Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 6975–6980, July 1996
Biochemistry

The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins

(transcription/RNA processing/two-hybrid assay/serine/arginine-rich proteins)

ANTON YURYEV*†, MEERA PATTURAJAN*, YING LITINGTUNG, RAVI V. JOSHI‡, CRISTL GENTILE, MAHA GEBARA, AND JEFFRY L. CORDEN§

Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, 725 North Wolfe Street Baltimore, MD 21205-2185

Communicated by Hamilton O. Smith, Johns Hopkins University School of Medicine, Baltimore, MD, February 28, 1996 (received for review January 31, 1996)

ABSTRACT **Although transcription and pre-mRNA processing are colocalized in eukaryotic nuclei, molecules linking these processes have not previously been described. We have identified four novel rat proteins by their ability to interact with the repetitive C-terminal domain (CTD) of RNA polymerase II in a yeast two-hybrid assay. A yeast homolog of one of the rat proteins has also been shown to interact with the CTD. These CTD-binding proteins are all similar to the SR (serine/arginine-rich) family of proteins that have been shown to be involved in constitutive and regulated splicing. In addition to alternating Ser-Arg domains, these proteins each contain discrete N-terminal or C-terminal CTD-binding domains. We have identified SR-related proteins in a complex that can be immunoprecipitated from nuclear extracts with antibodies directed against RNA polymerase II. In addition, *in vitro* splicing is inhibited either by an antibody directed against the CTD or by wild-type but not mutant CTD peptides. Thus, these results suggest that the CTD and a set of CTD-binding proteins may act to physically and functionally link transcription and pre-mRNA processing.**

The C-terminal domain of the largest subunit of RNA polymerase II (CTD) consists of tandem repeats of the consensus sequence Tyr-Ser-Pro-Thr-Ser-Pro-Ser (1, 2). Deletion studies demonstrated that the CTD is essential for cell growth (3–6), but the nature of this essential function is not known. The CTD is only found on RNA polymerase II (pol II), suggesting that it plays a unique role in mRNA biogenesis (7). While most experiments to date have focused on a potential role in transcription initiation, other functions have not been excluded.

Several potentially important CTD interactions have been reported. (*i*) The CTD has been suggested to bind DNA in a nonspecific interaction that may function to remove inhibitory proteins from the promoter (8). (*ii*) Another potential interaction partner is the TATA-binding protein (TBP) (9). Pol IIA (pol II hypophosphorylated on the CTD) can interact directly with TBP, while pol IIO (pol II hyperphosphorylated on the CTD) is incapable of this interaction, a result that is consistent with the proposed role for CTD phosphorylation (for review, see ref. 10). The CTD has also been shown to bind the mediator complex that contains SRB [suppressor of RNA pol II (B) mutations] proteins and other transcription factors (11, 12). The mediator complex associates with pol II to form a holoenzyme that can respond to activator proteins *in vitro* (11, 12), but the component(s) of the mediator that directly contact the CTD are not known. Finally, crosslinking studies suggest that subunits of transcription factors TFIIE and TFIIF interact with the CTD (13). The crosslinking agent in these experiments was not located in the heptapeptide repeat domain, however.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Despite identification of interaction partners, the role of the CTD in transcription remains unclear. The CTD is not required for either basal (14, 15) or activated (16, 17) transcription of some genes *in vitro*. Furthermore, inhibition of CTD kinase does not block *in vitro* transcription from the adenovirus major late promoter or from a GAL4 VP16-activated promoter (18, 19). Thus, these results indicate that the CTD is not essential for specific initiation at some promoters.

CTD function may be required for postinitiation steps in the biogenesis of mRNA. O'Brien *et al.* (20) have demonstrated that several genes contain paused pol IIA complexes that can reenter the elongation mode coincident with CTD phosphorylation. In yeast, CTD-truncated pol II synthesizes an excess of GAL4-induced promoter proximal transcripts (D. L. Bentley, personal communication). Thus, these results argue that the CTD plays an important role subsequent to initiation. While the CTD has previously been proposed to function in pre-mRNA processing (refs. 7 and 21 and H. Rienhoff and J. Boeke, personal communication), no experimental data have yet supported these models.

We used the yeast two-hybrid system (22) to identify proteins that interact with the CTD. This unbiased approach did not yield proteins that are expected to be involved in transcription initiation, like TBP or the SRBs, but rather a set of proteins similar to RNA processing factors. In this paper we report the identification and characterization of four rat cDNAs and a yeast homolog that encode CTD-binding proteins whose similarity to serine/arginine-rich (SR) proteins suggests a role for the CTD in processing of pre-mRNA.

MATERIALS AND METHODS

Plasmids. The GAL4–CTD fusion plasmids contain CTD sequences fused in-frame to the GAL4 DNA-binding domain [GAL4 (DB)] (amino acids 1–147) of pPC62 as described (23). The mouse CTD sequence used extends from an *Spe*I restriction site in repeat 36 to the C terminus (24). This segment of the mouse CTD was also cloned into pY1at (25) and this plasmid was used to transform yeast. A plasmid shuffle assay was then used to test the ability of the murine CTD repeats to provide CTD

Abbreviations: CTD, C-terminal domain of the largest subunit of RNA polymerase II; pol II, RNA polymerase II; pol IIA, pol II hypophosphorylated on the CTD; pol IIO, pol II hyperphosphorylated on the CTD; TBP, TATA-binding protein; GST, glutathione *S*-transferase; GAL4(DB), GAL4 DNA-binding domain; GAL4(TA), GAL4 transcription activation domain; HA, influenza virus hemagglutinin.

Data deposition: The sequences reported in this paper have been deposited in GenBank database (accession nos. U49055, U49056, U49057, and U49058).

*A.Y. and M.P. contributed equally to this paper.
†Present address: CIBA–Geigy Co., 556 Morris Avenue, Summit, NJ 07901.
‡Present address: Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139.
§To whom reprint requests should be addressed.

function in yeast (25). Wild-type and mutant yeast CTD sequences used to construct GAL4 fusions are as described (25).

Library Screening. The GAL4–mouse CTD fusion was used to screen a 14.5-day mouse embryo yeast two-hybrid library (23). Positive clones were isolated by two rounds of screening and retested by isolation of the library plasmid and retransformation into yeast. Positive clones were further tested for their inability to activate with a set of heterologous baits including RAG-1, RAG-2, DHL4, and yeast TYA or with the GAL4(DB) alone. RAG GAL4(DB) fusions were provided by Steve Desiderio (Johns Hopkins University) and expressed the RAG genes (26). The DHL4 fusion expresses a novel helix–loop–helix transcription factor and was provided by Greg Kato (Johns Hopkins University). The yeast TYA fusion was obtained from C. B. Brachmann and J. D. Boeke (Johns Hopkins University).

Full-length cDNAs for the CTD-binding proteins were isolated from a rat hippocampus library [a gift from Anthony Lanahan (27)] using the short mouse clones as probe. Full-length or nearly full-length clones of rA1, rA4, rA8, and rA9 were sequenced.

RNA Pol II Binding Assay. The CTD-binding domain of rA1 was cloned in pGEX (Pharmacia) and expressed as a glutathione *S*-transferase (GST) fusion protein. Affinity beads containing either the rA1–GST fusion or GST alone were prepared by passing bacterial extracts over a glutathione-agarose column (Pharmacia). The columns retained rA1–GST at ≈1 mg/ml or GST at 1.5 mg/ml. Yeast RNA pol II (1.0 μg) was labeled with hemagglutinin (HA)-epitope-tagged cdc2 kinase and [γ-^{32}P]ATP (28). The kinase in this reaction was retained on monoclonal antibody (mAb) 12CA5 beads during the reaction. Pol IIA was prepared by labeling for 5 min with 10 μCi of ATP (3000 Ci/mmol; 1 Ci = 37 GBq) to a stoichiometry of less than 1. In a second reaction the labeled ATP was chased for an additional 5 min with 0.2 mM ATP. This chase is sufficient to shift the mobility of the largest subunit to that expected for pol II0. Labeled pol II was removed from the kinase reaction by centrifugation of the kinase beads and incubated with 25 μl of rA1–GST or GST beads in a final volume of 100 μl of binding buffer (0.1 M NaCl/10 mM Tris·HCl, pH 7.9). After binding for 15 min at 23°C, the reaction mixture was centrifuged at 5000 × *g* for 30 s. The beads were then washed two more times with 1.0 ml of binding buffer. The final pellet was resuspended in 2× SDS sample buffer, boiled for 90 s, electrophoresed on a 5% SDS gel, and autoradiographed. Virtually all of the labeled pol II was retained on the rA1–GST column. For peptide competition experiments, labeled yeast pol IIA was incubated with rA1–GST or GST on beads but with the addition of 10 μg of synthetic CTD peptide. After the binding reaction, the reaction mixtures were diluted and washed as above in the absence of peptide, and material was eluted and electrophoresed on a 5% SDS gel.

Peptides used in this study are as follows: WT, $(SPTSPSY)_8$; A2, $(APTSPSY)_8$; A5, $(SPTAPSY)_8$. Peptides were synthesized from *t*-butoxycarbonyl-labeled amino acids and their concentrations were determined as described (29).

Immunoprecipitation of Pol II Complexes. Nuclear extracts were prepared from wild-type HeLa cells and from a HeLa strain transfected with an α-amanitin-resistant HA-epitope-tagged mouse pol II large subunit gene (M. L. West and J.L.C., unpublished results). Nuclear extract (50 μg) was incubated with 4 μg of purified anti-CTD mAb 8WG16 (14) or anti-HA-tag mAb 12CA5 (30) in 500 μl of modified RIPA buffer (1 mM Tris·HCl, pH 7.8/150 mM NaCl/2% Triton X-100/0.01% SDS/2 mM pepstatin/0.6 mM leupeptin/2 mM benzamidine hydrochloride/1 mM phenylmethylsulfonyl fluoride/2.5% BSA) at 4°C on a rotator for 4 hr. Protein A-Sepharose beads (20 μl, Pharmacia), equilibrated with modified RIPA buffer, was added to the above mixture and further incubated at 4°C on a rotator for 1 hr. Beads were then washed three times with 1 ml of ice-cold modified RIPA buffer for 30 min. The final bead pellet was extracted in SDS sample buffer and subjected to SDS/PAGE followed by Western blot. The blots were probed with anti-SR protein mAb 16H3 (31) and the immunoreactive proteins were detected using ECL (Amersham).

***In Vitro* Splicing.** Capped and uniformly labeled [^{32}P]E1-IVS-E2 RNA was synthesized from pPIP85 (32) using T7 RNA polymerase (Promega). The RNA product was separated on a denaturing 8% polyacrylamide gel and eluted by crushing the gel and incubating with elution buffer containing 0.3 M sodium acetate, 1/3 vol of phenol, and tRNA (20 μg/ml) for 2 hr at 30°C on a rotator. The eluted RNA was separated from gel pieces by a Millipore Ultrafree-MC filter unit (0.45 μm), precipitated using 2 vol of ethanol, and dissolved in nuclease-free water (20–50 μl). Splicing reactions (12.5 μl) containing uniformly labeled pre-mRNA (1 to 5 × 10^5 cpm) were performed at 30°C in a solution of 24% HeLa nuclear extract (Promega), 2 mM $MgCl_2$, 60 mM KCl, 1 mM ATP, 20 mM creatine phosphate, and RNasin (Promega) at 1 unit/ml (32). After incubation for 2 hr, RNAs were extracted and subjected to denaturing 15% polyacrylamide gel electrophoresis followed by autoradiography.

RESULTS

Design of the Two-Hybrid Screen. Initial attempts at using the CTD as bait in the two-hybrid system failed because the CTD itself acts as an activator when fused to a DNA-binding domain. Several groups (33, 34) have reported similar observations and have argued that the ability to activate transcription is an important CTD function. We constructed a series of GAL4(DB)–CTD fusions as potential bait, including a series of mutant CTDs that had been tested for their ability to function in place of the wild-type CTD (25). Table 1 describes the CTD fusions tested for activation. The ability to activate transcription in a one-hybrid assay does not correlate with the ability to sustain viability when in the context of the pol II largest subunit gene in yeast. All of the phosphorylation site mutant CTDs activate, although they do not support viability. Transcription activation provided by GAL4(DB)–CTD fusion proteins is not an essential function of the CTD because the 16 most distal repeats of the mouse CTD are functional in the context of yeast pol II but do not activate transcription as a GAL4(DB) fusion. Phosphorylation of the CTD fusion proteins *in vivo* (data not shown) apparently creates an acidic activation domain. Multiple positively charged amino acids in the nonconsensus part of mouse CTD may neutralize this acidic activation potential. The failure of the distal repeats of the mouse CTD to activate allowed us to use this domain as bait in a two-hybrid screen.

Isolation of CTD-Interacting Clones. We screened a mouse embryo cDNA two-hybrid library (23) and isolated 46 positive clones out of 1.4 × 10^6 colonies. None of these positive clones interacted with a series of control proteins described above. Positive clones were sequenced and placed into groups based on

Table 1. Activation and RNA pol II functions of CTD mutants

CTD mutant	Activation in one-hybrid assay	CTD function in yeast pol II
WT yeast CTD	+	+
(YAPTSPS)18	+	–
(YEPTSPS)15	+	–
(YSPTAPS)15	+	–
(YSPTEPS)18	+	–
(FSPTSPS)10	+	–
(YAPTAPS)10	+	–
(YPSTSPS)15	+	–
Mouse CTD	–	+

Series of mutant CTDs shown were cloned and tested either for their ability to activate transcription as GAL4-(DB) fusions or for their ability to replace the wild-type CTD as part of the *RPB1* gene in yeast (25). The sequence of the repeated heptapeptide is given within parentheses with the mutated residue(s) indicated in boldface type. The number of such repeats is noted outside the parentheses. The mouse CTD construct contains repeats 36–52.

sequence similarity. The largest group consists of 35 clones containing four related sequences (mA1, mA5, mA7, and mA9). The second group of positive clones consists of two related sequences (mA4 and mA8) that are distinct from the first group. Mouse A series clones were also tested by reversing the two-hybrid partners. In this arrangement the CTD is fused to a GAL4 transcription activation domain [GAL4(TA)] and the mouse A series clone is fused to the GAL4(DB). In this orientation both the mA1 and mA8 clones were shown to interact with both mouse and yeast CTDs, indicating that their CTD-binding domains also recognize the consensus CTD.

Six of the 46 positive clones did not fall into the above groups. These clones contain three different sequences encoding proteins that interacted poorly with the CTD when tested as GAL4(DB) fusions. We have not pursued these clones further. Notably, this set does not contain TBP, SRBs, or any general transcription factor. We tested GAL4(DB)–TBP fusions (full-length human TBP or the N- or C-terminal halves) for their ability to interact with the CTD. No CTD interaction was observed (data not shown), suggesting that the CTD and TBP do not interact *in vivo*. The positive mouse CTD-interacting clones that represent the first two interacting groups were selected for further study.

Full-Length CTD-Binding Proteins. Using the short mouse two-hybrid clones as probes, we were able to isolate full-length clones for four different CTD-interacting proteins and have named these rat genes rA1, rA4, rA8, and rA9. The rat clones were retested in the two-hybrid system and all four activated in a CTD-dependent fashion (data not shown). We have sequenced all four clones and determined that they are novel cDNAs encoding proteins that are similar to SR proteins (35, 36).

Fig. 1*A* shows the sequence organization of four rat CTD-binding proteins. One common feature, the RS (Arg-Ser) domain, contains a concentration of Arg-Ser diamino acid sequences similar to that seen in SR proteins known to be involved in splicing (Fig. 1*B*). These RS domains also have related diamino acid sequences conforming to the general consensus Arg(Lys)-Ser(Thr, Asp, Glu). When the Ser (or Thr) residues are phosphorylated (35, 36), all of these diamino acid sequences display alternating positive and negative charges.

Two different CTD-binding domains are represented in this set of four proteins. rA1 and rA9 contain a very similar ≈80-amino acid sequence that is shown in Fig. 1*C*. This conserved region lies at the C terminus and is nearly identical to the mouse clones (mA1 and mA9, respectively) used as the probe to identify the rat clones. We have not yet isolated a rat clone that is the equivalent of the mouse A5 clone. Several experiments suggested that the conserved ≈80-amino acid sequence is the CTD-binding domain. (*i*) Deletion of sequences to either side of this homology region in the mouse mA1 clone did not eliminate interaction with the CTD in the two-hybrid assay (Fig. 1*A*, regions a and b). (*ii*) We have cloned a pair of rA9 splice variants that contain different C-terminal sequences and no CTD-binding homology region. These clones do not interact with the CTD in the two-hybrid assay (data not shown).

rA4 and rA8 contain similar N-terminal CTD-interacting domains that are different than the rA1/A9 CTD-binding sequence. The conserved ≈120-amino acid homology regions from rA4 and rA8 that are shown in Fig. 1*D* aligned with the sequences of the original mouse clones mA4 and mA8, respectively. The sequence of the longest rA4 clone differs from its mouse homologue in that a 50-bp segment is deleted, causing a frame shift in the CTD-binding domain (Fig. 1*D*). Conservation of the partial rA4 CTD-binding sequence and the existence of a long open reading frame 3′ of the deletion suggests that this cDNA derives from an active gene. As was observed for rA9, we suspect that the deletion of sequences in rA4 results from alternative splicing.

rA4 and rA8 contain other sequence motifs that are common to RNA-binding proteins. Immediately downstream of the RS domain both proteins contain a similar RNA recognition motif (37). Database searches using the program BLAST (38) indicate

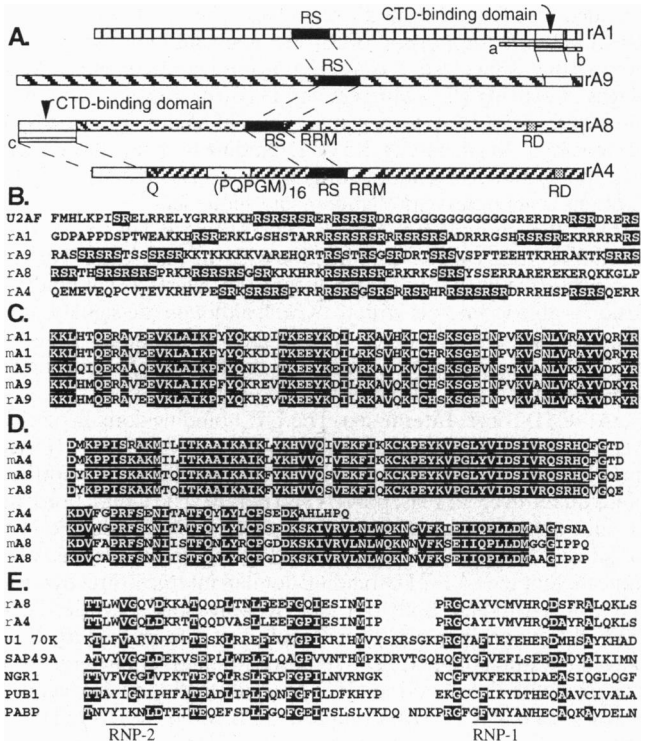

FIG. 1. Sequence alignments of rat CTD-binding proteins. (*A*) Sequence organization of rat CTD-binding proteins rA1, rA9, rA4, and rA8. The following domains are indicated; RS, Arg/Ser rich; RRM, RNA recognition motif; RD, Arg/Asp rich; Q, Gln rich; (PQPGM), repeated pentapeptide in rA4. Lines under the rA1 and rA8 clones marked a, b, and c represent the smallest mouse clones that interact with the CTD in the two-hybrid assay. The sizes of the expected rat proteins are as follows: rA1, 129 kDa; rA9, 146 kDa; rA8, 144 kDa; rA4, 130 kDa. The rA4 protein size is based on a fusion of mouse and rat sequences as the longest rat cDNA clone contains a short deletion that leads to a frame shift in the CTD-binding domain. (*B*) Conserved RS domains. The RS domains of rA1, rA9, rA4, and rA8 are compared with a previously identified SR protein (U2AF, GenBank accession no. X64044). Black-boxed residues represent Arg-Ser and Ser-Arg dipeptides. The SR proteins alignments were taken from the output of BLAST searches. Because of the repetitive nature of the homology, multiple alignments are possible. (*C*) rA1/A9 CTD-binding domain. The rat sequences are aligned with the homologous mouse clones and conserved residues are in black boxes. (*D*) rA4/A8 CTD-binding domain. The rat clones are aligned with the homologous mouse clones and conserved residues are in black boxes. (*E*) rA4/A8 RNA recognition motifs. The RNA recognition motifs of rA4 and rA8 are aligned with a set of related RNA-binding proteins (accession nos. in parentheses) including the human U1 70-kDa (S03048) and SAP49A (L35013) splicing factors and yeast RNA-binding proteins Ngr1p (P32831), Pub1p (P32588), and Pab1p (P04147). Black-boxed residues are identical among at least four of the sequences. In addition to these residues, there are many similarities that are not boxed. The positions of the two motifs, RNP-1 and RNP-2 that make up the RRM are underlined.

that rA4 and rA8 are related to a number of RNA-binding proteins (Fig. 1*E*), some of which, like the U1 70-kDa protein, are involved in splicing (39). The strongest region of homology to other RNA-binding proteins is in the RNP-2 motif and adjacent sequences. The similarity to the RNP-1 motif is not as strong. In addition to the RS and RRM motifs, both rA4 and rA8 contain glutamine-rich regions at their N termini and a motif rich in Arg-Asp diamino acid sequences (RD) toward the C terminus. rA4 is unique in having a pentapeptide sequence (Pro-Gln-Pro-Gly-Met) that is repeated 16 times.

A Yeast CTD-Binding Protein Homolog. BLAST searches on the National Center for Biotechnology Information network (37)

revealed a sequence related to rA4/A8. The yeast *NRD1* gene, identified as a suppressor of an intronic mutation, encodes a protein that is involved in transcription elongation or the earliest stages of splicing (E. Steinmetz and D. Brow, personal communication). *NRD1* contains several regions of homology to the rA8 protein (Fig. 2*A*) including the CTD-binding domain, alternating Arg motif, and RRM motif. In addition to RE repeats, *NRD1* contains several Ser-Arg diamino acid sequences.

The N-terminal residues of *NRD1* were fused to the GAL4(TA) and tested for interaction with the mouse CTD in the two-hybrid assay. Fig. 2*B* shows that the *NRD1* CTD-binding fusion is able to interact with the CTD, although the signal is not as strong as seen with rA8. This weaker signal may be due to competition with endogenous Nrd1p or to a less efficient interaction with the nonconsensus murine heptapeptides.

rA1–CTD Direct Interaction. The CTD-binding domain of rA1 was expressed as a GST fusion protein and binding to pol II was assayed. Fig. 3*A* demonstrates that the rA1–GST fusion protein binds quantitively to yeast pol IIA and pol II0. The interaction can be substantially competed by a wild-type CTD peptide and less effectively by mutant CTD peptides (Fig. 3*B*). Thus, these results indicate that the rA1 CTD-binding domain interacts directly with the CTD. Although both IIA and II0 were bound by the rA1–GST fusion protein, we cannot rule out the possibility that phosphorylation of the CTD is required for interaction as our IIA was phosphorylated on the CTD to a substoichiometric level.

RNA Pol II–SR Protein Complex. Pol II complexes from HeLa cell nuclear extracts were immunoprecipitated with a mAb directed against pol II (14) or with mAb 12CA5, which recognizes an epitope tag in the largest subunit of pol II. Proteins in antibody–antigen complexes were separated by SDS/PAGE and examined by immunoblot analysis with mAb 16H3, which recognizes alternating Arg regions in a subset of SR proteins and inhibits splicing (31). Fig. 4 shows that antibodies directed against pol II coimmunoprecipitate a set of three 16H3-reactive proteins of ≈100, 140, and 200 kDa. A similar set of three proteins is also precipitated from nuclear extract prepared from rat testis, a tissue that contains high levels of several CTD-binding protein mRNAs (data not shown). We have also used affinity chromatography to demonstrate that SR-like proteins of apparent molecular masses of 70 kDa and 140 kDa interact with the CTD (data not shown).

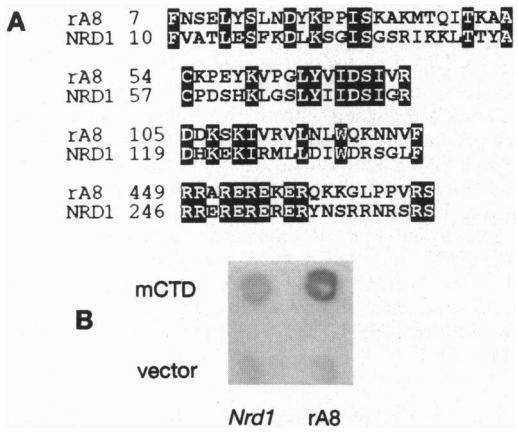

FIG. 2. Nrd1p is a CTD-binding protein. (*A*) The yeast *NRD1* gene shares regions of homology with CTD-binding protein rA8. Identical residues are indicated by black boxes. The top two homologies are in the CTD-binding domain, while the third makes up the RS/RE motif. In addition, *NRD1* and rA8 contain similarly positioned RRM motifs (not shown). (*B*) A *NRD1* fusion protein interacts with the mouse CTD in a yeast two-hybrid assay. The *NRD1* CTD-binding domain was obtained by PCR, using yeast genomic DNA as template, cloned into GAL4(TA) fusion vector, and assayed for β-galactosidase as described (23). The interaction of the rA8 GAL4(TA) fusion protein with the GAL4(DB)–CTD (mouse) fusion protein (mCTD) and the GAL4(DB) vector alone are shown for comparison.

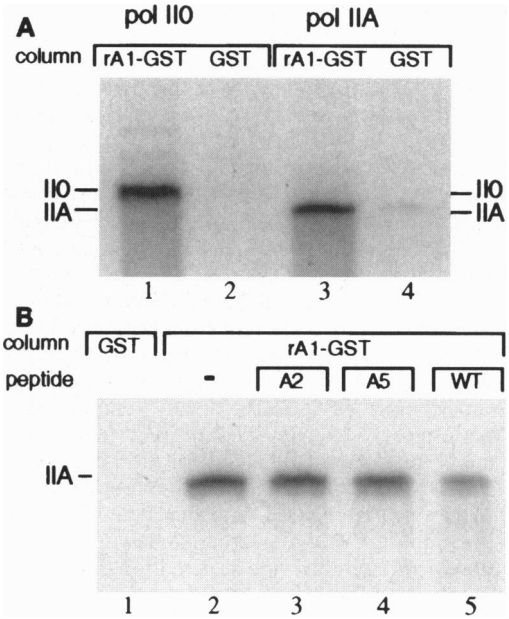

FIG. 3. Direct interaction between rA1 and RNA pol II. (*A*) Retention of RNA pol II on an rA1 CTD-binding domain column. Yeast pol II0 (lanes 1 and 2) or pol IIA (lanes 3 and 4) were incubated with glutathione-agarose beads containing bound GST–rA1 CTD-binding domain fusion protein (lanes 1 and 3) or GST alone (lanes 2 and 4). Bound proteins were eluted and electrophoresed on an SDS gel. The positions of the labeled II0 and IIA subunits are indicated. (*B*) RNA pol II binding is inhibited by CTD peptide. Binding of pol IIA to GST–rA1 beads in the presence of 10 μg of peptides WT, A2, and A5.

The ≈140-kDa band is similar in size to the 16H3-immunoreactive protein immunoprecipitated with anti-pol II mAb 8WG16 and is close to the size of a predominant 16H3-reactive protein reported by Neugebauer *et al.* (31). While we have not shown that these proteins are CTD-binding proteins, they are close to the expected sizes of full-length rA1 and rA8 proteins.

The CTD Is Involved in Splicing. To directly test for CTD function in splicing, we monitored the effect of adding anti-CTD mAb 8WG16 or CTD peptides to *in vitro* splicing reactions. Fig.

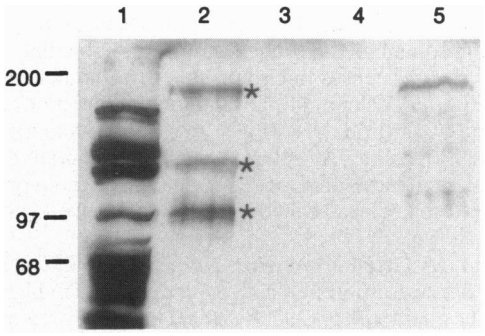

FIG. 4. Interaction between SR proteins and RNA pol II. Pol II complexes were immunoprecipitated from HeLa nuclear extracts and examined by immunoblot analysis with anti-SR protein mAb 16H3. This antibody recognizes a family of SR proteins in HeLa nuclear extract (lane 1). Anti-SR protein mAb 16H3 immunoreactive bands of ≈100 kDa, 140 kDa, and 180 kDa (asterisks) are precipitated with anti-CTD mAb 8WG16 from HeLa extracts (lane 2). The same bands are not seen when heterologous mAb 12CA5 or no antibody is used in the HeLa cell nuclear extract (lanes 3 and 4). A similar set of SR-related proteins are seen in 12CA5 immunocomplexes when the HeLa cell nuclear extract is derived from cells transfected with an HA-epitope-tagged α-amanitin-resistant pol II large subunit gene (lane 5).

5*A* shows that mAb 8WG16 inhibits splicing in a concentration-dependent manner but the control anti-HA-tag mAb 12CA5 does not. The concommitant loss of the reaction product and reaction intermediates indicates that the anti-CTD antibody inhibits an early step in splicing. Fig. 5*B* shows that a peptide consisting of eight consensus CTD repeats is also able to inhibit splicing of the adenovirus 2 leader exons. The failure to observe intermediates at the highest concentration of peptide indicates that an early step in splicing is blocked. At lower concentrations we observe an accumulation of lariet-containing intermediates and free exon 1 (data not shown), suggesting that the cleavage and splicing reactions have been uncoupled. Two mutant peptides in which potential phosphorylation sites have been changed to Ala are not inhibitory, although they are substrates for Cdc2 phosphorylation (Fig. 5*C*).

DISCUSSION

A two-hybrid screen was used to identify a family of proteins that interact with the CTD. The similarity of these CTD-binding proteins to SR proteins strongly suggests that they play a role in splicing, perhaps interacting simultaneously with the CTD and the nascent transcript to link the synthesis and processing of pre-mRNA. Support for this model is based on (*i*) observed interactions between a set of CTD-binding proteins and the CTD in a two-hybrid assay, (*ii*) the immunoprecipitation of a pol II complex containing SR-related proteins, and (*iii*) inhibition of *in vitro* splicing with an anti-CTD antibody or wild-type CTD peptide.

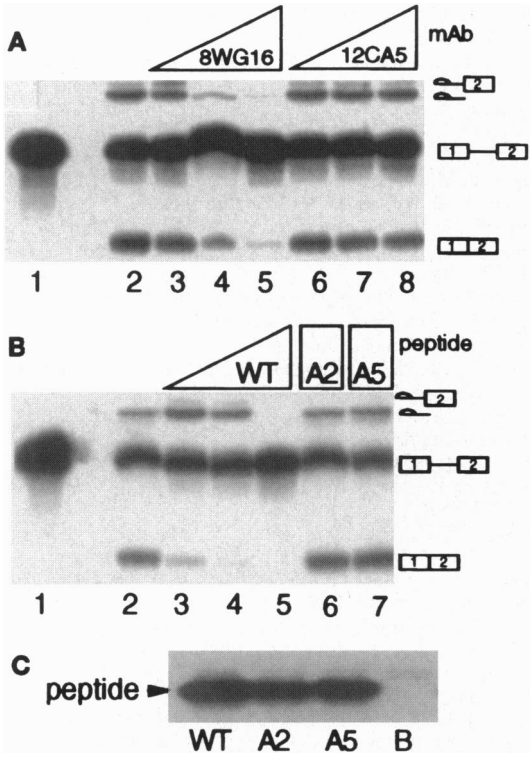

FIG. 5. Inhibition of *in vitro* splicing by mAb 8WG16 and CTD peptide. (*A*) mAb 8WG16 inhibits *in vitro* pre-mRNA splicing. Uniformly labeled pre-mRNA was incubated with HeLa cell nuclear extract. Lanes: 1, unspliced precursor RNA; 2, splicing reaction incubated for 2 hr; 3–5, same as lane 2 but with 2, 4, or 6 μg of 8WG16, respectively; 6–8, same as lane 2 but with 2, 4, or 6 μg of 12CA5, respectively. The nature of the splicing products are depicted on the right. (*B*) CTD peptide inhibits pre-mRNA splicing *in vitro*. Uniformly labeled pre-mRNA was incubated with HeLa cell nuclear extract and CTD peptides WT (0.625, 1.25, 2.5 μg in lanes 3, 4, and 5, respectively), A2 (2.5 μg, lane 6), and A5 (2.5 μg, lane 7) under splicing conditions and the resulting products were purified and analyzed. (*C*) Peptide phosphorylation by Cdc2 kinase. Each kinase reaction contained 4 μg of the peptides used in *B* in 10 μl of 50 mM Tris·HCl, pH 7.9/100 mM KCl/10 mM $MgCl_2$/10 μCi of [γ-^{32}P]ATP. Reactions were initiated by addition of 1 μl of sea star Cdc2 kinase (Upstate Biotechnology) and were incubated at 23°C for 30 min. An equal volume of 2× SDS sample buffer was added and the samples were electrophoresed on a 12.5% polyacrylamide/SDS gel. The wet gel was exposed directly for autoradiography. The mobility of labeled peptide is shown at the left. Peptides are as indicated above the gel; lane B contains only Cdc2 kinase.

While yeast two-hybrid interactions are potentially artifactual, the identification of several different but related sets of CTD-interacting clones offers strong support for their biological relevance. If a single clone or set of clones had been identified, then it could be argued that a conserved homology domain might interact fortuitously with the CTD. The observation of two sets of CTD-binding domains, both contained in mammalian SR-like proteins, is unlikely to have been obtained by chance. Rather, the occurrence of two different CTD-binding domains strongly supports the biological importance of the CTD interaction. The presence of one of these CTD-binding domains in the yeast *NRD1* gene, which contains other motifs in common with SR proteins, argues for the evolutionary conservation of CTD-binding proteins. While little is known about *NRD1*, its identification as a suppressor of a mutation located in the intron of a selectable marker gene suggests that it plays a role in mRNA biosynthesis. The similarity of this protein to SR proteins suggests a possible role in pre-mRNA processing.

Localization studies have shown that splicing occurs on nascent pre-mRNA transcripts (refs. 40 and 41; for review, see ref. 42). More recently, Bregman *et al.* (43) have shown that a form of pol II phosphorylated on the CTD colocalizes with splicing components in the nucleus. If the CTD is involved in linking transcription and splicing, then we would expect to find splicing proteins complexed with pol II *in vivo*. Coimmunoprecipitation from nuclear extracts of pol II and a set of SR-related proteins offers strong evidence for such a complex. Similar results have been independently obtained by S. L. Warren and colleagues (Yale University, personal communication). While we have not identified specific CTD-binding proteins in these complexes, the proteins identified in the two-hybrid screen are excellent candidates based on their size and the presence of alternating Arg motifs.

SR proteins were initially defined by their reactivity with a mAb that binds lampbrush chromosomes, which are active sites of pol II transcription (44). This antibody, mAb104, recognizes an epitope consisting of alternating phosphoserine and arginine residues present in a family of six proteins that are highly conserved among metazoans (35). In addition to the arginine/serine-rich RS domain, these proteins contain RNA recognition motifs (37) and have been demonstrated to bind specific sites on RNA (45–47). CTD-binding proteins differ from known SR proteins; they are higher in molecular weight than the set of six conserved SR proteins (44) and may be more closely related to sets of high molecular weight SR proteins reported by Blencowe *et al.* (48, 49) and Neugebauer *et al.* (31).

The SR proteins detected by mAb 16H3 are detected in the earliest spliceosomal complex E (31). Inhibition of splicing by this mAb indicates that SR proteins are involved in the early steps of splicing. Because this antibody also detects proteins bound to pol II, we looked for evidence that the CTD might also be functionally involved in early splicing steps. Inhibition of splicing with either an anti-CTD antibody or CTD peptides (Fig. 5) offers the first direct evidence that the CTD plays a role in pre-mRNA processing. Anti-CTD mAb 8WG16 may interfere with binding of splicing components to the CTD or RNA or may antagonize splicing by binding the CTD in close proximity to splicing components. CTD peptides may inhibit by competing with splicing complexes for essential CTD-binding proteins.

Although there is no previous experimental evidence suggesting a role for the CTD in pre-mRNA processing, several obser-

vations are consistent with such a role. (*i*) The CTD is present only on RNA polymerases that synthesize pre-mRNA (7). Pol II transcripts are often very long and are spliced by mechanisms that differ from intron removal in pol I or pol III transcripts (39). Pol II transcripts engineered to be produced by pol III are not spliced, indicating that splicing is dependent on pol II (50). (*ii*) The length of the CTD varies in different organisms with more complicated multicellular organisms having more heptapeptide repeats (7). Smaller eukaryotes such as yeast, which have shorter CTDs, also contain fewer interrupted genes and these genes typically contain only one short intron. The longer more varied CTD of mammalian pol II is consistent with a role in splicing multiple long introns. (*iii*) Finally, splicing components colocalize with sites of pre-mRNA synthesis (42, 43).

CTD-binding proteins may act to directly link the nascent transcript to the CTD or they may interact with other SR proteins that bind the transcript independently of the CTD. Such interactions may accelerate the splicing reaction by drawing splicing components together. In this sense the CTD would act as a linear splicing platform. Tight coupling of the nascent transcript to the CTD could also contribute to ordered splicing and, if the binding of SR proteins to the CTD is regulated, then alternate splicing could be controlled. In this context CTD phosphorylation might play a role in either constitutive or alternative splicing. The linkage of splicing and transcription by CTD-binding proteins could also act as part of a feedback mechanism that could shut off transcription if splicing is interrupted. Functional studies of CTD-binding proteins will be important in determining the role(s) of the CTD in linking the processes of transcription and splicing.

We thank Dr. Michael Sayre for a gift of RNA pol II, Dr. Clark Riley and the Biopolymer Lab for DNA sequencing, Dr. Wu-schyong Liu for peptide synthesis, Dr. Emily Barron-Casella for HeLa nuclear extract, and Dr. Ty Lanahan for the rat hippocampus cDNA library. We also thank Dr. Pierre Chevray for the mouse embryo two-hybrid library and advise on its use, Dr. Melissa Moore for splicing reagents and helpful advise on their use, and Dr. Mark Roth for mAb 16H3. Finally, we thank Drs. Eric Steinmetz and David Brow (University of Wisconsin) and Drs. Ben Blencowe and Phil Sharp (Massachusetts Institute of Technology) for sharing data prior to publication and Drs. Jef Boeke and Tom Kelly (Johns Hopkins University) for helpful comments on the manuscript. This work was supported by the Howard Hughes Medical Institute.

1. Corden, J. L., Cadena, D. L., Ahearn, J., Jr., & Dahmus, M. E. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 7934–7938.
2. Allison, L. A., Moyle, M., Shales, M. & Ingles, C. J. (1985) *Cell* **42,** 599–610.
3. Nonet, M., Sweetser, D. & Young, R. A. (1987) *Cell* **50,** 909–915.
4. Bartolomei, M. S., Halden, N. F., Cullen, C. R. & Corden, J. L. (1988) *Mol. Cell. Biol.* **8,** 330–339.
5. Allison, L. A., Wong, J. K., Fitzpatrick, V. D., Moyle, M. & Ingles, C. J. (1988) *Mol. Cell. Biol.* **8,** 321–329.
6. Zehring, W. A., Lee, J. M., Weeks, J. R., Jokerst, R. S. & Greenleaf, A. L. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 3698–3702.
7. Corden, J. L. & Ingles, C. J. (1992) in *Transcriptional Regulation*, eds. McKnight, S. L. & Yamamoto, K. R. (Cold Spring Harbor Lab. Press, Plainview, NY), Vol. 1, pp. 81–108.
8. Peterson, C. L., Kruger, W. & Herskowitz, I. (1991) *Cell* **64,** 1135–1143.
9. Usheva, A., Maldonado, E., Goldring, A., Lu, H., Houbavi, C., Reinberg, D. & Aloni, Y. (1992) *Cell* **69,** 871–881.
10. Dahmus, M. E. (1995) *Biochim. Biophys. Acta* **1261,** 171–182.
11. Kim, Y. J., Bjorklund, S., Li, Y., Sayre, M. H. & Kornberg, R. D. (1994) *Cell* **77,** 599–607.
12. Koleske, A. J. & Young, R. A. (1994) *Nature (London)* **368,** 466–469.
13. Kang, M. E. & Dahmus, M. E. (1995) *J. Biol. Chem.* **270,** 23390–23397.
14. Thompson, N. E., Steinberg, T. H., Aronson, D. B. & Burgess, R. R. (1989) *J. Biol. Chem.* **264,** 11511–11520.
15. Kim, W. Y. & Dahmus, M. E. (1989) *J. Biol. Chem.* **264,** 3169–3176.
16. Zehring, W. A. & Greenleaf, A. L. (1990) *J. Biol. Chem.* **265,** 8351–8353.
17. Buratowski, S. & Sharp, P. A. (1990) *Mol. Cell. Biol.* **10,** 5562–5564.
18. Serizawa, H., Conaway, J. W. & Conaway, R. C. (1993) *Nature (London)* **363,** 371–374.
19. Jiang, Y. & Gralla, J. D. (1994) *Nucleic Acids Res.* **22,** 4958–4962.
20. O'Brien, T., Hardin, S., Greenleaf, A. & Lis, J. T. (1994) *Nature (London)* **370,** 75–77.
21. Greenleaf, A. L. (1993) *Trends Biochem. Sci.* **18,** 117–119.
22. Fields, S. & Song, O. (1989) *Nature (London)* **340,** 245–246.
23. Chevray, P. M. & Nathans, D. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 5789–5793.
24. Ahearn, J., Jr., Bartolomei, M. S., West, M. L., Cisek, L. J. & Corden, J. L. (1987) *J. Biol. Chem* **262,** 10695–10705.
25. West, M. L. & Corden, J. L. (1995) *Genetics* **140,** 1223–1233.
26. Schatz, D. G., Oettinger, M. A. & Baltimore, D. (1989) *Cell* **59,** 1035–1048.
27. Lyford, G. L., Yamagata, K., Kaufmann, W. E., Barnes, C. A., Sanders, L. K., Copeland, N. E., Gilbert, D. J., Jenkins, N. A., Lanahan, A. A. & Worley, P. F. (1995) *Neuron* **14,** 433–445.
28. Desai, D., Gu, Y. & Morgan, D. O. (1992) *Mol. Biol. Cell* **3,** 571–582.
29. Cagas, P. M. & Corden, J. L. (1995) *Proteins* **21,** 149–160.
30. Wilson, I. A., Niman, H. L., Houghten, R. A., Cherenson, A. R., Connolly, M. L. & Lerner, R. A. (1984) *Cell* **37,** 767–778.
31. Neugebauer, K. M., Stolk, J. A. & Roth, M. B. (1995) *J. Cell Biol.* **129,** 899–908.
32. Moore, M. J. & Sharp, P. A. (1992) *Science* **256,** 992–994.
33. Seipel, K., Georgiev, O., Gerber, H. P. & Schaffner, W. (1993) *Nucleic Acids Res.* **21,** 5609–5615.
34. Xiao, H., Lis, J. T., Xiao, H., Greenblatt, J. & Friesen, J. D. (1994) *Nucleic Acids Res.* **22,** 1966–1973.
35. Zahler, A. M., Lane, W. S., Stolk, J. A. & Roth, M. B. (1992) *Genes Dev.* **6,** 837–847.
36. Fu, X. D. (1995) *RNA* **1,** 663–680.
37. Birney, E., Kumar, S. & Krainer, A. R. (1993) *Nucleic Acids Res.* **21,** 5803–5816.
38. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
39. Moore, M. J., Query, C. C. & Sharp, P. A. (1993) in *The RNA World*, eds. Gestland, R. & Atkins, J. (Cold Spring Harbor Lab. Press, Cold Spring Harbor, NY), pp. 303–357.
40. Beyer, A. L. & Osheim, Y. N. (1988) *Genes Dev.* **2,** 754–765.
41. LeMaire, M. F. & Thummel, C. S. (1990) *Mol. Cell. Biol.* **10,** 6059–6063.
42. Mattaj, I. W. (1994) *Nature (London)* **372,** 727–728.
43. Bregman, D. B., Du, L., van der Zee, S. & Warren, S. L. (1995) *J. Cell Biol.* **129,** 287–298.
44. Roth, M. B., Zahler, A. M. & Stolk, J. A. (1991) *J. Cell Biol.* **115,** 587–596.
45. Zuo, P. & Manley, J. L. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 3363–3367.
46. Kohtz, J. D., Jamison, S. F., Will, C. L., Zuo, P., Luhrmann, R., Garcia-Blanco, M. A. & Manley, J. L. (1994) *Nature (London)* **368,** 119–124.
47. Heinrichs, V. & Baker, B. S. (1995) *EMBO J.* **14,** 3987–4000.
48. Blencowe, B. J., Nickerson, J. A., Issner, R., Penman, S. & Sharp, P. A. (1994) *J. Cell Biol.* **127,** 593–607.
49. Blencowe, B. J., Issner, R., Kim, J. S., McCaw, P. & Sharp, P. A. (1995) *RNA* **1,** 852–865.
50. Sisodia, S. S., Sollner-Webb, B. & Cleveland, D. W. (1987) *Mol. Cell. Biol.* **7,** 3602–3612.