CHARACTERIZATION OF THE ESSENTIAL PRE-MRNA SPLICING FACTOR PSF: INVESTIGATION OF RNA BINDING SPECIFICITY AND SPLICING-RELATED COMPLEX FORMATION

By

Rui Peng

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Biological Sciences

August, 2005

Nashville, Tennessee

Approved:

James G. Patton

Charles Singleton

Wallace M. LeStourgeon

Laurence J. Zwiebel

Ronald B. Emeson

ACKNOWLEDGEMENTS

During the past six years at Vanderbilt, there have been many people who have helped and supported me in my training. First and foremost, I would like to thank my advisor, Dr. James G. Patton, for continual support and encouragement. Thank you for always be there whenever I need help and guidance, and thank you for letting me a great measure of independence. I would never ever forget the support and understanding from you during my difficult times. I would also like to thank the members of my thesis committee, Dr. Charles Singleton, chair, Dr. Wallace M. LeStourgeon, Dr. Laurence J. Zwiebel, and Dr. Ronald B. Emeson, for their helpful insights and discussions. In addition, I would like to thank our collaborator, Dr. Andy Link, and the members of the Link lab for their assistance.

I am very grateful to be a member of the Patton lab, in which I have worked with so many wonderful persons. My appreciations to all the members of the Patton lab, past and present, thank you all for providing a warm, joyful, and friendly environment that makes lab life easier and truly enjoyable. In particular, I would like to thank Daron Barnard, Billy Dye, Jun Li, Izzy Pérez, and Brooke Thompson, for their direct contributions to the project. Additionally, I would like to say thank you once more to Daron Barnard, who had not only taught me a variety of techniques, but also set an excellent example of a good graduate student. It was very lucky for me to work with you during my rotation and the first several months in the lab. Jun Li, my best friend in and out of the lab, thank you again for many helpful discussions, as well as all the laughter and happy moments you had brought. Brian Griffith, Christian Gocke, Barry Crawford, Robin Ryther, Ian Hawkins, Alex Flynt, and Nikki Shariat, it has been a great

experience to work with you, and thank you for numerous favors, helping me with my English, and great music you introduced. Elizabeth Thatcher and Nan Li, the new blood, you often remind me of myself in my first year, and I wish you both have a successful and happy graduate life.

I would also like to thank the administrative staff of the Department of Biological Sciences, particularly Roz Johnson and Nancy Jackson, for their diligent help throughout the years.

Lastly, I would like to thank my family, my husband, my parents, and my sister. I would never be here without their love and support, which are and will forever remain my greatest asset.

This work was supported by research grants from the National Science Foundation (MCB-997452) and the National Institutes of Health to James G. Patton (GM62487).

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

α-TM	α-Tropomyosin
BP	Branch point
CBP	Cap binding protein
CRS	cAMP-responsive sequence
CTD	C-terminal domain
DBD	DNA binding domain
DBHS	*Drosophila* behavior, human splicing domain
DTT	Dithiothreitol
EDTA	Ethylenediamine tetraacetic acid
ESE	Exonic splicing enhancer
G	Glycine
GST	Glutathione-S-transferase
HBV	Hepatitis B virus
HDAC	Histone deacetylase
his	Histidine
HIV	Human immunodeficiency virus
hnRNP	Heterogeneous nuclear ribonucleoprotein
HtH	Helix-turn-helix motif
IGFRE	Insulin-like growth factor response element
IL1	Internal loop 1
IL2	Internal loop 2

INS	Instability element
IP	Immunoprecipitation
MN	Micrococcal nuclease
mRNA	Messenger RNA
mut	Mutant
N	Asparagine
NE	Nuclear extract
NMD	Nonsense mediated decay
nt	Nucleotide
NTC	Prp19 complex
P	Proline
PCC	PSF-containing complex
PSF	PTB-associated splicing factor
PSP1	Paraspeckle protein 1
PSP2	Paraspeckle protein 2
PTB	Polypyrimidine-tract binding protein
PTC	Premature stop codon
Py	Polypyrimidine-tract
Q	Glutamine
RNA Pol II	RNA polymerase II
RNF4	Ring finger protein 4
RRM	RNA recognition motif
SELEX	Systematic evolution of ligands by exponential enrichment

SF1	Splicing factor 1
snRNA	Small nuclear RNA
snRNP	Small nuclear ribonucleoprotein particle
TREX	Transcription/export complex
U2AF	U2 snRNP auxiliary factor
wt	Wild type
ytRNA	Yeast transfer RNA

CHAPTER I

INTRODUCTION

The mechanism of pre-mRNA splicing

Most eukaryotic genes are transcribed into pre-mRNA containing coding sequences (exons) disrupted by intervening sequences (introns). Prior to translation, the introns must be efficiently and accurately removed and the exons joined together in a process known as pre-mRNA splicing (Burge et al., 1999; Hastings and Krainer, 2001). It is very important to maintain the fidelity of pre-mRNA splicing, an error as small as a single nucleotide can completely shift the reading frame leading to the production of truncated proteins.

To define an intron, three cis-acting elements are required: the 5'splice site, the 3' splice site, and the branch point/polypyrimidine-tract just upstream of the 3' splice site. The 5' and 3' splice sites define the exon/intron boundaries while the branch point provides an adenosine for nucleophilic attack on the 5' splice site. Splicing occurs in two sequential transesterification reactions, the first of which involves cleavage of the 5' exon and attack of the 5' splice site by the branch point adenosine, generating a lariat intron-3' exon intermediate. The second step ligates the exons together and releases the lariat intron (Fig. 1).

Although the basic chemistry of splicing is extremely conserved and quite straightforward, splicing is indeed a very complicated task, given the small size of the cis-elements (splice sites and the branch point) and the size of the eukaryotic genome.

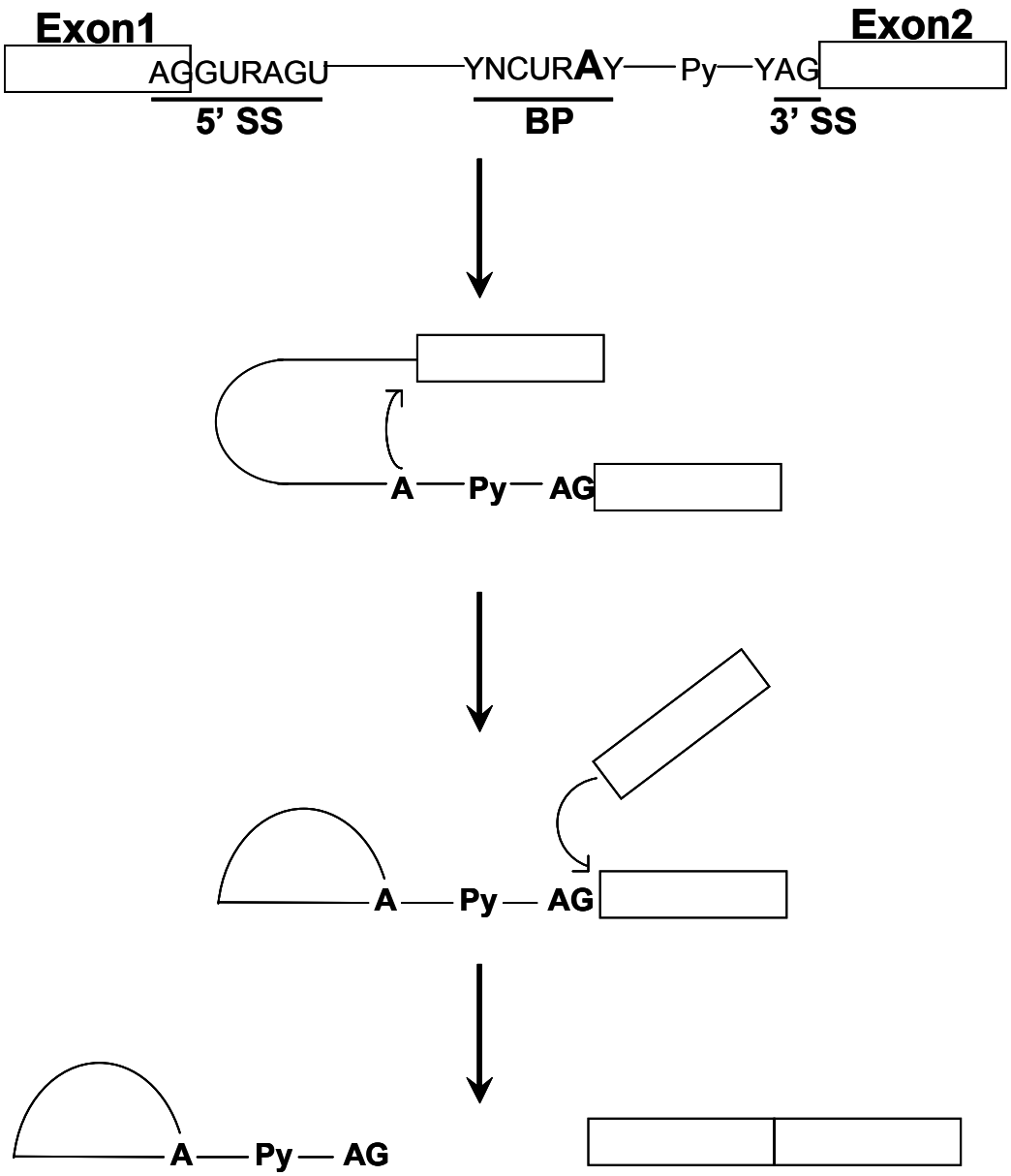


Figure 1. Two catalytic steps of pre-mRNA splicing. Diagram of two exons and the intervening intron. The consensus sequences for higher eukaryotic 5' splice site (5' SS), 3'splice site (3' SS), branch point (BP) and the polypyrimidine-tract (Py) are indicated (N = any nucleotide, Y = pyrimidine, R = purine). The two sequential transesterification steps are shown.

The process is simplified somewhat in lower eukaryotes such as *S. cerevisiae* where introns are generally small and the cis-elements are highly conserved. However, in higher eukaryotes, the typical length of an intron can be as large as 200,000 nucleotides (nt) while the average size for an exon is less than 200 nt (Berget, 1995). Furthermore, splice site sequences in higher eukaryotes are not highly conserved (see Fig. 1), making it even more difficult for precise exon/intron definition. Given these constraints, the accuracy of splicing is quite amazing.

Splicing is carried out within a large complex referred to as the spliceosome. The spliceosome is composed of five small nuclear ribonucleoprotein particles (U1, U2, U4/U6, and U5 snRNPs; Lührmann, 1988) and other non-snRNP protein factors (reviewed in Jurica and Moore, 2003). Each snRNP contains one or two specific small nuclear RNAs (snRNA) bound by several protein factors that play important roles in both spliceosome assembly and the two catalytic steps of splicing. Dynamic base pairing interactions between the snRNAs and the pre-mRNA and between the snRNAs themselves are crucial for accurate definition of the exon/intron boundaries (Nilsen, 1998). U1 snRNA base pairs with the 5' splice site while U2 pairs with the branch point. Upon association with U4/U6.U5, a series of rearrangements occur, including dissociation of U1 from the 5' splice site and dissociation of U4 from U6, allowing different regions of U6 to pair with both the 5' splice site and with U2 snRNA. During these changes, U5 appears to play an important role in tethering the two exons to juxtapose them for catalysis (Alvi et al., 2001; McConnell and Steitz, 2001; Newman, 1997; O'Keefe and Newman, 1998; also see Fig. 2 and spliceosome assembly below).

To ensure the accuracy of splicing, these RNA rearrangements take place in a highly ordered and well regulated manner (Reed, 2000). Although the mechanism of this is not completely understood, two groups of protein factors are thought to be involved. One contains members of a superfamily of ATPases, including the DEAD and DEAH box families. These ATPases are proposed to facilitate RNA rearrangements by unwinding RNA duplexes or by disrupting RNA-protein interactions (Staley and Guthrie, 1998). The second group of factors represents members of a family of RNA binding proteins containing RNA recognition motifs (RRMs; Burd and Dreyfuss, 1994; Swanson, 1995). The RRMs of these proteins are thought to mediate binding to single-stranded RNA, but RRMs also participate in protein-protein interactions, allowing this group of factors to play distinct roles in annealing critical RNA-RNA pairs, stabilizing single-stranded RNAs, and facilitating exon/intron definition (Fu, 1995; Ghetti et al., 1995; Valcárcel et al., 1996).

Spliceosome Assembly

During the past twenty years, many efforts have been dedicated to understand the assembly and composition of the spliceosome. Based on studies using cell extracts and *in vitro* splicing, most of the current models of spliceosome formation suggest a stepwise assembly pathway that requires ATP and substrate pre-mRNAs. The process begins with recognition of the 5' splice site by U1 snRNP and the polypyrimidine-tract by the U2 snRNP auxiliary factor (U2AF) heterodimer, initiating the formation of an ATP-independent, discrete early complex (E complex). Binding of U2AF to the polypyrimidine-tract recruits U2 snRNP to the branch point sequence, forming complex

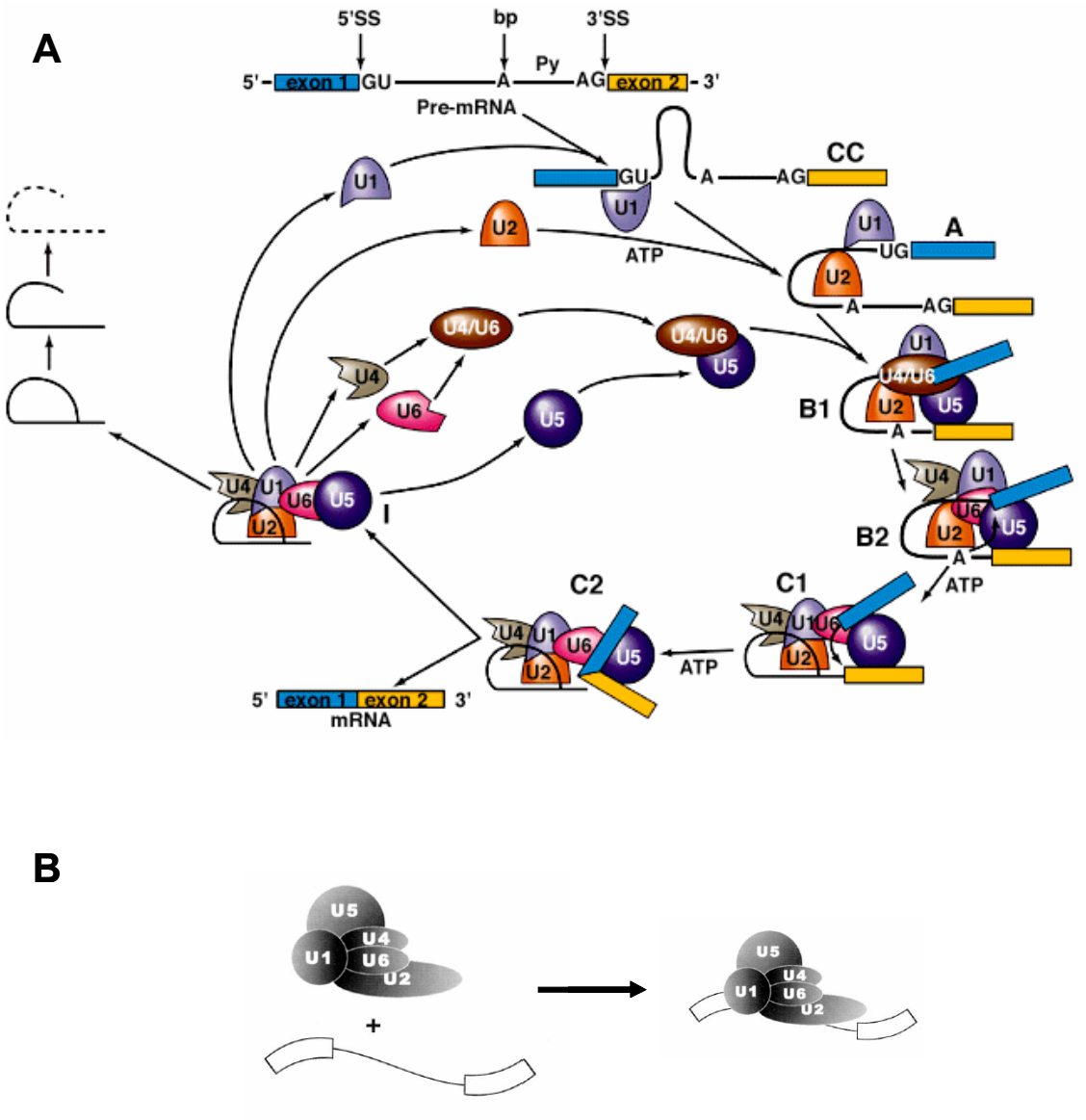


Figure 2. Models for spliceosome assembly. **(A)** The model adapted from Weaver (1999) showing the spliceosome cycle (stepwise assembly of the spliceosome complexes and recycling of snRNPs for the next round of splicing). **(B)** The model adapted from Stevens et al. (2002) showing the yeast penta-snRNP assembled onto the pre-mRNA substrate as a pre-formed particle.

A. The B complex is then formed when the U4/U6.U5 tri-snRNP joins the complex. A series of dynamic RNA/protein rearrangements then converts the B complex to the C complex, the short-lived, active spliceosome (reviewed in Reed and Palandjian, 1997; Fig. 2A). These short-lived spliceosomal complexes can be distinguished by their snRNP composition and different mobilities in native gels.

Over the years, these individual complexes have been purified and their contents analyzed by different strategies (reviewed in Jurica and Moore, 2003). The common theme linking these studies has been the demonstration that the spliceosome is a dynamic entity. However, to what extent this model reflects the *in vivo* situation is not entirely clear. In 2002, two different endogenous complexes were isolated from yeast. One of them, the Cwf/Cwc complex was purified from both *S. cerevisiae* and *S. pombe* (Ohi et al., 2002). Although it is very similar to the *in vitro*-assembled mammalian C complex (Jurica et al., 2002) in both structure and contents (U2, U5, U6 snRNPs as well as other second step splicing factors), the fact that the Cwf complex remains stable in splicing-deficient strains questions its functional relevance. Nevertheless, the identification of this complex raises the question as to whether spliceosomes might also exist as larger pre-formed complexes. Consistent with this possibility was the discovery of a 45S penta-snRNP from *S. cerevisiae* which contains all of the five splicing snRNPs and over 60 pre-mRNA splicing factors. Strikingly, the penta-snRNP is capable of splicing when supplied with micrococcal nuclease treated nuclear extracts and therefore functions as a pre-formed particle (Stevens et al., 2002). Thus, at least in the yeast S. cerevisiae there may be preformed spliceosomes rather than stepwise assembly on demand (Fig. 2B). In this view, *in vitro* assembly intermediate complexes may just

reflect multiple dynamic conformational changes during maturation of largely pre-formed spliceosomes (Nilsen, 2003). It remains unclear whether pre-formed spliceosomes are limited to yeast since no such particles have been purified from mammalian extracts.

The composition of the spliceosome

Despite the two perhaps contradictory models of spliceosome assembly, it has been agreed for years that the spliceosome is a large, complicated particle (Moore et al., 1993). Early experiments in both mammalian and yeast systems identified the five U snRNPs and a variety of other proteins as essential splicing factors. Recent powerful combinations of affinity purification coupled to large-scale protein identification by mass spectrometry has led to a fairly comprehensive proteomic analyses of different spliceosomal complexes, providing a much broader view of the spliceosome. Currently, the list of spliceosome proteins has expanded to nearly 300 candidate proteins with the challenge to determine the function of individual components.

Eighteen different spliceosomal complexes purified by different groups have been reported, and their core components can be assorted into six groups (Table 1; reviewed in Jurica and Moore, 2003). The first group is snRNP related proteins, including the Sm proteins and proteins specific to U1, U2, U4/U6, and U5 snRNPs. The second group is a group of miscellaneous splicing factors, most of which have been known for years as spliceosomal components and essential for splicing. These include Splicing Factor 1 (SF1), U2AF65, U2AF35, cap binding proteins (CBP), and second step factors.

Table 1. Proteins identified in purified spliceosomal complexes. Derived from Jurica and Moore (2003). Proteins identified in any of the following complexes are indicated (●): **M:** mixed spliceosome complexes assembled *in vitro* from human nuclear extract (Neubauer et al., 1998; Rappsilber et al., 2002; Zhou et al., 2002); **A, B*, C:** *in vitro* assembled A complex (Hartmuth et al., 2002), B* complex (Makarov et al., 2002), C complex (Jurica et al., 2002) from human nuclear extract, respectively; **Cw:** endogenous Cwf/Cwc complex from *S. cerevisiae* and *S. pombe* (Ohi et al., 2002); **P:** endogenous penta-snRNP from *S. cerevisiae* (Stevens et al., 2002); **35S:** 35S U5 snRNP from human extracts (Makarov et al., 2002); **U2:** 12S U2 snRNP from human nuclear extracts (Will et al., 2002); **Tri:** U4/U6.U5 tri-snRNP from *S. cerevisiae* (Gottschalk et al., 1999; Stevens and Abelson, 1999); **H:** H complex assembled *in vitro* from human nuclear extracts (Jurica et al., 2002); **TAP:** splicing factor containing complexes from *S. cerevisiae* (Gavin et al., 2002), supplemental data S3, complexes 128, 129, 155, 158, 160, and 161). Proteins found in 1-2 out of the 6 TAP complexes are marked with(○).

Human Name	*S.cerevisiae* Name	M	A	B*	C	Cw	P	35S	U2	Tri	H	TAP
Core snRNP proteins												
SmB/B'	Smb1	●	●	●	●	●	●	●	●	●	●	●
SmD1	Smd1	●	●	●	●	●	●	●	●	●		●
SmD2	Smd2	●	●	●	●	●	●	●	●	●		●
SmD3	Smd3	●	●	●	●	●	●	●	●	●	●	●
SmE1	Sme1	●	●	●	●		●	●	●	●		
SmF1	Smf1	●	●	●	●	●	●	●	●	●		○
SmG1	Smx2	●	●	●	●	●	●	●	●	●		
LSM2	Lsm2	●		●	●		●			●		○
LSM3	Lsm3	●		●	●		●					○
LSM4	Lsm4	●					●			●		●
LSM5	Lsm5									●		○
LSM6	Lsm6	●					●			●		○
LSM7	Lsm7	●					●			●		○
LSM8	Lsm8	●					●			●		
U1 snRNP specific proteins												
U1 70kD	Snp1	●	●				●					●
U1 A	Mud1	●	●				●				●	●
U1 C	Yhc1	●	●				●					●
	Prp39						●					●
FBP11	Prp40	●					●					●
	Snu56						●					●
	Nam8						●					○
	Snu71						●					●
	Snu65						●					●

Table 1, continued

Human Name	*S.cerevisiae* Name	M	A	B*	C	Cw	P	35S	U2	Tri	H	TAP
U2 snRNP specific proteins												
U2 A'	Lea1	●	●	●	●	●	●		●		●	●
U2 B"	Msl1	●	●	●	●	●	●		●			●
SF3a60	Prp9	●	●	●	●	●	●		●		●	●
SF3a66	Prp11	●	●	●	●	●	●		●			●
SF3a120	Prp21	●	●	●	●	●	●		●		●	●
SF3b49	Hsh49	●	●	●	●	●	●		●			○
SF3b145	Cus1	●	●	●	●	●	●		●		●	●
SF3b130	Rse1	●	●	●	●	●	●		●		●	●
SF3b155	Hsh155	●	●	●	●	●	●		●		●	●
SF3b10		●	●	●					●			
SF3b14b	Rds3p	●	●	●					●			
p14	Snu17	●	●	●			●		●			
U5 snRNP specific proteins												
PRP8	Prp8	●	●	●	●	●	●	●		●	●	●
U5-200kDa	Brr2	●	●	●	●	●	●	●		●	●	
U5-116kDa	Snu114	●	●	●	●	●	●	●		●	●	●
U5-102kDa	Prp6	●	●	●	●		●			●	●	●
U5-100kDa	Prp28	●	●		●						●	○
U5-40kDa		●		●	●	●		●			●	
U5-15kDa	Dib1	●					●			●		●
U4/U6 snRNP specific proteins												
HPRP3	Prp3	●					●			●		●
HPRP4	Prp4	●			●		●			●		●
RY-1		●										
USA-Cyp	Cpr1	●										
15.5tri-snRNP	Snu13	●					●			●		
Miscellaneous splicing factors												
U2AF65	Mud2	●	●						●		●	
U2AF35		●	●						●			
SF1	Msl5	●										
CBP20	Cbc2	●	●	●	●							○
CBP80	Sto1	●	●	●	●						●	●
PRP31	Prp31	●					●			●		●
	Snt309					●	●					●
PRP17	Prp17	●			●	●		●				○
SLU7	Slu7	●			●	●						○

Table 1, continued

Human Name	*S.cerevisiae* Name	M	A	B*	C	Cw	P	35S	U2	Tri	H	TAP
PRP18	Prp18											○
PSF		●		●							●	
	Prp38						●			●		○
PRP24	Prp24											○
DDX3	Ded1	●					●					
	Npl3						●					○
SKIP	Prp45 (Cwf13)	●		●	●	●	●	●			●	○
ECM2	Ecm2 (Cwf5)	●		●	●	●	●	●			●	●
SART1	Snu66	●	●	●	●		●			●		●
p68	Dbp2	●	●	●	●						●	
SPF30		●	●						●			
FLJ31121	Snu23			●			●			●		○
SAD1	Sad1	●					●					
LUC7	Luc7 (Luc7)	●					●					●
	Spp381						●			●		
PUF60		●			●				●		●	
PRP4 kinase		●										
Proteins containing a DEAD/H box motif (RNA unwindases)												
DDX16	Prp2	●			●							○
PRP5	Prp5	●	●						●			
PRP16	Prp16	●			●							
PRP22	Prp22	●		●	●	●						○
PRP43	Prp43	●	●	●	●				●		●	●
UAP56	Sub2	●										●
HDB/DICE1		●		●								
Abstrakt		●			●							
eIF4a3		●		●	●							
DDX35					●			●				
DDX9		●	●									
KIAA0052		●			●						●	
p72		●			●						●	
Proteins associated with the Prp19 complex												
CDC5	Cef1 (Cdc5)	●		●	●	●	●	●			●	○
PRP19	Prp19	●	●		●	●	●	●			●	●
ISY1	Isy1 (Cwf12)	●		●	●	●	●	●			●	○
SYF1	Syf1 (Cwf3)	●		●	●	●	●	●				○
CRN	Clf1 (Cwf4)	●		●	●	●	●	●				●
GCIP-IP	Syf2(C3E7.13C)	●		●	●	●	●	●				

Table 1, continued

Human Name	*S.cerevisiae* Name	M	A	B*	C	Cw	P	35S	U2	Tri	H	TAP
PRL1	Prp46 (Cwf1)	●		●	●	●	●	●			●	●
BCAS2	(Cwf7)	●		●		●		●				
	Ntc20					●	●					
	Cwc2 (Cwf2)					●	●					●
Auxiliary splicing factors												
SPF45		●	●						●			
PTB		●									●	
EWS		●			●							
SR proteins												
SRm300		●			●							
SRm160		●			●							
SC35		●	●									
SRp40		●	●									
SRp55		●	●									
SRp75		●			●							
SRp30c		●	●	●								
9G8		●	●	●	●							
SRp54		●										
SFRS10		●			●							
SRp20		●	●	●								
ASF/SF2		●	●		●				●			
Proteins with roles in RNA metabolism processes linked to splicing												
Aly/REF	Yra1	●	●	●	●						●	
RNPS1		●			●						●	
Y14		●			●							
MAGOH		●			●							
hTHO2	Rlr1	●										
hHPR1		●			●						●	
HsKin17	Rts2p	●		●								
ASR2B		●	●									
KIAA0983		●										
C21orf66		●										
PAB2		●		●	●						●	
CF I-68kD		●										
CF I-25kD		●										
CPSF 160K		●		●								
Tat-SF1	Cus2	●					●					

The third group belongs to a superfamily of ATPases, consisting of DEAD/H box proteins, some of which are known RNA unwindases. In yeast, eight such proteins (Prp2p, Prp16p, Prp22p, Prp43p, Brr2p, Prp5p, Prp28p, and Sub2p) have been shown to function in pre-mRNA splicing, and each one of them is currently linked to a specific ATP-driven structural transition step in the splicing cycle (reviewed in Staley and Guthrie, 1998). For example, Prp28/U5-100kDa has been suggested to disassociate U1 from the 5' splice site (Chen et al., 2001; Staley and Guthrie, 1999). Interestingly, several DEAD/H proteins have been identified in mammalian spliceosomal complexes that are unrelated to yeast proteins suggesting additional complexity in mammals (Jurica and Moore, 2003).

The fourth group of proteins represents members of the Prp19 complex (NTC; Chan et al., 2003; Chen et al., 2002; Ohi and Gould, 2002; Tarn et al., 1994), most of which have been shown to be required for splicing by genetic and functional analysis. In yeast, the NTC complex joins the spliceosome subsequent to U4 snRNP dissociation, and is required for stable association of U5 and U6 with the activated spliceosome (Chan et al., 2003). In mammals, hPrp19 was found to associate with CDC5 in a similar complex (Ajuh et al., 2000), and this CDC5/Prp19 complex has been implicated to function after joining of the U4/U6.U5 tri-snRNP but prior to or during the first and second catalytic steps of splicing (Ajuh et al., 2000; Makarov et al., 2002).

The fifth group consists of auxiliary splicing factors which function to facilitate recognition of non-conserved splice sites by the splicing machinery. The best characterized of these factors are the SR protein family (reviewed in Cáceres et al., 1997; Fu, 1995; Gravely, 2000). The presence of these proteins in purified spliceosomal

complexes varies from substrate to substrate and also depends on the particular purification strategy, probably due to loose association with other spliceosome components.

The last group, which might not participate in pre-mRNA splicing directly, is composed of protein factors involved in other nuclear functions, such as mRNA 3'-processing factors (cleavage factor I, CF I; cleavage and polyadenylation specificity factor, CPSF 160), components of exon junction complex (EJC), and the transcription/export (TREX) complex. The TREX complex has been proposed to link transcription elongation to mRNA export (Reed and Cheng, 2005; Strasser et al., 2002), while the EJC complex functions in mRNA export and nonsense mediated decay (NMD) (Le Hir et al., 2001; Maquat, 2005). The association of these proteins with the spliceosome provides further evidence supporting the functional coupling between splicing and other nuclear mRNA processing events (Maniatis and Reed, 2002; Reed, 2003).

Alternative splicing

As noted above, splicing is a complicated process, involving multiple RNA and protein factors. Besides the complexity caused by long introns and weakly conserved splice sites, the presence of more than one intron in a given gene creates an even more complicated situation. The splicing machinery does not always remove all introns to generate mature mRNA (constitutive splicing). Instead, splicing and splice site choice are often highly regulated. Alternative splicing generates different mRNAs encoding distinct protein products from a single gene, greatly increasing protein diversity from a

more limited genome (Patton and Smith, 2001; Smith and Valcarcel, 2000). Originally considered a rare event, the completion of human genome and recent application of bioinformatic tools have revealed that up to 74% of human genes are alternatively spliced (Johnson et al., 2003). In addition to protein diversity, alternative splicing also functions in the regulation of gene expression. When coupled with NMD, around 18-25% of the alternatively spliced exons are predicted to regulate transcript abundance (reviewed in Stamm et al., 2005).

Alternative splicing is regulated through the combinatorial interplay of variable strength cis-acting elements and the binding of trans-acting proteins (reviewed in Matlin et al., 2005; Smith and Valcarcel, 2000). Because the splice site sequences are not highly conserved, a variety of other elements have been identified that help define intron and exon boundaries. The best characterized of these cis-acting regulatory elements are referred to as splicing enhancers and silencers. They are recognized by different alternative splicing factors (usually hnRNPs and SR proteins), which in turn facilitate or inhibit the recruitment of splicing machinery to a particular splice site. In many cases, the regulation of alternative splicing is not just direct promoting or inhibiting the usage of one splice site by individual regulators, but a combinatorial effect of multiple interactions between positive and negative regulators with enhancers and silencers. Thus, the cell- or tissue-specific regulation of splicing is usually controlled by the overall ratio of different alternative splicing factors present, not simply depends on the amount of one particular factor.

Many tissue-specific protein isoforms that result from alternative splicing have been identified (Fig. 3). One good example of tissue-specific alternative splicing is the

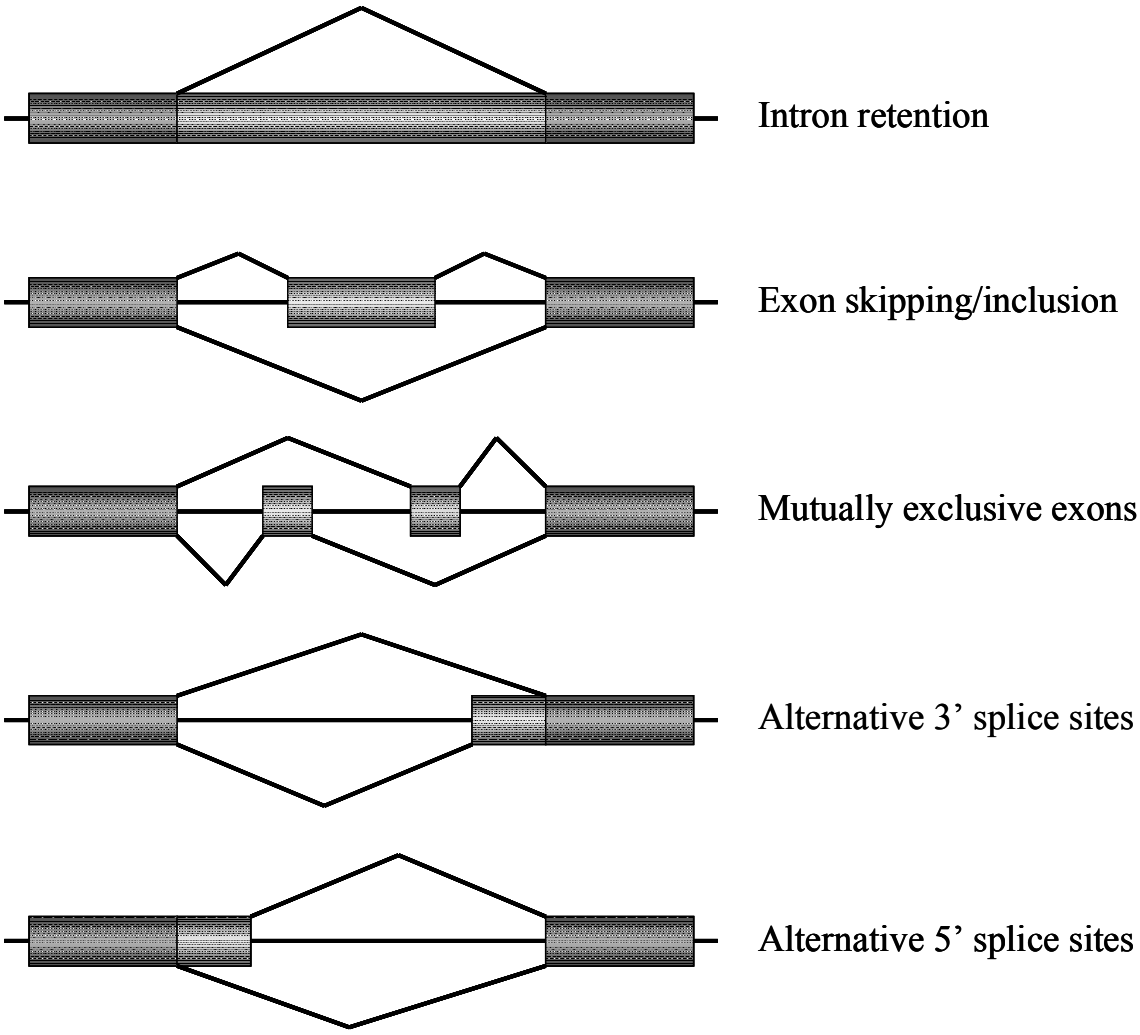


Figure 3. Patterns of alternative splicing. Constitutive and alternative exons are represented by dark and light grey boxes, respectively. Introns are represented by straight lines. Five basic patterns of alternative splicing are shown. The simplest ones are the retention of an intron and the skipping or inclusion of an exon. The third one represents a pair (or a set) of exons that are never spliced together (mutually exclusive), while the last two involves joining of one splice site to one of the several possible splice sites (alternative 5' and 3' splice sites).

splicing of α- tropomyosin (α-TM), exons 2 and 3, a pair of mutually exclusive exons (Fig. 4). In most cell types, and focusing on just the first four exons of α-TM, exons 1-3-4 are spliced together while in smooth muscle cells, exons 1-2-4 are joined (Smith and Nadal-Ginard, 1989). Inclusion of exon 2 involves multiple cis-acting elements and splicing factors, including repression of exon 3 by polypyrimidine tract-binding protein (PTB) and Raver1 (Gromak et al., 2003) and activation of exon 2 by the SR protein 9G8 through an exonic splicing enhancer (ESE) (our lab, unpublished data).

The Coupled Nuclear Network in Gene Expression

Though the gene expression pathway can be described as simple as in the Central Dogma of Molecular Biology: DNA→RNA→Protein, it is indeed a very complicated multi-step process. Eukaryotic genes are transcribed in the nucleus into pre-mRNAs which undergo three major posttranscriptional processing steps (5'-capping, splicing, and 3'-processing/polyadenylation) followed by nuclear export and translation in the cytoplasm. Over the years, distinct nuclear machineries have been discovered and assigned to individual steps. However, more and more recent studies have suggested that they are functionally coupled (reviewed in Maniatis and Reed, 2002; Fig. 5).

Coupling between transcription and RNA processing

Generally, 5'-capping, splicing, and polyadenylation are all coupled to transcription. Several transcription elongation factors and the carboxy-terminal domain (CTD) of the RNA polymerase II (RNA Pol II) large subunit have been demonstrated to function in the coupling. Distinct segments of the CTD interact with different sets of protein factors

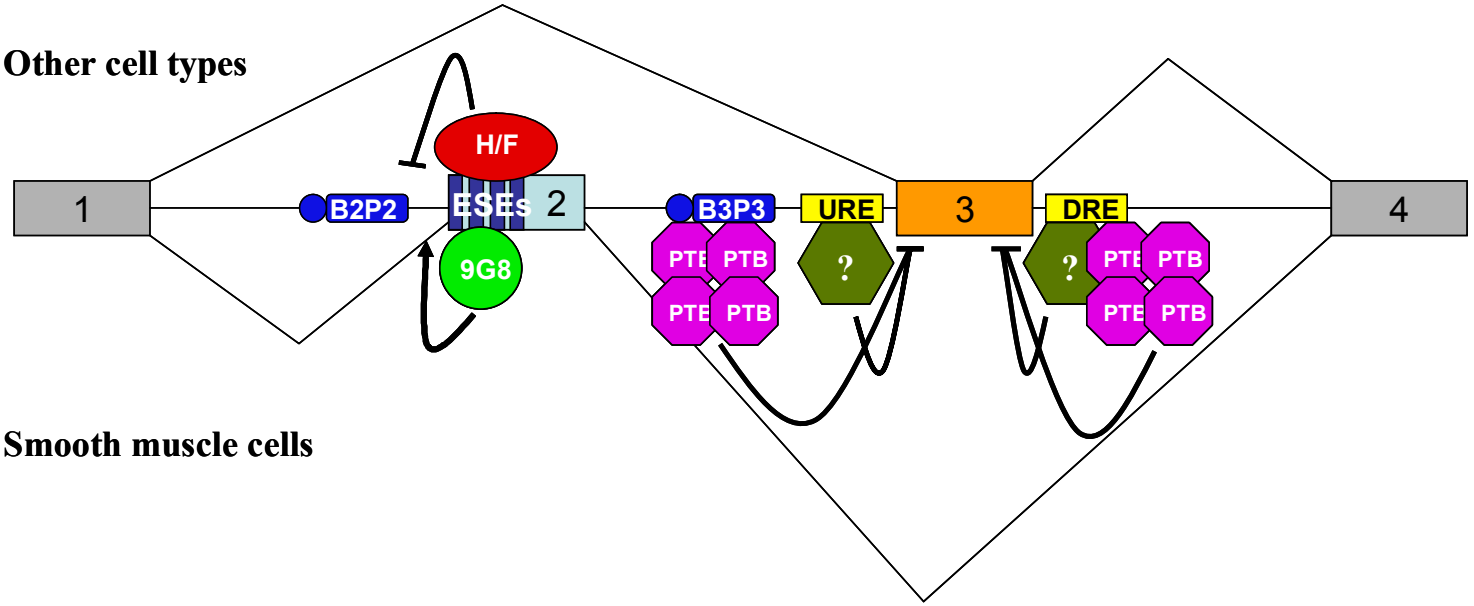


Figure 4. Regulation of α- tropomyosin splicing. Constitutive exons (1 and 4) and mutually exclusive exons (2 and 3) are represented by grey and colored boxes, respectively. Introns are represented by straight lines. Regulatory elements and alternative splicing factors are indicated. ESE: exonic splicing enhancer; B2P2: branch point/polypyrimidine-tract of exon 2; B3P3: branch point/polypyrimidine-tract of exon 3; URE: Upstream Regulatory element; DRE: Downstream Regulatory element; H/F: hnRNP H and hnRNP F; ?: unkown protein factor.

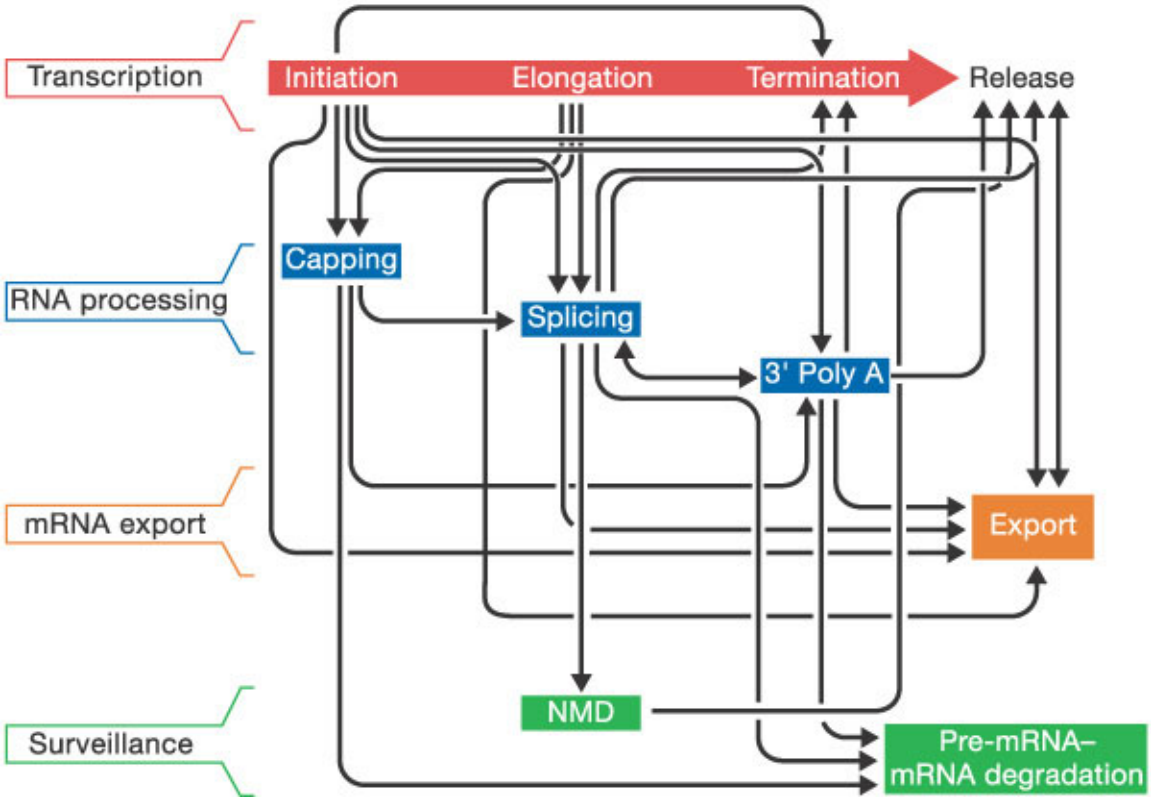


Figure 5. A coupled nuclear network of gene expression. Different stages during transcription are indicated along the top red arrow, and steps in posttranscriptional processing, mRNA export, and mRNA surveillance are shown below. Physical and/or functional coupling between individual steps is indicated by black arrows (adapted from Maniatis and Reed, 2002).

involved in capping, splicing, or 3'-processing (Fong and Bentley, 2001), which are co-transcriptionally recruited to the hyperphosphorylated CTD, and then transfer to the nascent pre-mRNA (Komarnitsky et al., 2000; Vagner et al., 2000). While the CTD serves as a platform for assembly of pre-mRNA processing machineries, two transcription elongation factors, P-TEFB and TAT-SF1, have also been shown to recruit capping and splicing factors to the nascent pre-mRNA (Fong and Zhou, 2001; Shatkin and Manley, 2000).

Since coupling of individual pre-mRNA processing events to transcription is potentially a more efficient mechanism to control overall gene expression, it is not surprising that these different processing steps are coupled as well. Besides efficiency, coupling opens more opportunities for regulation. For example, transcription from different promoters or the rate of elongation can alter the pre-mRNA splicing pattern (Cramer et al., 1999; de la Mata et al., 2003; Howe et al., 2003; Kornblihtt, 2005), whereas splicing factors can promote transcription elongation (Bres et al., 2005). Similarly, splicing of the last intron can promote downstream 3'-end cleavage and polyadenylation (Lou et al., 1996; McCracken et al., 2002; Vagner et al., 2000).

Coupling of pre-mRNA splicing to mRNA export and NMD

Two interesting observations have linked splicing to mRNA export and NMD, respectively. First, splicing is required for efficient mRNA export (Luo and Reed, 1999). Second, mRNAs containing nonsense codons located more than 50-55 nt upstream of the last exon-exon junction are targeted for degradation (Maquat and Carmichael, 2001; Nagy and Maquat, 1998). Though seemingly unrelated at first

glance, these two processes are actually linked by the EJC complex (Le Hir et al., 2001). The EJC is positioned approximately 20-24 nt upstream of exon-exon junctions upon completion of splicing and contains proteins involved in mRNA export (UAP56, TAP, REF/Aly) and NMD (Upf2, Upf3, RNPS1, Y14 and MAGOH) (reviewed in Maquat, 2004). Formation of the EJC targets spliced mRNA for export. Some EJC components, such as Y14 and Magoh, remain associated with mRNA until removed by passage of ribosomes during the pioneer round of translation (Ishigaki et al., 2001; Le Hir et al., 2001). This acts as a checkpoint for pre-mature stop codons (PTC). Retained EJC components on an mRNA after the pioneer round of translation serve to recruit other NMD factors and target that mRNA to NMD (reviewed in Maquat, 2005). Taken together, the EJC couples pre-mRNA splicing to mRNA export and NMD, serving as a crucial checkpoint for accurate splicing but, in addition, may also link splicing and translation since it may also promote mRNA polysome association (Nott et al., 2004).

PSF: A Multifunctional Nuclear Factor

PSF and its related proteins

Human PTB-associated splicing factor (PSF; Patton et al., 1993) is a 100 kDa protein that co-localizes with splicing factors in nuclear speckles (Dye and Patton, 2001), and is comprised of an N-terminal glycine-rich domain, a proline/glutamine (P/Q) rich domain, two RRMs, and a C-terminal region with two nuclear localization signals (Fig. 6). Though first identified in a complex with PTB (Patton et al., 1993), the majority of PSF does not associate with PTB (Meissner et al., 2000), but is often found as a

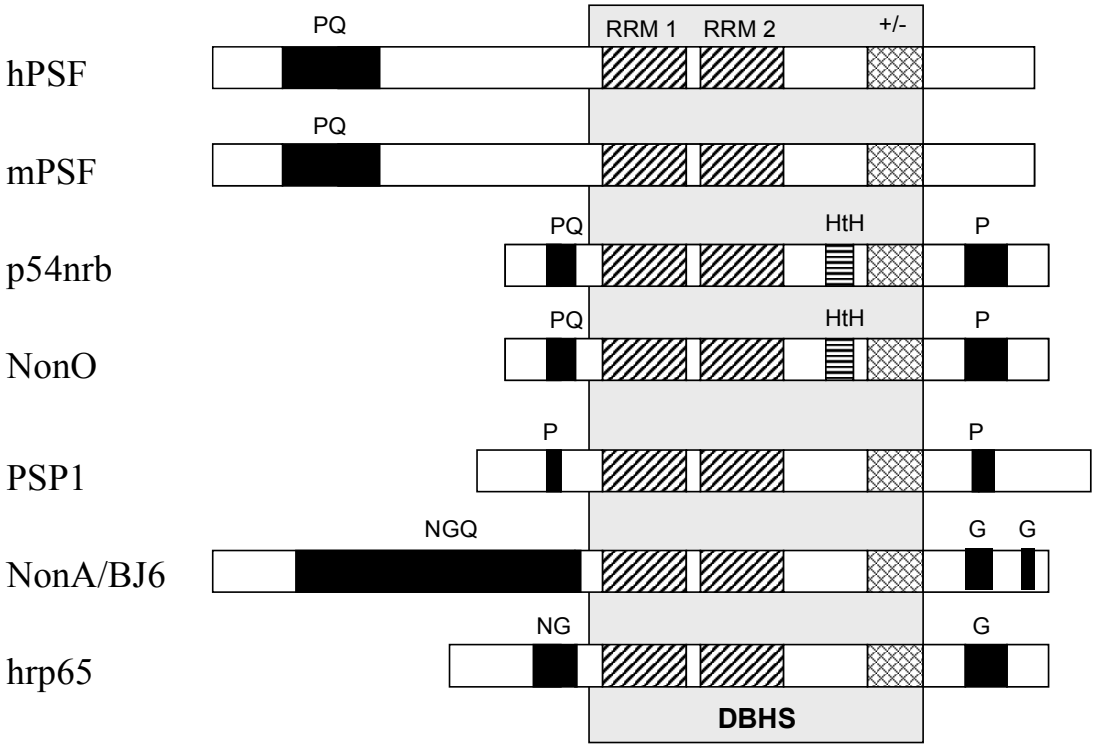


Figure 6. Structure comparison of PSF and its related proteins. Human PSF, $p54^{nrb}$, PSP1, mouse PSF, NonO, *Drosophila* NonA/BJ6, and *C. tentans* hrp65 are highly homologous over a 320 amino acid region referred to as the *Drosophila* Behavior, Human Splicing domain (DBHS; grey box; Dong et al., 1993). This region includes the two RNA Recognition Motifs (RRM1 and RRM2, diagonally hatched boxes), a helix-turn-helix motif (HtH, horizontally hatched boxes), and a basic/acidic region (+/-, cross-hatched boxes). Additional similarity exists in regions rich in glycine (G), asparagine (N), proline (P), and glutamine (Q) residues (black boxes).

heterodimer with p54^{nrb} (Peng et al., 2002; Straub et al., 1998; Zhang et al., 1993), a closely related protein originally identified based on cross-reactivity to antibodies raised against the yeast U5 snRNP-associated second-step splicing factor Prp18 (Dong et al., 1993). p54^{nrb} is 71% identical to PSF over a region of 320 amino acids that encompasses their RRMs (referred to as the *Drosophila* Behavior, Human Splicing domain, DBHS; Dong et al., 1993; see Fig. 6), and has been implicated in multiple functions involving both RNA and DNA binding (Basu et al., 1997; Peng et al., 2002; Sewer et al., 2002; Straub et al., 1998; Straub et al., 2000; Zhang et al., 1993; Zhang and Carmichael, 2001).

Recently, another DBHS domain containing protein, Paraspeckle Protein 1 (PSP1), was identified in humans (Andersen et al., 2002; Fox et al., 2002). PSP1 localizes in the paraspeckle, a novel nuclear domain that is usually located adjacent to splicing speckles. Interestingly, p54^{nrb}, but not PSF, was also found in paraspeckles (Fox et al., 2002). Upon actinomycin D induced transcriptional inhibition, both PSF and the paraspeckle proteins (p54^{nrb}, PSP1, PSP2) relocalize to discrete caps at the nucleolar periphery (Andersen et al., 2002; Dye and Patton, 2001; Fox et al., 2002; Shav-Tal et al., 2005). More recent proteomic studies of nucleolar dynamics revealed that after actinomycin D treatment, the amounts of nucleolar PSF, p54^{nrb}, and PSP1 increased by 5, 3, and 2 fold, respectively, with even more accumulation of this family of proteins in the nucleolus after treatment with 5,6-dichlorobenzimidazole riboside, which selectively inhibits RNA Pol II but not Pol I (Andersen et al., 2005).

Other homologs of PSF include mPSF (mouse PSF; Shav-Tal et al., 2001), NonO (mouse p54^{nrb}; Yang et al., 1993), NonA/BJ6 (*Drosophila*; important in *Drosophila*

visual acuity and male courtship song; Besser et al., 1990; Jones and Rubin, 1990), and hrp65 (*C. tentans*), a component of nuclear fibers associated with specific pre-mRNPs (Miralles et al., 2000; Wurtz et al., 1996).

PSF functions in pre-mRNA splicing.

PSF was cloned as a protein associated with the alternative splicing factor PTB. Unlike PTB, PSF has been proposed to be an essential splicing factor. Immunodepletion of PSF from HeLa nuclear extract first suggested that PSF might play a role in early spliceosome formation (Patton et al., 1993) but subsequent depletion/repletion experiments suggested that it might be rate-limiting for the second step(Gozani et al., 1994). Studies of the B complex identified both PSF and a degradation form of PSF (referred to as SAP102 and SAP68, respectively) as components of spliceosome (Bennett et al., 1992) and specific association of PSF with the C complex was discovered two years later (Gozani et al., 1994). Recently, several more advanced analyses of spliceosomal complexes using different purification strategies detected PSF in both early (H, B*) and late (C) complexes (Bennett et al., 1992; Gozani et al., 1994; Jurica et al., 2002; Makarov et al., 2002; Rappsilber et al., 2002).

Besides its presence in spliceosomal complexes, PSF binds U5 snRNA specifically (Peng et al., 2002; see Chapter II), consistent with its co-purification with U4/U6.U5 tri-snRNP (Teigelkamp et al., 1997) and its interaction with the U5 snRNP specific protein hPrp8 (U5-220 kDa) (G. Moreau and M. Moore, personal communication). Furthermore, PSF was identified as part of a snRNP-free U1A complex, which plays a potential role in both splicing and polyadenylation (Lutz et al., 1998). Konarska and

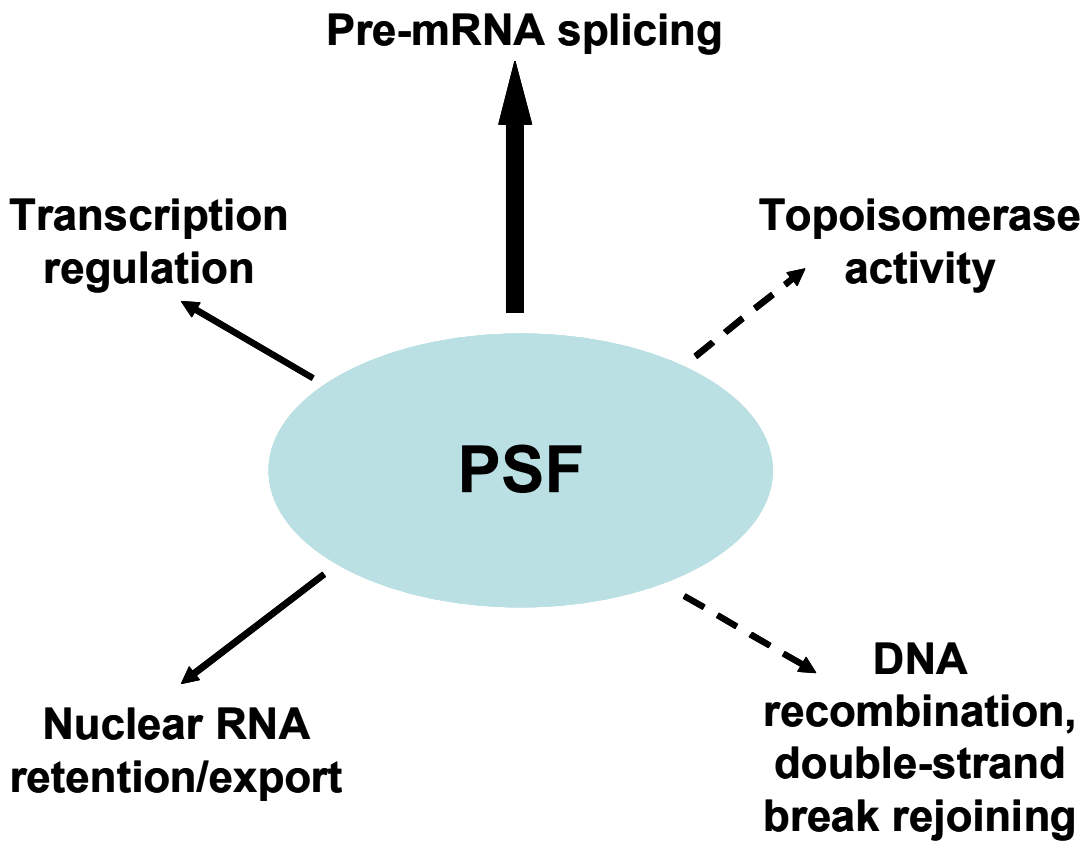


Figure 7. Multiple functions have been proposed for PSF (see text for details).

colleagues recently showed that both PSF and p54^{nrb} directly contact the 5' splice site within large transcription/splicing complexes (Kameoka et al., 2004). Overall, the data described above are consistent with a role for PSF in both early and late steps of splicing but exactly how it functions unknown.

Potential multi nuclear functions for PSF

Besides splicing, other nuclear functions have been proposed for PSF (Fig. 7). In several cases, PSF has been implicated as a transcription regulator. By itself, PSF inhibits the transcription of the porcine P-450 cholesterol side-chain cleavage enzyme gene (P450scc) through its interaction with the insulin-like growth factor response element (IGFRE) (Urban and Bodenburg, 2002; Urban et al., 2000). Binding of the mouse VL30 retrotransposon RNA to PSF disassociates PSF from IGFRE, turning on expression of the P450scc gene (Song et al., 2004). PSF and p54^{nrb} have also been shown to bind the DNA binding domain (DBD) of nuclear hormone receptors and form a complex with the repressor Sin3A, inducing transcription silencing through the recruitment of class I histone deacetylases (HDAC) to the receptor DBD (Mathur et al., 2001). However, in the case of human CYP17 gene, the negative effects of PSF and p54^{nrb} on transcription can be reversed by cAMP stimulation. Increased levels of cAMP activate phosphatase(s), which dephosphorylate PSF/p54^{nrb}/SF-1, releasing Sin3A/HDAC from the complex. In the absence of Sin3A/HDAC, more PSF/p54^{nrb}/SF-1 binds to cAMP-responsive sequences (CRS), and promotes transcription of the hCYP17 gene (Sewer and Waterman, 2002; Sewer et al., 2002). The transcriptional repressing activity of PSF is associated with its N-terminal region (Mathur et al., 2001; Urban and

Bodenburg, 2002). Recently, both PSF and $p54^{nrb}$ were shown to interact with the CTD of RNA Pol II (Emili et al., 2002). It remains unclear whether this indicates involvement in coupling transcription and splicing or simply reflects participation in transcription regulation.

In addition to transcriptional regulation, PSF has been proposed to regulate gene expression posttranscriptionally (Zhang and Carmichael, 2001; Zolutukhin et al., 2002). PSF has been shown to interact with cis-acting instability elements (INS) in Human Immunodeficiency Virus type 1 (HIV) gag/pol and env mRNAs (Zolutukhin et al., 2002). Binding of PSF to the INS counteracts the Rev-dependent export of these viral mRNAs leading to decreased virus production. Besides HIV, PSF may play a role in retaining aberrant transcripts in the nucleus. PSF, $p54^{nrb}$ and matrin 3 were identified in a complex that binds specifically and cooperatively to hyperedited inosine-containing RNAs. Interestingly, these inosine-containing RNAs can be exported by Rev but are retained in the nucleus by the PSF/$p54^{nrb}$/matrin 3 complex (Zhang and Carmichael, 2001). In both cases, PSF seems to counteract the nuclear mRNA export by Rev but the exact mechanism has not been determined. Association of PSF with Hepatitis B Virus (HBV) RNA was also detected but the functional significance of such binding is not clear (T. Heise, personal communication).

Roles for PSF in facilitating topoisomerase I (Top I) activity, DNA recombination, and DNA double-strand break rejoining have also been proposed (Akhmedov and Lopez, 2000; Bladen et al., 2005; Straub et al., 1998; Straub et al., 2000). Unlike the regulatory activities described above, most of these functions were postulated based on

protein-protein interactions and *in vitro* functional assays. Whether such activities exist *in vivo* remains to be determined.

Summary

Pre-mRNA splicing is crucial for control of eukaryotic gene expression, and involves a large number of RNA and protein factors that ensure accuracy and efficiency. Although much progress has been achieved in our understanding of the assembly and the composition of the spliceosome as well as splicing regulatory elements and factors, many questions remain, especially in higher eukaryotes. For example, the exact function of individual components of the spliceosome and any structural changes and/or factor requirements during the two catalytic steps. Furthermore, the recent discovery of the yeast penta-snRNP raises the question as to whether pre-formed spliceosomes also exist in higher eukaryotes and if so, how is alternative splicing regulated. In an even broader view, splicing is coupled with other nuclear events, providing many feedback loops and checkpoints, making a potentially more efficient and highly regulated environment for gene expression. However, with the coupling network becoming more and more complex over time, how is overall coordination regulated?

Most of the essential splicing factors that have been identified are involved in spliceosome assembly whereas a smaller number have been proposed to function in the second catalytic step of splicing(Chua and Reed, 1999; Gozani et al., 1994; Horowitz and Krainer, 1997; Lindsey et al., 1995; Lindsey and Garcia-Blanco, 1998; Zhou and Reed, 1998). As one of the essential factors, PSF plays an important role in both early spliceosome assembly and in the second step, with evidence placing it in association with

the 5' splice site as well as with the U4/U6.U5 tri-snRNP. Nevertheless, exactly how it functions remains unknown. Therefore, a better understanding of the role of PSF in splicing would lead to further insight into the mechanism of splicing. To accomplish this goal, we started out with the identification of protein factors that interact with PSF and determined its RNA binding specificity. This led to the discovery of a PSF-containing multi-snRNP complex resembling the yeast penta-snRNP.

CHAPTER II

PSF AND P54^{NRB} BIND A CONSERVED STEM IN U5 SNRNA.

Introduction

Most eukaryotic genes are transcribed into pre-mRNAs containing coding sequences (exons) disrupted by intervening sequences (introns). Prior to translation, the introns must be efficiently and accurately removed and the exons joined together in a process known as pre-mRNA splicing (Burge et al., 1999; Hastings and Krainer, 2001). Splicing is carried out within the spliceosome, a large complex composed of small nuclear ribonucleoproteins (U1, U2, U4/U6, and U5 snRNPs; Lührmann, 1988) and other non-snRNP protein factors. The joining of exons requires two transesterification reactions, the first of which involves cleavage of the 5' exon and generation of a lariat intron-3' exon intermediate. The second step ligates the exons together and releases the intron. SnRNPs play important roles in both spliceosome assembly and the two catalytic steps of splicing. Dynamic base pairing interactions between the snRNAs and the pre-mRNA and between the snRNAs themselves are crucial for accurate definition of the exon/intron boundaries (Nilsen, 1998). U1 base pairs with the 5' splice site while U2 pairs with a region upstream of the 3' splice site referred to as the branch point. Upon association with U4/U6.U5, a series of rearrangements occur, including dissociation of U1 from the 5' splice site and dissociation of U4 from U6, allowing different regions of U6 to pair with both the 5' splice site and with U2 snRNA. During these changes, U5 appears to play an important role in tethering the two exons to

juxtapose them for catalysis (Alvi et al., 2001; McConnell and Steitz, 2001; Newman, 1997; O'Keefe and Newman, 1998).

To ensure the accuracy of splicing, these RNA rearrangements take place in a highly ordered and well regulated manner (Reed, 2000). Although the mechanism of this is not completely understood, two groups of protein factors are thought to be involved. One contains members of a superfamily of ATPases, including the DEAD and DEAH box families. These ATPases are proposed to facilitate RNA rearrangements by unwinding RNA duplexes or by disrupting RNA-protein interactions (Staley and Guthrie, 1998). The second group of factors are members of a family of RNA binding proteins containing RNA recognition motifs (RRMs; Burd and Dreyfuss, 1994; Swanson, 1995). The RRMs of these proteins are thought to mediate binding to single-stranded RNA, but RRMs also participate in protein-protein interactions, allowing this group of factors to play distinct roles in annealing critical RNA pairs, stabilizing single-stranded RNAs, and facilitating exon/intron definition (Fu, 1995; Ghetti et al., 1995; Valcárcel et al., 1996).

Most of the splicing factors that have been identified are involved in spliceosome assembly whereas a smaller number have been proposed to function in the second step of splicing (Chua and Reed, 1999; Gozani et al., 1994; Horowitz and Krainer, 1997; Lindsey and Garcia-Blanco, 1998; Zhou and Reed, 1998). Human PTB-associated splicing factor (PSF; Patton et al., 1993) is one of these proteins. PSF is a100 kDa protein that co-localizes with splicing factors in nuclear speckles (Dye and Patton, 2001), and is comprised of an N-terminal glycine-rich domain, a proline/glutamine (P/Q) rich domain, two RRMs, and a C-terminal region with two nuclear localization signals. It

was first identified due to its association with Polypyrimidine-Tract Binding protein (PTB; Patton et al., 1991), but it appears that only a fraction of PTB co-localizes with PSF (Meissner et al., 2000), and the functional relevance of this interaction has not been demonstrated. Immunodepletion of PSF from nuclear extract first suggested that PSF might play a role in early spliceosome formation (Patton et al., 1993), while subsequent depletion/repletion experiments suggested that PSF might be rate-limiting for the second step (Gozani et al., 1994). PSF co-purifies with U4/U6.U5 tri-snRNP (Teigelkamp et al., 1997), and, in addition, several studies using different purification strategies have detected PSF in both early (H, A, B) and late (C) spliceosomal complexes (Bennett et al., 1992; Gozani et al., 1994; Jurica et al., 2002; Makarov et al., 2002; Rappsilber et al., 2002). Overall, the data are consistent with a role for PSF in both early and late steps of splicing but exactly how it functions is unknown. That PSF could play multiple roles in splicing is perhaps not surprising since it has also been implicated in transcription, topoisomerase activity, nuclear RNA retention, DNA recombination, and DNA double-strand break rejoining (Akhmedov and Lopez, 2000; Bladen et al., 2005; Mathur et al., 2001; Sewer and Waterman, 2002; Sewer et al., 2002; Song et al., 2004; Straub et al., 1998; Straub et al., 2000; Urban and Bodenburg, 2002; Urban et al., 2000; Zhang and Carmichael, 2001; Zolutukhin et al., 2002). Given the fact that many nuclear events are coupled (Maniatis and Reed, 2002), understanding the multifunctional roles of PSF could provide insight into overall nuclear function.

Here, we performed yeast two-hybrid screens and identified $p54^{nrb}$ as a PSF-interacting factor. This interaction was confirmed both by co-immunoprecipitation from HeLa nuclear extract, and by the association of recombinant proteins *in vitro*. We

also used iterative selection techniques (SELEX; Systematic Evolution of Ligands by Exponential Enrichment; Szostak, 1992; Tuerk and Gold, 1990) to identify the optimal RNA binding sites for PSF and $p54^{nrb}$. These experiments resulted in the selection of RNAs matching a sequence found in the 3' side of a conserved stem in U5 snRNA. Both filter binding assays and RNA affinity experiments using biotinylated U5 snRNAs demonstrated that the two proteins bind U5 snRNA with both the sequence and structure of stem 1b contributing to binding specificity. Sedimentation analyses confirmed their association with spliceosomes and with U4/U6.U5 tri-snPNP.

Materials and Methods

Recombinant protein expression

Full-length human $p54^{nrb}$ was expressed from the pGex-2T vector as an N-terminal fusion with GST and from the pET 28A vector (Novagen) as a his-tagged fusion. The GST-$p54^{nrb}$ deletions described in Fig. 8 (GST-$p54^{nrb}\Delta$17-220, GST-$p54^{nrb}\Delta$71-220, GST-$p54^{nrb}\Delta$17-369, GST-$p54^{nrb}\Delta$71-464, and GST-p54Δ226-464) were generated by reverse PCR using pGex-2T-$p54^{nrb}$ as the template (Coolidge and Patton, 1995; Imai et al., 1991). Detailed cloning strategies and primer sequences are available upon request. All clones were verified by sequencing.

GST-$p54^{nrb}$ and deletion mutants were expressed in the *E. coli* strain HB101, and pET 28A-$p54^{nrb}$ was expressed in BL21 (DE3)-pLysS. Cells were grown in LB media at 37°C until an $O.D._{600}$ of 0.5 was attained. Expression was induced for 3 h by the addition of IPTG to a final concentration of 1 mM. GST-$p54^{nrb}$ fusion proteins were

purified by passage over glutathione agarose (Sigma), eluted with a glutathione gradient, and dialyzed against buffer D (20 mM Tris-HCl, pH 7.9, 100 mM KCl, 0.2 mM EDTA, 0.5 mM DTT, 5% glycerol). His-tagged $p54^{nrb}$ was purified by passage over Ni-NTA agarose (Qiagen) and further purified by chromatography on a Mono-S column (Pharmacia). His-tagged PSF was prepared as previously described (Patton et al., 1993) with additional purification over a Mono-S column. Average purity of the proteins was 95%. Purified proteins were dialyzed into either buffer A (Pérez et al., 1997) for use in SELEX assays or into buffer D for other experiments.

Yeast two-hybrid screens

Full-length PSF cDNA was cloned into the pBTM116 vector as a fusion with the LexA DNA binding domain and used to screen a 7-day mouse embryo cDNA library cloned into pVP16 and expressed as fusions with the VP16 transcriptional activation domain. Bait and prey vectors were co-transformed into the yeast strain L40 (his3Δ200, trp-901, leu2-3, 112, ade2, LYS2::$(lexAop)_4$-HIS3, URA3::$(lexAop)_8$-LacZ, partial genotype), allowing selection of positive interactions based on growth on his^- plates and β-galactosidase activity. Initial selection for positive clones was performed on his^- plates containing a final concentration of 15 mM 3-amino-triazole. β-galactosidase assays were performed as previously described (Kaiser et al., 1994).

Immunoprecipitations

HeLa nuclear extract (approximately 50 μg; Abmayr et al., 1988) was diluted to a volume of 400 μl with buffer D containing 0.5% NP-40 and incubated with 1.2 μg

anti-PSF antibody (Patton et al., 1993) and 15 µl (bead volume) protein G-Sepharose (Pharmacia) for 2 h at 4°C. Following 3 washes with buffer D containing 0.5% NP-40, beads were resuspended in 20 µl 2X Laemmli loading buffer (Laemmli, 1970). Following separation on 9% SDS-PAGE gels, immunoprecipitated proteins were transferred to PVDF and $p54^{nrb}$ was detected by Western blot analysis using a monoclonal antibody raised against GST-$p54^{nrb}$.

PSF-$p54^{nrb}$ interaction assays

In vitro translated, ^{35}S-labeled PSF was prepared using the TNT® T7 Coupled Reticulocyte Lysate System (Promega). The radiolabeled protein was diluted in PP-300 buffer (20 mM Tris-HCl, pH 8.0, 300 mM NaCl, 0.5 mM EDTA, 0.2% NP-40 and 0.5 mM DTT) and mixed with equal amounts of glutathione-agarose beads to which GST-$p54^{nrb}$ or one of its deletion mutant derivatives was pre-bound. Following incubation in a total volume of 225 µl at room temperature for 1 h, the beads and bound proteins were precipitated by brief centrifugation, and the supernatant was discarded. The beads were then washed 3 times with 500 µl of PP-300 and resuspended in 30 µl of 2X Laemmli loading buffer. Bound proteins were resolved by SDS-PAGE on 9% gels, and radiolabeled PSF visualized by phosphorimager analysis.

Systematic Evolution of Ligands by Exponential Enrichment (SELEX)

SELEX assays were performed using 7.5 µg of each recombinant protein in the binding steps, as described (Pérez et al., 1997; Tuerk and Gold, 1990). Briefly, a pool of DNA (5'-GCGTCTCGAGAAGCTTCC(N_{20})AGTCGGGAATTCGGATCCCtatagtgagtcgtatta

–3') was synthesized containing a randomized 20-nucleotide sequence (N_{20}) flanked by anchor sequences that served as primer annealing sites for PCR, and for T7 RNA polymerase. The amplified DNA pool was then transcribed and resulting RNAs were incubated with his-tagged PSF and/or his-tagged $p54^{nrb}$ in binding buffer (10 mM Tris-HCl, pH 7.5, 100 mM KCl, 2.5 mM $MgCl_2$, 0.1% Triton X-100, 0.1 mg/ml yeast tRNA). The proteins and bound RNAs were recovered with Ni-NTA agarose beads (Qiagen) and RNAs were recovered by phenol extraction and ethanol precipitation. Selected RNAs were then reverse transcribed and amplified by PCR to produce an enriched pool of DNA. After 8 rounds of selection, reverse transcription, and amplification, the final pool was cloned and sequenced. Prior to selection, each pool of RNA was incubated with Ni-NTA agarose to eliminate nonspecific binding to the beads. Consensus sequences were generated using the GCG DNA analysis software.

***In vitro* transcription**

The 5', 3', and 5'-3' human U5 snRNA mutants described in Fig. 11D were generated by site directed mutagenesis from pHU5a2 (wild type; Patton, 1991) as described (Coolidge and Patton, 1995; Imai et al., 1991). All mutants were confirmed by sequencing, and their secondary structures analyzed by both Mfold3 (Mathews et al., 1999; Zuker et al., 1999; *http*://bioinfo.math.rpi.edu/%7Emfold/rna/form1.cgi) and the Vienna package (Hofacker et al., 1994; *http*://www.tbi.univie.ac.at/cgi-bin/RNAfold.cgi). Wild type, ^{32}P-CTP-labeled, U5 snRNA, unlabeled competitor RNAs and biotinylated RNAs were transcribed from templates (Bfa I-linearized pHU5a2) using the SP6 MEGAscript™ *in vitro* transcription kit (Ambion). Template DNAs were removed by

digestion with RNase-free DNase I. Biotin was uniformly incorporated by the addition of biotin-14-CTP (Invitrogen), as described (Dye et al., 1998). An adenovirus-derived splicing substrate RNA (AdML) was transcribed as described (Michaud and Reed, 1993).

Gel shift assays

1.3 pmol radiolabeled wild-type U5 snRNA was mixed with 5.8 pmol his-tagged PSF and/or 5.8 pmol his-tagged $p54^{nrb}$ in reactions containing 1 mM ATP, 2 mM $MgCl_2$, 0.5 mM DTT, 0.4 mg/ml yeast tRNA, and enough buffer D to reach 60 mM KCl in a total volume of 10 µl. Reactions were incubated at 30°C for 15 min, and then stored on ice for additional 5 min. Samples were separated on 5% non-denaturing polyacrylamide gels (50 mM Tris, 50 mM Glycine; Konarska and Sharp, 1987). Gels were dried and protein-RNA complexes visualized by phosphorimager analysis.

Nitrocellulose filter binding assays

Filter binding assays were performed as described (Lynch and Maniatis, 1995). 5 pmol radiolabeled wild-type U5 snRNA was incubated with 2 pmol his-tagged PSF and/or 2 pmol his-tagged $p54^{nrb}$ in reactions containing 1 mM ATP, 2 mM $MgCl_2$, 0.5 mM DTT, 0.4 mg/ml yeast tRNA, enough buffer D to reach 60 mM KCl, and the indicated amount of unlabeled competitor RNA in a total volume of 10 µl. The amount of labeled RNA bound to filters was monitored by control reactions in which Buffer D was substituted for proteins. Reactions were incubated at 30°C for 15 min, and then stored on ice for additional 5 min. Samples were diluted with 90 µl wash buffer (12 mM Tris-HCl, pH 7.9, 60 mM KCl, 0.12 mM EDTA) and immediately filtered through a

0.45 Micron pore nitrocellulose sandwiched in a Hybri-Slot^{TM} manifold (Gibco-BRL). Filters were washed once with 200 µl wash buffer, air-dried, and the amount of retained, labeled U5 was quantitated.

RNA affinity assays

15 µl HeLa nuclear extract (approximately 75 µg) was incubated with 1 µg biotinylated RNA in each 50 µl reaction containing 0.5 mM ATP, 20 mM phosphocreatine, 2 mM $MgCl_2$, 0.5 mM DTT, 15 µl buffer D and 0.3 mg/ml yeast tRNA. Reactions were incubated for 15 min at 30°C, chilled on ice for 5 minutes, and then mixed with 20 µl streptavidin-agarose beads (Sigma), pre-equilibrated in 130 µl wash buffer (see above) containing 1mM PMSF, 50 µg/ml yeast tRNA, 0.01% NP-40. All samples were placed in 50 ml tubes packed with ice, and rocked on a tilt board at 4°C for 30 min. Beads were then washed twice with 200 µl ice-cold wash buffer containing 0.005% NP-40 and resuspended in 2X Laemmli loading buffer. Following SDS-PAGE on 10% gels (or 5% gels for U5-200 kDa and U5-116 kDa), proteins were transferred to nitrocellulose and Western blots were performed using either anti-PSF (Patton et al., 1993), anti-p54^{nrb}, anti-U5-116 kDa (Fabrizio et al., 1997), anti-U5-200 kDa (Lauber et al., 1996), anti-snRNP (from human patients), anti-SRrp86 (Barnard and Patton, 2000), or polyclonal anti-U2AF^{65} antibodies.

Sucrose gradient sedimentation of snRNPs and splicing complexes

In vitro splicing reactions using AdML pre-mRNA were carried out as previously described (Barnard and Patton, 2000). For sedimentation analyses, 200 µl splicing

reactions were incubated at 30°C for 15 min, and then layered onto 10%-30% sucrose gradients containing 2 mM $MgCl_2$, 0.5 mM DTT, 60% buffer D, 1 mM PMSF, 5 mM NaF and 0.01% NP-40. After centrifugation in a Beckman SW-60 rotor at 25,000 rpm for 14 h at 4°C, fractions from the gradients were collected. RNAs were recovered from each fraction by phenol/$CHCl_3$ extraction and separated on 8M urea-15% PAGE. Splicing products and snRNAs were visualized by phosphorimager analysis and silver staining, respectively. Proteins from each fraction were analyzed by Western blot analysis using different antibodies as described above.

Results

PSF interacts with $p54^{nrb}$.

To further define the functional role of PSF, yeast two-hybrid protein-protein interaction screens (Fields and Song, 1989) were performed to identify factors that interact with PSF. A cDNA clone encoding PSF was fused to the LexA DNA binding domain and used to screen a mouse 7-day embryo cDNA two-hybrid library fused to the VP16 transactivation domain. Sixty-three positive clones were sequenced and 29 were found to encode the mouse protein NonO (Yang et al., 1993). NonO is the mouse homolog of human $p54^{nrb}$ (also referred to as nmt 55; Dong et al., 1993; Traish et al., 1997), which is very similar in sequence to PSF (Fig. 8A). For the rest of the clones, only the mouse RING Finger protein 4 (RNF4, accession # AF169300) and an as yet unidentified protein were identified five and six times, respectively, whereas none of the remaining clones were detected more than twice. Interestingly, separate two-hybrid

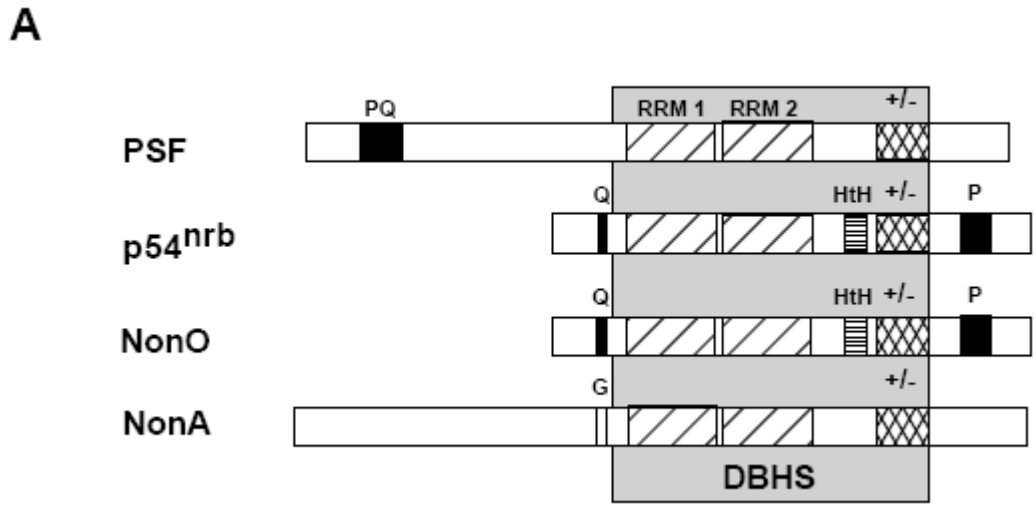
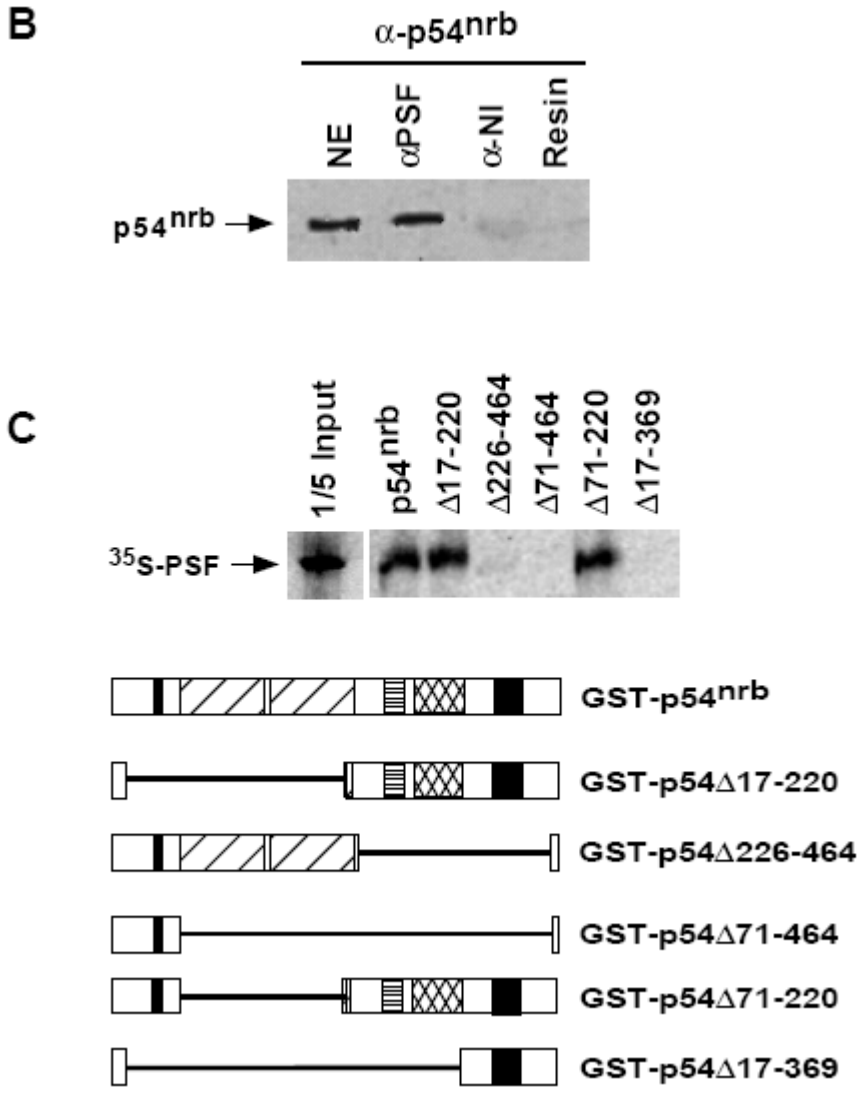


Figure 8. PSF-p54^{nrb} interaction. **(A)** Human PSF, p54^{nrb}, mouse NonO, and *Drosophila* NonA are highly homologous over a 320 amino acid region referred to as the *Drosophila* Behavior, Human Splicing domain (DBHS; shaded box; Dong et al., 1993). This region includes the two RNA Recognition Motifs (RRM1 and RRM2, diagonally hatched boxes), a helix-turn-helix motif (HtH, horizontally hatched boxes), and a basic/acidic region (+/-, cross-hatched boxes). Additional similarity exists in regions rich in proline (P) and glutamine (Q) residues (black boxes). **(B)** Immunoprecipitation of HeLa nuclear extract (NE) was performed with either an anti-PSF antibody (α-PSF), non-immune serum (α-NI), or protein G beads alone (Resin). Precipitated proteins were analyzed by Western blot analysis with an anti-p54^{nrb} antibody. **(C)** *In vitro* translated, ^{35}S-labeled PSF was incubated with the indicated GST-p54^{nrb} constructs in the presence of glutathione-agarose beads. After washing, bound proteins were eluted and resolved by SDS-PAGE. Deleted regions in the GST-p54^{nrb} mutants are represented by a thin line, with the numbers indicating the amino acid positions of the deleted residues. Domains of interest are represented as in (**A**).


B
α-p54nrb
NE
αPSF
α-NI
Resin
p54nrb
C
1/5 Input
p54nrb
Δ17-220
Δ226-464
Δ71-464
Δ71-220
Δ17-369
35S-PSF
GST-p54nrb
GST-p54Δ17-220
GST-p54Δ226-464
GST-p54Δ71-464
GST-p54Δ71-220
GST-p54Δ17-369

screens have shown that zinc finger proteins, such as RNF4, can interact with the second RRM in PSF and may be involved in nuclear localization of PSF (Dye and Patton, 2001).

The interaction between PSF and NonO/p54^{nrb} is especially interesting since the two proteins are so similar. Originally identified as a protein that cross-reacts with an antibody raised against the yeast U5 snRNP-associated second-step splicing factor Prp18, p54^{nrb} is 71% identical to PSF over a region of 320 amino acids that encompasses their RRMs (Fig. 8A; Dong et al., 1993). Other homologs of PSF and p54^{nrb} include NonA/BJ6 from *Drosophila*, which has been shown to be important in *Drosophila* visual acuity and male courtship song (Besser et al., 1990; Jones and Rubin, 1990), hrp65 from *C. tentans*, a component of nuclear fibers associated with specific pre-mRNPs (Miralles et al., 2000; Wurtz et al., 1996), and PSP1 from humans, a paraspeckle protein of unknown function (Andersen et al., 2002; Fox et al., 2002).

To verify that p54^{nrb} interacts with PSF and to identify the regions of p54^{nrb} that are required for binding, co-immunoprecipitation experiments and *in vitro* interaction assays were performed. As shown in Fig. 8B, polyclonal anti-PSF antibodies were capable of co-immunoprecipitating p54^{nrb} from HeLa nuclear extract, whereas p54^{nrb} was not precipitated in control reactions using non-immune serum or protein G beads alone. The converse experiment using antibodies against p54^{nrb} also resulted in co-immunoprecipitation (data not shown). To identify which sequences of p54^{nrb} are required for binding PSF, a series of glutathione-S-transferase-p54^{nrb} fusion proteins (GST-p54^{nrb}) were assayed for their ability to precipitate PSF using glutathione-agarose pull down assays (Fig. 8C). Full-length GST-p54^{nrb} and two deletion mutants (GST-p54$^{nrb}\Delta$17-220 and GST-p54$^{nrb}\Delta$71-220), both of which contain the putative

helix-turn-helix motif and basic/acidic region (Yang et al., 1993), were capable of binding PSF. In contrast, GST fusions that lack either or both of these regions (GST-p54^{nrb}Δ226-464, GST-p54^{nrb}Δ71-464 and GST-p54^{nrb}Δ17-369) could not precipitate PSF. Thus, it appears that the C-terminus of p54^{nrb} is required for its interaction with PSF. These results were corroborated by yeast two-hybrid assays, which showed that both p54^{nrb}Δ17-220 and p54^{nrb}Δ71-220 interacted with PSF in the two-hybrid system, whereas p54^{nrb}Δ226-464, p54^{nrb}Δ71-464, and p54^{nrb}Δ17-369, did not (data not shown). When the same mapping experiments were performed with a series of PSF deletion constructs, only full length PSF was capable of interacting with p54^{nrb} (data not shown). It appears that multiple contacts, or a precise tertiary structure, are needed for PSF to interact with p54^{nrb}.

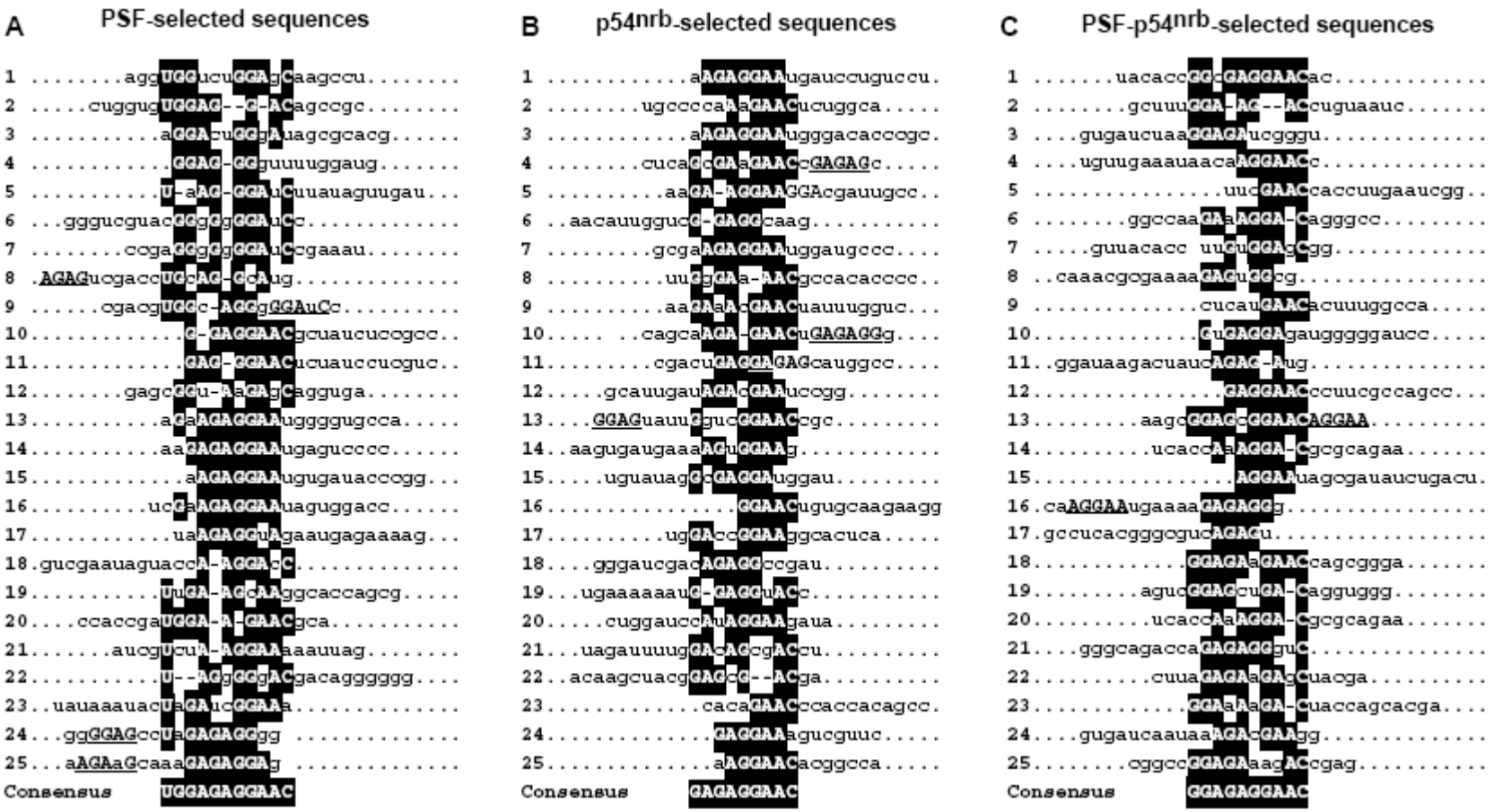
Determination of the optimal RNA binding sites of PSF and p54^{nrb}

Iterative selection assays (Szostak, 1992; Tuerk and Gold, 1990) were used to determine the optimal RNA binding sequence for PSF and p54^{nrb}. A pool of *in vitro* transcribed RNAs representing over 10^{12} different sequences was incubated with recombinant, hexahistidine-tagged (his-tagged) proteins, and bound RNAs were recovered by co-purification over Ni-NTA agarose. Sequencing of 20 independent clones from the initial pool showed that the randomized region contained roughly equal amounts of each nucleotide (data not shown). Prior to incubation with recombinant proteins, each RNA pool was pre-incubated with Ni-NTA agarose to prevent enrichment of nonspecific RNAs. Following selection and amplification for eight sequential rounds, individual clones were sequenced. For PSF, simple calculation of the purine and

pyrimidine content of the selected sequences showed enrichment for purines, from roughly 50% to 67% (data not shown). When the selected sequences were compared using multiple alignment algorithms, a consensus sequence was identified consisting of 5'- UGGAGAGGAAC -3' (Fig. 9A). Genomic data base searches with this sequence were not particularly useful since its length is less than the number of bases needed to represent a unique sequence in the human genome. However, since PSF had been shown to co-purify with U4/U6.U5 tri-snRNP, we compared the consensus sequence to these snRNAs. Strikingly, the selected sequence is identical to a region of U5 snRNA referred to as stem 1b (Fig. 9D, 9E).

All U5 snRNAs have two stem-loop structures (Fig. 9E). The overall secondary structure and particularly the 5' loop sequence are highly conserved among all eukaryotes while the rest of the primary sequence is less conserved (Branlant et al., 1983; Guthrie and Patterson, 1988). The sequence of stem 1b, including the PSF consensus sequence, is remarkably conserved from flies to humans suggesting functional importance (Fig. 9D, 9E). Interestingly, this sequence is not found in the corresponding region of yeast U5 snRNA, correlating with the absence of any identified PSF homologs in *S. cerevisiae* and thus far, *S. pombe*.

Given the high degree of homology between the RRMs of PSF and $p54^{nrb}$, the co-immunoprecipitation data shown in Fig. 8, and previous biochemical data suggesting that these proteins exist as a heterodimer (Straub et al., 1998; Zhang et al., 1993), it was of interest to determine the optimal RNA binding sites for $p54^{nrb}$ and mixtures of PSF-$p54^{nrb}$. As with PSF, selection with $p54^{nrb}$ or mixtures of PSF-$p54^{nrb}$ yielded sequences that also aligned well with the 3' side of U5 snRNA stem 1b (Fig. 9B, 9C, 9D)


A
PSF-selected sequences
B
p54nrb-selected sequences
C
PSF-p54nrb-selected sequences
Consensus
Consensus
Consensus

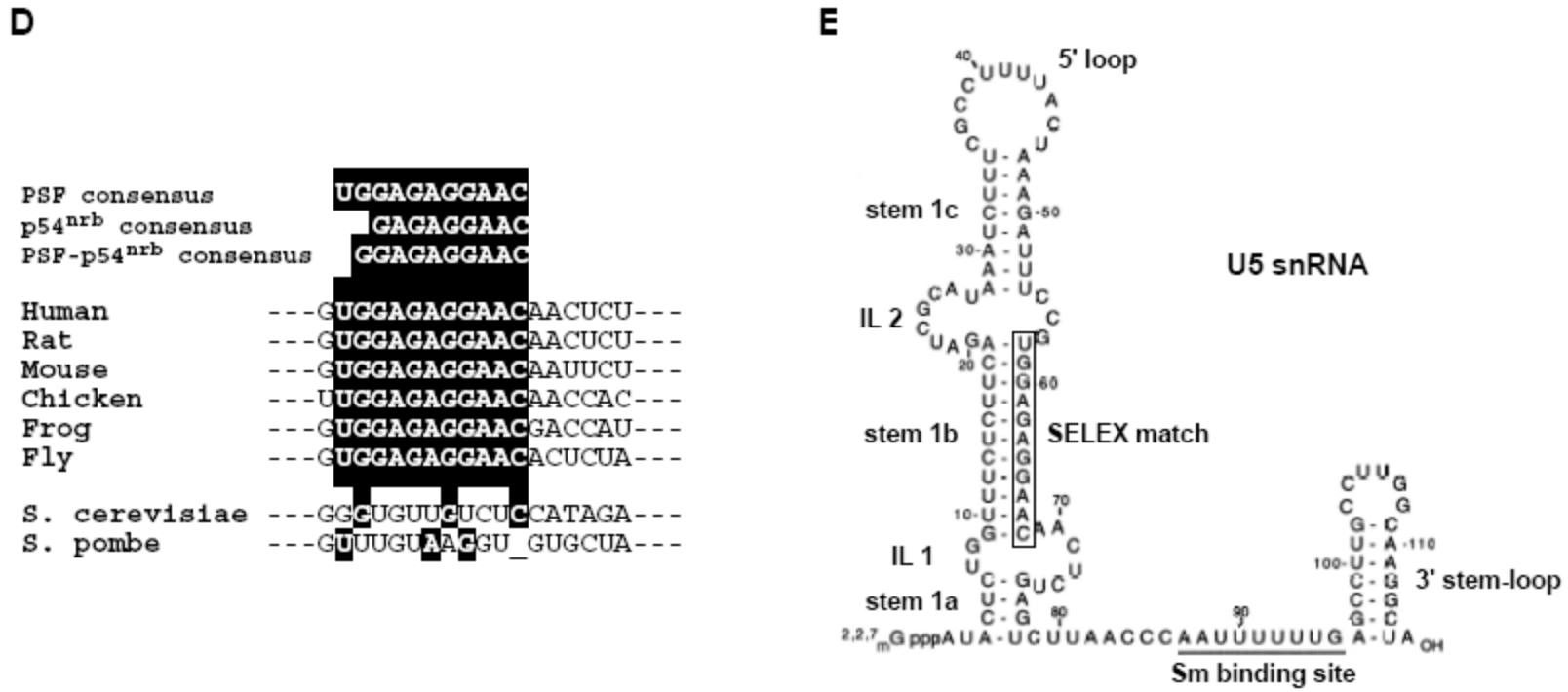


Figure 9. Identification of optimal RNA binding sites for PSF and p54^{nrb}. Iterative selection assays were performed using a pool of RNAs containing 20 randomized nucleotides and either recombinant, his-tagged PSF (**A**), recombinant, his-tagged p54^{nrb} (**B**), or a 1:1 mixture of PSF-p54^{nrb} (**C**). Eight rounds of transcription, selection, and amplification were performed before individual clones were isolated and sequenced. Alignments of the selected sequences (numbered) and the derived consensus sequences are as shown. Aligned nucleotides are displayed as capital letters on a black background, and other nucleotides that form additional alignments with the consensus sequence are in capital letters and underlined. **(D)** Alignment of PSF and p54^{nrb} consensus sequences to U5 stem 1b. The PSF and p54^{nrb} consensus sequences are shown aligned with the 3' strand of U5 snRNA stem 1b sequences from human, rat (*Rattus norvegicus*), mouse (*Mus musculus*), chicken (*Gallus gallus*), frog (*Xenopus laevis*), fly (*Drosophila melanogaster*), and yeast (*Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*) (Guthrie and Patterson, 1988). **(E)** The secondary structure of human U5 snRNA is depicted with the PSF RNA binding consensus sequence boxed (adapted from Ségault et al., 1999).

This suggests that the optimal RNA binding sites for PSF and p54^{nrb}, and/or mixtures thereof, are extremely similar, if not identical (Fig. 9C, 9D).

PSF and p54^{nrb} form a complex with U5 snRNA *in vitro*.

It is possible that the RNA sequences selected by mixtures of PSF-p54^{nrb} (Fig. 9C) were derived by interaction with individual proteins, or by interaction with PSF-p54^{nrb} complexes. We therefore examined binding of these proteins to U5 snRNA using gel shift assays. *In vitro* transcribed, uniformly ^{32}P-labeled, wild type U5 RNA was incubated with his-tagged PSF and his-tagged p54^{nrb} (either individually or in combination), in the presence of excess yeast tRNA as non-specific competitor, and subjected to 5% native gel electrophoresis. Upon incubation of recombinant PSF with wild type U5, a low mobility complex was observed (Fig. 10, lane 1). This complex could be competed away by unlabeled wild type U5 snRNA, but not yeast tRNA (data not shown), suggesting specific interaction between PSF and U5. When recombinant p54^{nrb} was incubated with U5 by itself, no complex was detected (Fig. 10, lane 2). In contrast, incubation of 1:1 mixture of recombinant PSF and p54^{nrb} with wild type U5 resulted in a complete shift of the PSF/U5 complex to a lower mobility complex (Fig. 10, lane 3), indicating that the two proteins form a complex with U5 snRNA simultaneously.

Binding of PSF and p54^{nrb} to U5 snRNA is dependent on both the sequence and structure of stem 1b.

Proteins with RRM domains, such as PSF and p54^{nrb}, are thought to primarily bind single stranded RNA. The fact that PSF and p54^{nrb} selected single stranded sequences that match the 3' side of U5 stem 1b, and that they form a complex with U5 *in vitro*,

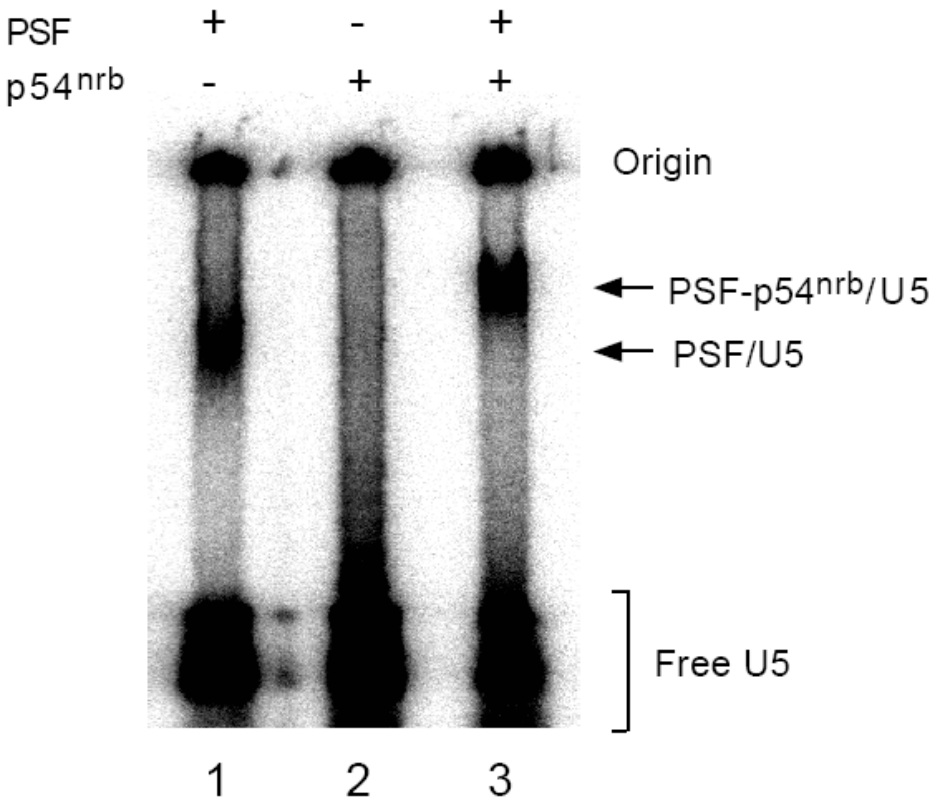
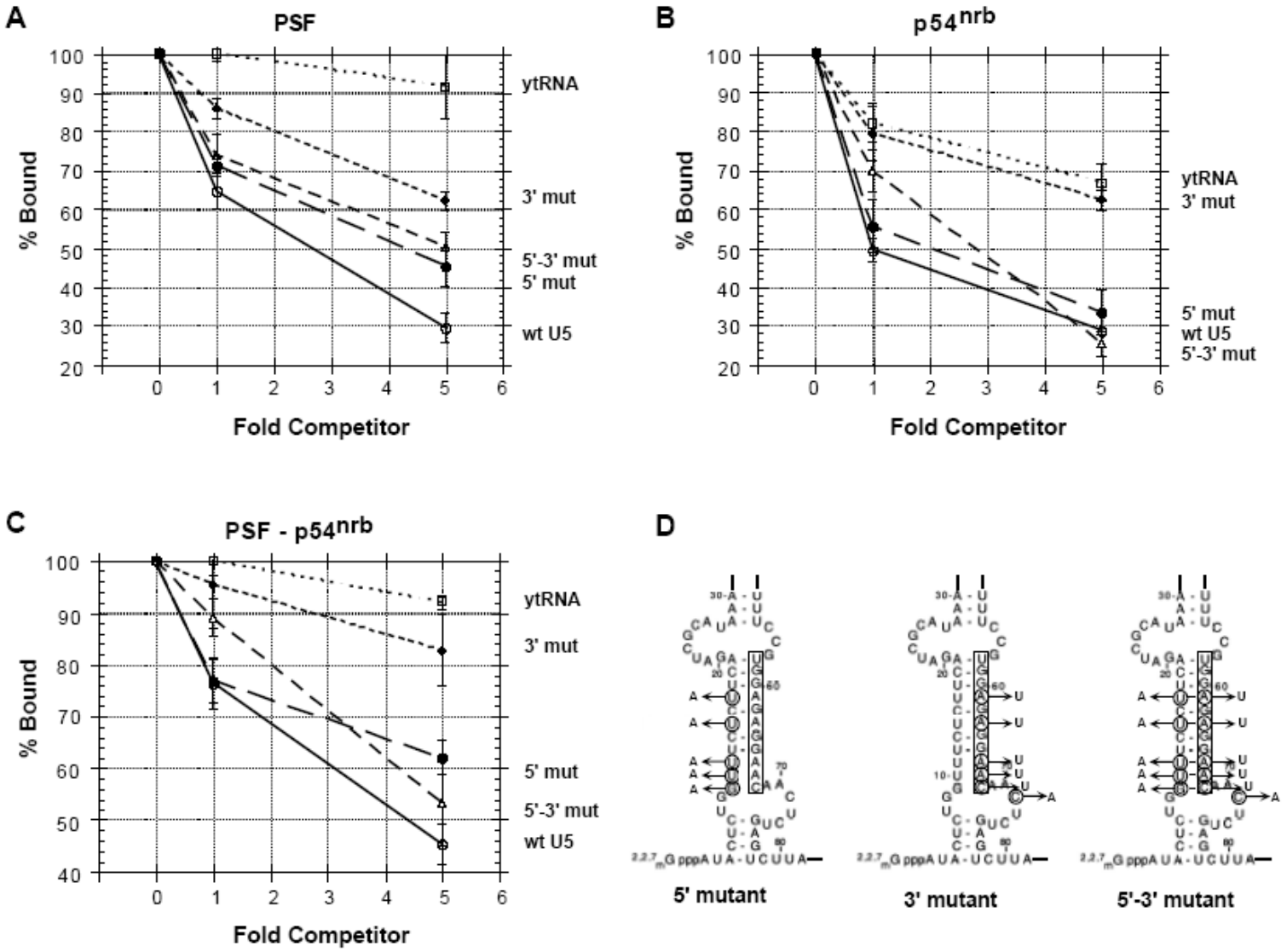


Figure 10. PSF and p54^{nrb} form a complex with U5 snRNA *in vitro*. *In vitro* transcribed wild type U5 RNA was incubated with his-tagged PSF (lane 1), his-tagged p54^{nrb} (lane 2), or 1:1 mixture of his-tagged PSF-p54^{nrb} (lane 3) in the presence of excess yeast tRNA. The reactions were incubated at 30^{o}C for 15 min, and complex formation was analyzed by 5% native gel electrophoresis. Free RNA and protein-RNA complexes are indicated by arrows.

raised the question as to whether such interaction requires melting of stem 1b or whether the stem remains paired. To analyze their interaction with U5 snRNA, filter binding assays were used. *In vitro* synthesized, uniformly ^{32}P-labeled, wild type U5 snRNA was incubated with recombinant, his-tagged PSF or p54^{nrb} in the presence or absence of competitor RNAs. Protein-RNA complexes were then separated from free RNA by filtration through nitrocellulose and the level of retained U5 was quantitated. Competitor RNAs consisted of wild type U5, U5 snRNAs containing mutations in stem 1b, or yeast tRNA as a nonspecific control. The different U5 snRNA mutants included a 5' mutant, in which the sequence of the 5' strand of stem 1b was changed to disrupt the base pairing of stem 1b without altering the optimal PSF-p54^{nrb} binding site, a 3' mutant in which the sequence of the 3' strand of stem 1b was changed so that both the base pairing and the PSF-p54^{nrb} binding site were disrupted, and a 5'-3' double mutant in which compensatory changes recreated stem 1b but with altered sequence on both strands (Fig. 11D).

Incubation of recombinant PSF with labeled wild type U5 led to retention of U5 snRNA on the nitrocellulose filters whereas only small amounts of U5 were retained in the absence of protein (about 2% of total input). To ensure specificity, unlabeled wild type U5 snRNA was included and found to effectively compete with labeled wild type U5, reducing the binding to 30% of maximum at a 5-fold molar excess of competitor RNA (Fig. 11A). In contrast, yeast tRNA showed little competition. Thus, it appears that the filter binding assay monitors specific interaction between PSF and U5. To test whether stem 1b is required for interaction with PSF, competition binding experiments were performed with the 5', 3' and 5'-3' U5 snRNA mutants. Compared to wild type

Figure 11. PSF and p54^{nrb} bind U5 snRNA with both the sequence and structure of stem 1b contributing to binding specificity. Filter binding experiments were performed with radiolabeled U5 snRNA in the presence of either his-tagged PSF (**A**), his-tagged p54^{nrb} (**B**), or a mixture of his-tagged PSF-p54^{nrb} (**C**). Increasing amounts of unlabeled competitor RNAs were included and the decrease in binding is shown with the indicated standard deviation ($n \geq 3$). Competitor RNAs: ○, wild type (wt U5); ●, the 5' mutant (5' mut); ◆, the 3' mutant (3' mut); △, the 5'-3' mutant (5'-3' mut); □, yeast tRNA (ytRNA). **(D)** U5 snRNA mutants used in filter binding experiments and RNA affinity assays. Mutated nucleotides (circled) were changed to the bases indicated by the arrows. Only the region surrounding stem 1b is shown, and no additional mutations were made outside of this region. The predicted PSF-p54^{nrb} binding site is boxed.


A
PSF
% Bound
Fold Competitor
ytRNA
3' mut
5'-3' mut
5' mut
wt U5
B
p54nrb
% Bound
Fold Competitor
ytRNA
3' mut
5' mut
wt U5
5'-3' mut
C
PSF - p54nrb
% Bound
Fold Competitor
ytRNA
3' mut
5' mut
5'-3' mut
wt U5
D
5' mutant
3' mutant
5'-3' mutant

U5, the 5' mutant could compete for binding but was not as effective a competitor as wild type U5. The 3' mutant showed some specificity for PSF binding but was consistently the poorest competitor at all concentrations. Surprising results were obtained with the 5'-3' mutant. We predicted that this RNA would not compete since the PSF binding site was completely changed, but competition assays showed that it displayed a competitive efficiency similar to the 5' mutant. Folding algorithms suggest that the compensatory mutations recreate stem 1b implying that the ability of this RNA to compete for PSF binding is due to the overall structure of the RNA rather than the sequence of the 3' strand of stem 1b. This may explain why the 5' mutant competed less efficiently than wild type U5. Combined, the three mutant RNAs suggest that optimal binding of PSF to U5 requires both the sequence and structure of stem 1b.

For $p54^{nrb}$, yeast tRNA was able to partially compete with labeled wild type U5 snRNA (Fig. 11B). Also, the 5' mutant and the 5'-3' mutant RNAs exhibited similar competitive abilities as wild type U5 snRNA, while the 3' mutant was no more effective than yeast tRNA. These results suggest that in the absence of the original stem 1b structure (compare the 5' mutant with the 3' mutant), the wild type sequence on the 3' side of the stem 1b is required for specific binding of $p54^{nrb}$ to U5 snRNA. However, in the presence of the original stem structure, the stem sequence is apparently less important for binding specificity (compare the 5'-3' mutant with wild type U5). Together, the binding experiments suggest that $p54^{nrb}$, like PSF, binds to U5 stem 1b, and that both the sequence and structure of stem 1b contribute to binding specificity.

Filter binding assays were also performed with the combination of PSF and $p54^{nrb}$ at approximately a 1:1 ratio (Fig. 11C). Like PSF, yeast tRNA was unable to

out-compete binding between the PSF-p54^{nrb} complex and wild type U5 snRNA. The 5'-3' and the 5' mutant RNAs were both effective competitor RNAs though not as efficient as wild type U5. With the 3' mutant, there was a noticeable difference in competitive efficiency using the PSF-p54^{nrb} complex as opposed to the individual proteins. For the combination, the 3' mutant competed only slightly better than did yeast tRNA. This suggests that the PSF-p54^{nrb} complex, compared to each individual protein, binds U5 most efficiently when the sequence of the 3' side of stem 1b is wild type. However, since the 5'-3' mutant also competes, the structure of stem 1b is also important for binding.

PSF and p54^{nrb} bind U5 snRNA in nuclear extract.

The gel shift assays and filter binding assays described above were performed using relatively simple *in vitro* systems. In contrast, assembly of nuclear U5 snRNA into U5 snRNP and U4/U6.U5 tri-snRNP requires that multiple proteins interact with the RNA. To determine whether PSF and p54^{nrb} contact U5 snRNA in the presence of snRNP proteins and other factors, RNA affinity selection assays were performed using biotinylated wild type and mutant U5 snRNAs. Following incubation of biotinylated U5 snRNAs in HeLa nuclear extract under splicing conditions, associated proteins were captured by passage over streptavidin-agarose and subjected to Western blot analysis.

As shown in Fig. 12, Sm proteins B/B' assembled onto the different biotinylated U5 snRNAs with similar efficiency. U5 snRNP-specific proteins (U5-200 kDa and U5-116 kDa) also associated with all U5 snRNAs with only slightly variable efficiencies between the different U5 snRNAs. Thus, both wild type and mutant U5 snRNAs were assembled

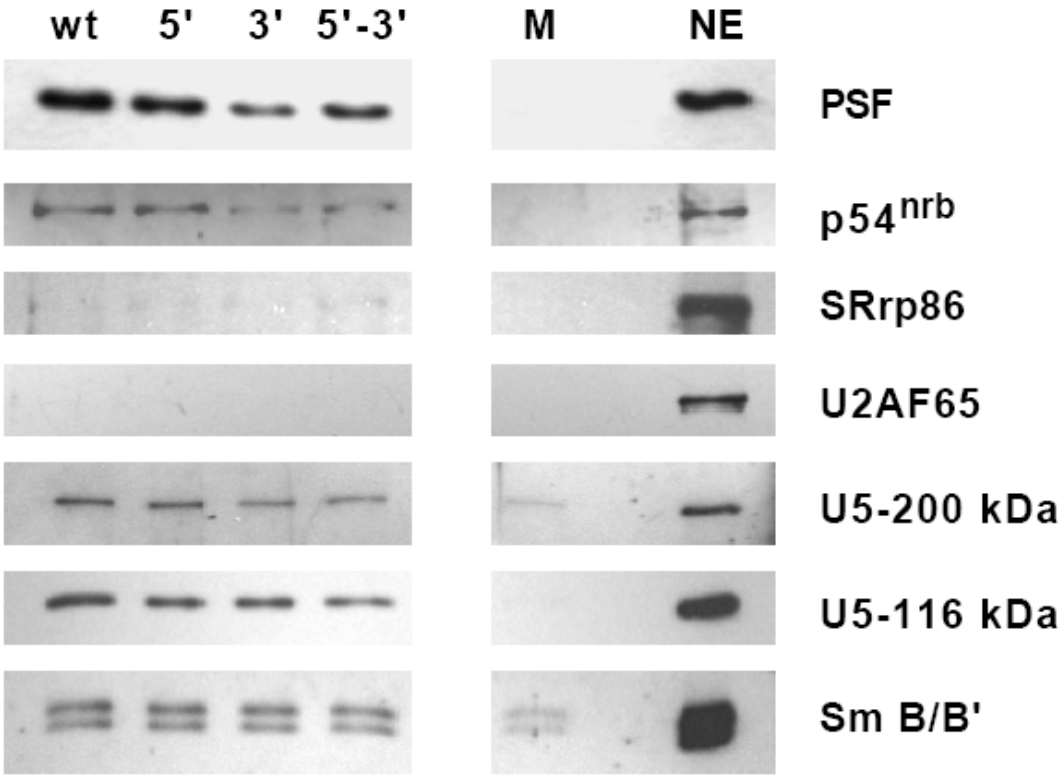


Figure 12. RNA affinity assays. Wild-type U5 snRNA (U5), the 5' mutant (5'), the 3' mutant (3'), and the 5'-3' double mutant (5'-3') (shown in Fig. 11D) were transcribed in the presence of biotin-14-CTP. Following incubation with HeLa nuclear extract (NE), biotinylated RNAs and associated proteins were captured by passage over streptavidin agarose. After extensive washing, bound proteins were eluted and resolved by SDS-PAGE. Mock reactions (M) were carried out in an identical manner without biotinylated RNA. Western blot analysis was performed using antibodies against the indicated proteins.

into snRNPs under these conditions. For PSF and p54^{nrb}, both proteins associated with U5 snRNA in nuclear extract (Fig. 12). PSF bound wild type U5 snRNA most efficiently with slightly less association with the 5' mutant, lesser association with the 5'-3' mutant, and the least association with the 3' mutant. These results mirror the filter binding results shown in Fig. 11. Likewise, the association of p54^{nrb} with the different U5 snRNAs exhibited a similar pattern: wild type > the 5' mutant > the 5'-3' mutant > the 3' mutant. As controls, neither U2AF^{65} (Zamore et al., 1992) nor SRrp86 (Barnard and Patton, 2000) associated with any of the biotinylated U5 snRNAs. These data indicate that both PSF and p54^{nrb} can bind to U5 snRNA in the presence of multiple RNA-binding proteins in nuclear extract and that such interaction is dependent on the sequence and structure of stem 1b.

PSF and p54^{nrb} associate with U4/U6.U5 tri-snRNP and splicing complexes.

Since the gel shift assays suggested that PSF and p54^{nrb} might bind U5 as a complex, we next analyzed the distribution of PSF and p54^{nrb} during splicing in HeLa nuclear extract by sucrose gradient sedimentation. Splicing reactions using labeled pre-mRNA were allowed to proceed for 15 min before separation on 10%-30% sucrose gradients. RNAs were extracted from each fraction and separated on a 15% denaturing gel to identify the location of both labeled splicing RNAs as well as snRNAs. Protein components of each fraction were assayed by Western blot analysis. As shown in Fig. 13A and 13B, pre-mRNA, mRNA, and splicing intermediates accumulated near the bottom of the gradient along with all five U snRNAs (fractions 17-20). While the different snRNAs sedimented across the gradient, defined RNA and protein peaks could

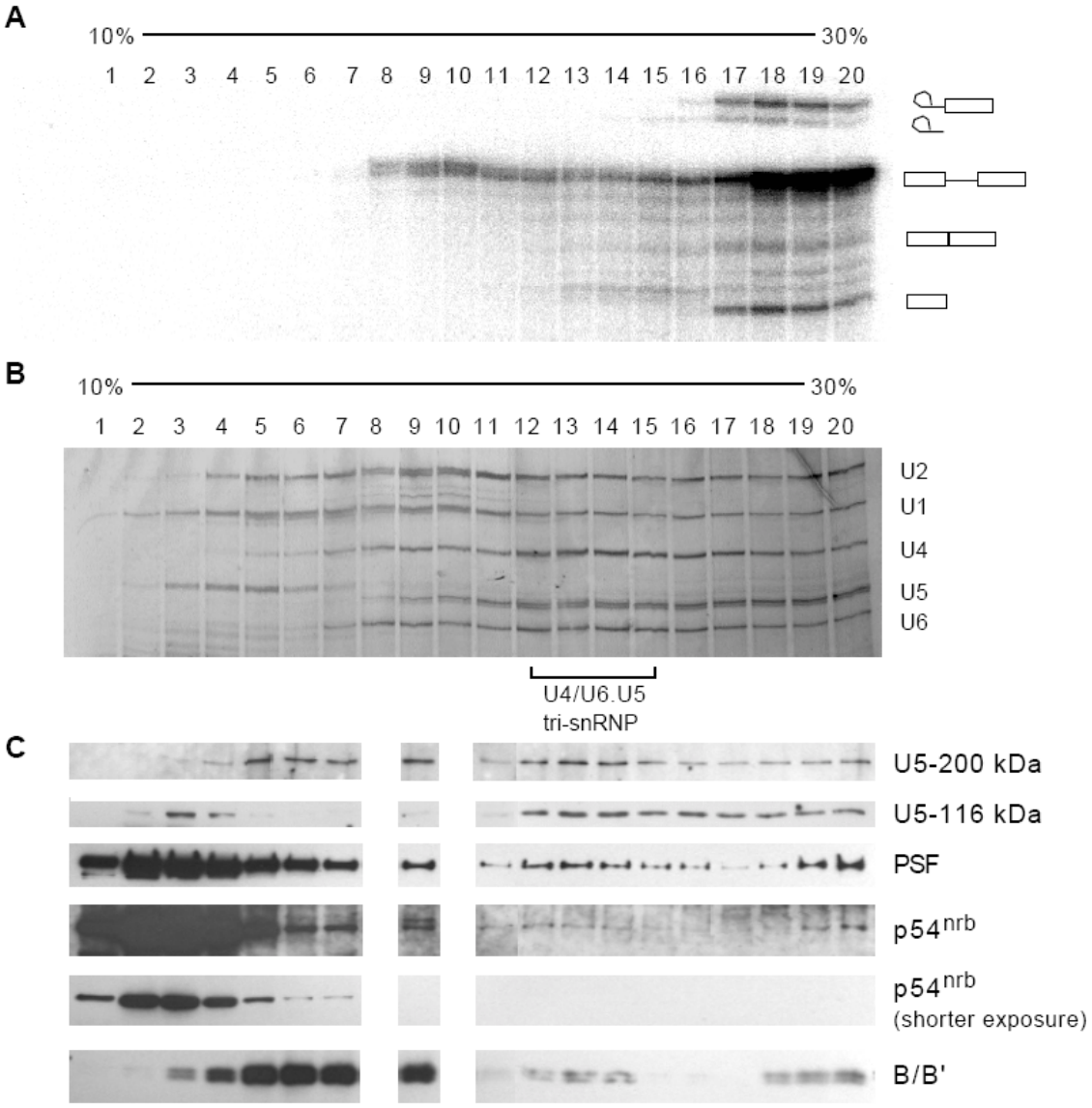


Figure 13. PSF and p54^{nrb} associate with U4/U6.U5 tri-snRNP and splicing complexes. 200 µl splicing reactions using AdML pre-mRNA were separated on 10%-30% sucrose gradients. RNAs from each fraction (numbered) were recovered by phenol/$CHCl_3$ extraction and separated on 8M urea-15% PAGE. Splicing products and snRNAs were visualized by phosphorimager analysis (**A**) and silver staining (**B**), respectively. Splicing RNAs and snRNAs are indicated on the right. The position of U4/U6.U5 tri-snRNP is shown. **(C)** Proteins from each fraction above (except fraction 8 and 10) were analyzed by Western blot analysis using different antibodies as indicated.

be detected that consist of free snRNPs and U4/U6.U5 tri-snRNP (fractions 12-15; Fig. 13B, 13C). The majority of PSF and p54^{nrb} sedimented near the top of the gradient (fractions 2-4; Fig. 13C). However, a portion of PSF and p54^{nrb} co-migrated with U4/U6.U5 tri-snRNP, and another subfraction co-sedimented with splicing complexes (fractions 19 and 20). These results confirm that PSF and p54^{nrb} associate with splicing complexes but that only a fraction of each protein co-sediments with such complexes.

Discussion

PSF and p54^{nrb} are multifunctional interacting splicing factors.

We have demonstrated that PSF interacts with p54^{nrb} and that both of these proteins, individually and in combination, select the same optimal RNA binding sequence from random pools of RNA. PSF and p54^{nrb} share 71% identity over a 320 amino acid region encompassing their RRMs (Dong et al., 1993) and multiple functions have been ascribed to each. PSF co-purifies with U4/U6.U5 tri-snRNP preparations (Teigelkamp et al., 1997) and a variety of biochemical experiments indicate that PSF plays an important role in pre-mRNA splicing (Gozani et al., 1994; Patton et al., 1993). In addition, roles for PSF in transcription, topoisomerase activity, nuclear RNA retention, DNA recombination, and DNA double-strand break rejoining have also been postulated (Akhmedov and Lopez, 2000; Bladen et al., 2005; Mathur et al., 2001; Sewer and Waterman, 2002; Sewer et al., 2002; Song et al., 2004; Straub et al., 1998; Straub et al., 2000; Urban and Bodenburg, 2002; Urban et al., 2000; Zhang and Carmichael, 2001; Zolutukhin et al., 2002). Similarly, multiple functions involving both RNA and DNA binding have been proposed

for p54^{nrb} (Basu et al., 1997; Sewer et al., 2002; Straub et al., 1998; Straub et al., 2000; Zhang et al., 1993; Zhang and Carmichael, 2001). Several lines of evidence indicate that p54^{nrb} is also involved in splicing. First, p54^{nrb} was originally isolated in screens designed to identify proteins that cross-react with antibodies against Prp18, a yeast second step splicing factor associated with U4/U6.U5 tri-snRNP and U5 snRNP (Dong et al., 1993; Horowitz and Abelson, 1993; Vijayraghavan and Abelson, 1990). Second, GFP fusions and antibodies raised against p54^{nrb} have been used to demonstrate localization to nuclei in a speckled pattern albeit somewhat more diffuse than some splicing factors (not shown). Third, over-expression of Spi-1/PU.1, an Ets-related transcription factor, blocks p54^{nrb} RNA binding which correlates with an effect on *in vitro* splicing (Hallier et al., 1996). Lastly, in this study, PSF and p54^{nrb} were found to co-sediment with splicing complexes (Fig. 13). As to how these two proteins might function in splicing, early experiments showed that immunodepletion of PSF from splicing extracts affected both early spliceosome assembly and the second catalytic step (Gozani et al., 1994; Patton et al., 1993). More recently, PSF and p54^{nrb} have been found to associate with the 5' splice site (Kameoka et al., 2004), consistent with a recent report that U4/U6.U5 tri-snRNP associates with the 5' splice site (Maroney et al., 2000). It is possible that PSF and p54^{nrb} interact with the 5' splice site early in spliceosome assembly and that this association is maintained throughout later stages of splicing, such as when the 5' splice site is positioned in proximity to U5 snRNP. Thus, PSF and p54^{nrb} appear to function in multiple steps of the splicing pathway, one of which may be rate-limiting for the second catalytic step. Given recent experiments suggesting a link between transcription and splicing (Fong and Bentley, 2001; Fong and Zhou, 2001;

Maniatis and Reed, 2002; Monsalve et al., 2000), it is also possible that the multiple roles proposed for PSF and p54^{nrb} indicate a potentially exciting link between these important nuclear functions.

PSF and p54^{nrb} bind a conserved stem in U5 snRNA.

SELEX experiments using PSF, p54^{nrb}, or the combination thereof, identified a purine-rich sequence identical to the 3' side of U5 snRNA stem 1b as their optimal RNA binding sites (Fig. 9). Previous iterative selection assays using fewer (five) rounds of selection with p54^{nrb} also identified a purine rich sequence, 5'-AGGGA-3' (Basu et al., 1997). Although this sequence is somewhat different from our results, closer examination of the selected sequences from these experiments show that they align well with U5 stem 1b. For example, 29 out of 30 clones had one or more similar GGAG or GGAA motifs, while 10 of the clones contained the sequence 5'-UGG_GAGGAA-3'. Together, the SELEX experiments suggest that both PSF and p54^{nrb} bind stem 1b.

Binding of PSF and p54^{nrb} to U5 was first verified by gel shift assays, then analyzed by filter binding assays and affinity selection experiments (Figs. 3-5). In gel shift assays, PSF/U5 complex was completely super-shifted by the addition of p54^{nrb}, resulting in a PSF-p54^{nrb}/U5 complex (Fig. 10). Although we cannot exclude the possibility that the two proteins have individual binding sites on U5 snRNA, the fact that they selected the same sequence (the 3' side of U5 snRNA stem 1b) suggests they bind to U5 stem 1b as a complex, consistent with previous reports (Straub et al., 1998; Zhang et al., 1993). In agreement with this, nuclear PSF and p54^{nrb} showed similar distribution patterns after sucrose gradient sedimentation (Fig. 13). In contrast, no p54^{nrb}/U5 complexes were

observed in gel shift assays but such interactions were clearly detectable in filter binding experiments. Thus, $p54^{nrb}$ by itself may bind U5 snRNA weakly, but such interaction might be stabilized by the presence of PSF.

Binding specificity for both proteins appears to depend on multiple elements, including the predicted binding sequence on the 3' side of stem 1b. The importance of the 3' strand of the stem 1b was first identified in SELEX assays and confirmed using the 5' and 3' mutant U5 snRNAs. In both the filter binding assays and the biotinylated-RNA selection experiments, PSF and $p54^{nrb}$ bound to wild type U5 and the 5' mutant RNA more efficiently than to the 3' mutant. However, the importance of the stem structure was shown using the 5'-3' double mutant. Since the 5'-3' mutant lacks the optimal binding sequence for PSF-$p54^{nrb}$, it seems that the structure of stem 1b contributes to binding specificity as well. While we have not experimentally verified whether the 5'-3' mutant adopts a structure identical to wild type U5, computer folding algorithms predict the same secondary structure. Whether other factors also contribute to binding specificity remains unclear. Secondary structure predictions of the 5' and 3' mutants suggest that these mutations primarily affect the structure of stem 1b with minor disruption of immediately adjacent structures (IL1 and IL2). The weak but detectable binding of PSF and $p54^{nrb}$ to the 3' mutant implies that regions other than stem 1b might also be involved in binding.

Association of PSF and $p54^{nrb}$ with U5 snRNA during splicing

Double-stranded RNAs adopt the A-form conformation which precludes base-specific interaction with protein side chains in the deep major groove (Steitz, 1999).

Therefore, the question arises as to how PSF and p54^{nrb} interact with an A-form helix while apparently maintaining sequence specificity. One speculative possibility is that PSF and p54^{nrb} could bind U5 in two ways: to the intact stem, or to the 3' side of the stem after melting. At least two ATPases (U5-100 kDa and U5-200 kDa) and one homolog of EF-2 GTPase (U5-116 kDa) have been found to associate with U5 snRNP and may function to mediate unwinding to facilitate the multiple RNA-RNA rearrangements that are central to splicing (Fabrizio et al., 1997; Laggerbauer et al., 1998; Teigelkamp et al., 1997). Interestingly, hPrp8 (U5-220 kDa), which makes multiple contacts with the pre-mRNA and with U5 (Chiara et al., 1997; MacMillan et al., 1994; Reyes et al., 1996), forms a stable complex with three U5 proteins (U5-200, U5-116, and U5-40; Achsel et al., 1998). IL2, the internal loop between stem 1b and 1c, is required for efficient association of hPrp8 and U5-116 kDa with U5 (Hinz et al., 1996; Ségault et al., 1999). Given that the U5-200 kDa protein is a putative unwindase, it seems reasonable to propose that stem 1b of U5 might also undergo a conformational change during spliceosome assembly or during the two catalytic steps. Interestingly, two-hybrid screens using fragments of hPrp8 have detected interaction with PSF (G. Moreau and M. Moore, personal communication), consistent with association of these two proteins to U5 in the vicinity of stem 1b.

As to how PSF and p54^{nrb} might function in this process, there are at least two possibilities. First, PSF and p54^{nrb} may initially bind to double stranded stem 1b but then bind more tightly to the 3' side of the stem upon unwinding. Such binding could stabilize the unpaired strands coincident with U4/U6.U5 tri-snRNP contacting the 5' splice site, or perhaps during the second step of splicing. Second, given that PSF has

been demonstrated to promote the annealing of DNA (Akhmedov and Lopez, 2000), PSF and p54^{nrb} could facilitate the reannealing of U5 stem 1b subsequent to melting, mediated by one or more unwindases. Either way, the interaction could be transient, supported by non-stoichiometric association of PSF with U4/U6.U5 tri-snRNP (Teigelkamp et al., 1997), and the fact that only a small amount of PSF and p54^{nrb} co-sedimented with U4/U6.U5 tri-snRNP and with splicing complexes (Fig. 13).

The high degree of conservation of U5 stem 1b in vertebrates and flies implies that it plays a critical role in U5 function. Ségault et al. (1999) examined the ability of several human U5 snRNA mutants to function in splicing by reconstituting U5-depleted nuclear extract with *in vitro* transcribed mutant U5 snRNAs. Although the rescue of splicing was prevented by deletion of IL2, a stem 1b mutant (sub-stem 1b) was still partially functional, perhaps suggesting that this region is not important after all. However, the sub-stem 1b mutant maintained a purine-rich sequence of 5'-CAGAGAGAAGU-3' on the 5' side of the stem. Comparison of this sequence with the 3' strand of the original stem (the optimal PSF-p54^{nrb} binding site) showed that all the changes are transitions, while most of our changes are transversions. Furthermore, about 50% of the SELEX sequences we identified contain at least one AGAG or GAAG motifs (Figs. 9A, 9B, 9C). Thus, it is possible that the sub-stem 1b mutant fortuitously maintained a binding site for PSF and p54^{nrb} on the 5' side of stem 1b.

Unlike higher eukaryotes, the PSF-p54^{nrb} binding site is not present in the corresponding regions of yeast U5 snRNA (Fig. 9D). This lack of conservation, together with the apparent absence of PSF and p54^{nrb} homologs in yeast, suggests either that yeast U5 functions differently from that of higher eukaryotes, or that yeast contain an as yet

unidentified functional homolog of PSF and $p54^{nrb}$. Regardless, the data presented in this paper show that PSF and $p54^{nrb}$ interact with stem 1b of U5 snRNA, requiring both the sequence and structure for binding. Detailed analysis of the effects that PSF and $p54^{nrb}$ have on U5 snRNA will be required to determine the exact role these proteins play in early/late steps of splicing, and any potential links between transcription and splicing.

CHAPTER III

THE SPLICING FACTOR PSF IS PART OF A MAMMALIAN MULTI-SNRNP COMPLEX.

Introduction

Most eukaryotic genes are transcribed into pre-mRNA containing coding sequences (exons) disrupted by intervening sequences (introns). Prior to translation, the introns must be efficiently and accurately removed and the exons joined together in a process known as pre-mRNA splicing (Burge et al., 1999; Hastings and Krainer, 2001). Splicing occurs in two sequential transesterification reactions, the first of which involves cleavage of the 5' exon and generation of a lariat intron-3' exon intermediate. The second step ligates the exons together and releases the lariat intron. Splicing is carried out within the spliceosome, a large complex composed of small nuclear ribonucleoprotein particles (U1, U2, U4/U6, and U5 snRNPs; Lührmann, 1988) and other non-snRNP protein factors (reviewed in Jurica and Moore, 2003). snRNPs play important roles in both spliceosome assembly and the two catalytic steps. The various interactions between the pre-mRNA, snRNPs, and other splicing factors are crucial for precise definition of the exon/intron boundaries (reviewed in Staley and Guthrie, 1998).

Based on studies using cell extracts and *in vitro* splicing, most of the current models of spliceosome formation suggest a stepwise assembly pathway that requires ATP and substrate pre-mRNAs. The process begins with recognition of the 5' splice site by U1 snRNP and the polypyrimidine-tract by the U2 snRNP auxiliary factor (U2AF)

heterodimer, initiating the formation of a discrete early complex (E complex). Binding of U2AF to the polypyrimidine-tract recruits U2 snRNP to the branch point sequence, forming complex A which is converted to the B complex upon the addition of the U4/U6.U5 tri-snRNP. A series of dynamic RNA/protein rearrangements then converts the B complex to the C complex, the short-lived, active spliceosome (reviewed in Reed and Palandjian, 1997). Over the years, individual spliceosomal complexes have been purified and their contents analyzed by different strategies (reviewed in Jurica and Moore, 2003). The common theme linking these studies has been the demonstration that the spliceosome is a dynamic entity. However, to what extent this model reflects the *in vivo* situation is not entirely clear. In 2002, two different endogenous complexes were isolated from yeast. The first complex, the penta-snRNP from *S. cerevisiae*, contains all five U snRNPs, and, while not functionally active by itself, appears to be a bona fide splicing complex since the addition of soluble factors to the complex restores activity (Stevens et al., 2002). The second one, the Cwf/Cwc complex, was purified from both *S. cerevisiae* and *S. pombe* and appears to be similar to mammalian C complexes (Jurica et al., 2002; Ohi et al., 2002). However, the Cwf/Cwc complex can be purified from splicing deficient strains raising questions as to its functional relevance (Ohi et al., 2002). Nevertheless, the identification of this complex again raises the question as to whether pre-formed spliceosomes might exist. In mammalian cells, large ribonucleoprotein particles containing all five U snRNPs have been purified by Sperling and colleagues (Muller et al., 1998), but these particles are highly likely to represent endogenous pre-mRNP/spliceosomes. Therefore, no particle representing pre-formed splieosome

complexes has been purified from mammalian extracts so far, it remains unclear whether the pre-assembly is limited to yeast only.

Human PTB-associated splicing factor (PSF; Patton et al., 1993) is a 100 kDa protein that co-localizes with splicing factors in nuclear speckles(Dye and Patton, 2001), and is comprised of an N-terminal glycine-rich domain, a proline/glutamine (P/Q) rich domain, two RNA recognition motifs (RRMs), and a C-terminal region with two nuclear localization signals. PSF has been proposed to be essential for both the early and late steps of pre-mRNA splicing (Gozani et al., 1994; Patton et al., 1993). PSF co-purifies with the U4/U6.U5 tri-snRNP (Teigelkamp et al., 1997), possibly through its interaction with U5 snRNA stem 1b (Peng et al., 2002; also see Chapter II). Different purification strategies have detected PSF in both early (H, A, B) and late (C) spliceosomal complexes (Bennett et al., 1992; Gozani et al., 1994; Jurica et al., 2002; Makarov et al., 2002; Rappsilber et al., 2002). Besides its role in splicing, PSF has also been implicated in transcription, topoisomerase activity, nuclear RNA retention, DNA recombination, and DNA double-strand break rejoining (Akhmedov and Lopez, 2000; Bladen et al., 2005; Mathur et al., 2001; Sewer and Waterman, 2002; Sewer et al., 2002; Song et al., 2004; Straub et al., 1998; Straub et al., 2000; Urban and Bodenburg, 2002; Urban et al., 2000; Zhang and Carmichael, 2001; Zolutukhin et al., 2002). Recently, both PSF and its highly related protein $p54^{nrb}$ were shown to interact with the C-terminal domain (CTD) of RNA Pol II (Emili et al., 2002) and contact the 5' splice site within large transcription/splicing complexes (Kameoka et al., 2004), suggesting potential involvement in the coupling between transcription and splicing.

Previously, we showed that PSF and $p54^{nrb}$ interact with U5 snRNA (Peng et al., 2002; also see Chapter II). During further analysis of this interaction using immunoprecipitation (IP) by PSF-specific antibodies, all five U snRNPs were found to co-IP with PSF when HeLa nuclear extracts were adjusted to splicing conditions, regardless the addition of pre-mRNA. Sedimentation analyses confirmed the formation of large PSF-containing multi-snRNP complexes (PCC complexes) in adjusted nuclear extracts. The PCC complex contains all of the five splicing snRNPs, and has a size close to the spliceosome. Interestingly, neither exogenous pre-mRNA nor ATP hydrolysis is required for the formation of the PCC complex. Mass spectrometry analysis revealed remarkably similar protein components, including many splicing-related protein factors, between complexes that formed in the absence or presence of pre-mRNA. This suggests that pre-formed spliceosomes and/or penta-snRNPs may also exist in mammalian cells.

Materials and Methods

In vitro transcription and splicing

In vitro transcription and splicing reactions were carried out as previously described (Barnard and Patton, 2000; Michaud and Reed, 1993). The adenovirus-derived splicing substrate RNA (AdML) was linearized with BamHI and transcribed with T7 RNA polymerase (Promega), whereas the antisense RNA was linearized with EcoRI and transcribed with SP6 RNA polymerase (Promega). *In vitro* splicing reactions were carried out in HeLa nuclear extracts (Abmayr et al., 1988) at 30℃. Products were

resolved on 8M urea-15% PAGE. Splicing complexes formed at different time points were analyzed by 4% native gel electrophoresis as previousely described (Patton et al., 1993).

Formation of the PCC

HeLa nuclear extracts were adjusted to splicing conditions (12 mM Tris-HCl, pH 7.9, 60 mM KCl, 0.12 mM EDTA, 0.5 mM ATP, 20 mM phosphocreatine, 2 mM $MgCl_2$, and 0.5 mM DTT) with or without addition of *in vitro* transcribed RNAs, and incubated on ice or at 30℃ for 0-15 mins as indicated in Results.

Depletion of endogenous poly(A) RNA from HeLa nuclear extarcts

300 μl HeLa nuclear extracts were adjusted to buffer 500 (20 mM Tris-HCl, pH 7.9, 500 mM KCl, 0.2 mM EDTA, 0.5 mM DTT, and 1 mM PMSF), and incubated with 10 mg oligo-dT conjugated cellulose at room temperature for 30 min. Reactions were then centrifuged and supernatant dialyzed against buffer D (20 mM Tris-HCl, pH 7.9, 100 mM KCl, 0.2 mM EDTA, 0.5 mM DTT, 5% glycerol) for 2h at 4℃.

Immunoprecipitation

Anti-PSF-coupled protein A sepharose was prepared as previously described (Harlow and Lane, 1988). 30 μl HeLa nuclear extracts (approximately 150 μg) were adjusted to splicing conditions (see above) to allow formation of the PCC. For control, 30 μl HeLa nuclear extracts without adjustment were diluted with 20 μl buffer D to maintain the same protein concentration. After incubation at indicated temperature and

time points, reactions were diluted with 50 µl IP buffer (60% buffer D containing 1 mM PMSF and 0.01% NP40), rotated at 4℃ for 30 min, and centrifuged at full speed for 30 sec to remove any aggregates. Reactions were further diluted to 200 µl with IP buffer, and then incubated with 25 µl anti-PSF-coupled protein A sepharose (or 25 µl protein A sepharose as control) at 4℃ for 1h, followed by 3 washes with either low salt wash buffer (20 mM Tris-HCl, pH 7.9, 100 mM KCl, 0.2 mM EDTA, and 0.005% NP-40) or high salt wash buffer (20 mM Tris-HCl, pH 7.9, 250 mM KCl, 0.2 mM EDTA, and 0.005% NP-40). Immunoprecipitated RNAs were recovered from beads by phenol/$CHCl_3$ extraction and separated on 8M urea-15% PAGE. Splicing products and snRNAs were visualized by phosphorimager analysis and silver staining, respectively. For large-scale purification of the PCC for mass spectrometry analysis, 90 µl HeLa nuclear extracts and 75 µl anti-PSF-coupled protein A sepharose (or protein A sepharose as control) were used for each reaction, immunoprecipitated proteins were eluted by 8M urea-50 mM Tris-HCl, pH 8.0.

Mass spectrometry analysis

Multidimensional liquid chromatography coupled with tandem mass spectrometry (Link et al., 1999) was used to identify protein composition of the PCC as described previously (Li et al., 2003). Acquired tandem mass spectral data were searched against a human subset of the National Center for Biotechnology Information nonredundant protein database. Data processing of the SEQUEST output files was performed as previously described (Link et al., 1999).

Gradient sedimentation of the PCC and spliceosomes

HeLa nuclear extracts adjusted to splicing conditions with or without addition of radiolabeled AdML pre-mRNA were incubated at 30℃ for 15 min to allow formation of the PCC. For sedimentation analyses, 200 µl of each reaction was layered onto 10%-30% sucrose gradients (or 5%-20% glycerol gradients) containing 2 mM $MgCl_2$, 0.5 mM DTT, 60% buffer D, 1 mM PMSF, 5 mM NaF and 0.01% NP-40. After centrifugation in a Beckman SW-60 rotor at 54,000 g for 3, 4, or 12.5 h (as indicated in Results) at 4℃, fractions from the gradients were collected. RNAs were recovered from each fraction by phenol/$CHCl_3$ extraction and separated on 8M urea-15% PAGE. Splicing products and snRNAs were visualized by phosphorimager analysis and silver staining, respectively. Proteins from each fraction were separated on 10% SDS-PAGE, and transferred to nitrocellulose. Western blots were performed using either anti-PSF (Patton et al., 1993) or anti-U5-116 kDa (Fabrizio et al., 1997) antibodies.

Results

All five U snRNPs co-immunoprecipitate with PSF in HeLa nuclear extracts adjusted to splicing conditions without addition of pre-mRNA.

In Chapter II, we showed that PSF and its closely related protein $p54^{nrb}$ bind the conserved stem 1b in U5 snRNA and that both proteins, together with other U5-specific factors, assemble onto biotinylated U5 snRNAs in Hela nuclear extracts under splicing conditions (12 mM Tris-HCl, pH 7.9, 60 mM KCl, 0.12 mM EDTA, 0.5 mM ATP, 20 mM phosphocreatine, 2 mM $MgCl_2$, and 0.5 mM DTT; Peng et al., 2002). When we utilized PSF-specific antibodies to further analyze its interaction with U5, we found that

all five snRNPs could be immunoprecipitated (Fig. 14A, lane 4). U1, U4, U5, and U6 were present in the IP pellet in approximately stoichiometric amounts while slightly more U2 was detected. In this experiment (shown in Fig. 14A, lane 4), immunoprecipitation was performed on reactions containing AdML pre-mRNA under splicing conditions. The ability to co-IP splicing substrates and intermediates with PSF (Fig. 14B) is consistent with co-sedimentation of PSF with splicing complexes (Chapter II Fig. 13). However, to our surprise, similar if not identical results were also obtained when immunoprecipitation reactions were performed under splicing conditions but without the addition of AdML pre-mRNA (Fig. 14A, compare lanes 3 and 4). In contrast, only background levels of U1 and U2 snRNPs were pulled down by PSF antibodies in nuclear extracts that were not adjusted to splicing conditions (compare lanes 1 and 2). It appears that when nuclear extracts are adjusted to splicing conditions, PSF associates with all snRNPs, regardless of the presence of exogenous pre-mRNA.

Formation of PSF-containing multi-snRNP complexes in HeLa nuclear extracts under splicing conditions without addition of pre-mRNA.

To determine whether PSF interacts with individual snRNPs or is part of a larger complex containing all 5 snRNPs, sedimentation analyses were performed. HeLa nuclear extracts or extracts adjusted to splicing conditions without addition of pre-mRNA were incubated at 30℃ for 15 minutes and then separated on 10%-30% sucrose gradients (54000g, 3h). In Hela extracts, both PSF and the U5 snRNP specific protein U5-116kDa sedimented near the top of the gradient (Fig. 15A, fractions 5 and above) with no association of PSF in complexes larger than the U4/U6.U5 tri-snPNP (fractions 4-5). However, when nuclear extracts were adjusted to splicing conditions minus

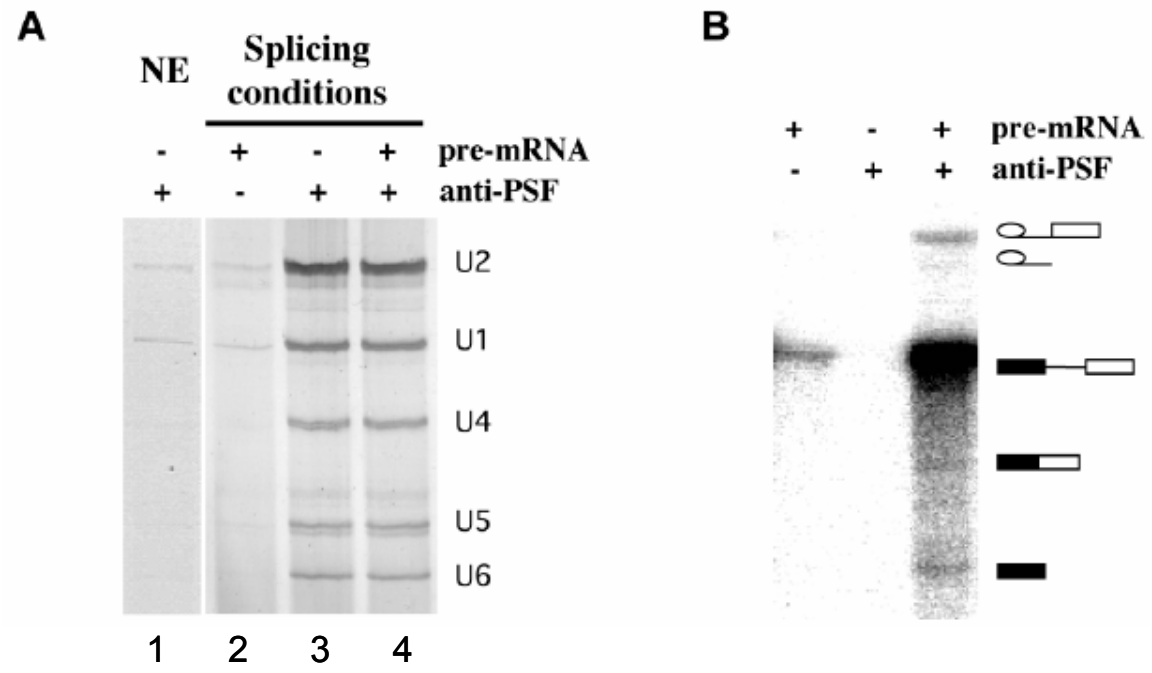


Figure 14. All of the five splicing snRNPs co-immunoprecipitate with PSF in HeLa nuclear extracts adjusted to splicing conditions without addition of pre-mRNA. HeLa nuclear extracts (NE) and nuclear extract adjusted to splicing conditions with (+) or without (-) addition of AdML pre-mRNA were incubated at 30℃ for 15 mins, and then subjected to immunoprecipitation by anti-PSF-coupled protein A sepharose. RNAs were recovered from IP pellets by phenol/$CHCl_3$ extraction and separated on 8M urea-15% PAGE. U snRNAs and splicing products were visualized by silver staining **(A)** and PhosphorImage analysis **(B)**, respectively.

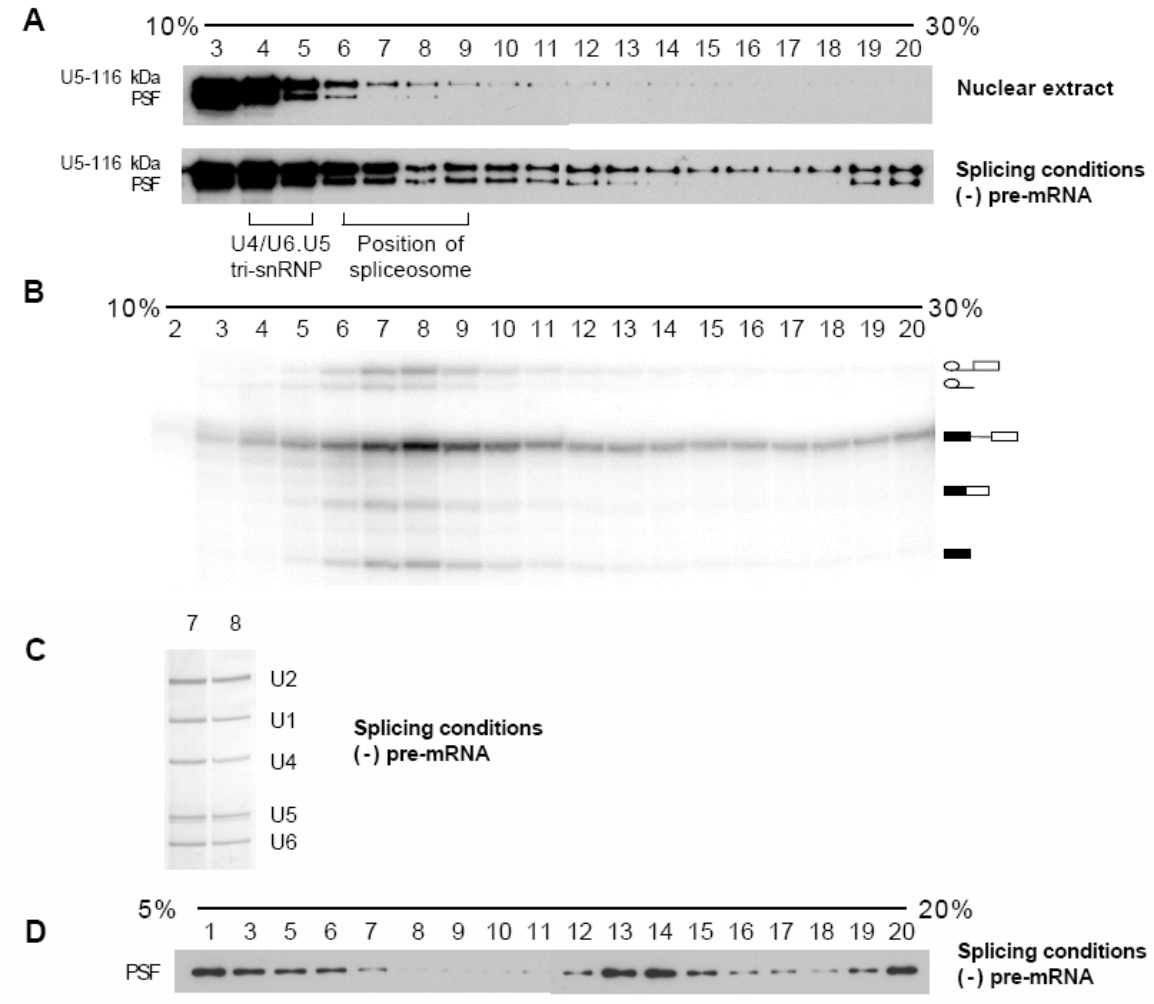


Figure 15. Formation of PSF-containing multi-snRNP complexes in adjusted HeLa nuclear extracts without addition of pre-mRNA. **(A)** HeLa nuclear extracts (with or without adjustment, as indicated) were incubated at 30℃ for 15 mins, and then separated on 10%-30% sucrose gradients at 20k rpm for 3h. Proteins from each fraction were analyzed by Western Blot analysis using different antibodies as indicated. Positions of U4/U6.U5 tri-snRNP and spliceosomes were indicated. **(B)** Splicing reactions using AdML pre-mRNA were incubated at 30℃ for 15 mins, and then separated on a parallel sucrose gradient. RNAs were recovered from each fraction by phenol/$CHCl_3$ extraction and separated on 8M urea-15% PAGE. Splicing products were visualized by PhosphorImage analysis. **(C)** RNAs from (A) (splicing conditions no pre-mRNA, fractions 7-8) were recovered as in (B), and visualized by silver staining. **(D)** Adjusted nuclear extracts as in (A) were incubated at 30℃ for 15 mins, followed by separation on a 5%-20% glycerol gradient at 20k rpm for 4h. Proteins from each fraction were analyzed by Western Blot analysis using PSF antibodies.

pre-mRNA substrate, the distribution of PSF and U5-116kDa changed. In addition to free protein at the top of the gradient, a fraction of both PSF and U5-116kDa sedimented across the gradient. The heaviest fraction pelleted to the bottom of the gradient and did so even when centrifugation speeds and times were significantly shortened, suggesting the formation of nonspecific aggregates (data not shown). However, fractions 6-11 contained PSF in complexes similar in size to mammalian spliceosomes (indicated by the peak of ADML pre-mRNA, mRNA, and splicing intermediates in parallel gradients, Fig. 15B, fractions 6-9). These fractions contain stoichiometric amounts of all 5 snRNPs (Fig. 15C), consistent with the IP results above. That these fractions represent a distinct complex was determined by further analysis using different sedimentation conditions which showed three distinct peaks (Fig. 15D): free protein and that associated with the tri-snRNP, a lighter fraction (fractions 13-14) we are naming PCC (PSF-Containing Complex), and a nonspecific aggregate. The PCC complex sediments at approximately 60S, similar in size to mammalian spliceosomes (Fig. 16).

Formation of the PCC complex is ATP-independent.

We next used immunoprecipitation experiments to further analyze the formation of the PCC complex in the presence or absence of pre-mRNA and at different time points. As determined by the presence of all 5 snRNPs, we could detect the formation of the PCC regardless of the presence or absence of pre-mRNA substrate or a negative control antisense RNA transcript. Interestingly, we could IP all 5 snRNPs even when reactions were assembled and kept on ice (time 0) and the amount and composition of the U snRNPs remained unchanged during continued incubation at 30℃ for 5-15 minutes (Fig.

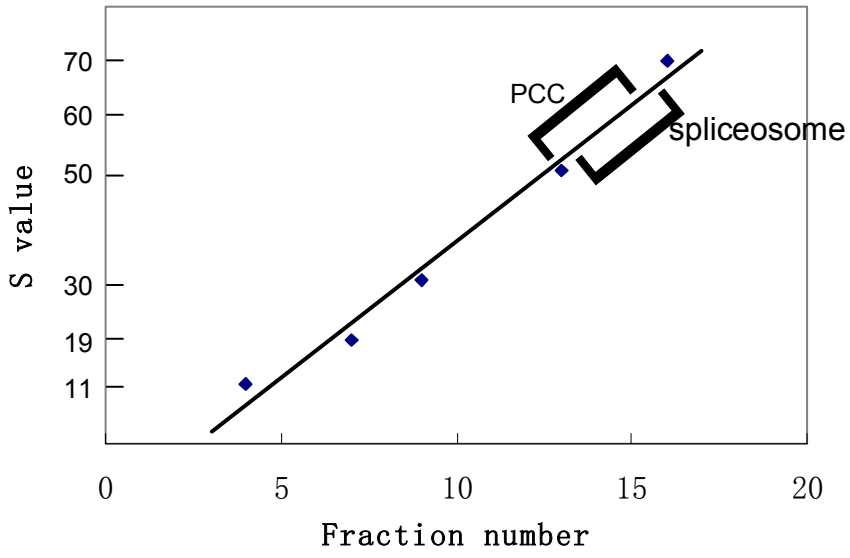


Figure 16. The PCC has a size around 60S. HeLa nuclear extracts were adjusted to splicing conditions with or without AdML pre-mRNA and incubated at 30℃ for 15 mins, followed by separation on 10%-30% sucrose gradients at 20k rpm for 12.5h. Size standards (catalase, 11S; thyroglobulin, 19S; small subunit of *E.coli* ribosome, 30S; large subunit of *E.coli* ribosome, 50S; 70S *E.coli* ribosome) were separated on parallel gradients.

17A). As there is no ATP hydrolysis in reactions kept on ice, the formation of the complex is ATP-independent.

For reactions assembled in the presence of pre-mRNA substrate, splicing did not affect the ability to IP all 5 snRNPs (Fig. 17B, C). For comparison, parallel splicing reactions were incubated at 30℃ and splicing complexes were analyzed on 4% native gels. As expected, H, A, and B/C complexes were readily detected at different time points, however, the pattern of snRNP IP with PSF antibodies across the same time points was the same. This raises the possibility that the PCC could be unrelated to splicing, that the PCC complex represents endogenous pre-mRNPs, or that the PCC exists as a pre-formed particle which then assembles onto AdML pre-mRNAs and functions in splicing.

The PCC complex forms in HeLa nuclear extracts depleted of endogenous polyA RNAs.

To address whether a pool of endogenous pre-mRNA is responsible for the complexes we observe, we treated HeLa nuclear extracts with oligo-dT conjugated beads to deplete any endogenous poly(A) RNAs. Formation of the PCC complex in depleted nuclear extracts was then analyzed by IP using PSF antibodies under splicing conditions. As shown in Fig. 18, all 5 snRNPs could be immunoprecipitated in poly(A)-depleted extracts, regardless of the presence or absence of subsequently added pre-mRNA substrate (Fig. 18A, lanes 4 and 5). To ensure that such depletion did not abolish splicing activity, we performed *in vitro* splicing in normal and depleted extracts (Fig. 18B). The efficiency of splicing was a bit less in treated extracts but this was consistent with the fact that the efficiency of snRNP immunoprecipitation was also slightly less in

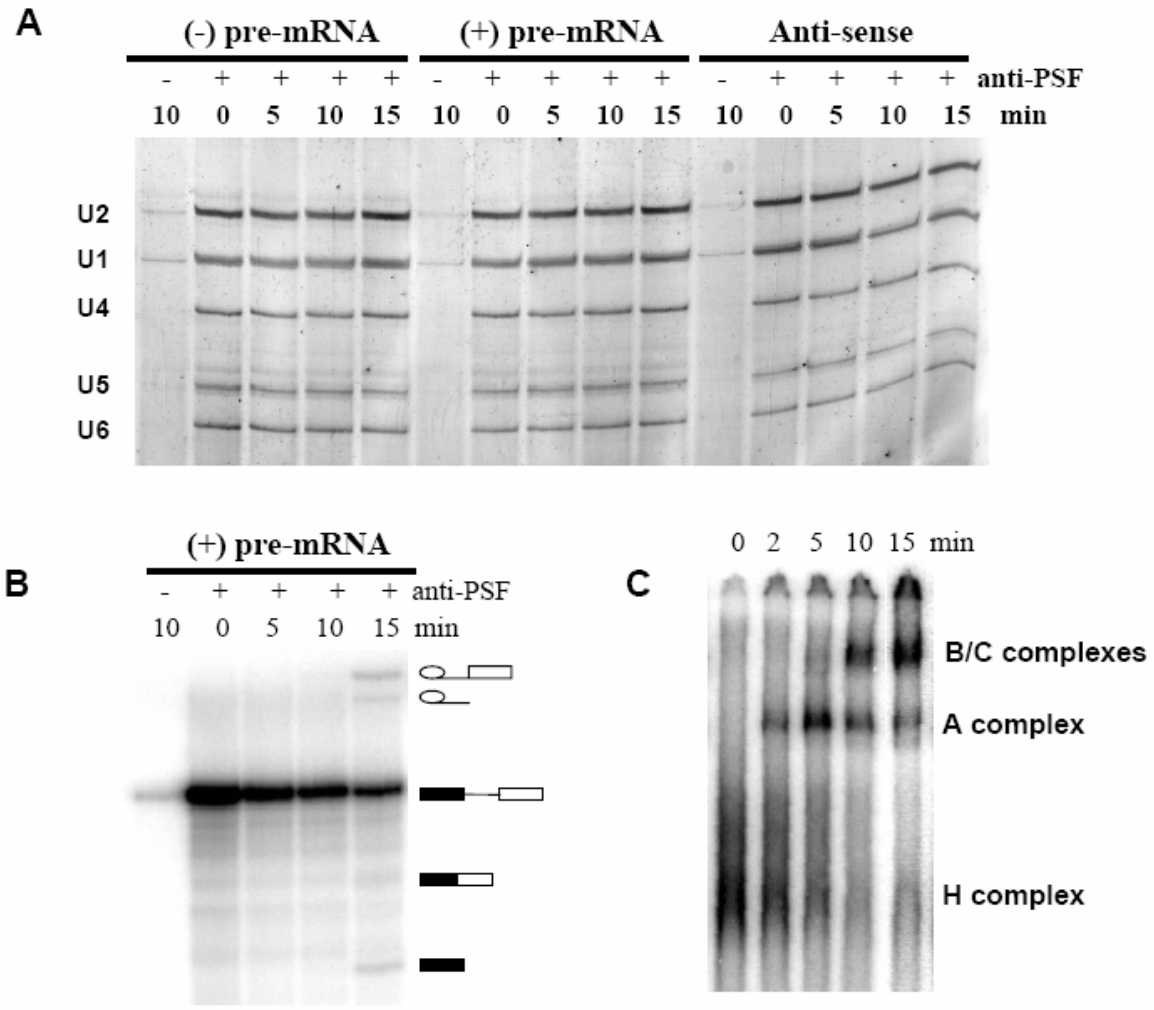


Figure 17. Formation of the PCC is ATP-independent. HeLa nuclear extracts adjusted to splicing conditions with (+) or without (-) addition of AdML pre-mRNA, or with anti-sense AdML (anti-sense), and nuclear extract (NE) were incubated at 30℃ for 0, 5, 10, 15 mins as indicated, and then subjected to immunoprecipitation by anti-PSF-coupled protein A sepharose. RNAs were recovered from IP pellets by phenol/$CHCl_3$ extraction and separated on 8M urea-15% PAGE. **(A)** U snRNAs were visualized by silver staining. **(B)** Splicing products from (+) pre-mRNA reactions (splicing reactions) were visualized by PhosphorImage analysis. **(C)** Parallel splicing reactions were incubated at 30℃, and splicing complexes formed at different time points (0, 5, 10, 15 mins as indicated) were analyzed by 4% native gel electrophoresis.

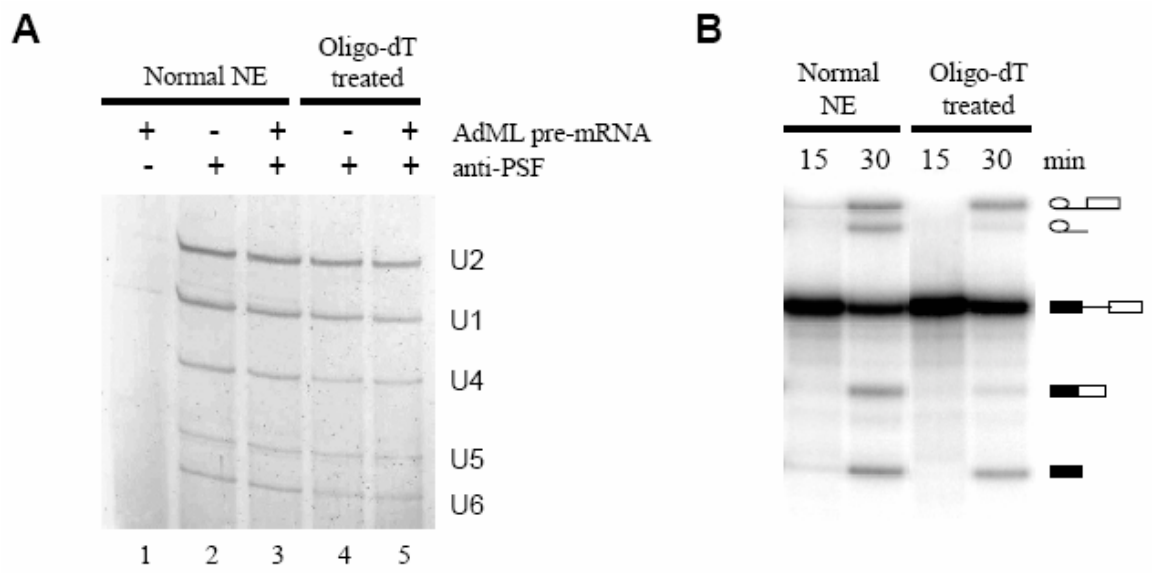


Figure 18. Depletion of endogenous poly(A) RNA did not affect the formation of the PCC. **(A)** Poly(A) RNA-depleted HeLa nuclear extracts (by treatment with oligo-dT conjugated cellulose) and normal extracts were adjusted to splicing conditions with or without addition of AdML pre-mRNA. Reactions were incubated at 30℃ for 15 mins before immunoprecipitation by anti-PSF-coupled protein A sepharose. RNAs were recovered from IP pellets by phenol/$CHCl_3$ extraction, separated on 8M urea-15% PAGE, and visualized by silver staining. **(B)** Splicing reactions using radiolabeled AdML pre-mRNA were carried out in either poly(A)-depleted or normal nuclear extracts for 15 and 30 min. RNAs were recovered as in (A), and visualized by PhosphorImage analysis.

the treated extracts. Nevertheless, the depletion of poly(A) RNA from nuclear extracts did not significantly affect the formation of the PCC complex with the caveat that low levels of endogenous pre-mRNA may still persist in the treated extracts.

The protein contents of PCC complexes formed in the absence or presence of pre-mRNA are remarkably similar.

The presence of five U snRNPs in the PCC complex suggests the possibility that it might be a "mammalian penta-snRNP". Just like the yeast penta-snRNP, the PCC complex is less stable than other splicing-related complexes that have been purified. The PCC can withstand up to 100 mM KCl washes and completely dissociates in 250 mM KCl (data not shown). Despite this, we decided to utilize multidimensional liquid chromatography coupled with tandem mass spectrometry (Link et al., 1999) to compare the composition of proteins immunoprecipitated with anti-PSF antibodies under splicing conditions and in the absence or presence of pre-mRNA substrate. Samples from mock IP (no PSF antibodies) and from IP of standard HeLa nuclear extracts were also included as controls. Not surprisingly, given the relatively nonstringent conditions, many proteins were detected in the IP pellets. To facilitate analysis, proteins detected in the mock IP were subtracted as nonspecific contaminants. Also, since the PCC complex does not form in standard nuclear extracts and dissociates in 250 mM KCl, proteins identified in IP pellets from nuclear extracts and from high salt wash conditions were also subtracted as non-complex-related proteins even though this resulted in the subtraction of more than ten hnRNPs, several known PSF interacting factors (such as matrin 3), heat shock proteins, 40S ribosomal proteins, and a variety of other proteins whose role in splicing or RNA processing is unknown.

From mass spectrometry, the immunoprecipitated proteins are summarized in Table 2, grouped into seven classes: snRNP proteins (U snRNP-specific proteins and Sm proteins), splicing factors, hnRNPs, RNA unwindases, transcription factors, Ser/Thr kinases, and factors involved in other RNA processing events (mRNA stability, mRNA export, mRNA 3'-processing, etc). Among the proteins, 70% are related to pre-mRNA splicing and the majority of these have also been identified in other purified splicing complexes, such as human prespliceosomes isolated by Luhrmann and colleagues (Hartmuth et al., 2002). A remarkable concordance was observed between the pattern of proteins in complexes formed in the presence or absence of pre-mRNA substrate. Several snRNP proteins, non-snRNP splicing factors, SR proteins, hnRNP proteins, and other RNA binding proteins involved in mRNA export and 3'-processing were detected in both complexes. Among four RNA unwindases identified, DDX9 (RNA Helicase A) and UAP56, which function in both spliceosome assembly and mRNA export (Fleckner et al., 1997; Luo et al., 2001) associated with only the (+) pre-mRNA preparation. Three Ser/Thr kinases were found only in the PCC complex formed in the absence of pre-mRNA, two of them, CK II and CaMK II, have been implicated in spliceosome assembly and pre-mRNA splicing (Parker and Steitz, 1997; Trembley et al., 2005).

From the above, we have shown that the PCC complex contains all of the five splicing snRNPs and other non-snRNP splicing factors. Considering that nearly 70% of the PCC proteins are known spliceosome components, or proteins related to splicing, it is reasonable to speculate that the PCC complex might be functionally related to splicing. The fact that neither the presence of exogenous pre-mRNA substrate nor ATP hydrolysis is required for its formation suggests that the PCC most closely resembles the yeast

Table 2. Proteins identified in PCC complexes. Proteins identified in PCC complexes that formed in the absence (-) and presence (+) of AdML pre-mRNA as listed. ●: proteins found only in the PCC; ○: proteins also present in IP pellet from nuclear extracts or high salt wash (see text for details).

Category/function	Protein name	(-) pre-mRNA	(+) pre-mRNA
snRNP proteins			
sm proteins	B/B'	○	○
	D1	○	○
	D2	○	○
	D3	○	
	E		●
	F	○	○
	G	●	
U1 snRNP	U1 70K	●	●
	U1 C		○
U2 snRNP	U2 A'	●	●
	U2 B''	●	
	SF3a120	●	●
	SF3a66		●
	Sf3a60		●
	SF3b150	●	●
	SF3b130	●	○
	SF3b50	●	●
	SF3b10	●	●
	SF3b14b	○	○
U4/U6 snRNP	hPrp4		●
U5 snRNP	hPrp8	●	
	U5-200		●
Splicing factors			
	PSF	○	○
	$p54^{nrb}$	○	○
	YB-1	●	
	SRm300	○	○
	TLS/FUS	○	○
	FBP-2	○	○
core SR proteins	ASF/SF2	●	●
	SRp20	○	
	9G8	●	●
hnRNPs			
	hnRNP C1/C2	●	●

Table 2, continued

Category/function	Protein name	(-) pre-mRNA	(+) pre-mRNA
	hnRNP G	●	●
	hnRNP H3	●	●
	hnRNP L	●	●
	hnRNP M	●	●
	hnRNP-homolog JKTBP	●	●
	Similar to hnRNP A3	●	●
	type A/B hnRNP		●
RNA unwindases			
	RH II/Gu	●	●
	DDX9		●
	UAP56		●
	p72	○	○
Transcription factors			
	TF II D-70	●	
	TIP49a		●
	TIP21		●
	mMED8		●
	ILF2/NF45	●	
	TLE3		●
	FBP-3		●
	TDP-43	●	
	TAT-SF1		●
	Pol III	●	
	Pol III C11	●	
Other RNA processing factors			
mRNA-decay	HuR	●	●
export?	NF90(ILF3)	●	●
export	Aly	○	
3'-processing	CFIm, 25 kDa subunit	●	●
	CFIm, 59 kDa subunit		●
snRNP	SMN	●	
Ser/Thr kinases			
	CK II	●	
	CaMK-II beta subunit	●	
	STK9	●	

penta-snRNP. Unfortunately, a variety of functional tests have proven intractable. Nevertheless, the discovery of a potential mammalian penta-snRNP and the intriguing possibility that it might represent a pre-formed spliceosome raises interesting questions for splicing regulation, especially control of alternative splicing.

Discussion

The PCC complex is a multi-snRNP complex.

We have shown that the PCC complex, a PSF-containing multi-snRNP complex, forms in HeLa nuclear extracts under splicing conditions. It has a size similar to that of the spliceosome, and contains all of the five splicing snRNPs as well as other non-snRNP splicing factors. Both sucrose gradient sedimentation analysis and co-IP experiments using PSF-specific antibodies demonstrated that neither ATP nor exogenous pre-mRNA is required for formation of the complex. Approximately 70 protein factors were identified as components of the PCC complex, and 70% of them are either known spliceosomal components or functionally related to splicing. Though similar, the PCC complex is clearly different from fully-assembled/active spliceosomes in two ways. First, assembly of active spliceosomes requires ATP hydrolysis. Second, many spliceosomal proteins, including cap binding proteins (CBP) and essential splicing factors such as U2AF65, second step factors Prp16, Prp17, and the recently characterized Prp19 complex proteins (Ajuh et al., 2000), were not identified in the particle. It is possible that some of these protein factors are lost during purification while others are not readily identified by mass spectrometry due to technical difficulties.

Spliceosomal complexes and the PCC

During the past two decades, many splicing complexes and splicing related particles have been studied by several groups (reviewed in Jurica and Moore, 2003). As discussed above, the presence of multiple snRNPs and the protein contents of the PCC complex make it distinct from other spliceosomal complexes (E, A, B, C complex, yeast Cwf/Cwc complexes, and the so called "super-spliceosome", lnRNP; Hartmuth et al., 2002; Jurica et al., 2002; Makarov et al., 2002; Michaud and Reed, 1991; Muller et al., 1998; Ohi et al., 2002). In 1988, Konarska and Sharp reported the formation of a pseudospliceosome complex in HeLa nuclear extracts in the absence of pre-mRNA (Konarska and Sharp, 1988). This complex, which contains U2, U4, U5, and U6 snRNPs, formed under high salt condition (250 mM salt), and was undetectable under the low salt conditions required for the PCC complex, thus, it is unlikely to be a part of the PCC complex. The most comparable splicing related complex to the PCC complex would be the recently described yeast penta-snRNP (Stevens et al., 2002). Both form under low salt conditions (50 mM NaCl for the penta-snRNP, and 60 mM KCl for the PCC complex) in the absence of pre-mRNA, contain all five splicing snRNPs, and lack some of the known spliceosomal protein factors. However, the PCC complex also contains a large number of hnRNPs as well as other RNA processing factors involved in transcription, mRNA 3' processing, and mRNA export, making it a distinct multi-snRNP complex.

Is the PCC complex functional?

A key unanswered question is whether the PCC functions during splicing. To address this, concentrated PCC complexes with high purity will be required for complementation experiments or direct functional analyses. A variety of purification strategies have been used, unfortunately, none were able to answer the question for a variety of technical reasons. First, complexes immunopurified using antibodies against either full length PSF or a short peptide were inefficiently immunoprecipitated and difficult to elute with PSF peptides (data not shown). Second, unlike the yeast penta-snRNP, the PCC complex is not very stable, and starts to disassociate while re-sedimentating in sucrose (or glycerol), resulting in poor separation from smaller complexes (see Fig. 15 and data not shown). Future work will be directed toward overcoming these obstacles to determine functionality. However, even if successful, interpreting the implications of such experiments will still be challenging. The composition of the PCC complex suggests it functions in pre-mRNA splicing but it is possible that it could also function in other nuclear events, consistent with proposed roles for PSF in a variety of pathways. For any of these, loss of one or more essential factors could render the complex nonfunctional. If supplied with soluble factors or other proteins fractions, for example, micrococcal nuclease (MN) treated nuclear extracts, functionality might or might not be restored depending on a variety of variables. Even in case of restored functionality, there still exists a possibility that the complex might not function as an intact/pre-formed particle.

Does pre-assembly also exist in mammals?

Three years ago, the discovery of the yeast penta-snRNP suggested that not all spliceosomes assemble de novo in a stepwise manner (Stevens et al., 2002). The similarity between the PCC complex and the yeast penta-snRNP opens the possibility that mammalian penta-snRNPs or preformed-spliceosomes may also exist in mammalian cells. This is particularly interesting when considering alternative splicing. The most straightforward models of splicing regulation posit that a variety of RNA elements and factors act combinatorially to precisely control splice site selection. Should pre-formed spliceosomes exist, one could envision two classes of introns. The first would be constitutively spliced introns with strong splicing signals and little cell- or tissue-specific regulation. Such introns could be spliced using a largely pre-formed machinery. In contrast, highly regulated splicing events might utilize a step-wise pathway allowing for precise control. Considering the high frequency of alternative splicing events in mammalian cells, it is possible that both pathways exist.

CHAPTER IV

SUMMARY AND CONCLUSIONS

Pre-mRNA splicing is a delicate process involving multiple steps in spliceosome assembly and splice site choice followed by accurate joining of exons. While many proteins have been identified as essential splicing factors, the exact functions of most of these factors are poorly understood. Among these factors is PSF which has been identified as a spliceosomal protein under a variety of purification conditions (Chapter I, also see Table 1) and plays an important role in early spliceosome assembly and the second step of splicing (Gozani et al., 1994; Patton et al., 1993). More recent evidence revealed its association with both the 5' splice site (Kameoka et al., 2004) and the U4/U6.U5 tri-snRNP (Teigelkamp et al., 1997) but exactly where and when it functions during splicing remains unclear.

Here, experiments were performed to better understand the RNA binding specificity of PSF and identify potential protein-protein interaction partners. The finding that PSF is found within a large complex containing all 5 snRNPs raises the interesting question as to whether pre-formed spliceosomes exist in mammalian cells. Such a finding would dramatically alter how spliceosome assembly is viewed, whether stepwise, largely pre-formed, or both. Future functional experiments will be needed to address this interesting question along with precisely defining the role that PSF plays in splicing.

PSF and p54^{nrb} Associate with U5 SnRNA during Splicing.

PSF and p54^{nrb} bind a conserved stem in U5 snRNA.

The high degree of conservation of U5 stem 1b in vertebrates and flies (Fig. 9D) implies that it plays a critical role in U5 function. Interestingly, we have demonstrated that PSF interacts with its highly related protein p54^{nrb} and that both of these proteins, individually and in combination, select the same purine-rich sequence located on the 3' side of U5 snRNA stem 1b as their optimal RNA binding sites (Chapter II). Binding of PSF and p54^{nrb} to U5 was first verified by gel shift assays, and then analyzed by filter binding assays as well as RNA affinity selection experiments. Binding specificity for both proteins appears to depend on multiple elements, including the sequence and structure of stem 1b. Furthermore, sedimentation analyses confirmed that both proteins associate with spliceosomes and with U4/U6.U5 tri-snPNP.

Unlike higher eukaryotes, the PSF-p54^{nrb} binding site is not present in the corresponding regions of yeast U5 snRNA (Fig. 9D). This lack of conservation, together with the apparent absence of PSF and p54^{nrb} homologs in yeast, suggests either that yeast U5 functions differently from that of higher eukaryotes, or that yeast contain an as yet unidentified functional homolog of PSF and p54^{nrb}. Nevertheless, this intriguing correlation indicates functional significance for the interaction between PSF-p54^{nrb} and U5 snRNA in higher eukaryotes.

Model for PSF-p54^{nrb} in potential U5 rearrangement

Since double-stranded RNAs adopt the A-form conformation which precludes base-specific interaction with protein side chains in the deep major groove (Steitz, 1999), the question arises as to how PSF and p54^{nrb} interact with an A-form helix while apparently maintaining sequence specificity. Two ATPases (U5-100 kDa and U5-200 kDa) and one homolog of EF-2 GTPase (U5-116 kDa) have been found to associate with U5 snRNP. U5-100kDa and its yeast homolog Prp28 have been proposed to function in disassociation of U1 from the 5' splice site (Chen et al., 2001; Staley and Guthrie, 1999), and U5-200kDa/Brr2 has been implicated to facilitate U4/U6 unwinding (Laggerbauer et al., 1998; Raghunathan and Guthrie, 1998). Interestingly, IL2, the internal loop between stem 1b and 1c, is required for efficient association of hPrp8 and U5-116 kDa with U5 (Hinz et al., 1996; Ségault et al., 1999), while hPrp8 forms a stable complex with U5-200kDa, U5-116kDa, and U5-40kDa (Achsel et al., 1998). Given the proximity between stem 1b and at least two putative RNA unwindases (U5-100 and U5-200), it seems reasonable to propose that stem 1b of U5 might also undergo a conformational change during spliceosome assembly or during the two catalytic steps. Therefore, we speculate that PSF and p54^{nrb} could bind U5 in two ways: to the intact stem, or to the 3' side of the stem after melting.

As to how PSF and p54^{nrb} might function in this process, there are at least two possibilities (Fig. 19). First, PSF and p54^{nrb} may initially bind to double stranded stem 1b but then bind more tightly to the 3' side of the stem upon unwinding. Such binding could stabilize the unpaired strands coincident with U4/U6.U5 tri-snRNP contacting the 5' splice site, or perhaps during the second step of splicing. Second, given that PSF has

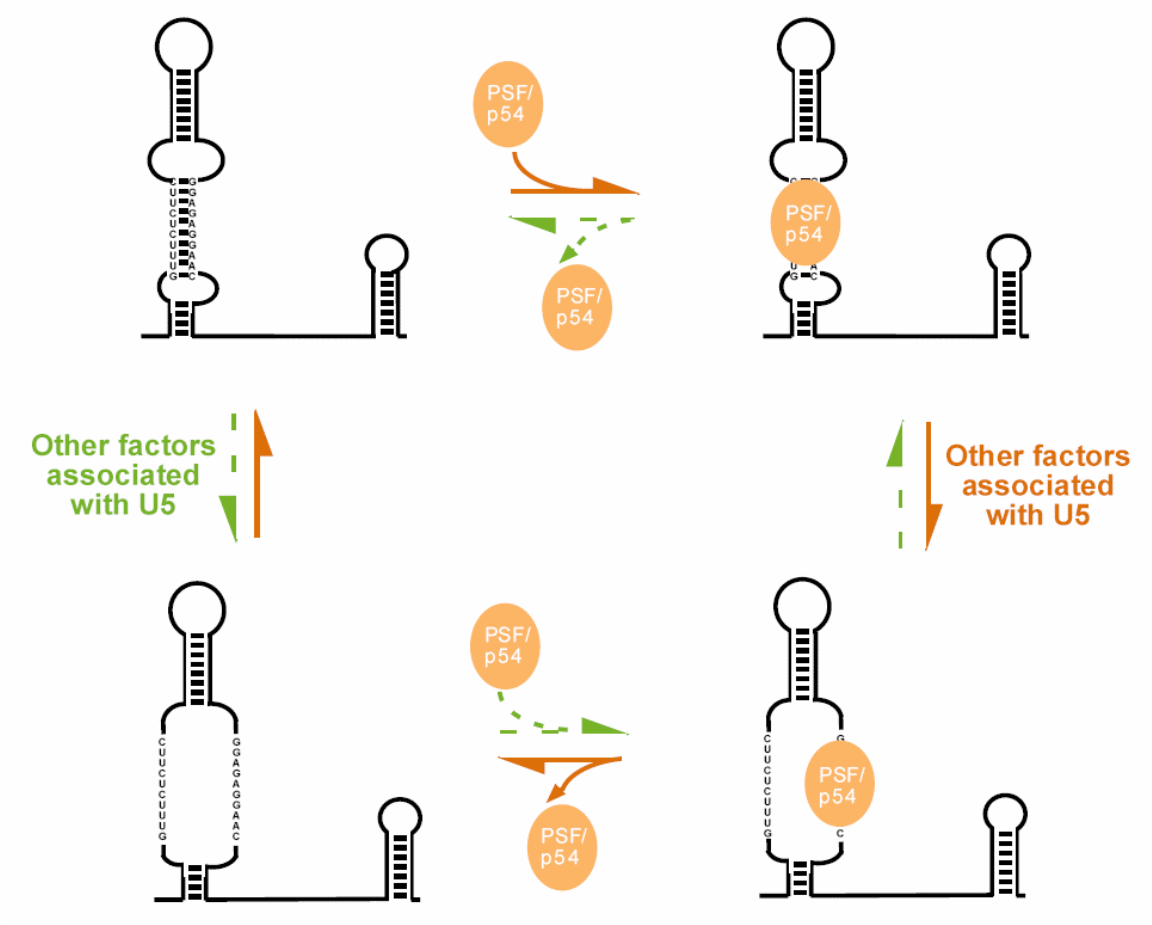


Figure 19. Model for PSF-p54^{nrb} in potential U5 rearrangement. During the spliceosome assembly or during the transition from the first catalytic step to the second step of splicing, association of PSF-p54^{nrb} with U5 snRNA may affect U5 structure in two possible ways. As shown in clockwise, after PSF-p54^{nrb} binds the double-stranded stem 1b of U5 snRNA, other protein factors associated with U5 (such as an RNA unwindase) unwind stem 1b. PSF-p54^{nrb} now binds the 3' side of the stem and helps to stabilize the opened U5 structure. Later disassociation of PSF-p54^{nrb} from U5 results in reconstruction of the original stem structure. Or, as shown in counter clockwise, stem 1b is first unwound by other protein factors such as an RNA unwindase associated with U5, while binding of PSF-p54^{nrb} to the 3' side of the stem induces reconstruction of the double-stranded stem. PSF-p54^{nrb} then disassociates, leaving U5 ready for next round of splicing.

been demonstrated to promote the annealing of DNA (Akhmedov and Lopez, 2000), PSF and $p54^{nrb}$ could facilitate the reannealing of U5 stem 1b subsequent to melting, mediated by one or more unwindases. Either way, the interaction could be transient, supported by non-stoichiometric association of PSF with the U4/U6.U5 tri-snRNP (Teigelkamp et al., 1997) and the fact that only a small amount of PSF and $p54^{nrb}$ co-sediments with U4/U6.U5 tri-snRNP and with splicing complexes (Chapter II, Fig. 13).

PSF as a Part of a Mammalian Multi-snRNP Complex

During our studies of PSF/U5 interactions, we discovered the PCC, a PSF-containing multi-snRNP complex in HeLa nuclear extracts adjusted to splicing conditions. The PCC contains all of the five splicing snRNPs and has a size similar to that of the spliceosome. Approximately 70 protein factors were identified as components of the PCC complex and 70% of them are either known spliceosomal components or functionally related to splicing (see Chapter III).

The PCC is a distinct multi-snRNP complex.

Both the presence of multiple snRNPs and the protein contents of the PCC make it distinct from other *in vitro* assembled spliceosomal complexes (E, A, B, C complex) and the endogenous Cwf/Cwc complexes from yeast (reviewed in Jurica and Moore, 2003) (compare Table 2 with Table 1). In addition, the low salt resistance of the PCC (100 mM KCl) suggests that it might not be related to the pseudospliceosome complex detected in HeLa nuclear extracts which contains U2, U4, U5, and U6 snRNPs but forms only under high salt conditions (250 mM salt; Konarska and Sharp, 1988). Furthermore,

the PCC is apparently different from fully-assembled or active endogenous spliceosomes in two ways. First, assembly of active spliceosomes requires ATP hydrolysis whereas neither ATP nor exogenous pre-mRNA is required for the formation of the PCC. Second, many spliceosomal proteins, including cap binding proteins (CBP) and essential splicing factors such as U2AF65, second step factors Prp16, Prp17, and the recently characterized Prp19 complex proteins, were not identified in the PCC.

The most comparable splicing related complex to the PCC would be the recently described yeast penta-snRNP (Stevens et al., 2002). Though different in that the PCC contains a large number of hnRNPs and other RNA processing factors involved in transcription, mRNA 3'-processing, and mRNA export, these two complexes clearly resemble each other in several ways. Both form under low salt conditions (50 mM NaCl for the penta-snRNP, and 60 mM KCl for the PCC complex) in the absence of pre-mRNA, contain all five splicing snRNPs, and lack several known spliceosomal protein factors.

Does pre-assembly also exist in mammals?

Given the similarity between the PCC and the yeast penta-snRNP, the discovery of which suggested that not all spliceosomes assemble de novo in a stepwise manner (Stevens et al., 2002), it is possible that mammalian penta-snRNPs or preformed-spliceosomes may also exist in mammalian cells. This is particularly interesting when considering alternative splicing. The most straightforward models of splicing regulation posit that a variety of RNA elements and factors act combinatorially to precisely control splice site selection. Should pre-formed spliceosomes exist, one

could envision two classes of introns. The first would be constitutively spliced introns with strong splicing signals and little cell- or tissue-specific regulation. Such introns could be spliced using a largely pre-formed machinery for efficient splicing. In contrast, highly regulated splicing events might utilize a step-wise pathway or a semi-pre-formed pathway allowing for precise control and multiple steps along the pathway. In the semi-pre-formed model originally proposed by Abelson and colleagues, U1 snRNP first recognizes the 5' splice site after which the U2/U4/U6/U5 tetra-snRNP is recruited to the branch point/3' splice site (Stevens et al., 2002). In the eukaryotic system, one can imagine splice site choice being regulated in a similar manner (Fig. 20). Considering the high frequency of alternative splicing events in mammalian cells, it is possible that both pathways exist. In support of this, mammalian complexes resembling both the yeast penta- and tetra- snRNPs have been discovered: the PCC complex and the U2/U4/U6/U5 pseudospliceosome complex by Konarska and Sharp (Konarska and Sharp, 1988), respectively. Unfortunately, neither has been proven functional in splicing so future work will be required to address this challenging but fascinating hypothesis.

A PSF coupled network?

Other than its role in pre-mRNA splicing, PSF has been proposed to function in multiple nuclear events including transcription, and mRNA retention/export, two processes known to be coupled to splicing (Reed, 2003; also see Chapter I). Interestingly, besides snRNPs and splicing factors, protein factors involved in transcription, 3'-processing, and mRNA export were also identified in the PCC, such as TAT-SF1, which has been demonstrated to couple transcription and splicing, and

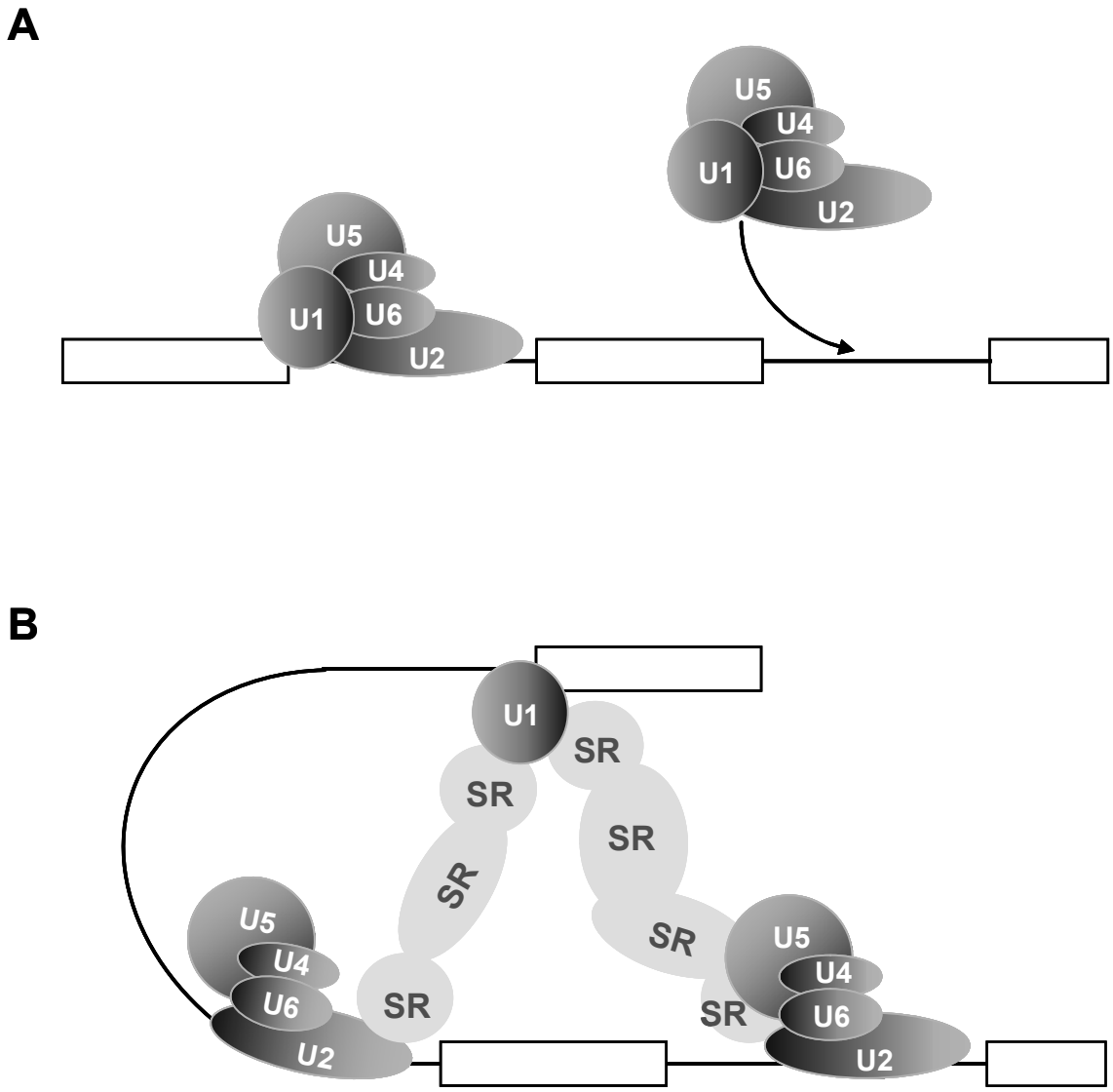


Figure 20. Model for potential penta-snRNP and tetra-snRNP in mammalian pre-mRNA splicing. **(A)** Constitutively spliced introns are recognized by penta-snRNPs in a pre-formed manner. **(B)** One example of how tetra-snRNPs may function in cell- or tissue-specific regulation of alternative splicing. 5' and 3' splice sites are recognized by U1 snRNP and tetra-snRNP, respectively. Splice site choice is therefore regulated by different sets of splicing factors that bridge U1 with tetra-snRNPs.

Aly/REF, a component of the EJC complex that couples splicing to NMD and mRNA export (Chapter I and III). Does the PCC simply represent those coupled nuclear machineries described above? Or does PSF serve to link the various processes in some way? We do not have an answer as yet, but at least two more lines of evidence suggest that PSF might be involved in coupling transcription and splicing. First, PSF and $p54^{nrb}$ were found in a large mammalian transcription/splicing complex (Kameoka et al., 2004) similar to the yeast TREX complex (Strasser et al., 2002). Second, both PSF and $p54^{nrb}$ were shown to interact with the CTD of RNA Pol II (Emili et al., 2002), an important player in coupling transcription and splicing.

As to the potential role(s) of PSF in this process, it is possible that PSF might function like TAT-SF1, which serves as a transcription coregulator and recruits splicing factors to the nascent pre-mRNA (Fong and Zhou, 2001). Given the fact that PSF interacts with both the 5' splice site (Kameoka et al., 2004) and U1A (Lutz et al., 1998), it may directly contact the 5' splice site and/or help recruit U1 after transcription initiation through its interaction with Pol II CTD thereby bringing splice sites into the proximity of splicing factors associated with Pol II CTD.

In summary, this work has demonstrated that PSF interacts with its highly related protein $p54^{nrb}$ and that both bind the conserved stem 1b of U5 snRNA. The interaction between U5 and PSF-$p54^{nrb}$ is critical for U5 function and might promote or stabilize possible U5 conformational rearrangements during splicing. Furthermore, we have identified the PCC, a PSF-containing multi-snRNP that resembles the yeast penta-snRNP. These data provide evidence for PSF function in pre-mRNA splicing and suggest

potential links between transcription and splicing although much more work will be needed to test these intriguing hypotheses.

REFERENCES

Abmayr, S. M., Workman, J. L., and Roeder, R. G. (1988). The pseudorabies immediate early protein stimulates in vitro transcription by facilitating TFIID:promoter interactions. Genes and Dev *2*, 542-553.

Achsel, T., Ahrens, K., Brahms, H., Teigelkamp, S., and Lührmann, R. (1998). The human U5-220kD protein (hPrp8) forms a stable RNA-free complex with several U5-specific proteins, including an RNA unwindase, a homologue of ribosomal elongation factor EF-2, and a novel WD-40 protein. Mol Cell Biol *18*, 6756-6766.

Ajuh, P., Kuster, B., Panov, K., Zomerdijk, J., Mann, M., and Lamond, A. (2000). Functional analysis of the human CDC5L complex and identification of its components by mass spectrometry. EMBO J *19*, 6569-6581.

Akhmedov, A. T., and Lopez, B. S. (2000). Human 100-kDa homologous DNA-pairing protein is the splicing factor PSF and promotes DNA strand invasion. Nuc Acids Res *28*, 3022-3030.

Alvi, R., Lund, M., and O'Keefe, R. T. (2001). ATP-dependent interaction of yeast U5 snRNA loop 1 with the 5' splice site. RNA *7*, 1013-1023.

Andersen, J., Lam, Y., Leung, A., Ong, S., Lyon, C., Lamond, A., and Mann, M. (2005). Nucleolar proteome dynamics. Nature *433*, 77-83.

Andersen, J. S., Lyon, C. E., Fox, A. H., Leung, A. K. L., Lam, Y. W., Steen, H., M., M., and I., L. A. (2002). Directed Proteomic Analysis of the Human Nucleolus. Current Biology *12*, 1-11.

Barnard, D. C., and Patton, J. G. (2000). Identification and characterization of a novel serine-arginine rich splicing regulatory protein. Mol Cell Biol *20*, 3049-3057.

Basu, A., Dong, B., Krainer, A. R., and Howe, C. C. (1997). The intracisternal A-particle proximal enhancer-binding protein activates transcription and is identical to the RNA- and DNA-binding protein p54nrb/NonO. Mol Cell Biol *17*, 677-686.

Bennett, M., Michaud, S., Kingston, J., and Reed, R. (1992). Protein components specifically associated with prespliceosome and spliceosome complexes. Genes and Dev *6*, 1986-2000.

Berget, S. M. (1995). Exon recognition in vertebrate splicing. J Biol Chem *270*, 2411-2414.

Besser, H. V., Schnabel, P., Wieland, C., Fritz, E., Stanewsky, R., and Saumweber, H. (1990). The puff-specific Drosophila protein Bj6, encoded by the gene no-on transient A, shows homology to RNA-binding proteins. Chromosoma *100*, 37-47.

Bladen, C., Udayakumar, D., Takeda, Y., and Dynan, W. (2005). Identification of the polypyrimidine tract binding protein-associated splicing factor.p54(nrb) complex as a candidate DNA double-strand break rejoining factor. J Biol Chem *280*, 5205-5210.

Branlant, C., Krol, A., Lazar, E., Haendler, B., Jacob, M., Galego-Dias, L., and Pousada, C. (1983). High evolutionary conservation of the secondary structure and of certain nucleotide sequences of U5 RNA. Nuc Acids Res *11*, 8359-8367.

Bres, V., Gomes, N., Pickle, L., and Jones, K. (2005). A human splicing factor, SKIP, associates with P-TEFb and enhances transcription elongation by HIV-1 Tat. Genes Dev *19*, 1211-1226.

Burd, C. G., and Dreyfuss, G. (1994). Conserved structures and diversity of functions of RNA-binding proteins. Science *265*, 615-621.

Burge, C. B., Tuschl, T., and Sharp, P. A. (1999). Splicing of precursors to mRNAs by spliceosomes., In The RNA World, R. F. Gesteland, T. R. Cech, and J. F. Atkins, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 525-560.

Cáceres, J. F., Misteli, T., Screaton, G. R., Spector, D. L., and Krainer, A. R. (1997). Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. J Cell Biol *138*, 225-238.

Chan, S., Kao, D., Tsai, W., and Cheng, S. (2003). The Prp19p-associated complex in spliceosome activation. Science *302*, 279-282.

Chen, C., Yu, W., Tsao, T., Wang, L., Chen, H., Lin, J., Tsai, W., and Cheng, S. (2002). Functional and physical interactions between components of the Prp19p-associated complex. Nucleic Acids Res *30*, 1029-1037.

Chen, J., Stands, L., Staley, J., Jackups, R. J., Latus, L., and Chang, T. (2001). Specific alterations of U1-C protein or U1 small nuclear RNA can eliminate the requirement of Prp28p, an essential DEAD box splicing factor. Mol Cell *7*, 227-232.

Chiara, M. D., Palandjian, L., Krämer, R. F., and Reed, R. (1997). Evidence that U5 snRNP recognizes the 3' splice site for catalytic step II in mammals. EMBO J *16*, 4746-4759.

Chua, K., and Reed, R. (1999). Human step II splicing factor hSlu7 functions in restructuring the spliceosome between the catalytic steps of splicing. Genes Dev *13*, 841-850.

Coolidge, C. J., and Patton, J. G. (1995). Run-around PCR: a novel way to create duplications using polymerase chain reaction. Biotechniques *18*, 763-764.

Cramer, p., Caceres, J. F., Cazalla, D., Kadener, S., Muro, A. F., Baralle, F. E., and Kornblihtt, A. R. (1999). Coupling of transcription with alternative splicing: RNA pol II promoters modulate SF2/ASF and 9G8 effects on an exonic splicing enhancer. Mol Cell *4*, 251-258.

de la Mata, M., Alonso, C., Kadener, S., Fededa, J., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D., and Kornblihtt, A. (2003). A slow RNA polymerase II affects alternative splicing in vivo. Mol Cell *12*, 525-532.

Dong, B., Horowitz, D. S., Kobayashi, R., and Krainer, A. R. (1993). Purification and cDNA cloning of HeLa cell p54^{nrb}, a nuclear protein with two RNA recognition motifs and extensive homology to human splicing factor PSF and *Drosophila* NONA/BJ6. Nuc Acids Res *21*, 4085-4092.

Dye, B. T., Buvoli, M., Mayer, S. A., Lin, C.-H., and Patton, J. G. (1998). Enhancer Elements Activate the Weak 3' Splice Site of α-Tropomyosin Exon 2. RNA *4*, 1523-1536.

Dye, B. T., and Patton, J. G. (2001). An RNA Recognition Motif (RRM) is Required for the Localization of PTB-Associated Splicing Factor (PSF) to Subnuclear Speckles. Exp Cell Res *263*, 131-144.

Emili, A., Shales, M., McCracken, S., Xie, W., Tucker, P. W., Kobayashi, R., Blencowe, B. J., and Ingles, C. J. (2002). Splicing and transcription associated proteins PSF and p54/NonO bind to the RNA polymerase II. RNA *8*, 1102-1111.

Fabrizio, P., Laggerbauer, B., Lauber, J., Lane, W. S., and Lührmann, R. (1997). An evolutionarily conserved U5 snRNP-specific protein is a GTP-binding factor closely related to the ribosomal translocase EF-2. EMBO J *16*, 4092-4106.

Fields, S., and Song, O.-k. (1989). A novel genetic system to detect protein-protein interactions. Nature *340*, 245-246.

Fleckner, J., Zhang, M., Valcarcel, J., and Green, M. R. (1997). U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. Genes Dev *11*, 1864-1872.

Fong, N., and Bentley, D. L. (2001). Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. Genes Dev *15*, 1783-1795.

Fong, Y. W., and Zhou, Q. (2001). Stimulatory effect of splicing factors on transcriptional elongation. Nature *414*, 929-933.

Fox, A. H., Lam, Y. W., Leung, A. K. L., Lyon, C. E., Andersen, J. S., M., M., and I., L. A. (2002). Paraspeckles; A Novel Nuclear Domain. Current Biology *12*, 13-25.

Fu, X.-D. (1995). The superfamily of arginine/serine-rich splicing factors. RNA *1*, 663-680.

Gavin, A., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J., Michon, A., Cruciat, C.*, et al.* (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature *415*, 141-147.

Ghetti, A., Company, M., and Abelson, J. N. (1995). Specificity of Prp24 binding to RNA: a role for Prp24 in the dynamic interaction of U4 and U6 snRNAs. RNA *1*, 132-145.

Gottschalk, A., Neubauer, G., Banroques, J., Mann, M., Lührmann, R., and Fabrizio, P. (1999). Identification by mass spectrometry and functional analysis of novel proteins of the yeast [U4/U6.U5] tri-snRNP. EMBO J *18*, 4535-4548.

Gozani, O., Patton, J. G., and Reed, R. (1994). A novel set of spliceosome-associated proteins (SAPs) and the essential splicing factor PSF bind stably to pre-mRNA prior to catalytic step II of the splicing reaction. EMBO J *13*, 3356-3367.

Gravely, B. R. (2000). Sorting out the complexity of SR protein functions. RNA *6*, 1197-1211.

Gromak, N., Rideau, A., Southby, J., Scadden, A., Gooding, C., Huttelmaier, S., Singer, R., and Smith, C. (2003). The PTB interacting protein raver1 regulates alpha-tropomyosin alternative splicing. EMBO J *22*, 6356-6364.

Guthrie, C., and Patterson, B. (1988). Spliceosomal snRNAs. Annu Rev Genet *22*, 387-419.

Hallier, M., Tavitian, A., and Moreau-Gachelin, F. (1996). The Transcriptional Factor Spi-1/PU.1 Binds RNA and Interferes with the RNA-binding Protein $p54^{nrb}$. J Biol Chem *271*, 11177-11181.

Harlow, E., and Lane, D. (1988). Antibodies (Cold Spring Harbor, NY: Cold Spring Harbor Press).

Hartmuth, K., Urlaub, H., Vornlocher, H., Will, C., Gentzel, M., Wilm, M., and Lührmann, R. (2002). Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. Proc Natl Acad Sci U S A *99*, 16719-16724.

Hastings, M. L., and Krainer, A. R. (2001). pre-mRNA splicing in the new millenium. Curr Opin Cell Biol *13*, 302-309.

Hinz, M., Moore, M. J., and Bindereif, A. (1996). Domain Analysis of Human U5 RNA. J Biol Chem *271*, 19001-19007.

Hofacker, I. L., Fontana, W., Stadler, P. F., Bonhoeffer, S., Tacker, M., and Schuster, P. (1994). Fast Folding and Comparison of RNA Secondary Structures. Monatshefte f Chemie *125*, 167-188.

Horowitz, D. S., and Abelson, J. (1993). A U5 small nuclear ribonucleoprotein particle protein involved in the second step of pre-mRNA splicing in S. cerevisiae. Mol Cell Biol *13*, 2959-2970.

Horowitz, D. S., and Krainer, A. R. (1997). A human protein required for the second step of pre-mRNA splicing is functionally related to a yeast splicing factor. Genes Dev *11*, 139-151.

Howe, K., Kane, C., and Ares, M. J. (2003). Perturbation of transcription elongation influences the fidelity of internal exon inclusion in Saccharomyces cerevisiae. RNA *9*, 993-1006.

Imai, Y., Matshushima, Y., Sugimura, T., and M., T. (1991). A simple and rapid method for generating a deletion by PCR. Nuc Acids Res *19*, 2785.

Ishigaki, Y., Li, X., Serin, G., and Maquat, L. (2001). Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. Cell *106*, 607-617.

Johnson, J., Castle, J., Garrett-Engele, P., Kan, Z., Loerch, P., Armour, C., Santos, R., Schadt, E., Stoughton, R., and Shoemaker, D. (2003). Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. Science *302*, 2141-2144.

Jones, K. R., and Rubin, G. M. (1990). Molecular analysis of no-on-transient A, a gene required for normal vision in Drosophila. Neuron *4*, 711-723.

Jurica, M., and Moore, M. (2003). Pre-mRNA splicing: awash in a sea of proteins. Mol Cell *12*, 5-14.

Jurica, M. S., Licklider, L. J., Gygi, S. R., Grigorieff, N., and Moore, M. J. (2002). Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. RNA *8*, 426-439.

Kaiser, C., Michaelis, S., and Mitchell, A. (1994). Methods in yeast genetics. (Plainview: Cold Spring Harbor Laboratory Press).

Kameoka, S., Duque, P., and Konarska, M. (2004). p54(nrb) associates with the 5' splice site within large transcription/splicing complexes. EMBO J *23*, 1782-1791.

Komarnitsky, P., Cho, E., and S., B. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes Dev *14*, 2452-2460.

Konarska, M. M., and Sharp, P. A. (1987). Interactions between small nuclear ribonucleoprotein particles in formation of spliceosomes. Cell *49*, 763-774.

Konarska, M. M., and Sharp, P. A. (1988). Association of U2, U4, U5, and U6 small nuclear ribonucleoproteins in a spliceosome-type complex in absence of precursor RNA. Proc Natl Acad Sci *85*, 5459-5462.

Kornblihtt, A. (2005). Promoter usage and alternative splicing. Curr Opin Cell Biol *17*, 262-268.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680-685.

Laggerbauer, B., Achsel, T., and Lührmann, R. (1998). The human U5-200kD DEXH-box protein unwinds U4/U6 RNA duplexes in vitro. Proc Natl Acad Sci *95*, 4188-4192.

Lauber, J., Fabrizio, P., Teigelkamp, S., Lane, W. S., Hartmann, E., and Lührmann, R. (1996). The HeLa 200 kDa U5 snRNP-specific protein and its homologue in S. cerevisiae are members of the DEXH-box protein family of putative RNA helicases. EMBO J *15*, 4001-4015.

Le Hir, H., Gatfield, D., Izaurralde, I., and Moore, M. J. (2001). The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense mediated mRNA decay. EMBO J *20*, 4987-4997.

Li, J., Hawkins, I. C., Harvey, C. D., Jenning, J. L., Link, A. J., and Patton, J. G. (2003). Regulation of alternative splicing by SRrp86 and its interacting proteins. Mol Cell Biol *23*, 7437-7447.

Lindsey, L. A., Crow, A. J., and Garcia-Blanco, M. A. (1995). A mammalian activity required for the second step of pre-messenger RNA splicing. J Biol Chem *270*, 13415-13421.

Lindsey, L. A., and Garcia-Blanco, M. A. (1998). Functional conservation of the human homolog of the yeast pre-mRNA splicing factor Prp17p. J Biol Chem *273*, 32771-32775.

Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M., and Yates, J. R. (1999). Direct analysis of protein complexes using mass spectrometry. Nature Biotech *17*, 676-682.

Lou, H., Gagel, R. F., and Berget, S. M. (1996). An intron enhancer recognized by splicing factors activates polyadenlylation. Genes Dev *10*, 208-219.

Lührmann, R. (1988). snRNP proteins, In Structure and function of major and minor small nuclear ribonucleoprotein particles, M. L. Birnstiel, ed. (Berlin: Springer-Verlag), pp. 71-99.

Luo, M.-J., and Reed, R. (1999). Splicing is required for rapid and efficient mRNA export in metazoans. Proc Natl Acad Sci *96*, 14937-14942.

Luo, M.-J., Zhou, Z., Magni, K., Christoforides, C., Rappsilber, J., Mann, M., and Reed, R. (2001). Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and Aly. Nature *413*, 644-647.

Lutz, C. S., Cooke, C., O'Conner, J. P., Kobayashi, R., and Alwine, J. C. (1998). The snRNP-free U1A (SF-A) complex(es): Identification of the largest subunit as PSF, the Polypyrimidine Tract Binding Protein-Associated Factor. RNA *4*, 1493-1499.

Lynch, K. W., and Maniatis, T. (1995). Synergistic interactions between two distinct elements of a regulated splicing enhancer. Genes Dev *9*, 284-293.

MacMillan, A. M., Query, C. C., Allerson, C. R., Chen, S., Verdine, G. L., and Sharp, P. A. (1994). Dynamic association of proteins with the pre-mRNA branch region. Genes Dev *8*, 3008-3020.

Makarov, E. M., Makarova, O. V., Urlaub, H., Gentzel, M., Will, C. L., Wilm, M., and Luhrmann, R. (2002). Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. Science *298*, 2205-2208.

Maniatis, T., and Reed, R. (2002). An extensive network of coupling among gene expression machines. Science *416*, 499-506.

Maquat, L. (2004). Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. Nat Rev Mol Cell Biol *5*, 89-99.

Maquat, L. (2005). Nonsense-mediated mRNA decay in mammals. J Cell Sci *118*, 1773-1776.

Maquat, L., and Carmichael, G. (2001). Quality control of mRNA function. Cell *104*, 173-176.

Maroney, P. A., Romfo, C. M., and Nilsen, T. W. (2000). Functional recognition of the 5' splice site by U4/U6.U5 tri-snRNP defines a novel ATP-dependent step in early spliceosome assembly. Mol Cell *6*, 317-328.

Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999). Expanded Sequence Dependence of Thermodynamic Parameters Improves Prediction of RNA Secondary Structure. J Mol Biol *288*, 911-940.

Mathur, M., Tucker, P. W., and H., S. H. (2001). PSF Is a Novel Corepressor That Mediates Its Effect through Sin3A and the DNA Binding Domain of Nuclear Hormone Receptors. Mol Cell Biol *21*, 2298-2311.

Matlin, A., Clark, F., and Smith, C. (2005). Understanding alternative splicing: towards a cellular code. Nat Rev Mol Cell Biol *6*, 386-398.

McConnell, T. S., and Steitz, J. A. (2001). Proximity of the invariant loop of U5 snRNA to the second intron residue during pre-mRNA splicing. EMBO J *20*, 3577-3586.

McCracken, S., Lambermon, M., and Blencowe, B. (2002). SRm160 splicing coactivator promotes transcript 3'-end cleavage. Mol Cell Biol *22*, 148-160.

Meissner, M., Dechat, T., Gerner, C., Grimm, R., Foisner, R., and Sauermann, G. (2000). Differential nuclear localization and nuclear matrix association of the splicing factors PSF and PTB. J Cell Biochem *76*, 559-566.

Michaud, S., and Reed, R. (1991). An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. Genes Dev *5*, 2534-2546.

Michaud, S., and Reed, R. (1993). A functional association between the 5' and 3' splice site is established in the earliest prespliceosome complex (E) in mammals. Genes Dev *7*, 1008-1020.

Miralles, F., Ofverstedt, L.-G., Sabri, Y., Hellman, U., Skoglund, U., and Visa, N. (2000). Electron tomography reveals posttranscriptional binding of pre-mRNPs to specific fibers in the nucleoplasm. J Cell Biol *148*, 271-282.

Monsalve, M., Wu, Z., Adelmant, G., Puigserver, P., Fan, M., and Spiegelman, B. M. (2000). Direct coupling of transcription and mRNA processing through the thermogenic coactivator PGC-1. Mol Cell *6*, 307-316.

Moore, M. J., Query, C. C., and Sharp, P. A. (1993). Splicing of precursors to mRNA by the spliceosome., In The RNA World, R. F. G. a. J. F. Atkins, ed. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 303-358.

Muller, S., Wolpensinger, B., Angenitzki, M., Engel, A., Sperling, J., and Sperling, R. (1998). A supraspliceosome model for large nuclear ribonucleoprotein particles based on mass determinations by scanning transmission electron microscopy. J Mol Biol *283*, 383-394.

Nagy, E., and Maquat, L. (1998). A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. Trends Biochem Sci *23*, 198-199.

Neubauer, G., King, A., Rappsilber, J., Calvio, C., Watson, M., Ajuh, P., Sleeman, J., Lamond, A., and Mann, M. (1998). Mass spectrometry and est-database searching allows characterization of the mult-protein spliceosome complex. Nature Gen *20*, 46-50.

Newman, A. (1997). RNA splicing: out of the loop. Current Biol *7*, R418-420.

Nilsen, T. (2003). The spliceosome: the most complex macromolecular machine in the cell? BioEssays *25*, 1147-1149.

Nilsen, T. W. (1998). RNA-RNA interactions in nuclear pre-mRNA splicing., In RNA Structure and Function., R. Simons, and M. Grunberg-Manago, eds. (Cold Spring Harbor: Cold Spring Harbor Laboratory Press), pp. 279-307.

Nott, A., Le Hir, H., and Moore, M. (2004). Splicing enhances translation in mammalian cells: an additional function of the exon junction complex. Genes Dev *18*, 210-222.

O'Keefe, R. T., and Newman, A. J. (1998). Functional analysis of the U5 snRNA loop 1 in the second catalytic step of yeast pre-mRNA splicing. EMBO J *17*, 565-574.

Ohi, M. D., and Gould, K. L. (2002). Characterization of interactions among the Cef1p-Prp19p-associated splicing complex. RNA *8*, 798-815.

Ohi, M. D., Link, A. J., Ren, L., Jennings, J. L., McDonald, W. H., and Gould, K. L. (2002). Proteomics analysis reveals stable multiprotein complexes in both fission and budding yeasts containing Myb-related Cdc5p/Cef1p novel pre-mRNA splicing factors and snRNAs. Mol Cell Biol *22*, 2011-2024.

Parker, A., and Steitz, J. (1997). Inhibition of mammalian spliceosome assembly and pre-mRNA splicing by peptide inhibitors of protein kinases. RNA *3*, 1301-1312.

Patton, J. G., Mayer, S. A., Tempst, P., and Nadal-Ginard, B. (1991). Characterization and molecular cloning of Polypyrimidine Tract Binding Protein: A component of a complex necessary for pre-mRNA splicing. Genes Dev *5*, 1237-1251.

Patton, J. G., Porro, E. B., Galceran, J., Tempst, P., and Nadal-Ginard, B. (1993). Cloning and characterization of PSF, a novel pre-mRNA splicing factor. Genes Dev *7*, 393-406.

Patton, J. G., and Smith, C. W. J. (2001). Alternative pre-mRNA Splicing: Prevalence, Mechanisms, and Disease (New York: John Wiley).

Patton, J. R. (1991). Pseudouridine modification of U5 RNA in ribonucleoprotein particles assembled in vitro. Mol Cell Biol *11*, 5998-6006.

Peng, R., Dye, B. T., Perez, I., and Patton, J. G. (2002). PSF and p54nrb bind a conserved stem in U5 snRNA. RNA *8*, 1334-1347.

Pérez, I., Lin, C.-H., McAfee, J. G., and Patton, J. G. (1997). Mutation of PTB Binding Sites Causes Misregulation of α-Tropomyosin Alternative 3' Splice Site Selection *In Vivo*. RNA *3*, 764-778.

Raghunathan, P., and Guthrie, C. (1998). RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. Curr Biol *8*, 847-855.

Rappsilber, J., Ryder, U., Lamond, A. I., and Mann, M. (2002). Large scale proteomic analysis of the human spliceosome. Genome Res *12*, 1231-1245.

Reed, R. (2000). Mechanisms of fidelity in pre-mRNA splicing. Curr Opin Cell Biol *12*, 340-345.

Reed, R. (2003). Coupling transcription, splicing and mRNA export. Curr Opin Cell Biol *15*, 326-331.

Reed, R., and Cheng, H. (2005). TREX, SR proteins and export of mRNA. Curr Opin Cell Biol *17*, 269-273.

Reed, R., and Palandjian, L. (1997). Spliceosome Assembly, In Eukaryotic mRNA Processing, A. R. Krainer, ed. (Oxford: IRL Press), pp. 103-129.

Reyes, J. L., Kois, P., Konforti, B. B., and Konarska, M. M. (1996). The canonical Gu dinucleotide at the 5' splice site is recognized by p220 of the U5 snRNP within the spliceosome. RNA *2*, 213-225.

Ségault, V., Will, C. L., Polycarpou-Schwarz, M., Mattaj, I. W., Branlant, C., and Lührmann, R. (1999). Conserved loop I of U5 small nuclear RNA is dispensable for both catalytic steps of pre-mRNA splicing in HeLa nuclear extracts. Mol Cell Biol *19*, 2782-2790.

Sewer, M., and Waterman, M. (2002). Adrenocorticotropin/cyclic adenosine 3',5'-monophosphate-mediated transcription of the human CYP17 gene in the adrenal cortex is dependent on phosphatase activity. Endocrinology *143*, 1769-1777.

Sewer, M. B., Nguyen, V. Q., Huang, C.-J., Tucker, P. W., Kagawa, N., and Waterman, M. R. (2002). Transcriptional activation of human CYP17 in H295R adrenocortical cells depends on complex formation among p54nrb/NonO, PSF, and SF-1, a complex that also participates in repression of transcription. Endocrinology *143*, 1280-1290.

Shatkin, A., and Manley, J. (2000). The ends of the affair: capping and polyadenylation. Nat Struct Biol *7*, 838-842.

Shav-Tal, Y., Blechman, J., Darzacq, X., Montagna, C., Dye, B., Patton, J., Singer, R., and Zipori, D. (2005). Dynamic sorting of nuclear components into distinct nucleolar caps during transcriptional inhibition. Mol Biol Cell *16*, 2395-2413.

Shav-Tal, Y., Cohen, M., Lapter, S., Dye, B., Patton, J. G., Vandekerckhove, J., and Zipori, D. (2001). Nuclear Relocalization of the pre-mRNA Splicing Factor PSF during Apoptosis involves hyperphosphorylation, masking of antigenic epitopes, and changes in protein interactions. Mol Biol Cell *12*, 2328-2340.

Smith, C. W. J., and Nadal-Ginard, B. (1989). Mutually exclusive splicing of α-tropomyosin exons enforced by an unusual lariat branch point location; implications for constitutive splicing. Cell *56*, 749-758.

Smith, C. W. J., and Valcarcel, J. (2000). Alternative pre-mRNA splicing: the logic of combinatorial control. TIBS *25*, 381-388.

Song, X., Sui, A., and Garen, A. (2004). Binding of mouse VL30 retrotransposon RNA to PSF protein induces genes repressed by PSF: effects on steroidogenesis and oncogenesis. Proc Natl Acad Sci U S A *101*, 621-626.

Staley, J., and Guthrie, C. (1999). An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. Mol Cell *3*, 55-64.

Staley, J. P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. Cell *92*, 315-326.

Stamm, S., Ben-Ari, S., Rafalska, I., Tang, Y., Zhang, Z., Toiber, D., Thanaraj, T., and Soreq, H. (2005). Function of alternative splicing. Gene *344*, 1-20.

Steitz, T. A. (1999). RNA recognition by proteins, In The RNA World. Second edition, R. F. Gesteland, T. R. Cech, and J. F. Atkins, eds. (Cold Spring Harbor: Cold Spring Harbor Laboratory Press), pp. 427-450.

Stevens, S., and Abelson, J. (1999). Purification of the yeast U4/U6.U5 small nuclear ribonucleoprotein particle and identification of its proteins. Proc Natl Acad Sci U S A *96*, 7226-7231.

Stevens, S. W., Ryan, D. E., Ge, H. Y., Moore, R. E., Young, M. K., Lee, T. D., and Abelson, J. (2002). Composition and functional characterization of the yeast spliceosomal penta-snRNP. Mol Cell *9*, 31-44.

Strasser, K., Masuda, S., Mason, P., Pfannstiel, J., Oppizzi, M., Rodriguez-Navarro, S., Rondon, A., Aguilera, A., Struhl, K., Reed, R., and Hurt, E. (2002). TREX is a conserved complex coupling transcription with messenger RNA export. Nature *417*, 304-308.

Straub, T., Grue, P., Uhse, A., Lisby, M., Knudsen, B. R., Tange, T. O., Westergaard, O., and Boege, F. (1998). The RNA-splicing factor PSF p54 controls DNA topoisomerase I activity by a direct interaction. J Biol Chem *273*, 26261-26264.

Straub, T., Knudson, B. R., and Boege, F. (2000). PSF/p54^{nrb} stimulates "jumping" of DNA topoisomerase I between separate DNA helices. Biochem *39*, 7552-7558.

Swanson, M. S. (1995). Funtions of nuclear pre-mRNA/mRNA binding proteins, In Pre-mRNA Processing, A. I. Lamond, ed. (Austin: R.G. Landes Company), pp. 17-34.
Szostak, J. W. (1992). *In vitro* genetics. TIBS *17*, 89-93.

Tarn, W., Hsu, C., Huang, K., Chen, H., Kao, H., Lee, K., and Cheng, S. (1994). Functional association of essential splicing factor(s) with PRP19 in a protein complex. EMBO J *13*, 2421-2431.

Teigelkamp, S., Mundt, C., Achsel, T., Will, C., and Lührmann, R. (1997). The human U5 snRNP-specific 100kD protein is an RS domain-containing, putative RNA helicase with significant homology to the yeast splicing factor Prp28p. RNA *3*, 1313-1326.

Traish, A. M., Huang, Y. H., Ashba, J., Pronovost, M., Pavao, M., McAneny, D. B., and Moreland, R. B. (1997). Loss of expression of a 55 kDa nuclear protein (nmt55) in estrogen receptor-negative human breast cancer. Diagn Mol Pathol *6*, 209-221.

Trembley, J., Tatsumi, S., Sakashita, E., Loyer, P., Slaughter, C., Suzuki, H., Endo, H., Kidd, V., and Mayeda, A. (2005). Activation of pre-mRNA splicing by human RNPS1 is regulated by CK2 phosphorylation. Mol Cell Biol *25*, 1446-1457.

Tuerk, C., and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science *249*, 505-510.

Urban, R. J., and Bodenburg, Y. (2002). PSF regulates growth factor stimulated gene expression in mammalian cells. Am J Physiol Endocrinol Metab *283*, E794-E798.

Urban, R. J., Bodenburg, Y., Kurosky, A., Wood, T. G., and Gasic, S. (2000). Polypyrimidine tract-binding protein-associated splicing factor is a negative regulator of transcriptional activity of the porcine p450scc insulin-like growth factor response element. Mol Endocrinol *14*, 774-782.

Vagner, S., Vagner, C., and Mattaj, I. (2000). The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF 65 to couple 3'-end processing and splicing. Genes Dev *14*, 403-413.

Valcárcel, J., Gaur, R. K., Singh, R., and Green, M. R. (1996). Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA. Science *273*, 1706-1709.

Vijayraghavan, U., and Abelson, J. (1990). PRP18, a protein required for the second reaction in pre-mRNA splicing. Mol Cell Biol *10*, 324-332.

Will, C., Urlaub, H., Achsel, T., Gentzel, M., Wilm, M., and Lührmann, R. (2002). Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p homologue and an SF3b DEAD-box protein. EMBO J *21*, 4978-4988.

Wurtz, T., Kiseleva, E., Nacheva, G., Alzhanova-Ericcson, A., Rosen, A., and Daneholt, B. (1996). Identification of two RNA-binding proteins in Balbiani ring premessenger ribonucleoprotein granules and presence of these proteins in specific subsets of heterogeneous nuclear ribonucleoprotein particles. Mol Cell Biol *16*, 1425-1435.

Yang, Y.-S., Hanke, J. H., Carayannapoulos, L., Craft, C. M., Capra, J. D., and Tucker, P. W. (1993). NonO, a Non-POU-Domain-Containing, Octamer-Binding Protein, Is the Mammalian Homolog of *Drosophila nonA*^{diss}. Mol Cell Biol *13*, 5593-5603.

Zamore, P. D., Patton, J. G., and Green, M. R. (1992). Cloning and domain structure of the mammalian splicing factor U2AF. Nature *355*, 609-614.

Zhang, W. W., Zhang, L. X., Busch, R. K., Farres, J., and Busch, H. (1993). Purification and characterization of a DNA-binding heterodimer of 52 and 100 kDa from HeLa cells. Biochem J *290*, 267-272.

Zhang, Z., and Carmichael, G. G. (2001). The fate of dsRNA in the nucleus: A p54nrb-containing complex mediates the nuclear retention of promiscuously A-to-I edited RNAs. Cell *106*, 465-475.

Zhou, Z., Licklider, L. J., Gygi, S. P., and Reed, R. (2002). Comprehensive proteomic analysis of the human spliceosome. Nature *419*, 182-185.

Zhou, Z., and Reed, R. (1998). Human homologs of yeast Prp16 and Prp17 reveal conservation of the mechanism for catalytic step II of pre-mRNA splicing. EMBO J *17*, 2095-2106.

Zolutukhin, A. S., Michalowski, D., Bear, J., Smulevitch, S. V., Traish, A. M., Peng, R., Patton, J. G., Shatsky, I., and Felber, B. K. (2002). p54nrb/PSF binds to the HIV-1 gag instability element and mediates mRNA downregulation. Submitted.

Zuker, M., Mathews, D. H., and Turner, D. H. (1999). Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide., In RNA Biochemistry and Biotechnology, J. Barciszewski, and B. F. C. Clark, eds. (Kluwer Academic Publishers), pp. 11-43.