
# Splicing and transcription-associated proteins PSF and p54nrb/nonO bind to the RNA polymerase II CTD.

Andrew Emili, Michael Shales, Susan McCracken, et al.

*RNA* 2002 8: 1102-1111

*RNA* (2002), *8*:1102–1111. Cambridge University Press. Printed in the USA.

DOI: 10.1017.S1355838202025037

**REPORT**

# Splicing and transcription-associated proteins PSF and p54$^{nrb}$/NonO bind to the RNA polymerase II CTD

**ANDREW EMILI,[1,2] MICHAEL SHALES,[1] SUSAN McCRACKEN,[1] WEIJUN XIE,[3] PHILIP W. TUCKER,[3] RYUJI KOBAYASHI,[4,5] BENJAMIN J. BLENCOWE,[1,2] and C. JAMES INGLES[1,2]**

[1]Banting and Best Department of Medical Research, C.H. Best Institute, University of Toronto, Toronto, Ontario M5G 1L6, Canada
[2]Department of Molecular and Medical Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada
[3]Department of Molecular Genetics and Microbiology, University of Texas at Austin, Austin, Texas 78705, USA
[4]Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA

## ABSTRACT

**The carboxyl-terminal domain (CTD) of the largest subunit of eukaryotic RNA polymerase II (pol II) plays an important role in promoting steps of pre-mRNA processing. To identify proteins in human cells that bind to the CTD and that could mediate its functions in pre-mRNA processing, we used the mouse CTD expressed in bacterial cells in affinity chromatography experiments. Two proteins present in HeLa cell extract, the splicing and transcription-associated factors, PSF and p54$^{nrb}$/NonO, bound specifically and could be purified to virtual homogeneity by chromatography on immobilized CTD matrices. Both hypo- and hyperphosphorylated CTD matrices bound these proteins with similar selectivity. PSF and p54$^{nrb}$/NonO also copurified with a holoenzyme form of pol II containing hypophosphorylated CTD and could be coimmunoprecipitated with antibodies specific for this and the hyperphosphorylated form of pol II. That PSF and p54$^{nrb}$/NonO promoted the binding of RNA to immobilized CTD matrices suggested these proteins can interact with the CTD and RNA simultaneously. PSF and p54$^{nrb}$/NonO may therefore provide a direct physical link between the pol II CTD and pre-mRNA processing components, at both the initiation and elongation phases of transcription.**

**Keywords: affinity chromatography; pre-mRNA processing; protein complexes; transcription**

## INTRODUCTION

Transcription and precursor (pre-)mRNA processing are highly coordinated events in gene expression and can influence one another. The carboxyl-terminal domain (CTD) of the largest subunit of RNA polymerase II (pol II) has important roles in coupling transcription to different steps in the processing of pre-mRNA, including formation of a 5′-end cap, splicing, and 3′-end formation (reviewed in Steinmetz, 1997; Hirose & Manley, 2000; Cramer et al., 2001). The CTD consists of tandem repeats of the heptapeptide consensus sequence YSPTSPS, of which there are 26 or 27 in the largest subunit of *Saccharomyces cerevisiae* RNA pol II (Allison et al., 1985; Nonet et al., 1987) and 52 in the mammalian enzyme (Corden et al., 1985). During initiation of transcription, RNA pol II with a hypophosphorylated CTD (pol IIa) is recruited to the promoter as part of a large multiprotein holoenzyme complex (Lu et al., 1991; Kim et al., 1994; Koleske & Young, 1994; Ossipow et al., 1995; Pan et al., 1997). The CTD is then hyperphosphorylated, primarily by the cyclin-dependent kinase (Cdk9) subunit of the elongation factor P-TEFb, resulting in a more processive, elongating form of RNA pol II (pol IIo) (reviewed in Dahmus, 1996; Conaway et al., 2000). The CTD has been implicated in the response of RNA pol II to sequence-specific DNA-binding transcrip-

Reprint requests to: C. James Ingles, Banting and Best Department of Medical Research, C.H. Best Institute, University of Toronto, 112 College Street, Toronto, Ontario M5G 1L6, Canada; e-mail: cj.ingles@utoronto.ca
[5]Present address: Department of Molecular Pathology, University of Texas, MD Anderson Cancer Center, 1515 Holcombe Blvd., Houston, Texas 77030, USA.

tional activator proteins (Allison & Ingles, 1989; Scafe et al., 1990; Gerber et al., 1995) and requires the mediator complex, with which it interacts, for this activated transcription (Kim et al., 1994).

In mammalian cells, truncation of the CTD prevents the efficient capping, splicing, and 3′ processing of nascent transcripts (McCracken et al., 1997a, 1997b) and reduces the accumulation of splicing factors at sites of transcription in vivo (Misteli & Spector, 1999). Experiments performed in vitro also argue for an important role of the CTD in splicing and polyadenylation. For example, anti-CTD antibodies can inhibit pre-mRNA splicing (Yuryev et al., 1996) and both RNA pol IIo (Hirose et al., 1999) and the CTD itself can stimulate splicing (Zeng & Berget, 2000). The 3′-end processing of nascent transcripts is also activated in vitro by the addition of purified RNA pol II or the CTD (Hirose & Manley, 1998). In an apparently reciprocal manner, splicing factors bound to nascent RNA can stimulate elongation of transcription, probably through interactions mediated by the elongation factor Tat-SF1, which appears to interact with both splicing snRNPs and P-TEFb (Fong & Zhou, 2001). Consistent with an important role for the CTD in mediating the coupling of transcription and pre-mRNA processing, splicing components (Mortillaro et al., 1996; Vincent et al., 1996; Yuryev et al., 1996; Kim et al., 1997; Morris & Greenleaf, 2000), 5′ capping enzymes (Yue et al., 1997; McCracken et al., 1997b; Ho et al., 1998), and 3′ processing factors (McCracken et al., 1997a; Barilla et al., 2001) have all been found to interact with the CTD.

Although a role for the CTD in promoting steps in pre-mRNA processing has been established, it is not known at which stage of the transcription cycle splicing components are recruited to RNA pol II complexes, nor which factors bound to the CTD coordinate transcription and splicing. Recent studies have provided evidence that recognition of promoter sequences by the transcriptional machinery can influence the ability of specific SR (serine/arginine repeat) family splicing factors to promote alternative exon inclusion in reporter transcripts (Cramer et al., 1999). Moreover, SR family proteins and other splicing factors have recently been detected in purified holoenzyme complexes containing RNA pol IIa (Robert et al., 2001). However, the relationship between these observations and previous work demonstrating a role for the CTD in pre-mRNA splicing is not known.

To learn more about the functions of the CTD, we have used affinity chromatography to identify proteins that bind to the CTD in mammalian cells. We demonstrate that polypyrimidine tract binding protein- (PTB) associated splicing factor (PSF; Patton et al., 1993) and p54$^{nrb}$/NonO (Dong et al., 1993; Yang et al., 1993), two proteins that share considerable amino acid sequence homology and that copurify in a complex (Zhang et al., 1993; Straub et al., 1998; Zhang & Carmichael, 2001), bind specifically and can be purified to near homogeneity from HeLa cell extracts by chromatography on an immobilized CTD matrix. Experiments suggesting that PSF and p54$^{nrb}$/NonO can bind the CTD and RNA simultaneously, and the demonstration that PSF and p54$^{nrb}$/NonO are associated with both the hypophosphorylated and hyperphosphorylated forms of RNA polymerase II in HeLa cell extracts, indicate that these two proteins could provide a direct physical link between RNA polymerase and other pre-mRNA processing components during the initiation and elongation phases of transcription.

## RESULTS

### Interaction of PSF and p54$^{nrb}$/NonO with the pol II CTD

To identify human proteins that interact with the CTD of the largest subunit of RNA polymerase II, we first expressed and purified a recombinant version of the mouse CTD with an N-terminal polyhistidine tag. This form of recombinant CTD was abundantly expressed in *Escherichia coli* cells in a soluble form and readily purified by $Ni^{2+}$-chelate affinity chromatography. A western blot analysis (Fig. 1A) showed that the recombinant CTD polypeptide was bound by the monoclonal antibody JEL352 (Moyle et al., 1989), a reagent that preferentially recognizes the nonphosphorylated form of the CTD. A series of affinity columns containing increasing concentrations of covalently coupled CTD ligand were loaded with HeLa whole-cell extract. After extensive washes with buffer containing 0.1 M NaCl, these columns were eluted with high salt and the eluates analyzed by SDS-PAGE and silver staining. As the concentration of immobilized CTD on the affinity columns was increased, there was a corresponding increase in the recovery of a number of proteins in the high salt eluates. Two of the more prominent proteins (seen in Fig. 1B as negatively staining broad bands) that bound preferentially to the CTD had apparent molecular masses of approximately 100 and 60 kDa; other CTD-binding proteins had masses of 180, 80, 40, 36, 34, and 32 kDa.

To identify proteins present in HeLa cell extracts that bound specifically to the CTD, the affinity chromatography purification scheme was scaled up. The input HeLa cell extract, first precleared by passage over an AffiGel-10 matrix containing no coupled protein, was then sequentially bound to and eluted from two CTD columns. The two most abundant polypeptides seen in Figure 1B, ~100 and ~60 kDa in size, were selectively purified to near homogeneity. After preparative-scale SDS gel electrophoresis of the high salt eluate from the second CTD column (Fig. 1C), the bands corresponding to these two proteins were subjected to proteolytic digestion in situ with trypsin. The resulting peptides

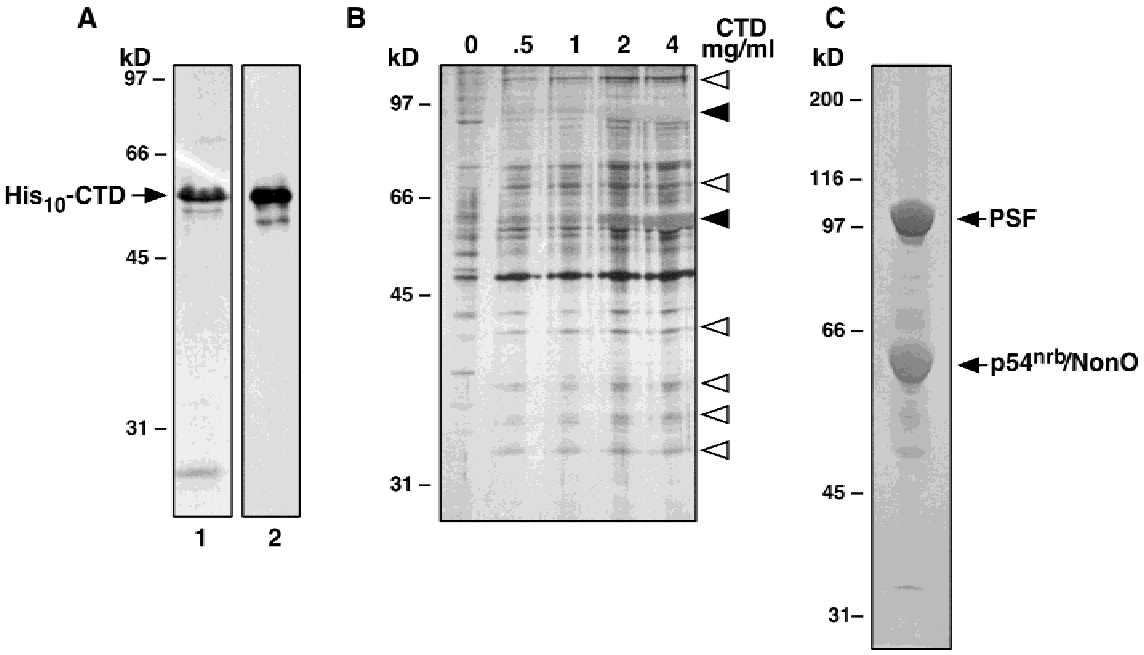


**FIGURE 1.** Identification of PSF and p54$^{nrb}$/NonO as pol II CTD-binding proteins. **A:** His-tagged recombinant mouse CTD was expressed in *E. coli,* purified by $Ni^{2+}$-chelate chromatography, and analyzed after SDS-PAGE by Coomassie blue staining (lane 1) and western blotting (lane 2) with the anti-CTD monoclonal antibody JEL352 (Moyle et al., 1989). **B:** Silver-stained SDS-gel of the high salt eluates of a series of 20-$\mu$L affinity columns containing the indicated concentration of immobilized mouse CTD. The columns were loaded with 300 $\mu$L HeLa whole-cell extract. Closed arrowheads indicate the positions of two overloaded and negatively staining polypeptides binding specifically to the CTD matrix; open arrowheads, other CTD-binding proteins. **C:** Coomassie blue-stained SDS gel showing the preparation of CTD-binding proteins subjected to microsequence analysis. Protein bands subsequently identified as PSF and p54$^{nrb}$/NonO are indicated.

were resolved by reverse phase HPLC and subjected to N-terminal (Edman) amino acid sequence determination. Peptides with the sequence YGEPGEVFINK and ERETPPRFAQHGTFEYEYSQR from the 100-kDa band and ALDR and EREQPPRFAQPGSFEYEYAMR from the 60-kDa band were obtained. A search of the GenBank protein sequence database identified perfect matches between the peptide sequences from the ~100-kDa protein and sequences within the human RNA/DNA-binding protein PSF (Patton et al., 1993). The peptides from the 60-kDa protein were identical to sequences within the related mammalian RNA/DNA-binding protein p54$^{nrb}$/NonO. These two proteins share 71% identity within a central 320-amino-acid region that includes two tandem RNA recognition motifs (RRMs), a helix-turn-helix (HTH) domain, and a domain that is rich in charged amino acid residues (Dong et al., 1993; Patton et al., 1993; Yang et al., 1993).

To verify that PSF and p54$^{nrb}$/NonO are capable of interacting with the CTD, we performed western blot analyses of high salt elutes from CTD and control affinity columns loaded with HeLa cell extract, using polyclonal antibodies raised against either PSF (Fig. 2A) or p54$^{nrb}$/NonO (data not shown). A single anti-PSF immunoreactive band with the same mobility as the form of this polypeptide present in HeLa cell extract, and as a recombinant PSF protein whose expression was induced in *E. coli* cells, was present in the high-salt eluate of the CTD column but not a control column lacking bound CTD. As both PSF and p54$^{nrb}$/NonO have RNA-binding motifs (Dong et al., 1993; Patton et al., 1993; Yang et al., 1993) and may be capable of interacting with either RNA or DNA (Basu et al., 1997; Yang et al., 1997), the interaction of these two proteins with the CTD might, in theory, be mediated through a nucleic acid intermediate. This possibility, however, seems unlikely, as extensive treatment of the input HeLa cell extract with either ribonuclease A or micrococcal nuclease did not impair the binding of PSF and p54$^{nrb}$/NonO to a CTD affinity column (data not shown; see also Fig. 2B). Furthermore, the eluates from our CTD columns did not contain significant amounts of nucleic acid as detected by labeling with T4 polynucleotide kinase (data not shown).

The apparent 1:1 stoichiometry (Fig. 1C) of the PSF and p54$^{nrb}$/NonO polypeptides in the CTD affinity column eluates suggested that they may form a stable heterodimeric complex. Indeed, such a 1:1 heterodimeric mixture of these same two proteins, first purified from HeLa cells as an activity binding to the upstream region of the human proliferation-associated antigen p120 gene (Zhang et al., 1993), has also been found in human A431 cell extracts associated with a third polypeptide, topoisomerase I (Straub et al., 1998). In an-

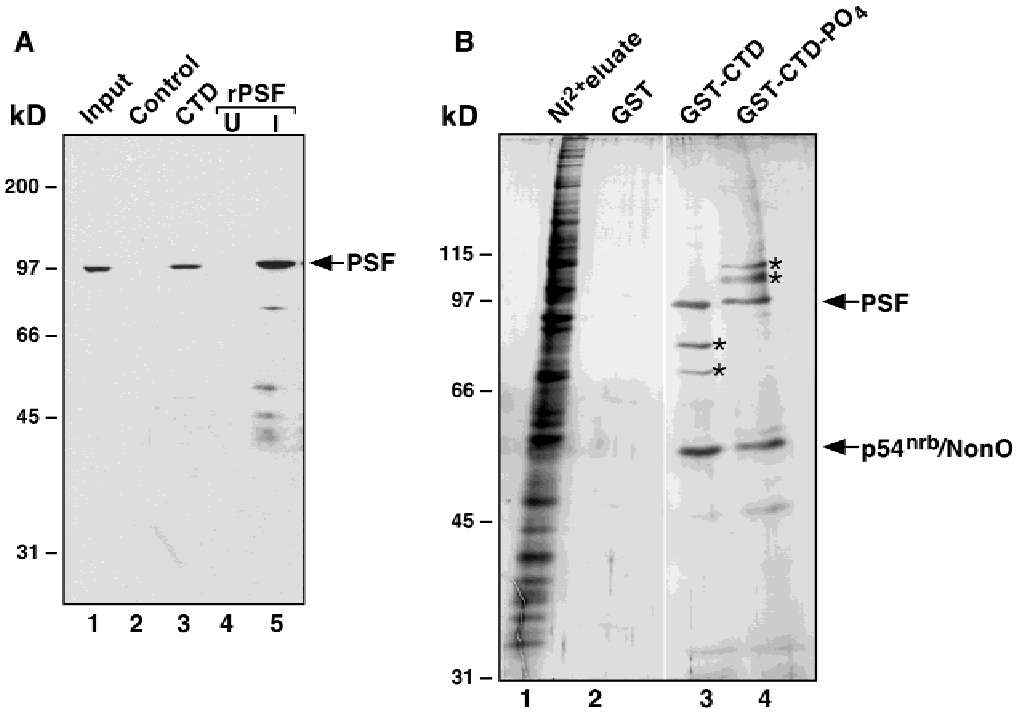


**FIGURE 2.** PSF and p54$^{nrb}$/NonO bind the RNA pol II CTD. **A:** Western blot of an SDS gel probed with polyclonal anti-PSF serum; HeLa cell extract (lane 1) was loaded on control and CTD columns; high salt eluates from the control and CTD affinity column (lanes 2 and 3); cell extracts from *E. coli* cells uninduced and induced to express recombinant human PSF (lanes 4 and 5). **B:** Silver-stained SDS-gel analysis of the high salt eluates from affinity columns containing ligands bound at 4 mg/mL; bound GST (lane 2) GST-CTD (lane 3), or the phosphorylated GST-CTD derivative (lanes 4). The identity of the indicated CTD-binding proteins was determined after proteolysis and MALDI-TOF mass spectrometry. Lane 1 is an aliquot of the HeLa cell extract prefractionated by $Ni^{2+}$-chelate chromatography (Straub et al., 1998) that was loaded on each column. The additional polypeptides, indicated by asterisks, were identified by MALDI-TOF mass spectrometry as GST-CTD-derived polypeptides that leached from the columns.

other study, PSF and p54$^{nrb}$/NonO were found to be present in a complex with the inner nuclear matrix protein, matrin 3, bound specifically to inosine-containing RNAs (Zhang & Carmichael, 2001). Neither the 91-kDa topoisomerase I nor the 125-kDa matrin 3 polypeptide were present with PSF and p54$^{nrb}$/NonO in the eluate of our CTD columns (Fig. 1C, see also Fig. 2B). The absence of matrin 3 in our eluates is consistent with the observation that it readily dissociates from PSF and p54$^{nrb}$/NonO during complex isolation (Zhang & Carmichael, 2001). Moreover, p54$^{nrb}$/NonO and topoisomerase I may have copurified because of their abilities to each bind the nickel–agarose affinity matrix (Straub et al., 1998; Zhang & Carmichael, 2001). The observation that the mixture of PSF and p54$^{nrb}$/NonO binds efficiently to a CTD affinity column upon rechromatography, and essentially no other polypeptides are detected in the eluates from this second CTD column (see Fig. 1C), indicates that the interaction between the CTD and one or both of these components is direct and is not mediated by any other protein in HeLa cell extract.

Our attempts to delineate a CTD-binding domain within either of these proteins were unsuccessful. Unlike the proteins purified from human cell extracts, neither recombinant PSF nor recombinant p54$^{nrb}$/NonO or a mixture of these two proteins expressed in *E. coli* cells bound efficiently to the CTD affinity matrix (data not shown). These recombinant proteins may therefore require one or more posttranslational modifications to bind the CTD efficiently. Consistent with this possibility, cellular forms of both PSF and p54$^{nrb}$/NonO are known to contain phosphorylated tyrosine residues (Otto et al., 2001). Such modifications could influence the ability of these proteins to interact with other cellular components.

### PSF and p54$^{nrb}$/NonO bind the phosphorylated and nonphosphorylated CTD

Our initial identification of PSF and p54$^{nrb}$/NonO as CTD binding proteins by affinity chromatography employed a nonphosphorylated CTD ligand. It was important, therefore, to determine the effect of phosphorylation of the CTD on this interaction because the elongating form of RNA polymerase II is multiply phosphorylated on the Ser-2 and Ser-5 positions within the YSPTSPS heptad repeat and this phosphorylation is important for the association of some of the factors that function in pre-mRNA processing (Mortillaro et al., 1996; Yue et al., 1997; McCracken et al., 1997b; Ho et al., 1998;

Patturajan et al., 1998b; Komarnitsky et al., 2000; Schroeder et al., 2000; Pei et al., 2001). For these experiments, we employed the immobilized GST-CTD ligands described earlier by McCracken et al. (1997a), including a GST–CTD derivative that was enzymatically phosphorylated by preincubation in nuclear extract. HeLa nuclear extract was pretreated with DNase and RNase, prefractionated by $Ni^{2+}$-chelate chromatography to enrich for PSF and p54$^{nrb}$/NonO (Straub et al., 1998), and then chromatographed on affinity columns containing GST, the GST-CTD or the phosphorylated GST-CTD derivative (Fig. 2B). The shift in mobility of the phosphorylated GST-CTD ligand (see Fig. 2B) indicated that it was phosphorylated to a degree similar to that of pol IIo. The high salt eluates from these columns were analyzed by SDS-PAGE and the major proteins bound to these CTD matrices were then identified by MALDI-TOF mass spectrometry. As shown in Figure 2B, of all the proteins present in the partially fractionated cell extract that was loaded on the columns, only PSF and p54$^{nrb}$/NonO bound efficiently to both the phosphorylated and nonphosphorylated GST-CTD ligands. Although the binding of the phosphorylated CTD could reflect an interaction with remaining nonphosphorylated heptad repeats, the results demonstrate that hyperphosphorylation of the CTD does not reduce binding of these two proteins. These chromatography experiments again demonstrate the remarkable selectivity of the CTD in binding a very limited number of proteins. Essentially only PSF and p54$^{nrb}$/NonO bound the CTD.

### PSF and p54$^{nrb}$/NonO associate with both the RNA pol IIa holoenzyme and phosphorylated RNA pol IIo

As potential components of pre-mRNA processing complexes, PSF and p54$^{nrb}$/NonO might have been expected to preferentially associate, like pre-mRNA capping enzymes (McCracken et al., 1997b; Yue et al., 1997; Ho et al., 1998), with the phosphorylated CTD present on elongating RNA pol IIo. The affinity chromatography experiment shown in Figure 2B, however, suggests that these two proteins might bind equally well to both nonphosphorylated and phosphorylated versions of the CTD heptad repeats and thus be associated with both pol IIa and pol IIo in mammalian cells. To explore this possibility, we first used the procedure described by Pan et al. (1997) of affinity chromatography on an immobilized human TFIIS matrix to isolate the hypophosphorylated RNA pol IIa-containing holoenzyme complex. We used western blot analyses to ascertain the relative abundance of PSF and p54$^{nrb}$/NonO in the isolated pol IIa holoenzyme. As shown in Figure 3A, these immunoblots indicated that both PSF and p54$^{nrb}$/NonO are present, along with the expected general initiation factors such as TFIIB and TFIIE, in the pol IIa holoenzyme complex. By comparison to recombinant standards, we were able to estimate the relative content of these proteins in the complex. Perhaps not surprisingly, given the presence of 52 heptad repeats within human pol II CTD, both PSF and p54$^{nrb}$/NonO were present in the pol II holoenzyme at stoichiometries considerably greater than 1:1. In contrast to the 56-kDa subunit of TFIIE and the TFIIB polypeptide, which were present in this preparation of holoenzyme in near equimolar amounts (0.8:1 and 1.3:1, respectively, relative to the pol II largest subunit), PSF and p54$^{nrb}$/NonO were present in this holoenzyme preparation in stoichiometries of 3.5:1 and 23:1, respectively. That we did not find equivalent amounts of PSF and p54$^{nrb}$/NonO in this holoenzyme preparation may reflect the different stabilities of the interactions of these two proteins with the CTD during isolation of the holoenzyme.

We also used coimmunoprecipitation to ask whether PSF and p54$^{nrb}$/NonO associate with the two different forms of RNA polymerase II. Consistent with the detection of both PSF and p54$^{nrb}$/NonO in purified holoenzyme containing pol IIa, antibodies to p54$^{nrb}$/NonO (Fig. 3B) and PSF (Fig. 3C) detected these proteins in coimmunoprecipitates prepared from HeLa and Cos7 nuclear extracts using the monoclonal antibody 8WG16, which preferentially recognizes the hypophosphorylated CTD in RNA pol IIa. To verify whether PSF also associates with the hyperphosphorylated pol IIo, we used mAb B3, which has been shown previously to specifically recognize only pol IIo, and not pol IIa, through binding to specific CTD phosphoepitopes (Mortillaro et al., 1996; Patturajan et al., 1998a). Pol IIo immunoprecipitates were collected from HeLa cell nuclear extract with mAb B3 and then immunoblotted with an antiserum specific for PSF. The immunoblot shown in Figure 3D indicates that PSF is associated with pol IIo in HeLa nuclear extract. This coimmunoprecipitation appears specific, as PSF is not detected in a control immunoprecipitate prepared with an antibody (mAb-B1C8) to the splicing coactivator SRm160. Taken together with the demonstration that PSF and p54$^{nrb}$/NonO bind to both the hypo- and hyperphosphorylated forms of the CTD and associate with the holoenzyme, our results indicate that these proteins interact with both the pol IIa and pol IIo forms of RNA polymerase.

### PSF and p54$^{nrb}$/NonO can facilitate the association of RNA with the CTD

To investigate whether PSF or p54$^{nrb}$/NonO has the potential to couple the CTD to pre-mRNA processing components, we asked whether these two proteins are able to facilitate the association of RNA with an immobilized CTD matrix. GST-CTD and GST columns, either preincubated or not with PSF and p54$^{nrb}$/NonO pro-

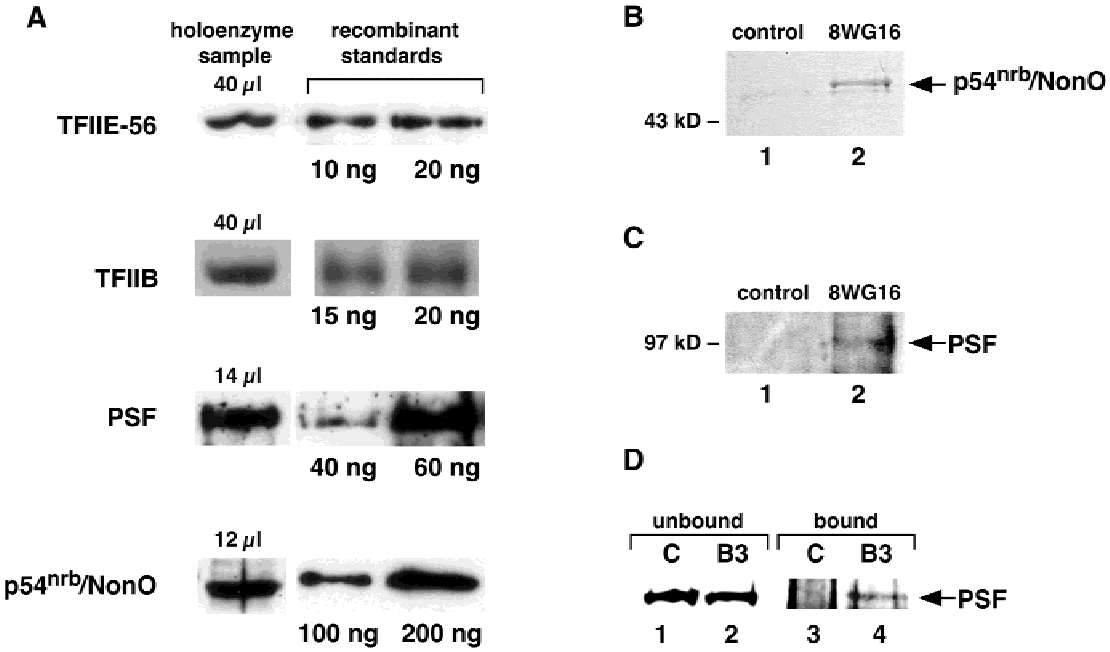


**FIGURE 3.** PSF and p54nrb/NonO are associated with both RNA pol IIa and pol IIo. **A:** Aliquots of a preparation of a TFIIS-affinity purified RNA pol II holoenzyme complex, and of preparations of purified recombinant TFIIE-56, TFIIB, PSF, and p54nrb/NonO, were fractionated by SDS-PAGE, transferred to nitrocellulose, and western blotted with the appropriate antibodies. **B:** HeLa nuclear extract was first immunoprecipitated with a control monoclonal antibody (anti-Foxp1; lane 1) or with the 8WG16 monoclonal antibody (lane 2), and the immunoprecipitates after SDS-PAGE fractionation were western blotted with anti-p54nrb/NonO polyclonal antibodies. **C:** Extracts from Cos7 cells transiently expressing human PSF were similarly immunoprecipitated with control (anti-Foxp1; lane 1) or 8WG16 antibodies (lane 2) and the fractionated immunoprecipitates were western blotted with an anti-PSF polyclonal antibody. **D:** HeLa nuclear extract was immunoprecipitated with either the phosphorylated CTD-specific monoclonal antibody B3 or a control antibody B1C8. An aliquot (5%) of the supernatants and one-half of the immunoprecipitated proteins were separated by SDS-PAGE, transferred to nitrocellulose and western blotted with anti-PSF antibodies. Immunoblotted proteins were visualized with an ECL system (Amersham Pharmacia).

teins, were assayed for their ability to retain a radiolabeled RNA transcript (see Materials and Methods for details). With the GST-CTD column preincubated with PSF and p54nrb/NonO, 68% of the RNA was retained on the column. Only 37% of the RNA was retained on the GST-CTD column not preincubated with PSF and p54nrb/NonO (Fig. 4). This increase in retention of RNA was specific, for when PSF and p54nrb/NonO were preincubated with a control GST column matrix prior to RNA addition, the level of RNA bound was similar to the background level retained on a column containing only GST alone (~10%; Fig. 4). The ability of the GST-CTD matrix to retain a fraction of RNA in the absence of PSF and p54nrb/NonO proteins suggests that the CTD itself may have RNA-binding activity, although we cannot exclude that low levels of copurifying proteins in the GST-CTD preparations might account for at least some of this binding activity. Nevertheless, the results demonstrate that a highly purified preparation of PSF and p54nrb/NonO proteins can facilitate the association of RNA with the CTD. Taken together with the results described above, our data are consistent with a role for these proteins in linking pre-mRNA processing events at different stages of transcription.

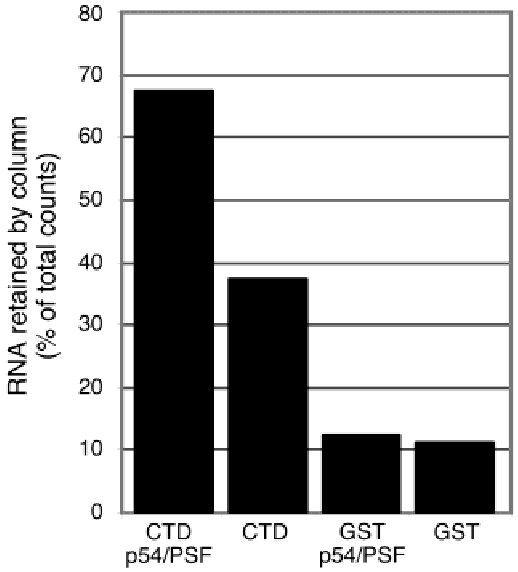


**FIGURE 4.** PSF and p54nrb/NonO increase RNA binding to a GST-CTD matrix. PSF and p54nrb/NonO (PSF/p54) were preincubated with immobilized GST-CTD or GST alone. Each column was then incubated with $^{32}$P-labeled RNA and washed. The amount of RNA retained by each column was determined. RNA retained by control GST and GST-CTD columns not incubated with PSF and p54nrb/NonO was also determined.

## DISCUSSION

Previous work has implicated PSF and p54$^{nrb}$/NonO at a number of steps in RNA processing. Both proteins preferentially associate with RNA sequences in pre-mRNA. PSF was first identified as a factor that associates with PTB, and was shown to be required at an early step in spliceosome formation (Patton et al., 1993). Subsequent data suggested that PSF may also play an important role in catalytic step II of splicing (Gozani et al., 1994). PSF was also identified as a component of a snRNP-free complex (SF-A) that contains the U1 snRNP-specific A protein and is implicated in the 3′-end cleavage of transcripts (Lutz et al., 1998). More recently, a complex consisting of PSF, p54$^{nrb}$/NonO, and matrin 3, an integral nuclear matrix protein that contains two RRMs as well as putative zinc fingers and other motifs, was shown to function in the nuclear retention of edited transcripts containing inosine residues (Zhang & Carmichael, 2001).

In addition to these roles in RNA processing, PSF and p54$^{nrb}$/NonO have also been implicated in transcriptional control. Both proteins have been shown to bind different DNA-binding nuclear hormone receptors and may modulate transcriptional activity (Basu et al., 1997; Yang et al., 1997; Urban et al., 2000; Mathur et al., 2001). Moreover, both proteins concentrate within speckle or speckle-associated domains in the nucleus that are enriched in splicing-related proteins as well as RNA pol IIo (Patton et al., 1993; Mortillaro et al., 1996; Kim et al., 1997; Dye & Patton, 2001; Fox et al., 2002). Our results demonstrating that PSF and p54$^{nrb}$/NonO bind to the CTD are consistent with these two proteins having roles in transcription as well as RNA processing and suggest that they may participate in the coupling of these processes in gene expression. Although these proteins could directly function in pre-mRNA processing when bound to the CTD, they might have a more indirect role, such as folding of the CTD into a form that can participate in pre-mRNA processing. Such a chaperone-like function for PSF and p54$^{nrb}$/NonO would be consistent with their identification in several different complexes. Both proteins could function in an analogous manner to the RRM-containing protein REF/Aly/BEF, a chaperone facilitating the DNA-binding activity of basic region-leucine zipper (bZIP) proteins (Virbasius et al., 1999), and in other contexts a factor participating in the coupling of splicing and mRNA export (Zhou et al., 2000; Rodrigues et al., 2001).

The identification of PSF and p54$^{nrb}$/NonO, in association with both pol IIa and pol IIo forms of RNA polymerase II, suggests that they could provide a link between RNA processing events at both the pol IIa-dependent steps of transcription initiation and the pol IIo-dependent elongation phases of transcription. Such a role might be effected directly by these proteins, or else indirectly through the interaction with one or more additional factors such as PTB, which binds to PSF and has multiple roles in the regulation of alternative splicing (Patton et al., 1993; reviewed in Wagner & Garcia-Blanco, 2001). Interestingly, several splicing factors, including SR family proteins, snRNP components, and the 65-kDa subunit of snRNP auxiliary factor (U2AF-65), were recently detected by western blot analysis in pol IIa holoenzyme complexes (Robert et al., 2001). Although it was not determined whether these splicing factors were binding to the CTD directly, it is possible that they, like PSF and p54$^{nrb}$/NonO, might participate in linking transcription to pre-mRNA processing prior to transcription initiation. Because the in vitro experiments of Hirose and Manley showed that purified pol IIo can promote presplicing complex formation in S100 extracts of HeLa cells that lack pol II (Hirose et al., 1999), a process that is also dependent on PSF (Patton et al., 1993), it is possible that pol IIo might be stimulating splicing complex formation through an interaction between PSF and its CTD. Similarly, as a component of the SF-A complex (Lutz et al., 1998), PSF, through binding to the CTD, may also link transcription to SF-A-dependent processing steps such as 3′-end cleavage. In conclusion, our identification of PSF and p54$^{nrb}$/NonO as CTD-binding proteins suggests that these proteins could provide a mechanism by which different steps in pre-mRNA synthesis and processing are temporally and spatially coupled within the mammalian cell nucleus.

## MATERIALS AND METHODS

### Recombinant proteins

The mouse RNA pol II CTD was expressed in *E. coli* cells as an N-terminally poly(10)histidine-tagged protein after subcloning the corresponding cDNA encoding the entire CTD from the plasmid pGCTD (Peterson et al., 1992; a gift of W. Dynan) into pET19b (Novagen). The mouse CTD was also expressed in *E. coli* cells as a GST-CTD derivative as described by McCracken et al. (1997a), using a plasmid that was a gift of D. Bentley. The *E. coli* expression plasmid for poly(10)histidine tagged PSF (Patton et al., 1993) and cDNA encoding p54$^{nrb}$/NonO (Dong et al., 1993) were obtained from J. Patton and A. Krainer, respectively. The three corresponding recombinant proteins were expressed as his-tagged products in BL21(DE3) cells and purified by $Ni^{2+}$-chelate chromatography (Emili & Ingles, 1995).

### Affinity chromatography

HeLa whole-cell extracts (10 mg protein/mL) were prepared and chromatographed on Affigel 10 (BioRad) matrices as previously described (Sopta et al., 1985), using affinity column buffer (ACB; 20 mM HEPES-NaOH, pH 7.9, 20% glycerol, 0.1 mM EDTA, 1 mM DTT, and either 0.1 M or 1.0 M NaCl).

### Protein purification and identification

Forty milliliters of HeLa cell extract were first chromatographed on a 2.0-mL column of Affigel containing no bound protein, then applied to a 1.0-mL column containing 2.0 mg/mL of immobilized-tagged CTD. After washing with 10 vol. of ACB containing 0.1 M NaCl, bound proteins were eluted by ACB containing 1.0 M NaCl, dialyzed and rechromatographed on a second CTD column. Peak fractions from the high salt eluate of this column were precipitated by addition of one-fifth of a volume of 0.15% deoxycholate and 70% TCA, separated by SDS-PAGE and stained with ultrapure Coomassie Brilliant Blue G-250 (Sigma) in 5% (v/v) acetic acid and 10% (v/v) methanol. The stained bands were excised and digested in situ with trypsin. Resulting peptides were fractionated by reverse phase HPLC and N-terminal sequences determined as described (Wang et al., 1996). Chromatography on GST-CTD matrices was performed as described by McCracken et al. (1997a). The proteins present in eluates from the GST-CTD columns (Fig. 2B) were digested in situ with trypsin and analyzed by MALDI-TOF mass spectroscopy. The RNA polymerase II holoenzyme complex was prepared from HeLa cell extracts by chromatography on a matrix of Affigel-10-coupled TFIIS as described by Pan et al. (1997). For western blots of PSF and p54$^{nrb}$/NonO, immune serum was provided by J. Patton and A. Krainer, respectively.

### Immunoprecipitation

Nuclear extract was prepared (Dignam et al., 1983) in 10 mM Tris, pH 7.6, 150 mM KCl, 0.1% NP40, 0.05% $NaN_3$, 2.5 mM $MgCl_2$, 5 mM KF, and 5 mM $\beta$-glycerophosphate from HeLa or Cos7 cells transiently expressing His-tagged PSF from the plasmid pCR3.1His-PSF. Immunoprecipitates were collected on protein A-Sepharose beads prebound with the monoclonal antibodies B1C8 (specific for the SRm160 splicing coactivator subunit; Blencowe et al., 1998); B3 (specific for pol IIo; Mortillaro et al., 1996), or 8WG16 (specific for pol IIa; Thompson et al., 1990). The immunoprecipitates were separated by SDS-PAGE and western blotted with anti-PSF antiserum or anti-p54$^{nrb}$/NonO antiserum.

### RNA binding assays

Fifty microliters of glutathione Sepharose beads, precoupled with GST-CTD (2.5 mg/mL) or GST (8 mg/mL), were equilibrated with IP100 (10 mM Tris, pH 7.6, 0.1% Nonidet P40, 2.5 mM $MgCl_2$, 5 mM KF, 5 mM $\beta$-glycerophosphate, 100 mM KCl) in micro Bio-Spin chromatography columns (BioRad). Approximately 200 ng each of p54$^{nrb}$/NonO and PSF (purified by GST-CTD affinity chromatography to near homogeneity; see Fig. 2) was applied to each column in 300 $\mu$L of IP100. PSF and p54$^{nrb}$/NonO were incubated on ice with the column matrices for 75 min with occasional mixing. Matching control columns were prepared with GST-CTD and GST but without PSF and p54$^{nrb}$/NonO. Columns were drained by gravity and were washed with 3 aliquots of 100 $\mu$L IP50 (50 mM KCl). A 320-nt $^{32}$P-RNA transcript, corresponding to the antisense strand from the middle of exon 4 to the 3′ end of the AAUAAA cleavage polyadenylation sequence of a reporter pre-mRNA derived from the Drosophila *doublesex* gene (McCracken et al., 2002), was diluted 1/100 in IP50. We applied 100 $\mu$L to each column matrix and incubated it for 10 min at room temperature with frequent gentle mixing prior to collecting fractions. The flow-through and wash fractions were collected by gravity flow. Columns were washed with 2 aliquots of IP50, 100 $\mu$L each. Flow-through, wash fractions, and the column matrices were assayed in a liquid scintillation counter by Cherenkov counting. "Percent counts retained" by each column was calculated as: column counts/(flow-through counts + wash counts + column counts) × 100.

## ACKNOWLEDGMENTS

We thank David Bentley for providing GST-CTD constructs, J. Patton and A. Krainer for antibodies against PSF and p54$^{nrb}$/NonO, respectively, Emanuel Rosonina for comments on the manuscript, and Jack Greenblatt and Brian Cox for MALDI-TOF mass spectrometry of proteins. This work was supported by a grant from the Medical Research Council/Canadian Institutes of Health Research (MRC/CIHR) (Canada) to CJI and grants from the CIHR and National Cancer Institute of Canada to BJB. RK was supported by the National Institutes of Health Grant CA45508, PWT by the National Institutes of Health Grant AI18016, and AE was a recipient of a MRC (Canada) Studentship.

*Received March 6, 2002; returned for revision March 18, 2002; revised manuscript received June 14, 2002*

## REFERENCES

Allison LA, Ingles CJ. 1989. Mutations in RNA polymerase II enhance or suppress mutations in GAL4. *Proc Natl Acad Sci USA 86*: 2794–2798.

Allison LA, Moyle M, Shales M, Ingles CJ. 1985. Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. *Cell 42*:599–610.

Barilla D, Lee BA, Proudfoot NJ. 2001. Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA 98*:445–450.

Basu A, Dong B, Krainer AR, Howe CC. 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.

Blencowe BJ, Issner R, Nickerson JA, Sharp PA. 1998. A coactivator of pre-mRNA splicing. *Genes & Dev 12*:996–1009.

Conaway JW, Shilatifard A, Dvir A, Conaway RC. 2000. Control of elongation by RNA polymerase II. *Trends Biochem Sci 25*:375–380.

Corden JL, Cadena DL, Ahearn JM Jr, Dahmus ME. 1985. A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. *Proc Natl Acad Sci USA 82*:7934–7938.

Cramer P, Caceres JF, Cazalla D, Kadener S, Muro AF, Baralle FE, Kornblihtt AR. 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.

Cramer P, Srebrow A, Kadener S, Werbajh S, de la Mata M, Melen G, Nogues G, Kornblihtt AR. 2001. Coordination between transcription and pre-mRNA processing. *FEBS Lett 498*:179–182.

Dahmus ME. 1996. Reversible phosphorylation of the C-terminal domain of RNA polymerase II. *J Biol Chem 271*:19009–19012.

Dignam JD, Lebovitz RM, Roeder RD. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res 11*:1475–1489.

Dong B, Horowitz DS, Kobayashi R, Krainer AR. 1993. Purification and cDNA cloning of HeLa cell p54nrb, a nuclear protein. *Nucleic Acids Res 21*:4085–4092.

Dye BT, Patton JG. 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, Ingles CJ. 1995. Promoter-dependent photocross-linking of the acidic transcriptional activator E2F-1 to the TATA-binding protein. *J Biol Chem 270*:13674–13680.

Fong YW, Zhou Q. 2001. Stimulatory effect of splicing factors on transcriptional elongation. *Nature 414*:929–933.

Fox AH, Lam YW, Leung AK, Lyon CE, Andersen J, Mann M, Lamond AI. 2002. Paraspeckles. A novel nuclear domain. *Curr Biol 12*: 13–25.

Gerber HP, Hagmann M, Seipel K, Georgiev O, West MA, Litingtung Y, Schaffner W, Corden JL. 1995. RNA polymerase II C-terminal domain required for enhancer-driven transcription. *Nature 374*: 660–662.

Gozani O, Patton JG, Reed R. 1994. A novel set of spliceosome-associated proteins 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.

Hirose Y, Manley JL. 1998. RNA polymerase II is an essential mRNA polyadenylation factor. *Nature 395*:93–96.

Hirose Y, Manley JL. 2000. RNA polymerase II and the integration of nuclear events. *Genes & Dev 14*:1415–1429.

Hirose Y, Tacke R, Manley JL. 1999. Phosphorylated RNA polymerase II stimulates pre-mRNA splicing. *Genes & Dev 13*:1234–1239.

Ho CK, Sriskanda V, McCracken S, Bentley D, Schwer B, Shuman S. 1998. The guanylyltransferase domain of mammalian mRNA capping enzyme binds to the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J Biol Chem 273*:9577–9585.

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

Kim YJ, Bjorklund S, Li Y, Sayre MH, Kornberg RD. 1994. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell 77*:599–608.

Koleske AJ, Young RA. 1994. An RNA polymerase II holoenzyme responsive to activators [see comments]. *Nature 368*:466–469.

Komarnitsky P, Cho EJ, Buratowski S. 2000. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes & Dev 14*:2452–2460.

Lu H, Flores O, Weinmann R, Reinberg D. 1991. The nonphosphorylated form of RNA polymerase II preferentially associates with the preinitiation complex. *Proc Natl Acad Sci USA 88*:10004–10008.

Lutz CS, Cooke C, O'Connor JP, Kobayashi R, Alwine JC. 1998. The snRNP-free U1A (SF-A) complex(es): Identification of the largest subunit as PSF, the polypyrimidine-tract binding protein-associated splicing factor. *RNA 4*:1493–1499.

Mathur M, Tucker PW, Samuels HH. 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.

McCracken S, Fong N, Rosonina E, Yankulov K, Brothers G, Siderovski D, Hessel A, Foster S, Shuman S, Bentley DL. 1997b. 5′-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes & Dev 11*:3306–3318.

McCracken S, Fong N, Yankulov K, Ballantyne S, Pan G, Greenblatt J, Patterson S, Wickens M, Bentley DL. 1997a. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature 385*:357–361.

McCracken S, Lambermon M, Blencowe BJ. 2002. SRm160 splicing coactivator promotes transcript 3′-end cleavage. *Mol Cell Biol 22*:148–160.

Misteli T, Spector DL. 1999. RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo. *Mol Cell 3*:697–705.

Morris DP, Greenleaf AL. 2000. The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J Biol Chem 275*:39935–39943.

Mortillaro MJ, Blencowe BJ, Wei X, Nakayasu H, Du L, Warren SL, Sharp PA, Berezney R. 1996. A hyperphosphorylated form of the large subunit of RNA polymerase II is associated with splicing complexes and the nuclear matrix. *Proc Natl Acad Sci USA 93*:8253–8257.

Moyle M, Lee JS, Anderson WF, Ingles CJ. 1989. The C-terminal domain of the largest subunit of RNA polymerase II and transcription initiation. *Mol Cell Biol 9*:5750–5753.

Nonet M, Sweetser D, Young RA. 1987. Functional redundancy and structural polymorphism in the large subunit of RNA polymerase II. *Cell 50*:909–915.

Ossipow V, Tassan JP, Nigg EA, Schibler U. 1995. A mammalian RNA polymerase II holoenzyme containing all components required for promoter-specific transcription initiation. *Cell 83*:137–146.

Otto H, Dreger M, Bengtsson L, Hucho F. 2001. Identification of tyrosine-phosphorylated proteins associated with the nuclear envelope. *Eur J Biochem 268*:420–428.

Pan G, Aso T, Greenblatt J. 1997. Interaction of elongation factors TFIIS and elongin A with a human RNA polymerase II holoenzyme capable of promoter-specific initiation and responsive to transcriptional activators. *J Biol Chem 272*:24563–24571.

Patton JG, Porro EB, Galceran J, Tempst P, Nadal-Ginard B. 1993. Cloning and characterization of PSF, a novel pre-mRNA splicing factor. *Genes & Dev 7*:393–406.

Patturajan M, Schulte RJ, Sefton BM, Berezney R, Vincent M, Bensaude O, Warren SL, Corden JL. 1998a. Growth-related changes in phosphorylation of yeast RNA polymerase II. *J Biol Chem 273*:4689–4694.

Patturajan M, Wei X, Berezney R, Corden JL. 1998b. A nuclear matrix protein interacts with the phosphorylated C-terminal domain of RNA polymerase II. *Mol Cell Biol 18*:2406–2415.

Pei Y, Hausmann S, Ho CK, Schwer B, Shuman S. 2001. The length, phosphorylation state, and primary structure of the RNA polymerase II carboxyl-terminal domain dictate interactions with mRNA capping enzymes. *J Biol Chem 276*:28075–28082.

Peterson SR, Dvir A, Anderson CW, Dynan WS. 1992. DNA binding provides a signal for phosphorylation of the RNA polymerase II heptapeptide repeats. *Genes & Dev 6*:426–438.

Robert F, Blanchette M, Maes O, Chabot B, Coulombe B. 2001. A human RNA polymerase II-containing complex associated with factors necessary for spliceosome assembly. *J Biol Chem 31*:31.

Rodrigues JP, Rode M, Gatfield D, Blencowe B, Carmo-Fonseca M, Izaurralde E. 2001. REF proteins mediate the export of spliced and unspliced mRNAs from the nucleus. *Proc Natl Acad Sci USA 98*:1030–1035.

Scafe C, Chao D, Lopes J, Hirsch JP, Henry S, Young RA. 1990. RNA polymerase II C-terminal repeat influences response to transcriptional enhancer signals. *Nature 347*:491–494.

Schroeder SC, Schwer B, Shuman S, Bentley D. 2000. Dynamic association of capping enzymes with transcribing RNA polymerase II. *Genes & Dev 14*:2435–2440.

Sopta M, Carthew RW, Greenblatt J. 1985. Isolation of three proteins that bind to mammalian RNA polymerase II. *J Biol Chem 260*: 10353–10360.

Steinmetz EJ. 1997. Pre-mRNA processing and the CTD of RNA polymerase II: The tail that wags the dog? *Cell 89*:491–494.

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

Thompson NE, Aronson DB, Burgess RR. 1990. Purification of eukaryotic RNA polymerase II by immunoaffinity chromatography. Elution of active enzyme with protein stabilizing agents from a polyol-responsive monoclonal antibody. *J Biol Chem 265*:7069–7077.

Urban RJ, Bodenburg Y, Kurosky A, Wood TG, 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.

Vincent M, Lauriault P, Dubois M, Lavoie S, Bensaude O, Chabot B. 1996. The nuclear matrix protein p255 is a highly phosphorylated form of RNA polymerase II largest subunit which associates with spliceosomes. *Nucleic Acids Res 24*:4649–4652.

Virbasius CM, Wagner S, Green MR. 1999. A human nuclear-localized chaperone that regulates dimerization, DNA binding, and transcriptional activity of bZIP proteins. *Mol Cell 4*:219–228.

Wagner EJ, Garcia-Blanco MA. 2001. Polypyrimidine tract binding protein antagonizes exon definition. *Mol Cell Biol 21*:3281–3288.

Wang R, Kobayashi R, Bishop JM. 1996. Cellular adherence elicits ligand-independent activation of the Met cell-surface receptor. *Proc Natl Acad Sci USA 93*:8425–8430.

Yang YS, Hanke JH, Carayannopoulos L, Craft CM, Capra JD, Tucker PW. 1993. NonO, a non-POU-domain-containing, octamer-binding protein, is the mammalian homolog of *Drosophila* nonAdiss. *Mol Cell Biol 13*:5593–5603.

Yang YS, Yang MC, Tucker PW, Capra JD. 1997. NonO enhances the association of many DNA-binding proteins to their targets. *Nucleic Acids Res 25*:2284–2292.

Yue Z, Maldonado E, Pillutla R, Cho H, Reinberg D, Shatkin AJ. 1997. Mammalian capping enzyme complements mutant *Saccharomyces cerevisiae* lacking mRNA guanylyltransferase and selectively binds the elongating form of RNA polymerase II [see comments]. *Proc Natl Acad Sci USA 94*:12898–12903.

Yuryev A, Patturajan M, Litingtung Y, Joshi RV, Gentile C, Gebara M, Corden JL. 1996. The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins. *Proc Natl Acad Sci USA 93*:6975–6980.

Zeng C, Berget SM. 2000. Participation of the C-terminal domain of RNA polymerase II in exon definition during pre-mRNA splicing. *Mol Cell Biol 20*:8290–8301.

Zhang WW, Zhang LX, Busch RK, Farres J, 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, Carmichael GG. 2001. The fate of dsRNA in the nucleus: A p54(nrb)-containing complex mediates the nuclear retention of promiscuously A-to-I edited RNAs. *Cell 106*:465–475.

Zhou Z, Luo MJ, Straesser K, Katahira J, Hurt E, Reed R. 2000. The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. *Nature 407*:401–405.