precipitated with a ^{32}P-labeled 54-mer oligonucleotide recovery standard prior to the extension reaction. Primers extensions were carried out exactly as previously described (Lai et al., 2005). The primer (5'-CGGAATGCCAAGCTGGTCG-3') was complementary to a region near the 5' end of the luciferase gene and yielded an expected 60-nucleotide extension product. Following gel electrophoresis, the extension products were detected by autoradiography.

Chromatin immunoprecipitation (ChIP)

ChIP was carried out by following a protocol provided by David Gilmour (Pennsylvania State University). Briefly, *Drosophila* S2 cells (either normal cells or cells stably transfected with constructs described above) were cross-linked by treatment with 1% formaldehyde for 10 minutes at room temperature, and the reaction was quenched by adding glycine to 240 mM. Following collection of the cells, the chromatin was sheared by sonication to produce DNA of about 500-1000 base pairs average length. The cleared

supernatant was mixed with an equal volume of 6 M urea and dialyzed overnight at 4°C against ChIP buffer (10mM Tris-HCl pH 8.0, 1mM EDTA, 0.5mM EGTA, 0.5mM PMSF, 10% glycerol, 0.1% sodium deoxycholate, 1.0% Triton X-100). The chromatin solution was pre-cleared by using protein A agarose beads (Pierce) then incubated overnight at 4°C with 4 µl of antiserum or pre-immune serum. Next, the solution was incubated with protein A beads for two hours at 4°C. Beads were then washed a total of thirteen times with various buffers at 4°C: six times with low salt wash buffer (0.1% SDS, 1% Triton X-100, 2mM EDTA, 20mM Tris-HCl, pH 8.1, 150mM NaCl), three times with high salt wash buffer (same but containing 500 mM NaCl), two times with lithium wash buffer (0.25 M LiCl, 1% NP-40, 1% sodium deoxycholate, 1mM EDTA, 10mM Tris-HCl pH 8.1) and two times with TE buffer (10 mM Tris-HCl pH 8.1, 1 mM EDTA). Precipitated chromatin was eluted with 1% SDS, 0.1M $NaHCO_3$, incubated at 65°C for 4 hr in 200 mM NaCl to reverse cross-links, then incubated at 45°C for 1 hr in the presence of 10 mM EDTA, 40 mM Tris-HCl pH 6.5, and 40 µg/ml proteinase K. The precipitated DNA was then purified using QIAquick PCR purification kits (QIAGEN).

The immunoprecipitated DNA was analyzed by PCR by using 2 µl of purified precipitated DNA, 200 ng each of upstream and downstream primers, and 45 µl of Platinum PCR SuperMix (Invitrogen). The PCR reactions were analyzed on 8% non-denaturing polyacrylamide gels. U1 promoter and negative control primers were selected based upon sequences of the cloned genes and available *D. melanogaster* genomic sequences in Flybase (www.flybase.net). The locations of the forward primer (5'-GTGTGGCATACTTATAGGGGTGCT-3') and of the reverse primer (5'-

GCTTTTCGATGCTCGGCAGCAG-3') used for ChIP analysis of the endogenous U1:95Ca gene are shown in Fig. 2.1A. For analysis of the stably transfected U1:95Ca-luciferase constructs, the same forward primer was employed but the reverse primer was identical to the one described above that was used for primer extension analysis of U1-luciferase transcripts (see preceding section). The negative control primers (5'-CCGAGAGTGATGAGCATTTGCC-3' and 5'-AACTTGCTCCCCTTTTGCGTGG-3') amplify a 177 bp region of the *Drosophila* genome located approximately 7000 base pairs 5' of the U1:95Ca gene. This distant upstream region contains no known transcriptional promoters as annotated in Flybase.

RESULTS

Detection of DmSNAP43 and DmSNAP50 at the endogenous U1:95Ca promoter in S2 cells.

To our knowledge, experiments that demonstrate DmSNAPc occupancy of *D. melanogaster* snRNA promoters *in vivo* have not been reported. Therefore, polyclonal antibodies against bacterially-produced DmSNAP43 and DmSNAP50 were prepared in rabbits and tested for suitability in ChIP assays by targeting the promoter of the endogenous U1:95Ca gene in S2 cells. Previous studies have indicated that the promoter of this gene is active both *in vitro* and *in vivo* (Lai et al., 2005; McNamara-Schroeder et al., 2001; Su et al., 1997; Zamrod et al., 1993). This gene and its promoter region are represented in Fig. 2.1A. PCR primers (Fig. 2.1A) that flank the U1:95Ca core promoter were used to examine ChIP enrichment of DNA that contained the PSEA. Negative

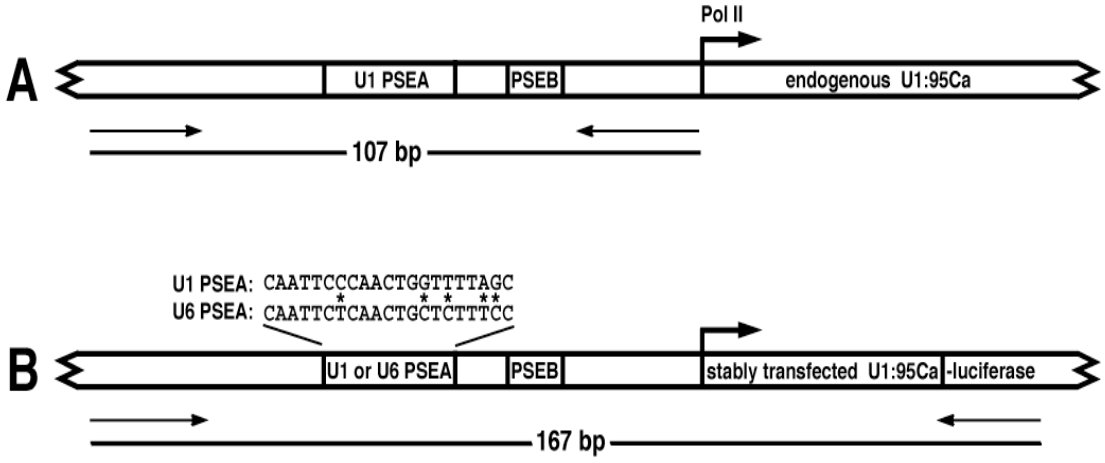


Figure 2.1. **(A) Genomic structure of the endogenous *D. melanogaster* U1:95Ca gene.** The locations of the two primers used for PCR amplification of chromatin immunoprecipitates are shown. The predicted PCR product is 107 bp in length. **(B) Structure of the U1-luciferase fusion constructs stably transfected into S2 cells.** Different constructs contained either a wild type U1 PSEA or a U6 PSEA that differed at only the 5 nucleotide positions indicated. The locations are shown of the pair of PCR primers that specifically amplify only DNA from the transfected constructs to produce a predicted 167 bp fragment.

control primers to verify the specificity of the reaction were chosen that amplify a segment of DNA located approximately 7000 bp upstream of the U1:95Ca gene (not shown).

S2 cells were grown in culture and fixed with formaldehyde; chromatin was then isolated and immunoprecipitated by using rabbit polyclonal antibodies raised against recombinant DmSNAP43 or DmSNAP50. Fig. 2.2 shows that DmSNAP43 and DmSNAP50 were specifically detected at the endogenous U1:95Ca promoter in S2 cells. ChIP using the anti-DmSNAP43 and anti-DmSNAP50 antibodies produced a strong signal when the U1 promoter primers were used for PCR amplification (Fig. 2.2, lanes 3 and 6). In contrast, only a weak background signal was observed when pre-immune antibodies were used for the ChIPs (lanes 2 and 5). The detection of DmSNAP43 and DmSNAP50 at the U1 promoter was specific because the negative control primers that amplified a segment of DNA located ~7000 bp upstream of the U1 promoter produced no detectable signal (lanes 4 and 7). Both sets of primers worked efficiently for PCR when total input DNA was used as template (lanes 1 and 8). These results validate the anti-DmSNAP43 and anti-DmSNAP50 antibodies as suitable for ChIP detection of DmSNAPc at the endogenous U1:95Ca promoter.

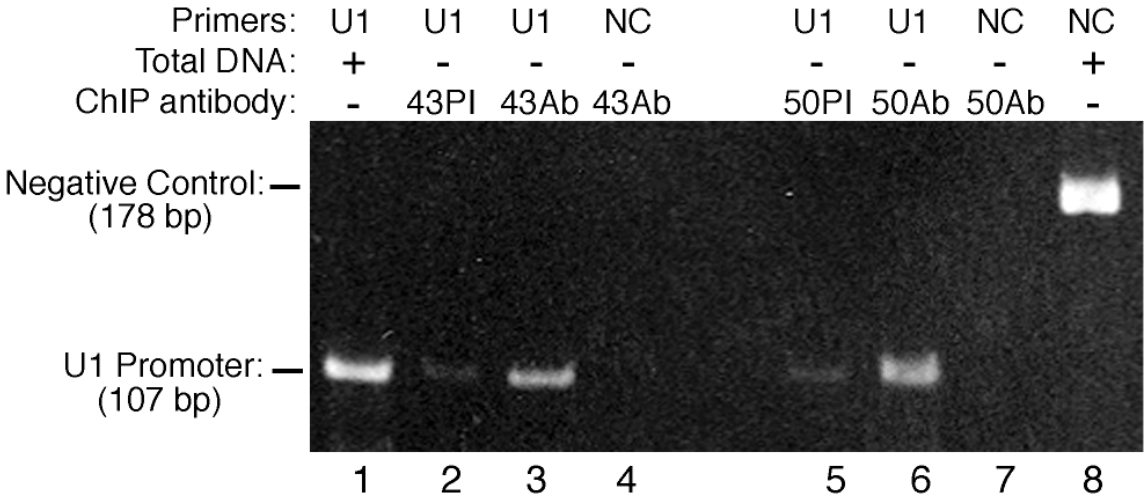


Figure 2.2. **Chromatin immunoprecipitation detects DmSNAPc at the endogenous U1:95Ca promoter in *D. melanogaster* S2 cells.** Antibodies against either DmSNAP43 (43Ab) or DmSNAP50 (50Ab) precipitated significant amounts of the endogenous U1 gene promoter (lanes 3 and 6 respectively). Use of pre-immune sera (43PI or 50PI) resulted in much weaker signals (lanes 2 and 5). The antibodies detected no DmSNAPc bound to a negative control (NC) region of DNA located ~7000 bp upstream of the U1:95Ca gene (lanes 4 and 7). As positive PCR controls, total DNA was amplified using primers specific to the U1 promoter (lane 1, 107 bp) or to the negative control region (lane 8, 178 bp).

U1 promoter activity and SNAPc promoter occupancy in stably transfected S2 cells

Previous work indicated that the U1:95Ca promoter could be used to express high levels of luciferase in transient transfection assays and that the luciferase mRNA was readily detectable by primer extension (Lai et al., 2005; McNamara-Schroeder et al., 2001). But surprisingly, a 5 bp substitution in the U1 PSEA that converted it to a U6 PSEA effectively inactivated the U1 promoter (Lai et al., 2005; McNamara-Schroeder et al., 2001) even though both the U1 and U6 PSEAs could bind DmSNAPc *in vitro* (Li et al., 2004; Wang and Stumph, 1998).

To investigate the binding of DmSNAPc to the promoter of the transfected U1-luciferase constructs *in vivo*, we prepared stably transfected S2 cell lines that contained the U1-luciferase constructs that we had previously used in transient transfection assays (Lai et al., 2005; McNamara-Schroeder et al., 2001). The constructs in the stably transfected cells contained either the wild-type U1 promoter or the U1 promoter with the PSEA switched to a U6 PSEA (Fig. 2.1B).

Following selection of the stably transfected cells with blasticidin, we determined Pol II promoter activity by using primer extension assays to measure U1-luciferase transcript levels (Fig.2.3). Cells that were stably transfected with the construct that contained the wild-type U1 promoter expressed high levels of U1-luciferase transcripts

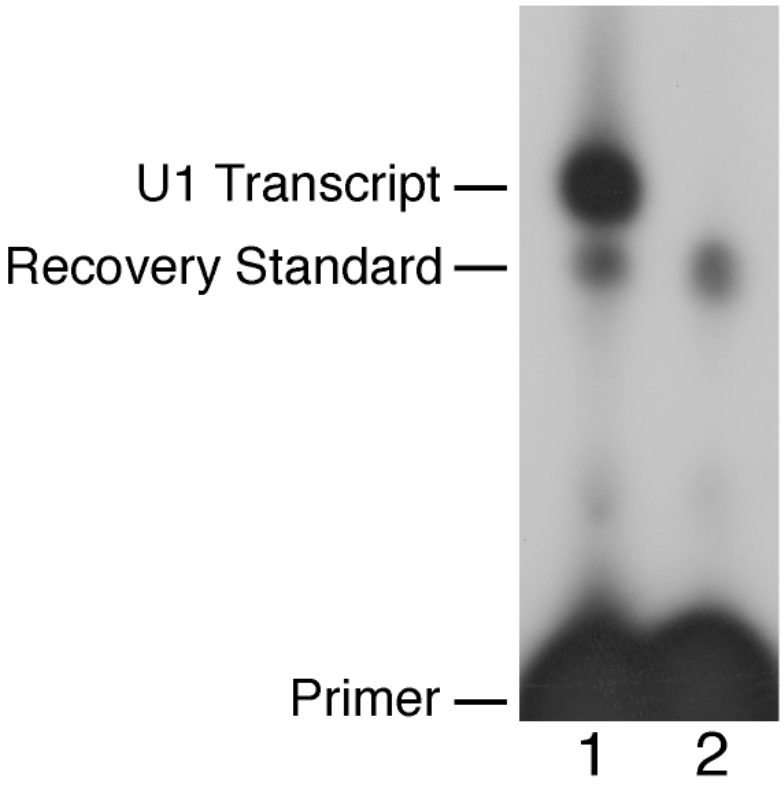


Figure 2.3. **Stably transfected U1-luciferase fusion constructs are actively expressed from the wild type U1 promoter (lane 1) but not from constructs that contain a U6 PSEA substitution (lane 2).** The autoradiogram shows the results of primer extension reactions using a luciferase gene-specific primer (Fig. 2.1B) annealed to RNA isolated from stably transfected S2 cells. The expected reverse transcript was 60 nucleotides in length. A 54 nucleotide internal recovery standard was included in each reaction.

(Fig. 2.3, lane 1). In contrast, cells transfected with the construct that contained the U6 PSEA substitution yielded no detectable transcripts (lane 2). Thus, the expression profiles of the two constructs in the stably transfected cells paralleled the results of previous transient expression assays.

We then used ChIP to examine DmSNAPc occupancy of the wild type and mutant U1 promoters in the stably transfected cells. Fig. 2.4 shows that the wild type U1 promoter, as expected, was precipitated by DmSNAP43 and DmSNAP50 antibodies (lanes 3 and 5) but not by the pre-immune antibodies (lanes 2 and 4). This indicates that DmSNAPc was occupying the wild type U1 promoter of the luciferase constructs in the stably transfected S2 cells. Interestingly, the antibodies against DmSNAP43 and DmSNAP50 also specifically precipitated the same region of the promoter from cells transfected with the construct that contained the U6 PSEA substitution (Fig. 2.4, lanes 8 and 10). Thus, DmSNAPc was bound *in vivo* to the U6 PSEA in the context of the U1 promoter even though the transfected construct was not being actively expressed.

TBP assembles on the endogenous U1 promoter *in vivo*

Because DmSNAPc was still able to occupy the inactive U1 promoter that contained the U6 PSEA, we next tested whether a later step in pre-initiation complex assembly might be disrupted, thereby rendering the promoter inactive. *In vitro* transcription experiments have indicated that the TATA box binding protein is required for snRNA transcription by Pol II in fruit flies (Zamrod et al., 1993) and humans

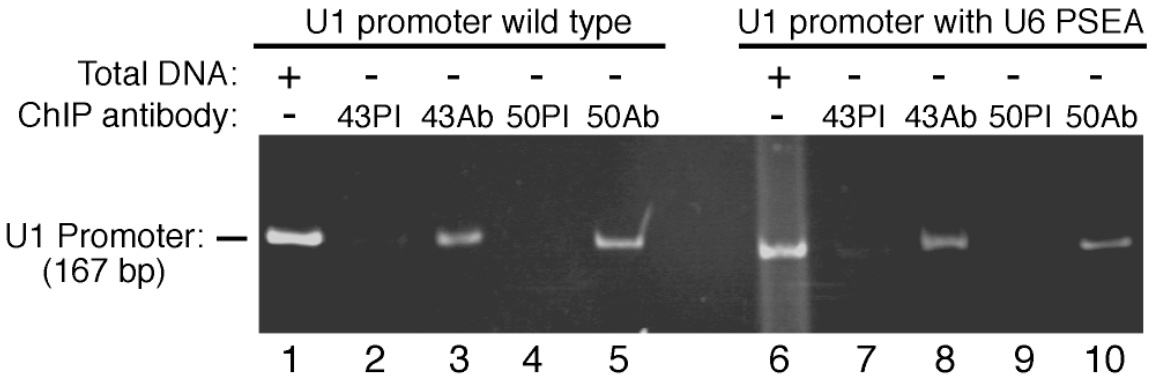


Figure 2.4. **ChIP indicates that DmSNAPc binds *in vivo* to the wild type promoter of actively expressed U1-luciferase constructs as well as to the inactive promoter of constructs that contain a U6 PSEA**. Chromatin from S2 cells stably transfected with one or the other of the two constructs illustrated in Fig. 1B was immunoprecipitated with antibodies against DmSNAP43 or DmSNAP50. In both cases, promoter-bound DmSNAPc was detected (lanes 3, 5, 8, and 10). On the other hand, pre-immune antibodies yielded signals that were considerably weaker or non-detectable (lane 2, 4, 7, and 9). Lanes 1 and 6 are PCR positive controls that used total DNA as template.

(Sadowski et al., 1993). However, we are not aware of any data that demonstrate TBP occupancy of the U1 promoter *in vivo*.

Therefore, we examined whether TBP could be detected at the endogenous U1:95Ca promoter by ChIP. Fig. 2.5 shows that the U1 promoter was efficiently precipitated by antibodies prepared against recombinant TBP (lane 3), whereas the pre-immune antibodies did not as efficiently precipitate the U1 promoter region (lane 2). Moreover, the antibodies against TBP did not precipitate the negative control region of the genome situated ~7000 bp upstream of the U1 promoter (lane 4). These results validated that the antibodies prepared against recombinant TBP could be used for ChIP detection of TBP on the U1:95Ca promoter in S2 cells.

TBP recruitment to the U1 promoter is disrupted by changing the U1 PSEA to a U6 PSEA

We next examined whether TBP could be detected on the U1 promoter in cells stably transfected with the luciferase expression constructs that contained either the wild type U1 promoter or the U1 promoter with the U6 PSEA substitution (Fig. 2.1B). Fig. 6 shows the results of ChIP assays designed to answer that question. Antibodies against TBP efficiently precipitated the promoter region of the stably transfected reporter construct that contained the wild type U1 PSEA (Fig.2.6, lane 3). In stark contrast, no signal was obtained with chromatin from the cells stably transfected with the construct

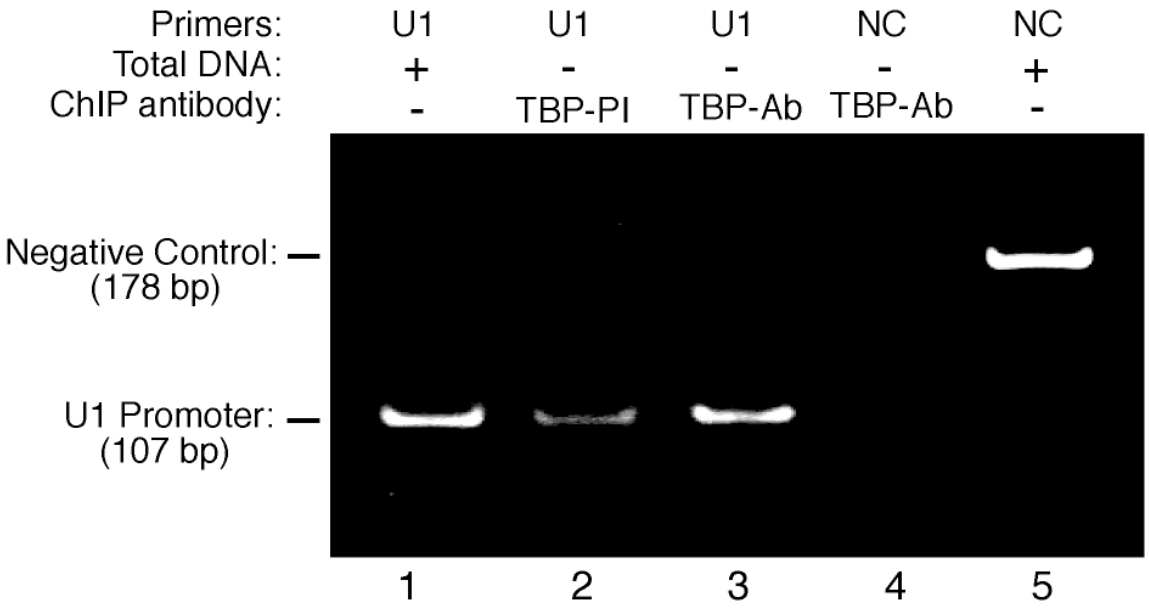


Figure 2.5. **ChIP detection of TBP at the endogenous U1:95Ca promoter in S2 cells**. Antibodies against TBP (TBPAb) efficiently precipitated the promoter region of the U1 gene (lane 3), whereas antibodies from pre-immune sera (PI) yielded a much reduced signal (lane 2). The TBP antibodies did not precipitate chromatin from the negative control region of the genome (lane 4).

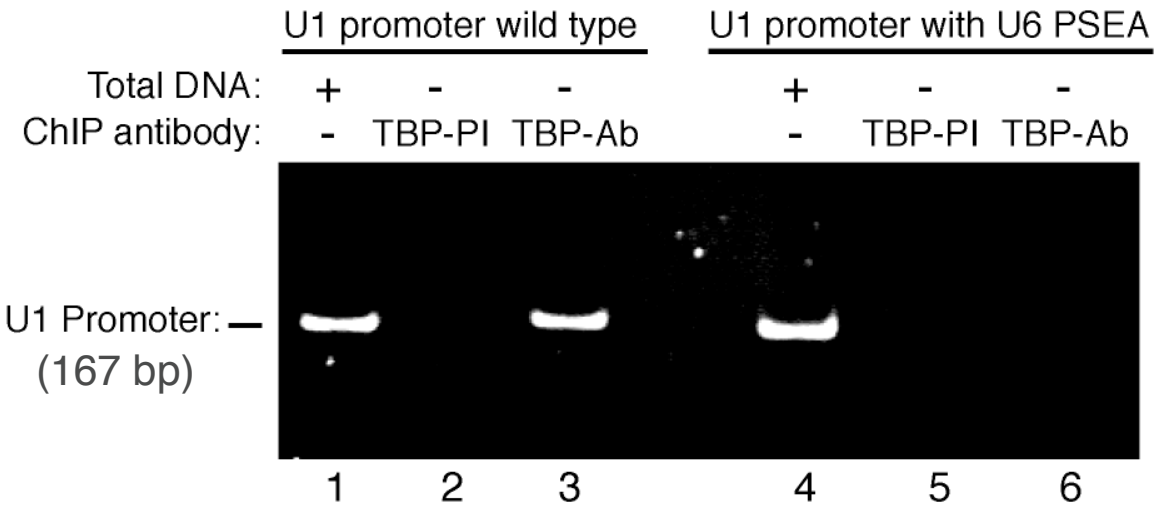


Figure 2.6. **Substitution of a U6 PSEA into the U1 promoter interferes with assembly of TBP *in vivo*.** Results from ChIP analyses are shown. The data indicate that TBP was assembled *in vivo* on the wild type promoter of a stably transfected U1-luciferase construct (lane 3). In contrast, TBP was not detected on the promoter of a similar construct that contained a 5 bp substitution that converted the U1 PSEA to a U6 PSEA (lane 6). Lanes 1, 2, 4, and 5 are control lanes similar to those shown in earlier figures.

that contained the U6 PSEA substitution (lane 6). Importantly, the total DNA positive control PCRs yielded the expected bands in each case (lanes 1 and 4). The precipitations were specific in that the pre-immune antibodies failed to precipitate the promoter DNA from either set of transfected cells (lanes 2 and 5). From these results, we conclude that TBP assembled efficiently on the transfected wild type U1 promoter *in vivo*. However, it failed to assemble on an otherwise identical promoter in which the U1 PSEA was changed to a U6 PSEA.

DISCUSSION

Previous studies in our lab indicated that the PSEA acts as a dominant element to determine the RNA polymerase specificity of *Drosophila* snRNA gene promoters. Exchanging the U1 and U6 PSEAs resulted in switching the RNA polymerase specificity of the promoters *in vitro* (Jensen et al., 1998; McNamara-Schroeder et al., 2001). *In vivo*, however, snRNA promoter activity was suppressed as a result of exchanging the U1 and U6 PSEAs (Lai et al., 2005; McNamara-Schroeder et al., 2001). It is perhaps not surprising that the requirements for snRNA expression could be more stringent *in vivo* than *in vitro*, potentially as a result of chromatin structure effects. One possibility was that the stable binding of DmSNAPc to the U1 promoter could be suppressed as a result of changing the U1 PSEA to a U6 PSEA. However, our ChIP results rule out that possibility: the ChIP assays shown in Fig. 2.4 reveal that DmSNAPc is still able to assemble on the U1 promoter that contains a U6 PSEA.

This finding suggested that a later stage in the assembly of a functional Pol II transcription complex might be disrupted when the U1 PSEA is changed to a U6 PSEA. Such later stage possibilities include the incorporation of TBP or of other Pol II general transcription factors (Kuhlman et al., 1999) or of Pol II itself into the pre-initiation complex (PIC). Alternatively, a complete Pol II PIC could assemble on the mutant promoter but the PIC could fail to productively initiate transcription for one reason or another.

Since TBP is required for transcription of *D. melanogaster* U1 genes (Zamrod et al., 1993) and in general plays a central early role in PIC assembly on Pol II promoters, we chose to investigate the TBP occupancy of the wild type and mutant U1 (U1 PSEA → U6 PSEA) promoters *in vivo*. The ChIP results shown in Figs. 5 and 6 demonstrate that TBP binds to the U1 promoter *in vivo*, but it fails to assemble if the U1 PSEA is switched to a U6 PSEA. This finding provides a reasonable explanation of why changing the U1 PSEA to a U6 PSEA abrogates the Pol II activity of the U1 promoter. Because TBP is not recruited to the mutant promoter, it follows that Pol II promoter activity would be lost.

Since DmSNAPc can bind to either the U1 PSEA or to the U6 PSEA in the context of the U1 promoter, why is TBP successfully recruited in one case but not in the other? We have previously demonstrated that DmSNAPc assumes different conformations when it binds to a U1 or U6 PSEA and that it differentially contacts the two different PSEAs as well as the DNA downstream of each (Lai et al., 2005; Li et al., 2004; Wang and Stumph, 1998). Most strikingly, the DmSNAP43 subunit contacts

approximately 20 bp of DNA downstream of the U1 PSEA but only about 5 bp downstream of the U6 PSEA. The extended region contacted by DmSNAP43 on the U1 promoter includes the PSEB (Li et al., 2004), and molecular modeling is consistent with the idea that both DmSNAP43 and TBP could co-occupy the PSEB (Lai et al., 2005). Thus, it is reasonable to assume that DmSNAPc, when bound to the wild type U1 promoter, participates in the recruitment of TBP to the PSEB. This is illustrated in the upper panel of the working model shown in Fig. 2.7. On the other hand, the data suggest that binding to the U6 PSEA induces a conformation in DmSNAPc that is incompatible with the recruitment of TBP (Fig. 2.7, lower panel) and the subsequent formation of a functional PIC. Further experiments will be required to fully test this model.

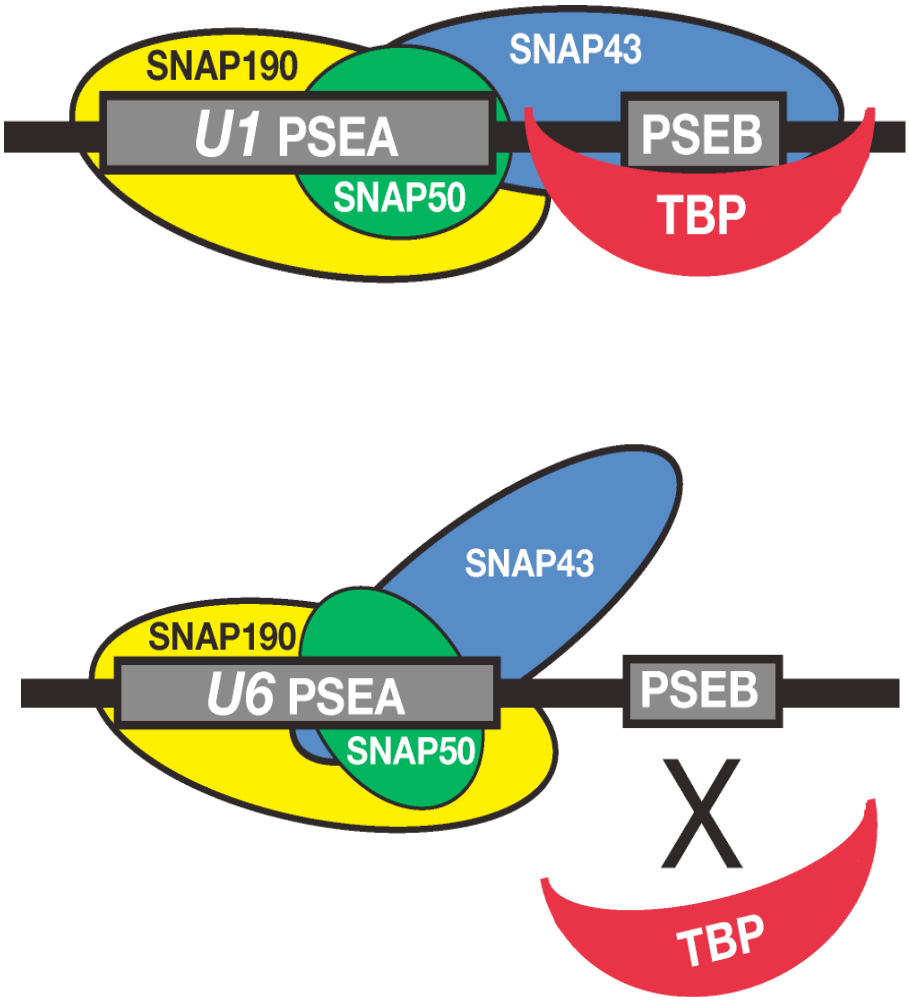


Figure 2.7. **Model for RNA polymerase II selectivity at the U1 snRNA promoter in fruit flies.** The regions of the U1 promoter DNA contacted by each of the DmSNAPc subunits are shown based upon *in vitro* site-specific protein-DNA photo-cross-linking data (Li et al., 2004; Wang and Stumph, 1998). When the promoter contains a U1 PSEA (upper diagram), the DmSNAP43 subunit contacts DNA far downstream of the PSEA, including the PSEB (Li et al., 2004). According to the working model, the close contacts of DmSNAP43 with the PSEB contribute to the recruitment of TBP to the U1 promoter *in vivo*. Substitution of a U6 PSEA does not allow the contacts between DmSNAP43 and the downstream DNA to occur (Li et al., 2004), and TBP subsequently fails to assemble (lower diagram)

REFERENCES

Carbon, P., and Krol, A. (1991). Transcription of the Xenopus laevis selenocysteine tRNA(Ser)Sec gene: a system that combines an internal B box and upstream elements also found in U6 snRNA genes. EMBO J *10*, 599-606.

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.

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

Das, G., Henning, D., Wright, D., and Reddy, R. (1988). Upstream regulatory elements are necessary and sufficient for transcription of a U6 RNA gene by RNA polymerase III. EMBO J *7*, 503-512.

Earley, J. M., III, Roebuck, K. A., and Stumph, W. E. (1984). Three linked chicken U1 RNA genes have limited flanking DNA sequence homologies that reveal potential regulatory signals. Nucleic Acids Res *12*, 7411-7421.

Gruber, A., Soldati, D., Burri, M., and Schümperli, D. (1991). Isolation of an active gene and of two pseudogenes for mouse U7 small nuclear RNA. Bio chim Bio phys Acta Gene Struct Expression *1088*, 151-154.

Guthrie, C. (1991). Messenger RNA splicing in yeast: clues to why the spliceosome is a ribonucleoprotein. Science *253*, 157-163.

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. (1992). Transcription of vertebrate snRNA genes and related genes, In Transcriptional Regulation, S. L. McKnight, and K. R. Yamamoto, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 281-313.

Hoffman, M. L., Korf, G. M., McNamara, K. J., and Stumph, W. E. (1986). Structural and functional analysis of chicken U4 small nuclear RNA genes. Mol Cell Biol *6*, 3910-3919.

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.

Kazmaier, M., Tebb, G., and Mattaj, I. W. (1987). Functional characterization of X. laevis U5 snRNA genes. EMBO J *6*, 3071-3078.

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.

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.

Lescure, A., Carbon, P., and Krol, A. (1991). The different positioning of the proximal sequence element in the Xenopus RNA polymerase II and III snRNA promoters is a key determinant which confers RNA polymerase III specificity. Nucleic Acids Res *19*, 435-441.

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.

Li, J. M., Haberman, R. P., and Marzluff, W. F. (1996). Common factors direct transcription through the proximal sequence elements (PSEs) of the embryonic sea urchin U1, U2, and U6 genes despite minimal sequence similarity among the PSEs. Mol Cell Biol *16*, 1275-1281.

Li, J. M., Parsons, R. A., and Marzluff, W. F. (1994). Transcription of the sea urchin U6 gene *in vitro* requires a TATA-like box, a proximal sequence element, and sea urchin USF, which binds an essential E box. Mol Cell Biol *14*, 2191-2200.

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

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.

Lobo, S. M., Ifill, S., and Hernandez, N. (1990). Cis-acting elements required for RNA polymerase II and III transcription in the human U2 and U6 snRNA promoters. Nucleic Acids Res *18*, 2891-2899.

McNamara, K. J., and Stumph, W. E. (1990). Site-directed mutational analysis of a U4 small nuclear RNA gene proximal sequence element: Localization and identification of functional nucleotides. J Biol Chem *265*, 9728-9731.

McNamara, K. J., Walker, R. J., Roebuck, K. A., and Stumph, W. E. (1987). Transcription signals of a U4 small nuclear RNA gene. Nucleic Acids Res *15*, 9239-9254.

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.

Morris, G. F., Price, D. H., and Marzluff, W. F. (1986). Synthesis of U1 RNA in a DNA-dependent system from sea urchin embryos. Proc Natl Acad Sci USA *83*, 3674-3678.

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.

Myslinski, E., Ame, J. C., Krol, A., and Carbon, P. (2001). An unusually compact external promoter for RNA polymerase III transcription of the human H1RNA gene. Nucleic Acids Res *29*, 2502-2509.

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

Parry, H. D., Tebb, G., and Mattaj, I. W. (1989b). The Xenopus U2 gene PSE is a single, compact, element required for transcription initiation and 3' end formation. Nucleic Acids Res *17*, 3633-3644.

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. (2005). Characterization of a multisubunit transcription factor complex essential for spliced-leader RNA gene transcription in Trypanosoma brucei. Mol Cell Biol *25*, 7303-7313.

Sharp, P. A. (1994). Split genes and RNA splicing. Cell *77*, 805-816.
Southgate, C., and Busslinger, M. (1989). *In vivo* and *in vitro* expression of U7 snRNA genes: cis- and trans-acting elements required for RNA polymerase II-directed transcription. EMBO J *8*, 539-549.

Steitz, J. A., Black, D. L., Gerke, V., Parker, K. A., Kramer, A., Frendewey, D., and Keller, W. (1988). Functions of the abundant U-snRNPs., In Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles, M. L. Birnsteil, ed. (Heidelberg, Federal Republic of Germany: Springer Verlag KG), pp. 115-154.

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.

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.

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.

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.

Wendelburg, B. J., and Marzluff, W. F. (1992). Two promoter elements are necessary and sufficient for expression of the sea urchin U1 snRNA gene. Nucleic Acids Res *20*, 3743-3751.

Woodhams, M. D., Stadler, P. F., Penny, D., and Collins, L. J. (2007). RNase MRP and the RNA processing cascade in the eukaryotic ancestor. BMC Evol Biol *7 Suppl 1*, S13.

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.

DISCUSSION

Establishment of a transcription pre-initiation complex on *Drosophila* snRNA genes starts when DmSNAPc binds to a PSEA. Earlier work in our lab indicated that the U1 or U6 PSEA is the primary determinant for RNA polymerase specificity and that the PSEAs are not interchangeable (Jensen et al., 1998; McNamara-Schroeder et al., 2001). Only 5 nucleotide positions out of 21 nucleotide positions of the PSEA are different between the U1 and U6 PSEAs that we use in our experiments, and by changing nucleotides near the 3′ end we can alter RNA polymerase specificity from one polymerase to the other *in vitro* (Jensen et al., 1998; McNamara-Schroeder et al., 2001). *In vivo*, however, snRNA promoter activity was suppressed as a result of exchanging the U1 and U6 PSEAs (Lai et al., 2005; McNamara-Schroeder et al., 2001). Several factors could play a role in this apparent difference between the *in vivo* and *in vitro* results.

First, the necessities for transcription are certainly more stringent *in vivo* than *in vitro*. The *in vitro* reactions contained only a single type of promoter template that was present in a large number of copies. A detectable signal can be produced *in vitro* even though only a small percentage of the template molecules are actively transcribed. *In vivo*, on the other hand, the transfected copies of a gene must compete against thousands of other genes within the cell. If the stability of the transcription pre-initiation complex is even partially compromised by the switch of the PSEAs, the introduced genes may not compete effectively with endogenous genes for available transcription factors.

Second, it is also valuable to think about the fact that the absence of detectable transcripts from U1 or U6 promoters that contain a switched PSEA may be due to post-

initiation events as well as to effects on transcription initiation. For example, if RNA polymerase III does initiate transcription of the U1 gene that contains the U6 PSEA substitution, these transcripts would very likely neither be terminated nor processed properly *in vivo*. Such aberrant transcripts would likely be subject to more rapid degradation than normal cellular RNAs. Likewise, potential RNA polymerase II-initiated transcripts that arise from U6 promoter constructs that contain the U1 PSEA may be rapidly turned over *in vivo*.

Finally, chromatin structure may also play a role in the suppression of transcription *in vivo* from constructs that contain the wrong PSEA. Positioned nucleosomes have been implicated in both the activation and repression of transcription from snRNA gene promoters (Boyd et al., 2000; Burnol et al., 1993; Gerlach et al., 1995; Stünkel et al., 1997). It is possible that the chromatin environment required for optimal RNA polymerase II transcription is different from that required for optimal RNA polymerase III transcription.

The photo-cross-linking studies in our lab further indicated that the cross-linking pattern of DmSNAPc to the U1 PSEA and U6 PSEA was different. This suggests that the conformation of the U1 and U6 protein-DNA-complexes are different. Additional results from our lab indicated that the U1 and U6 PSEAs are both bent by a similar degree toward the face of the DNA contacted by the DmSNAP43 subunit. Thus, we believe that the conformational differences in the U1 and U6 DNA-DmSNAP complexes exist primarily at the level of the protein rather than the level of the DNA. We believe

these conformational differences of DmSNAPc on U1 and U6 PSEAs result in the recruitment of RNA polymerase II basal transcription factors on a U1 promoter and RNA polymerase III basal transcription factors on U6 promoter. Earlier work in our lab has demonstrated a TBP requirement for U1 transcription in *Drosophila*. Human U6 transcription requires TBP, Bdp and a unique form of a TFIIB related factor termed Brf2. Surprisingly, TBP is not used for U6 transcription in fruit flies, instead, the TBP-related factor TRF1 in used for Pol III transcription in *Drosophila*. Based upon previous results in our lab, we suspected that DmSNAPc plays a direct role in recruiting TBP to the U1 promoter and TRF1 to U6 promoters.

Changing the U1 PSEA to a U6 PSEA could (1) alter the chromatin structure of the promoter, (2) disrupt pre-initiation complex (PIC) assembly, or (3) result in the synthesis of unstable transcripts (or a combination of all three). Based upon our knowledge of the system, it seemed most likely that disruption of PIC formation might occur. Therefore this is the aspect of the system on which I chose to focus my studies.

There are several stages at which PIC formation could be disrupted. For example, when the U1 PSEA is switched to a U6 PSEA, the binding of DmSNAPc to the U1 promoter might be destabilized. Alternatively, DmSNAPc may bind to the U6 PSEA in the context of the U1 promoter, but the recruitment of TBP may be disrupted. Other possibilities are that DmSNAPc and TBP might assemble, but TFIIB, TFIIE, TFIIF, or Pol II itself may fail to assemble. Finally, the entire PIC might form, yet Pol II for some reason or another would be unable to initiate transcription.

To investigate this matter, I carried out the ChIP investigations reported in chapter 2. The results of those experiments clearly indicate that DmSNAPc is still able to bind to the U1 promoter *in vivo* when the U1 PSEA is changed to a U6 PSEA. However, switching the U1 to a U6 PSEA resulted in the failure of TBP to bind to the altered promoter. This result supports our lab's model that the sequence of the PSEA allosterically affects the conformation of DmSNAPc, such that when DmSNAPc binds to a U6 PSEA, it is unable to recruit TBP to the promoter (Chapter 2, Figure 7).

The DmSNAP43 subunit contacts approximately 20 bp of DNA downstream of the U1 PSEA but only about 5 bp downstream of the U6 PSEA. The extended region contacted by DmSNAP43 on the U1 promoter includes the PSEB. We believe that when DmSNAPc binds to the wild type U1 promoter, DmSNAP43 participates in the recruitment of TBP to the PSEB. Our molecular modeling is consistent with the idea that both DmSNAP43 and TBP could co-occupy the PSEB. In future work, it would be interesting to map domains and amino acid residues of DmSNAPc that are involved in TBP recruitment.

I did not formally investigate whether TFIIB, TFIIE, TFIIF or Pol II assemble on the U1 promoter that contains a U6 PSEA. However, in general, the binding of TBP is an early event in the formation of a PIC. Based upon our knowledge of PIC assembly, it is probably safe to assume that these other factors will fail to bind to the promoter in the absence of TBP. However, this question could be investigated in the future if antibodies to these other factors became available in the lab.

According to our model for determining RNA polymerase specificity in insects (Jensen et al., 1998; Li et al., 2004; McNamara-Schroeder et al., 2001; Wang et al., 1997; Zamrod et al., 1993), we also believe that DmSNAPc binding to a U6 PSEA allows it to recruit TRF1 to the U6 TATA box. In the homologous human system, there is in fact direct evidence that two of the three conserved SNAPc subunits can interact directly with TBP (Henry et al., 1995; Hinkley et al., 2003; Ma and Hernandez, 2002; Yoon and Roeder, 1996) and that SNAP190 in particular can stabilize the binding of TBP to the U6 TATA box (Hinkley et al., 2003; Ma and Hernandez, 2002). We already know that switching the fly U6 PSEA to a U1 PSEA inactivates the U6 promoter *in vivo* (Lai et al., 2005; McNamara-Schroeder et al., 2001). Thus, in the *Drosophila* system it would be interesting to carry out the converse of the experiments reported in chapter 2 of this dissertation. That is, would switching the U6 PSEA to a U1 PSEA disrupt the recruitment of TRF1 to the U6 promoter?

Indeed, I attempted to do those experiments (not shown). I over-expressed fly TRF1 in *E. coli* and used the purified protein to make anti-TRF1 antibodies in rabbits. I then used those antibodies in ChIP assays to examine TRF1 occupancy of the endogenous U6 promoter in unmodified S2 cells. Even though the antibodies reacted with TRF1 by immunoblotting, they were not able to preferentially ChIP the U6 promoter DNA of the endogenous U6 genes (results not shown). Likewise, the TRF1 antibodies were unable to detect TRF1 on the promoters of transfected wild type U6 genes. Without positive result from these controls, it was not possible to complete the experiment to determine the effect of the PSEA switch on TRF1 recruitment to the U6 promoter.

It is not certain why the TRF1 ChIPs were unsuccessful. One possibility is that the TRF1 epitopes recognized by the antibody are specifically occluded on the U6 promoter *in vivo*. A second possibility is raised by the recent finding that, following transcription activation, TBP is not required for ongoing transcription at the promoters of certain genes (Tatarakis et al., 2008). It is conceivable that this could also be true for TRF1 at the U6 promoter.

Unfortunately, it is not clear how those possibilities could be distinguished experimentally. Whatever the reason for the inability to detect TRF1 by ChIP on the U6 promoter, my results are consistent with a ChIP-on-chip study from the Tjian and Sunduz Keles labs (Isogai et al., 2007). They used a *Drosophila* genome array to determine at high resolution the *in vivo* target genes of TRF1 and Brf. As expected, tRNA gene promoters were occupied by both TRF1 and Brf. Interestingly, although Brf was found at U6 promoters, those investigators were unable to detect TRF1 at the U6 promoters by their ChIP-on-chip analysis. Thus, my results and theirs are consistent in that neither I nor they were able to successfully ChIP TRF1 at *Drosophila melanogaster* U6 promoters.

REFERENCES

Boyd, D. C., Greger, I. H., and Murphy, S. (2000). *In vivo* footprinting studies suggest a role for chromatin in transcription of the human 7SK gene. Gene *247*, 33-44.

Burnol, A. F., Margottin, F., Huet, J., Almouzni, G., Prioleau, M. N., Méchali, M., and Sentenac, A. (1993). TFIIIC relieves repression of U6 snRNA transcription by chromatin. Nature *362*, 475-477.

Gerlach, V. L., Whitehall, S. K., Geiduschek, E. P., and Brow, D. A. (1995). TFIIIB placement on a yeast U6 RNA gene *in vivo* is directed primarily by TFIIIC rather than by sequence-specific DNA contacts. Mol Cell Biol *15*, 1455-1466.

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., and 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*, 18649-18657.

Isogai, Y., Takada, S., Tjian, R., and Keles, S. (2007). Novel TRF1/BRF target genes revealed by genome-wide analysis of *Drosophila* Pol III transcription. Embo J *26*, 79-89.

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.

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.

Stünkel, W., Kober, I., and Seifart, K. H. (1997). A nucleosome positioned in the distal promoter region activates transcription of the human U6 gene. Mol Cell Biol *17*, 4397-4405.

Tatarakis, A., Margaritis, T., Martinez-Jimenez, C. P., Kouskouti, A., Mohan, W. S., 2nd, Haroniti, A., Kafetzopoulos, D., Tora, L., and Talianidis, I. (2008). Dominant and redundant functions of TFIID involved in the regulation of hepatic genes. Mol Cell *31*, 531-543.

Wang, Z. X., Luo, T., and Roeder, R. G. (1997). Identification of an autonomously initiating RNA polymerase III holoenzyme containing a novel factor that is selectively inactivated during protein synthesis inhibition. Genes Dev *11*, 2371-2382.

Yoon, J. B., and Roeder, R. G. (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.

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.

Appendices

A. DNA constructs for expression of DmSNAP43, DmSNAP50 and TBP in *Escherichia coli.*

B. Purification of DmSNAP43 and DmSNAP50 from *Escherichia coli* by nickel chelate chromatography.

C. Purification of TBP from *Escherichia coli* by nickel chelate chromatography and S-protein affinity chromatography.

D. Protocol for chromatin immunoprecipitaion (ChIP) assay with urea denaturation.

E. Plasmid constructs for U1-luciferase fusion reporter constructs that contained either a wild type U1 PSEA or a U6 PSEA.

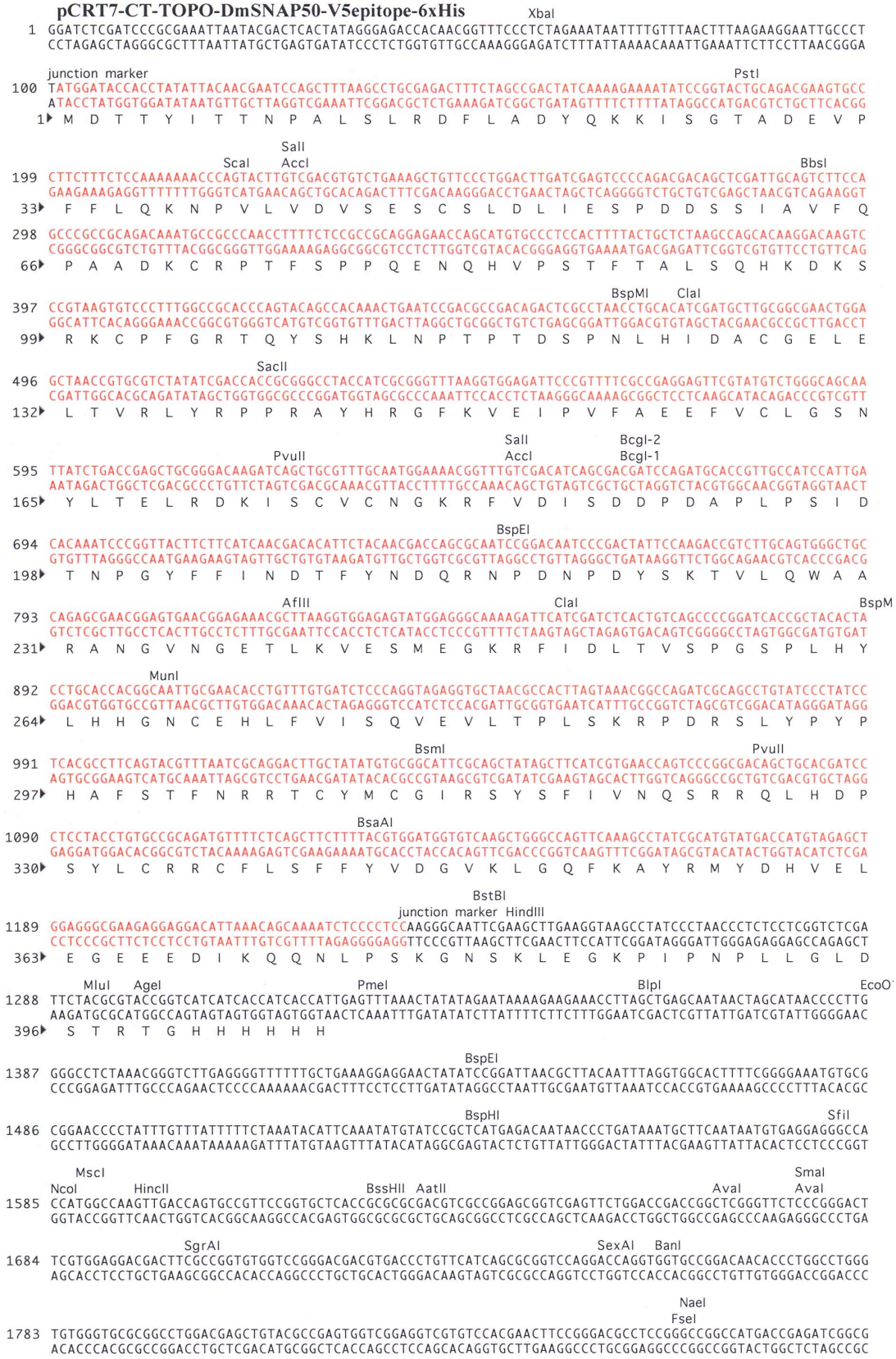
Appendix A:

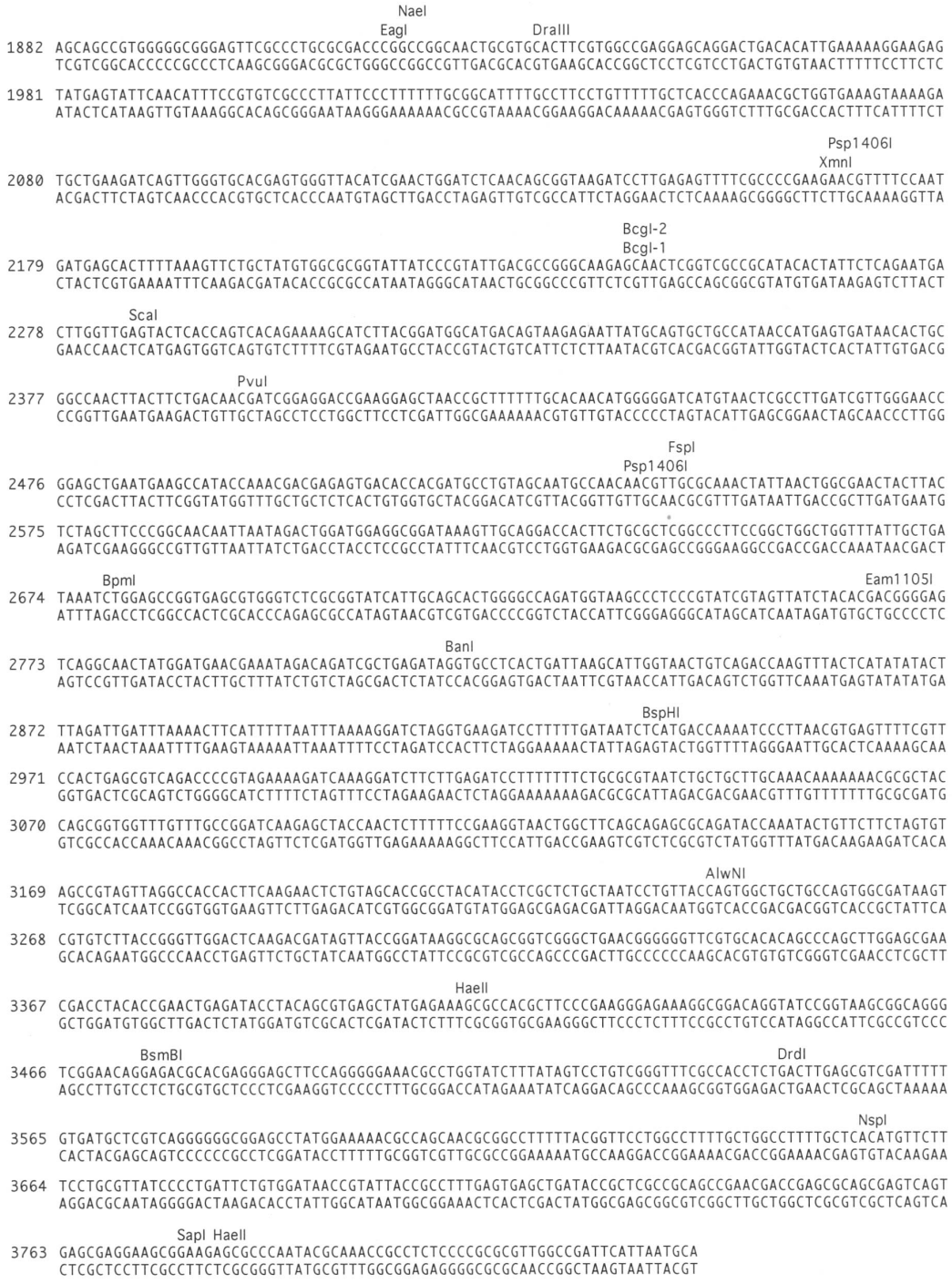
DNA constructs for expression of DmSNAP43, DmSNAP50 and TBP in *Escherichia coli.*

pCRT7/CT-TOPO-DmSNAP43-V5 epitope-6xHis

```
   1 GGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAATTGCCCT
     CCTAGAGCTAGGGCGCTTTAATTATGCTGAGTGATATCCCTCTGGTGTTGCCAAAGGGAGATCTTTATTAAAACAAATTGAAATTCTTCCTTAACGGGA

 100 TATGGAGCTGAATATCTTTGACGACTGCTGGGAGCTGGTGCAACGATTTCAGCGATTGGTTAATGATGGCGAAAACTGCGAGTTCGAGGTGTTCTGCCG
     ATACCTCGACTTATAGAAACTGCTGACGACCCTCGACCACGTTGCTAAAGTCGCTAACCAATTACTACCGCTTTTGACGCTCAAGCTCCACAAGACGGC
   1 M  E  L  N  I  F  D  D  C  W  E  L  V  Q  R  F  Q  R  L  V  N  D  G  E  N  C  E  F  E  V  F  C  R

 199 GTGCTGGCGAGAACTGCAGCTGCAGCACCTTTTCACTGCCCAGACGAACCACACAGAGGTGATAGCCACCACACTGGCGGCCCTGCATGTGGCCAAGCG
     CACGACCGCTCTTGACGTCGACGTCGTGGAAAAGTGACGGGTCTGCTTGGTGTGTCTCCACTATCGGTGGTGTGACCGCCGGGACGTACACCGGTTCGC
  33 C  W  R  E  L  Q  L  Q  H  L  F  T  A  Q  T  N  H  T  E  V  I  A  T  T  L  A  A  L  H  V  A  K  R

 298 ACTGTCGTGCTCCCGACGCACCACCGGGGACGTTTTCCCGGCATCTCGCGCTCAAAGGATCGGAGGTTTCTTTCTGCTCTACGTAATCTACTACAAGCA
     TGACAGCACGAGGGCTGCGTGGTGGCCCCTGCAAAAGGGCCGTAGAGCGCGAGTTTCCTAGCCTCCAAAGAAAGACGAGATGCATTAGATGATGTTCGT
  66 L  S  C  S  R  R  T  T  G  D  V  F  P  A  S  R  A  Q  R  I  G  G  F  F  L  L  Y  V  I  Y  Y  K  Q

 397 GCCCACGCACAACTTTATTAAGATCGAGGTCTCACCGCGCACTTGGCAAGAACTAACAGACTACGCTCTAGATCTGCGCAAGGATAGTCCGGAGCGGAA
     CGGGTGCGTGTTGAAATAATTCTAGCTCCAGAGTGGCGCGTGAACCGTTCTTGATTGTCTGATGCGAGATCTAGACGCGTTCCTATCAGGCCTCGCCTT
  99 P  T  H  N  F  I  K  I  E  V  S  P  R  T  W  Q  E  L  T  D  Y  A  L  D  L  R  K  D  S  P  E  R  K

 496 GGACACTCATCAGATCGCCTACATGCTGTGGCGCCTGACCCAGGAGCAGGCCTTCCGCTTCACCGCGCTCGACTATTGCCAGGGGTTGGACAATCTGGT
     CCTGTGAGTAGTCTAGCGGATGTACGACACCGCGGACTGGGTCCTCGTCCGGAAGGCGAAGTGGCGCGAGCTGATAACGGTCCCCAACCTGTTAGACCA
 132 D  T  H  Q  I  A  Y  M  L  W  R  L  T  Q  E  Q  A  F  R  F  T  A  L  D  Y  C  Q  G  L  D  N  L  V

 595 GGACTACGACCGTGTGGAGACCGTAGCGGGTGCCAAGGAACAGAGGCAGAGTGCCTTGATGCAGAAGCAACAGCGTGCGAACGGCGTCAGTCTCACATA
     CCTGATGCTGGCACACCTCTGGCATCGCCCACGGTTCCTTGTCTCCGTCTCACGGAACTACGTCTTCGTTGTCGCACGCTTGCCGCAGTCAGAGTGTAT
 165 D  Y  D  R  V  E  T  V  A  G  A  K  E  Q  R  Q  S  A  L  M  Q  K  Q  Q  R  A  N  G  V  S  L  T  Y

 694 CGAACTGGAGGGTCTGCGAGCACTGGACCAGGCAAGCCAGCCATTGTGTGAACTGGAAGCGGCATACAATGCCCAAAAGAAGCAATTGGCGGCTGGTCA
     GCTTGACCTCCCAGACGCTCGTGACCTGGTCCGTTCGGTCGGTAACACACTTGACCTTCGCCGTATGTTACGGGTTTTCTTCGTTAACCGCCGACCAGT
 198 E  L  E  G  L  R  A  L  D  Q  A  S  Q  P  L  C  E  L  E  A  A  Y  N  A  Q  K  K  Q  L  A  A  G  H

 793 TGAGCACGCTTTACCGCCCTCTCAAATATTCGGCCATTTGCGAGAAGTCTTTGCCGATATCCAAAGTGTTCTAGGAGCTAGAAAGAGTACTCCAGATGA
     ACTCGTGCGAAATGGCGGGAGAGTTTATAAGCCGGTAAACGCTCTTCAGAAACGGCTATAGGTTTCACAAGATCCTCGATCTTTCTCATGAGGTCTACT
 231 E  H  A  L  P  P  S  Q  I  F  G  H  L  R  E  V  F  A  D  I  Q  S  V  L  G  A  R  K  S  T  P  D  E

 892 GAAATGCACCACAACATCTACAGGCAACCAGTTGGAAGTGCGCCAGAGGGTGCGGAACAAGGCCATGTACGGCGTCGAGGAGCGGGAGCCGCAACACCA
     CTTTACGTGGTGTTGTAGATGTCCGTTGGTCAACCTTCACGCGGTCTCCCACGCCTTGTTCCGGTACATGCCGCAGCTCCTCGCCCTCGGCGTTGTGGT
 264 K  C  T  T  T  S  T  G  N  Q  L  E  V  R  Q  R  V  R  N  K  A  M  Y  G  V  E  E  R  E  P  Q  H  Q

 991 GACGGATGAACTAGAAGTGCAGCTGGAGGTCAACGAGACTTATCAACGCCGCATGTCCTCGGCCACCGTTTTCCAGAGGGAACTTCCAGAAGACGTGCA
     CTGCCTACTTGATCTTCACGTCGACCTCCAGTTGCTCTGAATAGTTGCGGCGTACAGGAGCCGGTGGCAAAAGGTCTCCCTTGAAGGTCTTCTGCACGT
 297 T  D  E  L  E  V  Q  L  E  V  N  E  T  Y  Q  R  R  M  S  S  A  T  V  F  Q  R  E  L  P  E  D  V  Q

1090 GCAAGAGTATGAGATGATTGAGTTTAGTGACGACGAGGAAATGGAAGTGGGTGAAAGCGAGGAGGTCACGGAAGAAGAACTCAAAGCTATTTTGGATAC
     CGTTCTCATACTCTACTAACTCAAATCACTGCTGCTCCTTTACCTTCACCCACTTTCGCTCCTCCAGTGCCTTCTTCTTGAGTTTCGATAAAACCTATG
 330 Q  E  Y  E  M  I  E  F  S  D  D  E  E  M  E  V  G  E  S  E  E  V  T  E  E  E  L  K  A  I  L  D  T

1189 CAAGGGCAATTCGAAGCTTGAAGGTAAGCCTATCCCTAACCCTCTCCTCGGTCTCGATTCTACGCGTACCGGTCATCATCACCATCACCATTGAGTTTA
     GTTCCCGTTAAGCTTCGAACTTCCATTCGGATAGGGATTGGGAGAGGAGCCAGAGCTAAGATGCGCATGGCCAGTAGTAGTGGTAGTGGTAACTCAAAT
 363 K  G  N  S  K  L  E  G  K  P  I  P  N  P  L  L  G  L  D  S  T  R  T  G  H  H  H  H  H  H

1288 AACTATATAGAATAAAAGAAGAAACCTTAGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAG
     TTGATATATCTTATTTTCTTCTTTGGAATCGACTCGTTATTGATCGTATTGGGGAACCCCGGAGATTTGCCCAGAACTCCCCAAAAAACGACTTTCCTC

1387 GAACTATATCCGGATTAACGCTTACAATTTAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATAT
     CTTGATATAGGCCTAATTGCGAATGTTAAATCCACCGTGAAAAGCCCCTTTACACGCGCCTTGGGGATAAACAAATAAAAAGATTTATGTAAGTTTATA

1486 GTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATGTGAGGAGGGCCACCATGGCCAAGTTGACCAGTGCCGTTCCGGTGCTCACCGCGC
     CATAGGCGAGTACTCTGTTATTGGGACTATTTACGAAGTTATTACACTCCTCCCGGTGGTACCGGTTCAACTGGTCACGGCAAGGCCACGAGTGGCGCG

1585 GCGACGTCGCCGGAGCGGTCGAGTTCTGGACCGACCGGCTCGGGTTCTCCCGGGACTTCGTGGAGGACGACTTCGCCGGTGTGGTCCGGGACGACGTGA
     CGCTGCAGCGGCCTCGCCAGCTCAAGACCTGGCTGGCCGAGCCCAAGAGGGCCCTGAAGCACCTCCTGCTGAAGCGGCCACACCAGGCCCTGCTGCACT

1684 CCCTGTTCATCAGCGCGGTCCAGGACCAGGTGGTGCCGGACAACACCCTGGCCTGGGTGTGGGTGCGCGGCCTGGACGAGCTGTACGCCGAGTGGTCGG
     GGGACAAGTAGTCGCGCCAGGTCCTGGTCCACCACGGCCTGTTGTGGGACCGGACCCACACCCACGCGCCGGACCTGCTCGACATGCGGCTCACCAGCC

1783 AGGTCGTGTCCACGAACTTCCGGGACGCCTCCGGGCCGGCCATGACCGAGATCGGCGAGCAGCCGTGGGGGCGGGAGTTCGCCCTGCGCGACCCGGCCG
     TCCAGCACAGGTGCTTGAAGGCCCTGCGGAGGCCCGGCCGGTACTGGCTCTAGCCGCTCGTCGGCACCCCCGCCCTCAAGCGGGACGCGCTGGGCCGGC

1882 GCAACTGCGTGCACTTCGTGGCCGAGGAGCAGGACTGACACATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTT
     CGTTGACGCACGTGAAGCACCGGCTCCTCGTCCTGACTGTGTAACTTTTTCCTTCTCATACTCATAAGTTGTAAAGGCACAGCGGGAATAAGGGAAAAA

1981 TGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACT
     ACGCCGTAAAACGGAAGGACAAAAACGAGTGGGTCTTTGCGACCACTTTCATTTTCTACGACTTCTAGTCAACCCACGTGCTCACCCAATGTAGCTTGA

2080 GGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATC
     CCTAGAGTTGTCGCCATTCTAGGAACTCTCAAAAGCGGGGCTTCTTGCAAAAGGTTACTACTCGTGAAAATTTCAAGACGATACACCGCGCCATAATAG

2179 CCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGA
     GGCATAACTGCGGCCCGTTCTCGTTGAGCCAGCGGCGTATGTGATAAGAGTCTTACTGAACCAACTCATGAGTGGTCAGTGTCTTTTCGTAGAATGCCT

2278 TGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCT
     ACCGTACTGTCATTCTCTTAATACGTCACGACGGTATTGGTACTCACTATTGTGACGCCGGTTGAATGAAGACTGTTGCTAGCCTCCTGGCTTCCTCGA
```

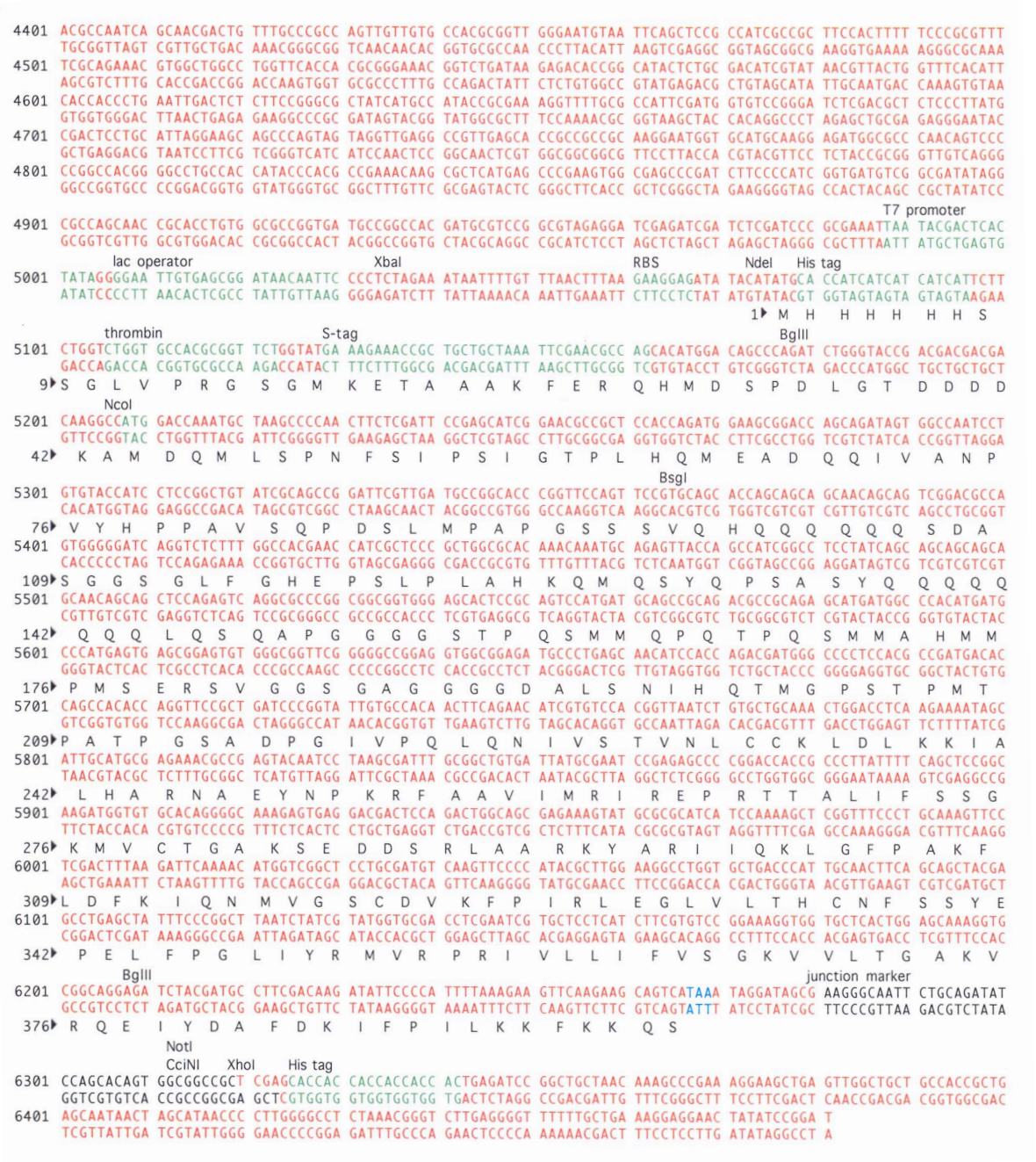
```
2377 AACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGAGTGACACCAC
     TTGGCGAAAAAACGTGTTGTACCCCCTAGTACATTGAGCGGAACTAGCAACCCTTGGCCTCGACTTACTTCGGTATGGTTTGCTGCTCTCACTGTGGTG

2476 GATGCCTGTAGCAATGCCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGA
     CTACGGACATCGTTACGGTTGTTGCAACGCGTTTGATAATTGACCGCTTGATGAATGAGATCGAAGGGCCGTTGTTAATTATCTGACCTACCTCCGCCT

2575 TAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGC
     ATTTCAACGTCCTGGTGAAGACGCGAGCCGGGAAGGCCGACCGACCAAATAACGACTATTTAGACCTCGGCCACTCGCACCCAGAGCGCCATAGTAACG

2674 AGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGAT
     TCGTGACCCCGGTCTACCATTCGGGAGGGCATAGCATCAATAGATGTGCTGCCCCTCAGTCCGTTGATACCTACTTGCTTTATCTGTCTAGCGACTCTA

2773 AGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTA
     TCCACGGAGTGACTAATTCGTAACCATTGACAGTCTGGTTCAAATGAGTATATATGAAATCTAACTAAATTTTGAAGTAAAAATTAAATTTTCCTAGAT

2872 GGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTC
     CCACTTCTAGGAAAAACTATTAGAGTACTGGTTTTAGGGAATTGCACTCAAAAGCAAGGTGACTCGCAGTCTGGGGCATCTTTTCTAGTTTCCTAGAAG

2971 TTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACGCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTT
     AACTCTAGGAAAAAAAGACGCGCATTAGACGACGAACGTTTGTTTTTTTGCGCGATGGTCGCCACCAAACAAACGGCCTAGTTCTCGATGGTTGAGAAA

3070 TTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGC
     AAGGCTTCCATTGACCGAAGTCGTCTCGCGTCTATGGTTTATGACAAGAAGATCACATCGGCATCAATCCGGTGGTGAAGTTCTTGAGACATCGTGGCG

3169 CTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATA
     GATGTATGGAGCGAGACGATTAGGACAATGGTCACCGACGACGGTCACCGCTATTCAGCACAGAATGGCCCAACCTGAGTTCTGCTATCAATGGCCTAT

3268 AGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAG
     TCCGCGTCGCCAGCCCGACTTGCCCCCCAAGCACGTGTGTCGGGTCGAACCTCGCTTGCTGGATGTGGCTTGACTCTATGGATGTCGCACTCGATACTC

3367 AAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGACGCACGAGGGAGCTTCCAGGGGGAAACGC
     TTTCGCGGTGCGAAGGGCTTCCCTCTTTCCGCCTGTCCATAGGCCATTCGCCGTCCCAGCCTTGTCCTCTGCGTGCTCCCTCGAAGGTCCCCCTTTGCG

3466 CTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAG
     GACCATAGAAATATCAGGACAGCCCAAAGCGGTGGAGACTGAACTCGCAGCTAAAAACACTACGAGCAGTCCCCCCGCCTCGGATACCTTTTTGCGGTC

3565 CAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGC
     GTTGCGCCGGAAAAATGCCAAGGACCGGAAAACGACCGGAAAACGAGTGTACAAGAAAGGACGCAATAGGGGACTAAGACACCTATTGGCATAATGGCG

3664 CTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCT
     GAAACTCACTCGACTATGGCGAGCGGCGTCGGCTTGCTGGCTCGCGTCGCTCAGTCACTCGCTCCTTCGCCTTCTCGCGGGTTATGCGTTTGGCGGAGA

3763 CCCCGCGCGTTGGCCGATTCATTAATGCA
     GGGGCGCGCAACCGGCTAAGTAATTACGT
```


pCRT7-CT-TOPO-DmSNAP50-V5epitope-6xHis
XbaI
junction marker
PstI
SalI
ScaI
AccI
BbsI
BspMI
ClaI
SacII
PvuII
SalI
AccI
BcgI-2
BcgI-1
BspEI
AflII
ClaI
BspM
MunI
BsmI
PvuII
BsaAI
BstBI
junction marker
HindIII
MluI
AgeI
PmeI
BlpI
EcoO
BspEI
BspHI
SfiI
MscI
SmaI
NcoI
HincII
BssHII
AatII
AvaI
AvaI
SgrAI
SexAI
BanI
NaeI
FseI


NaeI
EagI
DraIII
1882 AGCAGCCGTGGGGGCGGGAGTTCGCCCTGCGCGACCCGGCCGGCAACTGCGTGCACTTCGTGGCCGAGGAGCAGGACTGACACATTGAAAAAGGAAGAG
TCGTCGGCACCCCCGCCCTCAAGCGGGACGCGCTGGGCCGGCCGTTGACGCACGTGAAGCACCGGCTCCTCGTCCTGACTGTGTAACTTTTTCCTTCTC
1981 TATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGA
ATACTCATAAGTTGTAAAGGCACAGCGGGAATAAGGGAAAAAACGCCGTAAAACGGAAGGACAAAAACGAGTGGGTCTTTGCGACCACTTTCATTTTCT
Psp1406I
XmnI
2080 TGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAAT
ACGACTTCTAGTCAACCCACGTGCTCACCCAATGTAGCTTGACCTAGAGTTGTCGCCATTCTAGGAACTCTCAAAAGCGGGGCTTCTTGCAAAAGGTTA
BcgI-2
BcgI-1
2179 GATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGA
CTACTCGTGAAAATTTCAAGACGATACACCGCGCCATAATAGGGCATAACTGCGGCCCGTTCTCGTTGAGCCAGCGGCGTATGTGATAAGAGTCTTACT
ScaI
2278 CTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGC
GAACCAACTCATGAGTGGTCAGTGTCTTTTCGTAGAATGCCTACCGTACTGTCATTCTCTTAATACGTCACGACGGTATTGGTACTCACTATTGTGACG
PvuI
2377 GGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACC
CCGGTTGAATGAAGACTGTTGCTAGCCTCCTGGCTTCCTCGATTGGCGAAAAAACGTGTTGTACCCCCTAGTACATTGAGCGGAACTAGCAACCCTTGG
FspI
Psp1406I
2476 GGAGCTGAATGAAGCCATACCAAACGACGAGAGTGACACCACGATGCCTGTAGCAATGCCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTAC
CCTCGACTTACTTCGGTATGGTTTGCTGCTCTCACTGTGGTGCTACGGACATCGTTACGGTTGTTGCAACGCGTTTGATAATTGACCGCTTGATGAATG
2575 TCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGA
AGATCGAAGGGCCGTTGTTAATTATCTGACCTACCTCCGCCTATTTCAACGTCCTGGTGAAGACGCGAGCCGGGAAGGCCGACCGACCAAATAACGACT
BpmI
Eam1105I
2674 TAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAG
ATTTAGACCTCGGCCACTCGCACCCAGAGCGCCATAGTAACGTCGTGACCCCGGTCTACCATTCGGGAGGGCATAGCATCAATAGATGTGCTGCCCCTC
BanI
2773 TCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACT
AGTCCGTTGATACCTACTTGCTTTATCTGTCTAGCGACTCTATCCACGGAGTGACTAATTCGTAACCATTGACAGTCTGGTTCAAATGAGTATATATGA
BspHI
2872 TTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTT
AATCTAACTAAATTTTGAAGTAAAAATTAAATTTTCCTAGATCCACTTCTAGGAAAAACTATTAGAGTACTGGTTTTAGGGAATTGCACTCAAAAGCAA
2971 CCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACGCGCTAC
GGTGACTCGCAGTCTGGGGCATCTTTTCTAGTTTCCTAGAAGAACTCTAGGAAAAAAAGACGCGCATTAGACGACGAACGTTTGTTTTTTTGCGCGATG
3070 CAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGT
GTCGCCACCAAACAAACGGCCTAGTTCTCGATGGTTGAGAAAAAGGCTTCCATTGACCGAAGTCGTCTCGCGTCTATGGTTTATGACAAGAAGATCACA
AlwNI
3169 AGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGT
TCGGCATCAATCCGGTGGTGAAGTTCTTGAGACATCGTGGCGGATGTATGGAGCGAGACGATTAGGACAATGGTCACCGACGACGGTCACCGCTATTCA
3268 CGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAA
GCACAGAATGGCCCAACCTGAGTTCTGCTATCAATGGCCTATTCCGCGTCGCCAGCCCGACTTGCCCCCCAAGCACGTGTGTCGGGTCGAACCTCGCTT
HaeII
3367 CGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGG
GCTGGATGTGGCTTGACTCTATGGATGTCGCACTCGATACTCTTTCGCGGTGCGAAGGGCTTCCCTCTTTCCGCCTGTCCATAGGCCATTCGCCGTCCC
BsmBI
DrdI
3466 TCGGAACAGGAGACGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTT
AGCCTTGTCCTCTGCGTGCTCCCTCGAAGGTCCCCCTTTGCGGACCATAGAAATATCAGGACAGCCCAAAGCGGTGGAGACTGAACTCGCAGCTAAAAA
NspI
3565 GTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTT
CACTACGAGCAGTCCCCCCGCCTCGGATACCTTTTTGCGGTCGTTGCGCCGGAAAAATGCCAAGGACCGGAAAACGACCGGAAAACGAGTGTACAAGAA
3664 TCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGT
AGGACGCAATAGGGGACTAAGACACCTATTGGCATAATGGCGGAAACTCACTCGACTATGGCGAGCGGCGTCGGCTTGCTGGCTCGCGTCGCTCAGTCA
SapI HaeII
3763 GAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCA
CTCGCTCCTTCGCCTTCTCGCGGGTTATGCGTTTGGCGGAGAGGGGCGCGCAACCGGCTAAGTAATTACGT

pET30b(+)DmTBP-6xHis-S-tags

```
   1 TGGCGAATGG GACGCGCCCT GTAGCGGCGC ATTAAGCGCG GCGGGTGTGG TGGTTACGCG CAGCGTGACC GCTACACTTG CCAGCGCCCT AGCGCCCGCT
     ACCGCTTACC CTGCGCGGGA CATCGCCGCG TAATTCGCGC CGCCCACACC ACCAATGCGC GTCGCACTGG CGATGTGAAC GGTCGCGGGA TCGCGGGCGA
 101 CCTTTCGCTT TCTTCCCTTC CTTTCTCGCC ACGTTCGCCG GCTTTCCCCG TCAAGCTCTA AATCGGGGGC TCCCTTTAGG GTTCCGATTT AGTGCTTTAC
     GGAAAGCGAA AGAAGGGAAG GAAAGAGCGG TGCAAGCGGC CGAAAGGGGC AGTTCGAGAT TTAGCCCCCG AGGGAAATCC CAAGGCTAAA TCACGAAATG
 201 GGCACCTCGA CCCCAAAAAA CTTGATTAGG GTGATGGTTC ACGTAGTGGG CCATCGCCCT GATAGACGGT TTTTCGCCCT TTGACGTTGG AGTCCACGTT
     CCGTGGAGCT GGGGTTTTTT GAACTAATCC CACTACCAAG TGCATCACCC GGTAGCGGGA CTATCTGCCA AAAAGCGGGA AACTGCAACC TCAGGTGCAA
 301 CTTTAATAGT GGACTCTTGT TCCAAACTGG AACAACACTC AACCCTATCT CGGTCTATTC TTTTGATTTA TAAGGGATTT TGCCGATTTC GGCCTATTGG
     GAAATTATCA CCTGAGAACA AGGTTTGACC TTGTTGTGAG TTGGGATAGA GCCAGATAAG AAAACTAAAT ATTCCCTAAA ACGGCTAAAG CCGGATAACC
 401 TTAAAAAATG AGCTGATTTA ACAAAAATTT AACGCGAATT TTAACAAAAT ATTAACGTTT ACAATTTCAG GTGGCACTTT TCGGGGAAAT GTGCGCGGAA
     AATTTTTTAC TCGACTAAAT TGTTTTTAAA TTGCGCTTAA AATTGTTTTA TAATTGCAAA TGTTAAAGTC CACCGTGAAA AGCCCCTTTA CACGCGCCTT
 501 CCCCTATTTG TTTATTTTTC TAAATACATT CAAATATGTA TCCGCTCATG AATTAATTCT TAGAAAAACT CATCGAGCAT CAAATGAAAC TGCAATTTAT
     GGGGATAAAC AAATAAAAAG ATTTATGTAA GTTTATACAT AGGCGAGTAC TTAATTAAGA ATCTTTTTGA GTAGCTCGTA GTTTACTTTG ACGTTAAATA
 601 TCATATCAGG ATTATCAATA CCATATTTTT GAAAAAGCCG TTTCTGTAAT GAAGGAGAAA ACTCACCGAG GCAGTTCCAT AGGATGGCAA GATCCTGGTA
     AGTATAGTCC TAATAGTTAT GGTATAAAAA CTTTTTCGGC AAAGACATTA CTTCCTCTTT TGAGTGGCTC CGTCAAGGTA TCCTACCGTT CTAGGACCAT
 701 TCGGTCTGCG ATTCCGACTC GTCCAACATC AATACAACCT ATTAATTTCC CCTCGTCAAA AATAAGGTTA TCAAGTGAGA AATCACCATG AGTGACGACT
     AGCCAGACGC TAAGGCTGAG CAGGTTGTAG TTATGTTGGA TAATTAAAGG GGAGCAGTTT TTATTCCAAT AGTTCACTCT TTAGTGGTAC TCACTGCTGA
 801 GAATCCGGTG AGAATGGCAA AAGTTTATGC ATTTCTTTCC AGACTTGTTC AACAGGCCAG CCATTACGCT CGTCATCAAA ATCACTCGCA TCAACCAAAC
     CTTAGGCCAC TCTTACCGTT TTCAAATACG TAAAGAAAGG TCTGAACAAG TTGTCCGGTC GGTAATGCGA GCAGTAGTTT TAGTGAGCGT AGTTGGTTTG
 901 CGTTATTCAT TCGTGATTGC GCCTGAGCGA GACGAAATAC GCGATCGCTG TTAAAAGGAC AATTACAAAC AGGAATCGAA TGCAACCGGC GCAGGAACAC
     GCAATAAGTA AGCACTAACG CGGACTCGCT CTGCTTTATG CGCTAGCGAC AATTTTCCTG TTAATGTTTG TCCTTAGCTT ACGTTGGCCG CGTCCTTGTG
1001 TGCCAGCGCA TCAACAATAT TTTCACCTGA ATCAGGATAT TCTTCTAATA CCTGGAATGC TGTTTTCCCG GGGATCGCAG TGGTGAGTAA CCATGCATCA
     ACGGTCGCGT AGTTGTTATA AAAGTGGACT TAGTCCTATA AGAAGATTAT GGACCTTACG ACAAAAGGGC CCCTAGCGTC ACCACTCATT GGTACGTAGT
1101 TCAGGAGTAC GGATAAAATG CTTGATGGTC GGAAGAGGCA TAAATTCCGT CAGCCAGTTT AGTCTGACCA TCTCATCTGT AACATCATTG GCAACGCTAC
     AGTCCTCATG CCTATTTTAC GAACTACCAG CCTTCTCCGT ATTTAAGGCA GTCGGTCAAA TCAGACTGGT AGAGTAGACA TTGTAGTAAC CGTTGCGATG
1201 CTTTGCCATG TTTCAGAAAC AACTCTGGCG CATCGGGCTT CCCATACAAT CGATAGATTG TCGCACCTGA TTGCCCGACA TTATCGCGAG CCCATTTATA
     GAAACGGTAC AAAGTCTTTG TTGAGACCGC GTAGCCCGAA GGGTATGTTA GCTATCTAAC AGCGTGGACT AACGGGCTGT AATAGCGCTC GGGTAAATAT
1301 CCCATATAAA TCAGCATCCA TGTTGGAATT TAATCGCGGC CTAGAGCAAG ACGTTTCCCG TTGAATATGG CTCATAACAC CCCTTGTATT ACTGTTTATG
     GGGTATATTT AGTCGTAGGT ACAACCTTAA ATTAGCGCCG GATCTCGTTC TGCAAAGGGC AACTTATACC GAGTATTGTG GGGAACATAA TGACAAATAC
1401 TAAGCAGACA GTTTTATTGT TCATGACCAA AATCCCTTAA CGTGAGTTTT CGTTCCACTG AGCGTCAGAC CCCGTAGAAA AGATCAAAGG ATCTTCTTGA
     ATTCGTCTGT CAAAATAACA AGTACTGGTT TTAGGGAATT GCACTCAAAA GCAAGGTGAC TCGCAGTCTG GGGCATCTTT TCTAGTTTCC TAGAAGAACT
1501 GATCCTTTTT TTCTGCGCGT AATCTGCTGC TTGCAAACAA AAAAACCACC GCTACCAGCG GTGGTTTGTT TGCCGGATCA AGAGCTACCA ACTCTTTTTC
     CTAGGAAAAA AAGACGCGCA TTAGACGACG AACGTTTGTT TTTTTGGTGG CGATGGTCGC CACCAAACAA ACGGCCTAGT TCTCGATGGT TGAGAAAAAG
1601 CGAAGGTAAC TGGCTTCAGC AGAGCGCAGA TACCAAATAC TGTCCTTCTA GTGTAGCCGT AGTTAGGCCA CCACTTCAAG AACTCTGTAG CACCGCCTAC
     GCTTCCATTG ACCGAAGTCG TCTCGCGTCT ATGGTTTATG ACAGGAAGAT CACATCGGCA TCAATCCGGT GGTGAAGTTC TTGAGACATC GTGGCGGATG
1701 ATACCTCGCT CTGCTAATCC TGTTACCAGT GGCTGCTGCC AGTGGCGATA AGTCGTGTCT TACCGGGTTG GACTCAAGAC GATAGTTACC GGATAAGGCG
     TATGGAGCGA GACGATTAGG ACAATGGTCA CCGACGACGG TCACCGCTAT TCAGCACAGA ATGGCCCAAC CTGAGTTCTG CTATCAATGG CCTATTCCGC
1801 CAGCGGTCGG GCTGAACGGG GGGTTCGTGC ACACAGCCCA GCTTGGAGCG AACGACCTAC ACCGAACTGA GATACCTACA GCGTGAGCTA TGAGAAAGCG
     GTCGCCAGCC CGACTTGCCC CCCAAGCACG TGTGTCGGGT CGAACCTCGC TTGCTGGATG TGGCTTGACT CTATGGATGT CGCACTCGAT ACTCTTTCGC
1901 CCACGCTTCC CGAAGGGAGA AAGGCGGACA GGTATCCGGT AAGCGGCAGG GTCGGAACAG GAGAGCGCAC GAGGGAGCTT CCAGGGGGAA ACGCCTGGTA
     GGTGCGAAGG GCTTCCCTCT TTCCGCCTGT CCATAGGCCA TTCGCCGTCC CAGCCTTGTC CTCTCGCGTG CTCCCTCGAA GGTCCCCCTT TGCGGACCAT
2001 TCTTTATAGT CCTGTCGGGT TTCGCCACCT CTGACTTGAG CGTCGATTTT TGTGATGCTC GTCAGGGGGG CGGAGCCTAT GGAAAAACGC CAGCAACGCG
     AGAAATATCA GGACAGCCCA AAGCGGTGGA GACTGAACTC GCAGCTAAAA ACACTACGAG CAGTCCCCCC GCCTCGGATA CCTTTTTGCG GTCGTTGCGC
2101 GCCTTTTTAC GGTTCCTGGC CTTTTGCTGG CCTTTTGCTC ACATGTTCTT TCCTGCGTTA TCCCCTGATT CTGTGGATAA CCGTATTACC GCCTTTGAGT
     CGGAAAAATG CCAAGGACCG GAAAACGACC GGAAAACGAG TGTACAAGAA AGGACGCAAT AGGGGACTAA GACACCTATT GGCATAATGG CGGAAACTCA
2201 GAGCTGATAC CGCTCGCCGC AGCCGAACGA CCGAGCGCAG CGAGTCAGTG AGCGAGGAAG CGGAAGAGCG CCTGATGCGG TATTTTCTCC TTACGCATCT
     CTCGACTATG GCGAGCGGCG TCGGCTTGCT GGCTCGCGTC GCTCAGTCAC TCGCTCCTTC GCCTTCTCGC GGACTACGCC ATAAAAGAGG AATGCGTAGA
2301 GTGCGGTATT TCACACCGCA TATATGGTGC ACTCTCAGTA CAATCTGCTC TGATGCCGCA TAGTTAAGCC AGTATACACT CCGCTATCGC TACGTGACTG
     CACGCCATAA AGTGTGGCGT ATATACCACG TGAGAGTCAT GTTAGACGAG ACTACGGCGT ATCAATTCGG TCATATGTGA GGCGATAGCG ATGCACTGAC
2401 GGTCATGGCT GCGCCCCGAC ACCCGCCAAC ACCCGCTGAC GCGCCCTGAC GGGCTTGTCT GCTCCCGGCA TCCGCTTACA GACAAGCTGT GACCGTCTCC
     CCAGTACCGA CGCGGGGCTG TGGGCGGTTG TGGGCGACTG CGCGGGACTG CCCGAACAGA CGAGGGCCGT AGGCGAATGT CTGTTCGACA CTGGCAGAGG
2501 GGGAGCTGCA TGTGTCAGAG GTTTTCACCG TCATCACCGA AACGCGCGAG GCAGCTGCGG TAAAGCTCAT CAGCGTGGTC GTGAAGCGAT TCACAGATGT
     CCCTCGACGT ACACAGTCTC CAAAAGTGGC AGTAGTGGCT TTGCGCGCTC CGTCGACGCC ATTTCGAGTA GTCGCACCAG CACTTCGCTA AGTGTCTACA
2601 CTGCCTGTTC ATCCGCGTCC AGCTCGTTGA GTTTCTCCAG AAGCGTTAAT GTCTGGCTTC TGATAAAGCG GGCCATGTTA AGGGCGGTTT TTTCCTGTTT
     GACGGACAAG TAGGCGCAGG TCGAGCAACT CAAAGAGGTC TTCGCAATTA CAGACCGAAG ACTATTTCGC CCGGTACAAT TCCCGCCAAA AAAGGACAAA
2701 GGTCACTGAT GCCTCCGTGT AAGGGGGATT TCTGTTCATG GGGGTAATGA TACCGATGAA ACGAGAGAGG ATGCTCACGA TACGGGTTAC TGATGATGAA
     CCAGTGACTA CGGAGGCACA TTCCCCCTAA AGACAAGTAC CCCCATTACT ATGGCTACTT TGCTCTCTCC TACGAGTGCT ATGCCCAATG ACTACTACTT
2801 CATGCCCGGT TACTGGAACG TTGTGAGGGT AAACAACTGG CGGTATGGAT GCGGCGGGAC CAGAGAAAAA TCACTCAGGG TCAATGCCAG CGCTTCGTTA
     GTACGGGCCA ATGACCTTGC AACACTCCCA TTTGTTGACC GCCATACCTA CGCCGCCCTG GTCTCTTTTT AGTGAGTCCC AGTTACGGTC GCGAAGCAAT
2901 ATACAGATGT AGGTGTTCCA CAGGGTAGCC AGCAGCATCC TGCGATGCAG ATCCGGAACA TAATGGTGCA GGGCGCTGAC TTCCGCGTTT CCAGACTTTA
     TATGTCTACA TCCACAAGGT GTCCCATCGG TCGTCGTAGG ACGCTACGTC TAGGCCTTGT ATTACCACGT CCCGCGACTG AAGGCGCAAA GGTCTGAAAT
3001 CGAAACACGG AAACCGAAGA CCATTCATGT TGTTGCTCAG GTCGCAGACG TTTTGCAGCA GCAGTCGCTT CACGTTCGCT CGCGTATCGG TGATTCATTC
     GCTTTGTGCC TTTGGCTTCT GGTAAGTACA ACAACGAGTC CAGCGTCTGC AAAACGTCGT CGTCAGCGAA GTGCAAGCGA GCGCATAGCC ACTAAGTAAG
3101 TGCTAACCAG TAAGGCAACC CCGCCAGCCT AGCCGGGTCC TCAACGACAG GAGCACGATC ATGCGCACCC GTGGGGCCGC CATGCCGGCG ATAATGGCCT
     ACGATTGGTC ATTCCGTTGG GGCGGTCGGA TCGGCCCAGG AGTTGCTGTC CTCGTGCTAG TACGCGTGGG CACCCCGGCG GTACGGCCGC TATTACCGGA
3201 GCTTCTCGCC GAAACGTTTG GTGGCGGGAC CAGTGACGAA GGCTTGAGCG AGGGCGTGCA AGATTCCGAA TACCGCAAGC GACAGGCCGA TCATCGTCGC
     CGAAGAGCGG CTTTGCAAAC CACCGCCCTG GTCACTGCTT CCGAACTCGC TCCCGCACGT TCTAAGGCTT ATGGCGTTCG CTGTCCGGCT AGTAGCAGCG
3301 GCTCCAGCGA AAGCGGTCCT CGCCGAAAAT GACCCAGAGC GCTGCCGGCA CCTGTCCTAC GAGTTGCATG ATAAAGAAGA CAGTCATAAG TGCGGCGACG
     CGAGGTCGCT TTCGCCAGGA GCGGCTTTTA CTGGGTCTCG CGACGGCCGT GGACAGGATG CTCAACGTAC TATTTCTTCT GTCAGTATTC ACGCCGCTGC
3401 ATAGTCATGC CCCGCGCCCA CCGGAAGGAG CTGACTGGGT TGAAGGCTCT CAAGGGCATC GGTCGAGATC CCGGTGCCTA ATGAGTGAGC TAACTTACAT
     TATCAGTACG GGGCGCGGGT GGCCTTCCTC GACTGACCCA ACTTCCGAGA GTTCCCGTAG CCAGCTCTAG GGCCACGGAT TACTCACTCG ATTGAATGTA
3501 TAATTGCGTT GCGCTCACTG CCCGCTTTCC AGTCGGGAAA CCTGTCGTGC CAGCTGCATT AATGAATCGG CCAACGCGCG GGGAGAGGCG GTTTGCGTAT
     ATTAACGCAA CGCGAGTGAC GGGCGAAAGG TCAGCCCTTT GGACAGCACG GTCGACGTAA TTACTTAGCC GGTTGCGCGC CCCTCTCCGC CAAACGCATA
3601 TGGGCGCCAG GGTGGTTTTT CTTTTCACCA GTGAGACGGG CAACAGCTGA TTGCCCTTCA CCGCCTGGCC CTGAGAGAGT TGCAGCAAGC GGTCCACGCT
     ACCCGCGGTC CCACCAAAAA GAAAAGTGGT CACTCTGCCC GTTGTCGACT AACGGGAAGT GGCGGACCGG GACTCTCTCA ACGTCGTTCG CCAGGTGCGA
3701 GGTTTGCCCC AGCAGGCGAA AATCCTGTTT GATGGTGGTT AACGGCGGGA TATAACATGA GCTGTCTTCG GTATCGTCGT ATCCCACTAC CGAGATGTCC
     CCAAACGGGG TCGTCCGCTT TTAGGACAAA CTACCACCAA TTGCCGCCCT ATATTGTACT CGACAGAAGC CATAGCAGCA TAGGGTGATG GCTCTACAGG
3801 GCACCAACGC GCAGCCCGGA CTCGGTAATG GCGCGCATTG CGCCCAGCGC CATCTGATCG TTGGCAACCA GCATCGCAGT GGGAACGATG CCCTCATTCA
     CGTGGTTGCG CGTCGGGCCT GAGCCATTAC CGCGCGTAAC GCGGGTCGCG GTAGACTAGC AACCGTTGGT CGTAGCGTCA CCCTTGCTAC GGGAGTAAGT
3901 GCATTTGCAT GGTTTGTTGA AAACCGGACA TGGCACTCCA GTCGCCTTCC CGTTCCGCTA TCGGCTGAAT TTGATTGCGA GTGAGATATT TATGCCAGCC
     CGTAAACGTA CCAAACAACT TTTGGCCTGT ACCGTGAGGT CAGCGGAAGG GCAAGGCGAT AGCCGACTTA AACTAACGCT CACTCTATAA ATACGGTCGG
4001 AGCCAGACGC AGACGCGCCG AGACAGAACT TAATGGGCCC GCTAACAGCG CGATTTGCTG GTGACCCAAT GCGACCAGAT GCTCCACGCC CAGTCGCGTA
     TCGGTCTGCG TCTGCGCGGC TCTGTCTTGA ATTACCCGGG CGATTGTCGC GCTAAACGAC CACTGGGTTA CGCTGGTCTA CGAGGTGCGG GTCAGCGCAT
4101 CCGTCTTCAT GGGAGAAAAT AATACTGTTG ATGGGTGTCT GGTCAGAGAC ATCAAGAAAT AACGCCGGAA CATTAGTGCA GGCAGCTTCC ACAGCAATGG
     GGCAGAAGTA CCCTCTTTTA TTATGACAAC TACCCACAGA CCAGTCTCTG TAGTTCTTTA TTGCGGCCTT GTAATCACGT CCGTCGAAGG TGTCGTTACC
4201 CATCCTGGTC ATCCAGCGGA TAGTTAATGA TCAGCCCACT GACGCGTTGC GCGAGAAGAT TGTGCACCGC CGCTTTACAG GCTTCGACGC CGCTTCGTTC
     GTAGGACCAG TAGGTCGCCT ATCAATTACT AGTCGGGTGA CTGCGCAACG CGCTCTTCTA ACACGTGGCG GCGAAATGTC CGAAGCTGCG GCGAAGCAAG
4301 TACCATCGAC ACCACCACGC TGGCACCCAG TTGATCGGCG CGAGATTTAA TCGCCGCGAC AATTTGCGAC GGCGCGTGCA GGGCCAGACT GGAGGTGGCA
     ATGGTAGCTG TGGTGGTGCG ACCGTGGGTC AACTAGCCGC GCTCTAAATT AGCGGCGCTG TTAAACGCTG CCGCGCACGT CCCGGTCTGA CCTCCACCGT
```


T7 promoter
lac operator
XbaI
RBS
NdeI
His tag
thrombin
S-tag
BglII
NcoI
BsgI
junction marker
BglII
NotI
CciNI
XhoI
His tag

Appendix B:

Purification of DmSNAP43 and DmSNAP50 from *Escherichia coli* by nickel chelate chromatography.

Materials

- **Buffers**

6 M Guanidinium Lysis Buffer (pH 7.8): For 400 ml
6 M Guanidine HCl 299.27 gram Guanidine HCl
500 mM NaCl 11.66 gram NaCl
1.7 mM $NaH_2PO_4.2H_2O$ 0.10 gram $NaH_2PO_4.2H_2O$
18.3 mM Na_2HPO_4 1.03 gram Na_2HPO_4
Mix well and adjust the pH with 1N HCl or 1N NaOH as necessary.

Denaturing Binding Buffer (pH 7.8): For 100ml
8 M Urea 48.04 gram Urea
500 mM NaCl 2.92 gram NaCl
1.7 mM $NaH_2PO_4.H_2O$ 0.02 gram $NaH_2PO_4.H_2O$
18.3 mM Na_2HPO_4 0.25 gram Na_2HPO_4
Mix well and adjust the pH with 1N HCl or 1N NaOH as necessary.

Denaturing Wash Buffer I (pH 6.0): For 200ml
8 M Urea 96.09 gram Urea
500 mM NaCl 2.92 gram NaCl
17.5 mM $NH_2PO_4.H_2O$ 0.54 gram $NaH_2PO_4.H_2O$
2.5 mM Na_2HPO_4 0.07 gram Na_2HPO_4
Mix well and adjust the pH with 1N HCl or 1N NaOH as necessary.

Denaturing Wash Buffer II (pH 5.3):
8 M Urea
500 mM NaCl
17.5 mM $NH_2PO_4.H_2O$
2.5 mM Na_2HPO_4
Mix well and adjust the pH with 1N HCl or 1N NaOH as necessary.

Denaturing Elution Buffer (pH 4.0)	For 100ml
8 M Urea	48.04 gram Urea
500 mM NaCl	2.92 gram NaCl
20 mM NaH_2PO_4	0.31 gram Na_2HPO_4

Mix well and adjust the pH with 1N HCl or 1N NaOH as necessary.

Phosphate Dialysis Buffer to remove the Guanidine before SDS-PAGE (pH 7.8): For 1000ml	
8 M Urea	480.48 gram Urea
500 mM NaCl	29.22 gram NaCl
1.7 mM $NaH_2PO_4.H_2O$	0.26 gram $NaH_2PO_4.H_2O$
18.3 mM Na_2HPO_4	2.59 gram NaH_2PO_4

Mix well and adjust the pH with 1N HCl or 1N NaOH as necessary.

Remember :
1) For all the above buffers you should mix well and adjust the pH with 1N HCl or 1N NaOH as necessary.
2) Store buffers at room temperature.
3) Check pH immediately before use since the pH may change over time.

- **Additional materials**

1) ProBond ^{TM} resin in precharged with Ni^{2+} ions and appears blue in color. It is provided as a 50% slurry in 20% ehtanol. Binding capacity of ProBond resin is 1-5 mg of protein per ml of resin. (Part no. 46-0019- Invitrogen)
2) Poly-Prep Chromatography Columns – Catalog # 731-1550 - BIO-RAD.
3) BL21/dnaY competent cell for transformation reactions. dnaY plasmid contains a gene for tRNA that reads rare AGA and AGG arginine codons. It also provides resistance to Kanamycin.
4) DmSNAP43 and DmSNAP50 in pCRT7-TOPO expression vector, which provides resistance to Ampicillin.
5) LB agar plates and LB medium.
6) SOC medium.
7) IPTG to induce expression.

Expression of DmSNAP43 or DmSNAP50 in *Escherichia coli*

Day One

1. Thaw on ice, one vial of competent BL21/dnay.
2. Add 10 ng DNA (DmSNAP43 or DmSNAP50 expression plasmid) into the vial and mix it by stirring gently with pipette tip.
3. Incubate on in ice for 30 minutes.
4. Heat-Shock the cells for 30 seconds in 42°C water bath without shaking.
5. Immediately transfer the tube to ice. (2-10 minutes).
6. Add 250 µl of room temperature SOC medium.
7. Cap the tube tightly; tape the tube on its side (for better aeration) and incubate at 37°C for 30 minutes with shaking horizontally (200 rpm) in Innova 4335 incubator shaker.
8. Plate 50 µl to the first LB agar plate that contains 100 µl/ml Ampicillin and 50 µg/ml Kanamycin.
9. Plate 100 µl to the second LB agar plate that contains 100 µg/ml Ampicillin and 50 µg/ml Kanamycin.
10. Plate 150 µl to the third LB agar plate that contains 100 µg/ml Ampicillin and 50 µg/ml Kanamycin.
11. Incubate overnight (16 hrs) at 37°C.

Day Two

12. Use a sterile loop to pick one large isolated colony and make several streaks on a fourth plate that contains 100 µg/ml ampicillin and 50 µg/ml Kanamycin and incubate overnight (16 hrs) at 37°C. You could do that by touching a sterile loop to the bacteria and making several streaks in plate, the goal is to obtain isolated colonies on a large part of the agar surface as shown below.

13. Tightly wrap the other three plates prepared on day one with parafilm and store them in refrigerator at 4°C.

Day Three

14. Pick 5-10 colonies from the fourth plate from day two and add it to 100 ml LB medium containing100 µg/ml Ampicillin and 50 µg/ml Kanamycin.

15. Grow the 100 ml culture until OD_{600} reaches 0.3 to 1 .
16. Prepare two flasks with each flask having 1000 ml of LB medium containing 100 μg/ml Ampicillin and 50 μg/ml Kanamycin.
17. Add 30 ml of the 100 ml culture to each flask.
18. Read the OD_{600} for the two flasks until OD_{600} reaches 0.5, then add 0.244 g of IPTG per 1030 ml culture.
19. Crow the cells for three hours at 37°C (induced expression).

Preparation of ProBond™ Columns

(The work should be at room temperature)

20. While you are waiting for the culture to grow, check the pH (7.8) for denaturing binding buffer. Adjust as needed by using 1N HCL or 1N NaOH .
21. Start preparing the Ni column .You will use 8 ml of resuspended ProBond Resin,which has 4 ml settled resin and 4 ml of (20% ethanol + 80% water)
22. Resuspend the ProBond Resin (Part no. 46-0019- Invitrogen), in its bottle by inverting and gently tapping the bottle repeatedly.
    Do not use strong reducing agents such as DTT with ProBond columns. DTT reduces the nickel ions in the resin. In addition, do not use strong chelating agents such as EDTA or EGTA in the loading buffers or wash buffers, as these will strip the nickel from the columns.
23. Pipet 8 ml of reususpended resin into new clean 50 ml screw cap tube, allow the resin to settle completely by gently pelleting it by low speed centrifugation. Use Sorvall Legend RT Centrifuge at (800 x g). You will use two 50 ml screw cap tubes, one for centrifuging the resin and one to balance it. Gently aspirate the supernatant.
24. Add 12 ml of distilled water and resuspend the resin by using a rotating wheel for two min, repellet the resin by low speed centrifugation and aspirate the supernatant.
25. Add 12 ml of Denaturing Binding buffer and resuspend the resin by using a rotating wheel for two min, repellert the resin by low speed centrifugation and aspirate the supernatant.
26. Repeat step 25 with 12 ml of denaturing binding buffer, but leave 2 ml of the denaturing binding buffer on the top of the resin to avoid drying the resin while it stored at 4°C. Store prepared column at 4°C.
27. Equilibrate 320 ml of the Guanidinium lysis buffer (pH 7.8), 16 ml of denaturing binding buffer, 16 ml of denaturing wash buffer I (pH 6.0), 16 ml of denaturing wash buffer II (pH 5.3) and prepare 20 ml of denaturing elution buffer at room temperature.
28. Check the pH for each buffer and keep them at room temperature until you use them the next day.

Harvesting the cells

29. After three hours, harvest cells from 2060 ml culture by centrifugation (e.g. 5000 rpm for 15 minutes in SLA-3000 rotor) you will use eight bottles each one containing ≈ 260 ml of the cell culture.
30. Discard the supernatant and store the cell pellet in each tube at -80°C until next day.

Lysis of cells

Day four

31. Check the pH for the lysis buffer again (pH 7.8) and equilibrate it at 37°C by incubating the lysis buffer in 37°C water bath for 15 minutes.
32. Resuspend the cell pellet in each tube in 40 ml of lysis buffer (remember: you will have eight tubes, each tube will have 40 ml of cell lysate)
33. Slowly rock the cells for 15 minutes at room temperature to assure thorough cell lysis. Use Reliable Scientific Rocker at Speed 5.
34. Transfer the cell lysis to sixteen 50 ml screw cap tubes each tube containing 20 ml of the cell lysis.
35. Sonicate the cell lysate on ice with three 5-second pulses at high intensity in each tube. Branson Sonifier 450 at 7 output and 80 Duty cycles % and using the Micro-tip. Make sure that the tip of the sonicator is immersed in the whole cell lysate not just the top part of the cell lysate. Start sonicating tube #1 for 5-second and let it cool in ice then do the same with tube # 2 until you finish the sixteen tubes. Repeat the cycle two more times for all tubes.
36. Transfer the sonicated lysate to eight new clean tubes for SS-34 rotor.
37. Centrifuge the lysate at 2,988 xg (5000 rpm) in SS-34 rotor for 15 min to pellet the cellular debris. (Remember: You need the supernatant which is the lysate). Remove, combine, and save the lysates for denaturing purification on Ni colmn.
38. Remove 500 µl from the lysate as sample number one (lysate before Ni column sample) for SDS-PAGE analysis. (You will dialyze this sample against phosphate buffer containing no guanidine before SDS-PAGE analysis).

Purification under denaturing conditions

39. Bring the prepared resin (prepared in above steps 20 to 26) from 4°C and resuspend the resin in its tube by gently tapping the tube repeatedly. Transfer equal amounts ≈1 ml of the resuspended resin to eight clean 50 ml screw cap tubes. (So you will work with eight tubes each tube has ≈ 0.5 ml of settled Ni resin).
40. Use Sorvall Legend RT centrifuge at low speed (800 x g) to pellet the resin for one minute and gently aspirate the supernatant.

41. Add 40 ml of the lysate to each tube.
42. Bind for one hour at room temperature using a rotating wheel to keep the resin suspended in the lysate solution.
43. While you are waiting for the binding step, check the pH for each denaturing buffer again (see step 27).
44. Settle the resin by using Sorvall Legend RT centrifuge at low speed (800 x g) and carefully aspirate the supernatant. (Save 1ml from one tube as sample number two (flowthrough sample), you will dialyze this sample against phosphate buffer containing no guanidine)
45. Wash each tube with 0.5 ml of denaturing binding buffer by resuspending the resin by using rotating wheel for two minutes. Transfer the resuspended resion from each tube to the empty10-ml column (Poly-Prep Chromatography Columns – Catalog # 731-1550 - BIO-RAD)
46. Add 0.5 ml denaturing binding buffer to each 50 ml screw cap tube to wash any remaining resin in the tube and transfer it to the same column. You will have one column with 4 ml settled resin and ≈ 8 ml of binding buffer.
47. Resuspend the resin in Poly-Prep column by using rotating wheel for one minute. Settle the resin by using Sorvall Legend RT centrifuge at low speed (800 x g) to pellet the resin and carefully aspirate the supernatant. (Save 1ml as sample number three (Binding –sample one).
48. Wash the column with 8 ml of denaturing binding buffer by resuspending the resin by using rotating wheel for two minutes. Settle the resin by using Sorvall Legend RT centrifuge at low speed (800 x g) and carefully aspirate the supernatant. (Save 1 ml as sample number four (Binding- sample two)).
49. Repeat the wash step three times with 4 ml of wash buffer pH 6.0 and save 1 ml sample each time.
50. Repeat the wash step three times with 4 ml of wash buffer pH 5.3 and save 1 ml sample each time.
51. Clamp the column in vertical position and snap off the cap on the lower end. Elute the protein by adding 10 ml of denaturing elution buffer and collect 1ml fraction. (Do not resuspend the resin.)You will have ten fractions, 1ml each.
52. Repeat step 52 by eluting the protein with second 10 ml of the denaturing elution buffer and collect the second ten fractions. You will have total twenty elution fractions.
53. Remember to keep all elution fractions directly in ice and store all the twenty fractions in –80°C freezer right after collecting all the fractions.
54. Wash the column containing the used resin with 10 ml 20% ethanol twice and then add 10 ml 20% ethanol, cap it and store at 4°C with label for recycling.
55. Start preparing the phosphate dialysis buffer. To perform SDS-PAGE with samples in guanidinium lysis buffer, you need to dialyze the samples prior to SDS-PAGE to prevent the precipitation of SDS. (The first two samples need dialysis against phosphate buffer [containing no guanidine] for a few hours before SDS-PAGE analysis). The samples are: lysate before the column, and flow through.

Day Five

56. Start preparing two gels for SDS-PAGE analysis (Western-blot and Coomassie-Stained gel). You may see the protein band in all the twenty fractions in both Western-blot and Coomassie-Stained gel.

Second Purification: under Native Conditions to Re-nature protein and to remove Urea.

Materials

Adjusting buffer (pH 10.0)	For 20 ml
8 M urea	9.60 gram of Urea
500 mM NaCl	0.58 gram of NaCl
20 mM $Na_2HPO_4.H_2O$	0.05 gram of $Na_2HPO_4.H_2O$

After mixing and dissolving the above components, bring the pH of the adjusting buffer to 10.0 using NaOH. Store buffer at room temperature.

Native washing buffer (pH 8.0)	For 20 ml
500 mM NaCl	0.58 gram of NaCl
25 mM $Na_2HPO_4.H_2O$	0.07 gram of $Na_2HPO_4.H_2O$

After mixing and dissolving the above components, bring the pH to 8.0 using 1N NaOH or 1N HCl. Store buffer at 4°C.

Native eluting buffer (pH 8.0)	For 20 ml
500 mM Imidazole	0.61 gram of Imidazole
150 mM NaCl	0.17 gram of NaCl
25 mM $Na_2HPO_4.H_2O$	0.07 gram of $Na_2HPO_4.H_2O$

After mixing and dissolving the above components, bring the pH to 8.0 using 1N NaOH or 1N HCl. Store buffer at 4°C.

57. After the purification under denaturing conditions, pool as many fractions as have significant amount of protein as determined by Coomassie stained gels. (I pooled all twenty elution fractions with 20 ml total) from the denaturing purification.
58. Adjust the pH of the pooled elution fractions to pH 7.8 – 8.0 by using the adjusting buffer.
59. Prepare 1ml resin in one 50ml screw cap tube under denaturing condition similar to described in steps 21-26.
60. Apply the pooled elutions above to the resin at room temperature.

61. Let the pooled fractions from the denaturing purification (adjusted to pH 7.8-8.0) bind to the resin for 2 hours.

In 4°C cold room

62. Wash resin three times with 2 ml cold native washing (pH 8.0) buffer at 4°C.
63. Wash the resin one more time and transfer the re-suspended resin from the tube to empty 10-ml column.
64. Elute with 10 ml of the native elution buffer and collect 1 ml fractions. Store fractions at - 80°C after removing samples for SDS gel analysis.
65. Run SDS-PAGE gel. Stain with Coomassie blue.
66. Save chosen fractions after looking at Coomassie gel. Dialyze elution fractions twice against native elution buffer without imidazol to remove the imidazol. Three hours dialysis each time in cold room. (I pooled elution # 3 to elution # 7 with total 5 ml but I did not pool elution # 2 which is 1ml volume because it had most of the eluted protein; I dialyzed the pooled fractions and elution # 2 in two separate dialysis tubes).
67. Remove 50 μl sample for SDS-PAGE analysis and to determine protein concentration using Bradford assay.
68. Aliquot the rest of protein sample into screw cap 1.5 ml tubes, each tube 500 μl and store the tubes in - 80°C freezer.
69. Determine protein concentration by Bradford assay.
70. Run SDS gel to gauge purity of purified protein.
71. Typical results

    DmSNAP43: 7.0 mg protein in 1 ml for elution # 2 (7 mg/ml).
    2.3 mg protein in 4.2 ml for pooled elutions (0.5 mg/ml).

    DmSNAP50: 3.7 mg protein in 1 ml for elution #2 (3.7 mg/ml)
    4.9 mg protein in 4 ml for pooled elutions. (1.2 mg/ml)

Appendix C:

Purification of TBP from *Escherichia coli* by nickel chelate chromatography and S-protein affinity chromatography.

Materials

Bind/Wash buffer	For 100 ml
20 mM Tris-HCl (pH 7.5)	2 ml of 1M Tris-HCl (pH 7.5)
0.15 M NaCl	0.87 gram of NaCl
0.1% Triton X-100	0.1 ml of Triton X-100

Elution buffer (pH 2.0)	For 20 ml
0.2 M citrate	2 ml of 2 M citrate (pH 2.0)

2 M citrate (pH2.0), for 50 ml
19.21 gram of citric acid
Adjust pH with 10 M KOH to pH 2.0

Additional materials

1) S-protein Agarose (Novagen CAT # 69704-3).
2) Recombinant Enterokinase to cleave the S-Tag (Novagen CAT # 69066-3).
3) EKapture Agarose (Novagen CAT # 69068-3).
4) Rosetta2(DE3) competent cells for TBP expression which enhance expression of proteins that contain codons rarely used in *E.coli* (AGG,AGA,AUA,GUA,CCC,GGA). (CAT # 70950-3- Novagen).

Drosophila melanogaster TBP bacterial expression plasmid:
The *D. melanogaster* gene for TBP was purchased from Research Genetics and re-cloned into the pET-30b(+) vector (Novagen) that provided 6xHis and S tags at the amino terminus of the encoded protein.

Purification by nickel chelate chromatography

1. Purify TBP by two purifications by nickel chelate chromatography first under denaturing conditions and then under native conditions respectively from 2000 ml culture using Rosetta2(DE3) cells for TBP expression (described in Appendix B).

2. Pool the ten elution fractions (total 10 ml) from native purification by nickel chelate chromatography.

S-protein affinity chromatography

3. Dialyze the pooled fractions against 500 ml of bind/wash buffer in cold room overnight, then against another 500 ml of bind/wash buffer for three hours.
4. Filter the elution fractions after dialysis through a 0.22 micron membrane to prevent clogging of the resin. Save 30 μl sample for SDS-PAGE analysis.
5. Gently suspend the S-protein Agarose (Novagen CAT # 69704-3) by inversion and pipet 4 ml of the slurry (equivalent to 2 ml settled resin) to 15 ml screw cap tube. Centrifuge at low speed (500 x g) for 5 minutes and carefully aspirate the supernatant (Remember: you need to cut the tip of 1 ml pipet tip to transfer the resin).
6. Resuspend the S-protein Agarose in 2 ml of bind/wash buffer and centrifuge at low speed (500 x g) for 5 minutes and carefully aspirate the supernatant.
7. Add the filtered elution fractions to the prepared 2 ml settled resin.
8. Mix thoroughly and incubate at room temperature on an orbital shaker for 30 minutes. Do not shake vigorously as this will tend to denature protein.
9. Centrifuge the entire volume at low speed (500 x g) for 10 minutes and carefully aspirate the supernatant. Save 30 μl sample for SDS-PAGE analysis.
10. Resuspend the S-protein Agarose, which now contains bound tagged TBP in 10 ml bind/wash buffer. Mix by gently vortexing or by repeated inversion (avoid vigorous vortexing).
11. Repeat the washing twice to wash away unbound proteins. Save 30 μl sample from each wash for SDS-PAGE analysis.
12. Remove the final supernatant and elute the tagged TBP with 5 ml of 0.2 M citrate (pH 2.0) elution buffer by adding 5 ml of elution buffer and mix thoroughly and incubate at room temperature on an orbital shaker for 10 minutes.
13. Centrifuge the entire volume at low speed (500 x g) for 10 minutes and carefully aspirate the 5 ml elution sample.
14. Use NaOH to increase the pH to 7.0 right after elution by adding one diminutive drop of 10 N NaOH, mix, and check pH (accument BASIC-AB15 pH Meter). Repeat as necessary.
15. Save 30 μl sample for SDS-PAGE analysis.

S-Tag cleavage by Enterokinase

16. Dialyze the elution sample in cold room against 500 ml of bind/wash buffer twice, the first dialysis for over night and the second one for 4 hours.
17. Save 30 μl sample for SDS-PAGE analysis.

18. Add 20 units of recombinant Enterokinase to cleave the S-Tag (Novagen CAT # 69066-3) from the sample that contains tagged TBP in bind/wash buffer.
19. Incubate sample for up to 16 hours at room temperature.
20. After 16 hours incubation, save 30 μl sample for SDS-PAGE analysis.
21. Thoroughly resuspend the EKapture Agarose (Novagen CAT # 69068-3) by inversion and add 1ml of the slurry to the above Enterokinase reaction to remove the Enterokinase. Mix thoroughly and incubate for 10 minutes at room temperature on an orbital shaker.
22. Centrifuge at (500 x g) for 5 minutes and transfer the supernatant (which contains the TBP which no longer contains the S -Tag peptide) to fresh tube.
23. Wash the EKapture Agarose with an additional 1-2 ml of bind/wash buffer.
24. Centrifuge and pool the second supernatant with the previous supernatant. The final volume for the elution will be ≈ 6-7 ml.
25. Run SDS-PAGE gel with all the saved samples for analysis, stain with Coomassie blue stain.

Procedure for concentrating untagged TBP

26. Select two of Centricon YM-30 (Amicon – CAT # 4209 – clear) centrifugal devices (with molecular weight cut-off 30,000 Daltons).
27. Insert sample reservoir into filtrate vial according to manufacturer's instruction.
28. Add protein sample that contains the untagged TBP (2 ml maximum volume). Do not touch membrane with pipette tip.
29. Seal the device by covering the sample reservoir with parafilm in order to minimize sample evaporation. Then use needle to punch some holes on the parafilm to release the vacuum.
30. Place both covered devices and attached filtrate vial into an SS-34 centrifuge rotor: counterbalance with each other.
31. Spin Centricon YM-30 centrifugal devices at 1000-5000 X g (3000-6500 rpm) in a SS-34 rotor. Check the protein level every hour. When the protein left in sample reservoir is less than 500 μl,
32. Keep repeating the above step (# 31) by adding another 2ml from the original sample until you concentrate the entire original sample.
33. You should have around 800 μl as a final volume of concentrated protein from both devices. (This will take 3 hours). Do not exceed centrifugation limits described above.
34. Remove centrifugal filter assembly from centrifuge; then separate filtrate vial from sample reservoir.
35. Place retentate vial over sample reservoir and invert unit to recover the retentate. Centrifuge at 300-1000 Xg (1600-3000 rpm) in a SS-34 rotor for 2 minutes to transfer concentrate into retentate vial.
36. Remove device from centrifuge. Separate retentate vial from concentrator. The concentrated protein is in the retentate vial. Remove 30 μl sample for SDS-PAGE analysis and to determine protein concentration using Bradford assay.

37. Store the rest of concentrated sample into screw cap 1.5 ml tube in 80°C freezer to send it to a commercial vendor for polyclonal antibody production in rabbits.
38. After Bradford assay, you should be able to determine the protein concentration and determine how much of protein you should send.

Appendix D:

Protocol for Chromatin Immunoprecipitation assay (ChIP) with urea denaturation

(modified from a protocol obtained from David S. Gilmour: http://www.personal.psu.edu/faculty/d/s/dsg11/labmanual/Chromatin_structure/ChIP_for_*Drosophila*_cells_preferred.html)

Materials

Phosphate-buffered Saline (PBS) (pH 7.4)	For 1000 ml
137 mM NaCl	8 gram of NaCl
2.7 mM KCl	0.2 gram of KCl
10 mM Na_2HPO_4	1.44 gram of Na_2HPO_4
2 mM KH_2PO_4	0.24 gram of KH_2PO_4

Store at 4°C

Sonication buffer	For 30ml
10 mM Tris-HCl (pH 8.0)	0.3 ml of Tris-HCl (pH 8.0)
1 mM EDTA	0.06 ml of 0.5 M EDTA
0.5 mM EGTA	0.15 ml of 0.1 M EGTA

Store at 4°C

ChIP buffer	For 500 ml
10 mM Tris-HCl (pH 8.0)	5 ml of 1M Tris-HCl (pH 8.0)
1 mM EDTA	1 ml of 0.5M EDTA
0.5 mM EGTA	2.5 ml of 0.1M EGTA
0.5 mM PMSF (added right before use)	0.25 ml of 1M PMSF
10% glycerol	50 ml of glycerol
0.1% Na-deoxycholate	0.5 gram of Na-deoxycholate
1% Triton X-100	5 ml of Triton X-100

Store at 4°C.

Low salt wash buffer	For 200 ml
0.1% SDS	0.2 gram of SDS
1% Triton X-100	2 ml of Triton X-100
2 mM EDTA	0.8 ml of 0.5M EDTA
20 mM Tris-HCl (pH 8.1)	4 ml of 1MTris-HCl (pH 8.1)
150 mM NaCl	1.75 gram of NaCl

Store at 4°C

High salt wash buffer	For 200ml
0.1% SDS	0.2 gram of SDS
1% Triton X-100	2 ml of Triton X-100
2 mM EDTA	0.8 ml of 0.5M EDTA
20 mM Tris-HCl (pH8.1)	4 ml of 1M Tris-HCl (pH 8.1)
500 mM NaCl	5.8 gram of NaCl

Store at 4°C

Lithium wash buffer	For 200ml
0.25 M LiCl	2.11 gram of LiCl
1% NP-40	2 ml of NP-40
1% Na-deoxycholate	2 gram of Na-deoxycholate
1 mM EDTA	0.4 ml of 0.5 M EDTA
10 mM Tris-HCl (pH 8.1)	2 ml of 1 M Tris-HCl (pH 8.1)

Store at 4°C

TE buffer	For 200ml
10 mM Tris-HCl (pH 8.0)	2 ml of 1 M Tris-HCl (pH 8.0)
1 mM EDTA	0.4 ml of 0.5 M EDTA

Store at 4°C

Prepare fresh Formaldehyde cross-linking solution buffer right before use

Formaldehyde cross-linking solution	For 4 ml
11% Formaldehyde	1.19 ml of 37% Formaldehyde
0.1 M NaCl	0.08 ml of 5 M NaCl
1 mM EDTA	0.008 ml of 0.5 M EDTA
0.5 mM EGTA	0.02 ml of 0.1 M EGTA
50 mM Tris-HCl (pH 8.0)	0.2 ml of 1 M Tris-HCl (pH 8.0)

Prepare fresh ChIP elution buffer right before use. Keep at room temperature.

ChIP elution buffer	For 10 ml
1% SDS	0.1 gram of SDS
0.1 M NaHCO3	0.08 gram of NaHCO3

Siliconize the tubes

You will need 1.5ml conical screw cap tubes with Cap and O-ring/Sterilized.
(Fisher Scientific Cat # 02-681-373).
You will need Sigmacote (SL2-25ML – 085K4352). Use 1 ml of Sigmacote to siliconize the tubes by splashing the tubes and the screw caps, aspirating away, and air-drying the tubes for 24 hours. The next day, rinse the tubes with water, and let air-dry for 24 hours. After the tubes are completely dry, you can use them.

Remember: using siliconized tubes is critical to reduce non-specific background and to obtain a high yield of specific DNA.
Siliconized tubes are necessary to eliminate DNA bound nonspecifically to the tube's surface.

Sonication conditions to shear chromatin

Place 1ml of cell suspension in a 15 ml blue cap Falcon tube. Sonicate the suspension using the Branson Sonifier 450 with the microtip. (60% duty cycles, 1.5 output). Sonicate for 20 seconds on, 60 seconds off for 10 cycles. During sonication, you should keep the tube in an ice/salt bath at all times. Make sure that the tip of the sonicator is immersed in the cell suspension. Avoid the formation of air bubbles.

Determining the size of the DNA

(You don't need to determine the size of the DNA fragments, unless you have a problem with the sonication conditions).

1. Prepare a 2% agarose gel, use three combs together to make thick wells in the gel.
2. Transfer 1 ml of lysate after sonication to a fresh micro-tube. Centrifuge at 12000 rpm (Marathon micro A Centrifuge) for 10 minutes.
3. Transfer 50µl from the supernatant to a fresh micro-tube and add the following:
   2.5 µl of 2 µg/µl of Proteinase K
   5 µl of 0.5 µg/µl Rnase A
   5 µl of 10% SDS
   2 µl of 5 M NaCl
   Mix and incubate at RT for 30 mins.
4. Reverse crosslinking by incubation for 4 hrs at 65°C.
5. Mix 18 µl from the above lysate sample (after reversal of crosslinks) with 2 µl of 10X loading dye (fast dye only) to load into the gel.
6. To check DNA size distribution you will need DNA Ladder. Mix 2 µl of 1 KD ladder or 100 bp ladder with 2 µl of 10X loading dye and 16 µl water.
7. Run the gel at 100 Volt until the dye migrates to the bottom of the gel.
8. Stain with ethidium bromide for 30 minutes. The average length of the DNA fragments should be around 500 bp.

Day one

Formaldehyde crosslinking

1. Grow three tissue culture dishes of *Drosophila* S2 cells to 90% confluency. Pool all three dishes, and pipet 25 ml of pooled culture into a 50 ml blue cap Falcon tube.
2. Add to the cells 2.5 ml of freshly prepared 11% Formaldehyde cross-linking solution to give a final formaldehyde concentration of 1%. Incubate at room temperature for 10 minutes using rotating wheel (Labquake).
3. Quench the crosslinking by adding 2.5 M glycine to a final concentration of 240 mM (2.64 ml of 2.5 M glycine). Incubate at room temperature for 10 minutes on rotating wheel.
4. Collect cells by centrifugation at 700 xg (SS34-3000 rpm) for 10 minutes.
5. Resuspend and wash cells with 10ml cold PBS buffer. Collect cells by centrifugation at 700 xg for 10 minutes.
6. Resuspend cell pellets in 1 ml sonication buffer, supplemented with fresh protease inhibitors as follows: 1 μg/ml aprotinin, 1μg/ml pepstatin A, 0.5 mM PMSF. (1 μl of 1 mg/ml aprotinin, 1 μl of 1 mg/ml pepstatin and 0.5 μl of 1 mM PMSF).
7. Transfer cell suspension to a 15 ml blue cap Falcon tube. Mount tube in an ice/salt bath with the sonicator. Sonicate 10 cycles: 20 seconds on, 60 seconds off.

Work will be in 4°C cold room

8. Transfer cell lysate to a 1.5 ml micro-tube, Centrifuge at 13,000 rpm at 4°C for 10 minutes to remove debris.
9. Transfer supernatant to a 15 ml blue cap Falcon tube, and mix with an equal amount of 6 M Urea and mix well. The supernatant will be around 1 ml before adding the Urea. If you need to check the average length of the DNA fragments, save 50 μl sample to determine DNA size in agarose gel at this step before adding Urea. Average length of the DNA should be 500 bp
10. Dialyze overnight in 4°C cold room against 200 ml of ChIP buffer (10 mM Tris-HCl pH8.0, 1 mM EDTA, 0.5 mM EGTA, 0.5 mM PMSF, 10% glycerol, 0.1% Na-deoxycholate, 1.0% Triton X-100).

Day Two

Start using siliconized tubes with siliconized caps

11. Preparation of protein A agarose (PIERCE prod # 20333). Transfer (with a pipette whose tip is cut to widen the hole) 80µl resuspended protein A agarose beads (equivalent to 40 µl of settled beads) to a fresh siliconized micro-tube and do the wash steps below. (I washed each wash for few seconds, then spin at 3200 rpm for 1 minute).

    a. Twice with 1 ml of water.
    b. Twice with 1 ml of TE buffer (10 mM Tris-HCl pH8.0, 1 mM EDTA)
    c. Once with 1 ml of TE+BSA buffer (10 mM Tris-HCl pH8.0, 1 mM EDTA, 1 mg/ml BSA).
    d. Add 40 µl of TE+BSA buffer, and resuspend the beads.

    Remember: during washes, do not spin the beads at full speed because they can be crushed.

12. Centrifuge the dialyzed chromatin solution (from step 10) to remove insoluble debris.
13. Transfer the supernatant (chromatin solution) to the tube containing 80 µl prepared beads (40 µl settled beads) suspended in 10 mM Tris-HCl pH8.0, 1 mM EDTA, 1 mg/ml BSA (prepared in step 11).
14. Pre-clear the chromatin solution by mixing the solution with beads gently for 30 min in 4°C. Use a rotating wheel (Labquake) to keep the beads suspended in chromatin solution.

    Remember: Pre-clearing is essential to reduce nonspecific background.

15. Centrifuge at 3200 rpm for 5 minutes. Remove the supernatant (Pre-cleared chromatin solution) to fresh siliconized labeled tubes and store at -80°C for future use. Aliquot 150 µl into each tube for chromatin immunoprecipitation (ChIP). Save one 150 µl tube as total input DNA sample. At this point you could stop and save supernatants for future use, or continue with ChIP as follows:
16. Add 4 µl of antiserum or 4 µl pre-immune serum to 150 µl aliquot of pre-cleared chromatin solution. Incubate on rotating wheel overnight in 4°C cold room.

Day Three

17. Prepare 35 µl of Protein A agarose (17.5 µl settled beads) in labeled tube as in step (11) for each ChIP. To the 35 µl of bead suspension add the chromatin solution that was incubated overnight with antiserum or with pre-immune serum.

Mix the suspension and incubate it using a rotating wheel for two hours in 4°C cold room.

18. Centrifuge the beads at 3200 rpm for 5 minutes. The beads are the important fraction, but save the supernatants in -80°C in case they are needed in the future.
19. Wash the immunoprecipitates with buffers as indicated below. Incubate each wash at 4°C for at least 5 minutes on rotating wheel. (I incubated for 10 minutes). Then centrifuge at 3200 rpm for 1 minute and remove the supernatant.
20. Wash the beads six times with 1 ml of low salt wash buffer at 4°C.
21. Wash three times with 1 ml of high salt wash buffer at 4°C.
22. Wash two times with 1 ml of Lithium Wash Buffer on rotating wheel at 4°C. (I incubated first wash for two hours, then second one overnight).

Remember: Overnight wash is necessary to reduce nonspecific background.

Day Four

23. Wash beads once with 1 ml of cold TE buffer in 4°C.
24. Resuspend beads in 1 ml of cold TE buffer and transfer to a fresh siliconized tube at 4°C to eliminate DNA bound nonspecifically to the original tube's surface.
25. Elute the protein-DNA complexes by adding 250 µl of freshly prepared ChIP elution buffer to the pelleted protein A agarose/antibody/protein/DNA complexes. Vortex briefly to mix and incubate at room temperature for 15 minutes with rotation.
26. Spin-down the beads at 3200 rpm and carefully transfer the supernatant fraction (eluate) to another fresh siliconized tube. Elute once again with 250 µl of elution buffer. Combine eluates (total volume will be approximately 500 µl) In addition, prepare input DNA by adding 350 µl of elution buffer to 150 µl chromatin solution starting material/ input to make the final volume 500 µl.
27. Add 20 µl of 5 M NaCl to the eluates and incubate at 65°C for four hours to reverse formaldehyde crosslinks.
28. Add 10 µl of 0.5 M EDTA, 20 µl of 1 M Tris-HCl (pH 6.5) and 10 µl of 2 mg/ml proteinase K and incubate at 45°C for one hour.
29. Using QIAquick PCR Purification Kit (QIAGEN Cat. No. 28104), follow the steps on page 18 in the QIAquick Spin Handbook for purification. Use 2500 µl of PB buffer as 5 volumes per 1 volume of the sample (500 µl) in the binding step. Use 30 µl of TE buffer to elute in the eluting step. Store purified DNA at -20°C.
30. For PCR reaction you will need 45 µl of Platinum PCR SuperMix (Invitrogen – Cat. No. 11306-016), 200 ng of each primer (forward primer and reverse primer) and 2 µl of purified DNA.
    Run 27 cycles (ChIP64 under Nermeen's folder in the PCR machine).

94°C. for 2:00min
94°C.for 0:30 min. Denature
64°C for 0:45 min Anneal primers
72°C for 1:00 min Extension
Goto 2, 26 times.

72°C for 10:00 min
4°C forever

Remember: The conditions may need to be adjusted depending upon the efficiency of the reactions.

31. Run 18 µl of each PCR reaction with 2 µl of 10X loading dye (fast dye only) on 8% non-denaturing poly-acrylamide in 1X TBE at 150 Volt, until bromo phenol blue migrates to bottom of the gel. Mix 2 µl of 100 bp ladder with 2 µl of 10 X loading dye and 16 µl water to load into the same gel.
32. Stain with ethidium bromide for 20 minutes and de-stain for 20 minutes

Appendix E:

Plasmid constructs for U1-luciferase fusion reporter constructs that contained either a wild type U1 PSEA or a U6 PSEA.


KpnI
SacI
MluI
U1 wild type promoter fused with firefly luciferase gene
SphI
NcoI
DsaI
StyI
MluI
U1 start site
junction marker
fluci sta
SphI
StyI

```
      CAAAGGGGGACCTTCGAGGGAGCACGCGAGAGGACAAGGCTGGGACGGCGAATGGCCTATGGACAGGCGGAAAGAGGGAAGCCCTTCGCACCGCGAAAG
3664  TCAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTT
      AGTTACGAGTGCGACATCCATAGAGTCAAGCCACATCCAGCAAGCGAGGTTCGACCCGACACACGTGCTTGGGGGGCAAGTCGGGCTGGCGACGCGGAA
3763  ATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGG
      TAGGCCATTGATAGCAGAACTCAGGTTGGGCCATTCTGTGCTGAATAGCGGTGACCGTCGTCGGTGACCATTGTCCTAATCGTCTCGCTCCATACATCC
3862  CGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAA
      GCCACGATGTCTCAAGAACTTCACCACCGGATTGATGCCGATGTGATCTTCCTGTCATAAACCATAGACGCGAGACGACTTCGGTCAATGGAAGCCTTT
3961  AAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCA
      TTCTCAACCATCGAGAACTAGGCCGTTTGTTTGGTGGCGACCATCGCCACCAAAAAAACAAACGTTCGTCGTCTAATGCGCGTCTTTTTTTCCTAGAGT
4060  AGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCAC
      TCTTCTAGGAAACTAGAAAAGATGCCCCAGACTGCGAGTCACCTTGCTTTTGAGTGCAATTCCCTAAAACCAGTACTCTAATAGTTTTTCCTAGAAGTG
4159  CTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACC
      GATCTAGGAAAATTTAATTTTTACTTCAAAATTTAGTTAGATTTCATATATACTCATTTGAACCAGACTGTCAATGGTTACGAATTAGTCACTCCGTGG
4258  TATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGC
      ATAGAGTCGCTAGACAGATAAAGCAAGTAGGTATCAACGGACTGAGGGGCAGCACATCTATTGATGCTATGCCCTCCCGAATGGTAGACCGGGGTCACG
4357  TGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTT
      ACGTTACTATGGCGCTCTGGGTGCGAGTGGCCGAGGTCTAAATAGTCGTTATTTGGTCGGTCGGCCTTCCCGGCTCGCGTCTTCACCAGGACGTTGAAA
4456  ATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCAT
      TAGGCGGAGGTAGGTCAGATAATTAACAACGGCCCTTCGATCTCATTCATCAAGCGGTCAATTATCAAACGCGTTGCAACAACGGTAACGATGTCCGTA
4555  CGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGT
      GCACCACAGTGCGAGCAGCAAACCATACCGAAGTAAGTCGAGGCCAAGGGTTGCTAGTTCCGCTCAATGTACTAGGGGGTACAACACGTTTTTTCGCCA
4654  TAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCC
      ATCGAGGAAGCCAGGAGGCTAGCAACAGTCTTCATTCAACCGGCGTCACAATAGTGAGTACCAATACCGTCGTGACGTATTAAGAGAATGACAGTACGG
4753  ATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACG
      TAGGCATTCTACGAAAAGACACTGACCACTCATGAGTTGGTTCAGTAAGACTCTTATCACATACGCCGCTGGCTCAACGAGAACGGGCCGCAGTTATGC
4852  GGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATC
      CCTATTATGGCGCGGTGTATCGTCTTGAAATTTTCACGAGTAGTAACCTTTTGCAAGAAGCCCCGCTTTTGAGAGTTCCTAGAATGGCGACAACTCTAG
4951  CAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGC
      GTCAAGCTACATTGGGTGAGCACGTGGGTTGACTAGAAGTCGTAGAAAATGAAAGTGGTCGCAAAGACCCACTCGTTTTTGTCCTTCCGTTTTACGGCG
5050  AAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGG
      TTTTTTCCCTTATTCCCGCTGTGCCTTTACAACTTATGAGTATGAGAAGGAAAAAGTTATAATAACTTCGTAAATAGTCCCAATAACAGAGTACTCGCC
5149  ATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGCGCCCTGTAGCGGCGCATTAAG
      TATGTATAAACTTACATAAATCTTTTTATTTGTTTATCCCCAAGGCGCGTGTAAAGGGGCTTTTCACGGTGGACTGCGCGGGACATCGCCGCGTAATTC
5248  CGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTT
      GCGCCGCCCACACCACCAATGCGCGTCGCACTGGCGATGTGAACGGTCGCGGGATCGCGGGCGAGGAAAGCGAAAGAAGGGAAGGAAAGAGCGGTGCAA
5347  CGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGA
      GCGGCCGAAAGGGGCAGTTCGAGATTTAGCCCCCGAGGGAAATCCCAAGGCTAAATCACGAAATGCCGTGGAGCTGGGGTTTTTTGAACTAATCCCACT
5446  TGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAAC
      ACCAAGTGCATCACCCGGTAGCGGGACTATCTGCCAAAAAGCGGGAAACTGCAACCTCAGGTGCAAGAAATTATCACCTGAGAACAAGGTTTGACCTTG
5545  AACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAA
      TTGTGAGTTGGGATAGAGCCAGATAAGAAAACTAAATATTCCCTAAAACGGCTAAAGCCGGATAACCAATTTTTTACTCGACTAAATTGTTTTTAAATT
5644  CGCGAATTTTAACAAAATATTAACGTTTACAATTTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATT
      GCGCTTAAAATTGTTTTATAATTGCAAATGTTAAAGGGTAAGCGGTAAGTCCGACGCGTTGACAACCCTTCCCGCTAGCCACGCCCGGAGAAGCGATAA
5743  ACGCCAGCCCAAGCTACCATGATAAGTAAGTAATATTAAGGTACGTGGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAAC
      TGCGGTCGGGTTCGATGGTACTATTCATTCATTATAATTCCATGCACCTCCAAAATGAACGAAATTTTTTGGAGGGTGTGGAGGGGGACTTGGACTTTG
5842  ATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTT
      TATTTTACTTACGTTAACAACAACAATTGAACAAATAACGTCGAATATTACCAATGTTTATTTCGTTATCGTAGTGTTTAAAGTGTTTATTTCGTAAAA
                                                                                                    6026
5941  TTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATGGTACTGTAACTGAGCTAACATAA
      AAAGTGACGTAAGATCAACACCAAACAGGTTTGAGTAGTTACATAGAATACCATGACATTGACTCGATTGTATT
```


KpnI
SacI
MluI
U1 containg U6 PSEA promoter fused with firefly luciferase gene
SphI
NcoI
DsaI
StyI
MluI
U1 start site
junction marker
fluci sta
SphI
StyI

```
3664 TCAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTT
     AGTTACGAGTGCGACATCCATAGAGTCAAGCCACATCCAGCAAGCGAGGTTCGACCCGACACACGTGCTTGGGGGGCAAGTCGGGCTGGCGACGCGGAA

3763 ATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGG
     TAGGCCATTGATAGCAGAACTCAGGTTGGGCCATTCTGTGCTGAATAGCGGTGACCGTCGTCGGTGACCATTGTCCTAATCGTCTCGCTCCATACATCC

3862 CGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAA
     GCCACGATGTCTCAAGAACTTCACCACCGGATTGATGCCGATGTGATCTTCCTGTCATAAACCATAGACGCGAGACGACTTCGGTCAATGGAAGCCTTT

3961 AAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCA
     TTCTCAACCATCGAGAACTAGGCCGTTTGTTTGGTGGCGACCATCGCCACCAAAAAAACAAACGTTCGTCGTCTAATGCGCGTCTTTTTTTCCTAGAGT

4060 AGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCAC
     TCTTCTAGGAAACTAGAAAAGATGCCCCAGACTGCGAGTCACCTTGCTTTTGAGTGCAATTCCCTAAAACCAGTACTCTAATAGTTTTTCCTAGAAGTG

4159 CTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACC
     GATCTAGGAAAATTTAATTTTTACTTCAAAATTTAGTTAGATTTCATATATACTCATTTGAACCAGACTGTCAATGGTTACGAATTAGTCACTCCGTGG

4258 TATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGC
     ATAGAGTCGCTAGACAGATAAAGCAAGTAGGTATCAACGGACTGAGGGGCAGCACATCTATTGATGCTATGCCCTCCCGAATGGTAGACCGGGGTCACG

4357 TGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTT
     ACGTTACTATGGCGCTCTGGGTGCGAGTGGCCGAGGTCTAAATAGTCGTTATTTGGTCGGTCGGCCTTCCCGGCTCGCGTCTTCACCAGGACGTTGAAA

4456 ATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCAT
     TAGGCGGAGGTAGGTCAGATAATTAACAACGGCCCTTCGATCTCATTCATCAAGCGGTCAATTATCAAACGCGTTGCAACAACGGTAACGATGTCCGTA

4555 CGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGT
     GCACCACAGTGCGAGCAGCAAACCATACCGAAGTAAGTCGAGGCCAAGGGTTGCTAGTTCCGCTCAATGTACTAGGGGGTACAACACGTTTTTTCGCCA

4654 TAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCC
     ATCGAGGAAGCCAGGAGGCTAGCAACAGTCTTCATTCAACCGGCGTCACAATAGTGAGTACCAATACCGTCGTGACGTATTAAGAGAATGACAGTACGG

4753 ATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACG
     TAGGCATTCTACGAAAAGACACTGACCACTCATGAGTTGGTTCAGTAAGACTCTTATCACATACGCCGCTGGCTCAACGAGAACGGGCCGCAGTTATGC

4852 GGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATC
     CCTATTATGGCGCGGTGTATCGTCTTGAAATTTTCACGAGTAGTAACCTTTTGCAAGAAGCCCCGCTTTTGAGAGTTCCTAGAATGGCGACAACTCTAG

4951 CAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGC
     GTCAAGCTACATTGGGTGAGCACGTGGGTTGACTAGAAGTCGTAGAAAATGAAAGTGGTCGCAAAGACCCACTCGTTTTTGTCCTTCCGTTTTACGGCG

5050 AAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGG
     TTTTTTCCCTTATTCCCGCTGTGCCTTTACAACTTATGAGTATGAGAAGGAAAAAGTTATAATAACTTCGTAAATAGTCCCAATAACAGAGTACTCGCC

5149 ATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGCGCCCTGTAGCGGCGCATTAAG
     TATGTATAAACTTACATAAATCTTTTTATTTGTTTATCCCCAAGGCGCGTGTAAAGGGGCTTTTCACGGTGGACTGCGCGGGACATCGCCGCGTAATTC

5248 CGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTT
     GCGCCGCCCACACCACCAATGCGCGTCGCACTGGCGATGTGAACGGTCGCGGGATCGCGGGCGAGGAAAGCGAAAGAAGGGAAGGAAAGAGCGGTGCAA

5347 CGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGA
     GCGGCCGAAAGGGGCAGTTCGAGATTTAGCCCCCGAGGGAAATCCCAAGGCTAAATCACGAAATGCCGTGGAGCTGGGGTTTTTTGAACTAATCCCACT

5446 TGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAAC
     ACCAAGTGCATCACCCGGTAGCGGGACTATCTGCCAAAAAGCGGGAAACTGCAACCTCAGGTGCAAGAAATTATCACCTGAGAACAAGGTTTGACCTTG

5545 AACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAA
     TTGTGAGTTGGGATAGAGCCAGATAAGAAAACTAAATATTCCCTAAAACGGCTAAAGCCGGATAACCAATTTTTTACTCGACTAAATTGTTTTTAAATT

5644 CGCGAATTTTAACAAAATATTAACGTTTACAATTTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATT
     GCGCTTAAAATTGTTTTATAATTGCAAATGTTAAAGGGTAAGCGGTAAGTCCGACGCGTTGACAACCCTTCCCGCTAGCCACGCCCGGAGAAGCGATAA

5743 ACGCCAGCCCAAGCTACCATGATAAGTAAGTAATATTAAGGTACGTGGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAAC
     TGCGGTCGGGTTCGATGGTACTATTCATTCATTATAATTCCATGCACCTCCAAAATGAACGAAATTTTTTGGAGGGTGTGGAGGGGGACTTGGACTTTG

5842 ATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTT
     TATTTTACTTACGTTAACAACAACAATTGAACAAATAACGTCGAATATTACCAATGTTTATTTCGTTATCGTAGTGTTTAAAGTGTTTATTTCGTAAAA

                                                                                                    6026
5941 TTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATGGTACTGTAACTGAGCTAACATAA
     AAAGTGACGTAAGATCAACACCAAACAGGTTTGAGTAGTTACATAGAATACCATGACATTGACTCGATTGTATT
```