A Functional Interaction between the Carboxy-Terminal Domain of RNA Polymerase II and Pre-mRNA Splicing

Lei Du* and Stephen L. Warren

Department of Pathology and *Department of Genetics, Yale University School of Medicine, New Haven, Connecticut 06510

Abstract. In the preceding study we found that Sm snRNPs and SerArg (SR) family proteins co-immunoprecipitate with Pol II molecules containing a hyperphosphorylated CTD (Kim et al., 1997). The association between Pol IIo and splicing factors is maintained in the absence of pre-mRNA, and the polymerase need not be transcriptionally engaged (Kim et al., 1997). The latter findings led us to hypothesize that a phosphorylated form of the CTD interacts with pre-mRNA splicing components in vivo. To test this idea, a nested set of CTD-derived proteins was assayed for the ability to alter the nuclear distribution of splicing factors, and to interfere with splicing in vivo. Proteins containing heptapeptides 1-52 (CTD52), 1-32 (CTD32), 1-26 (CTD26), 1-13 (CTD13), 1-6 (CTD6), 1-3 (CTD3), or 1 (CTD1) were expressed in mammalian cells. The CTD-derived proteins become phosphorylated in vivo, and accumulate in the nucleus even though they lack a conventional nuclear localization signal. CTD52 induces a selective reorganization of splicing factors from discrete nuclear domains to the diffuse nucleoplasm, and significantly, it blocks the accumulation of spliced, but not unspliced, human β-globin transcripts. The extent of splicing factor disruption, and the degree of inhibition of splicing, are proportional to the number of heptapeptides added to the protein. The above results indicate a functional interaction between Pol II's CTD and pre-mRNA splicing.

THE preceding paper provides succinct background information about the COOH-terminal domain (CTD)[1] of RNA polymerase II (Pol II) (Kim et al., 1997 and references therein). Here, it is necessary to supplement this background with relevant genetic analyses of the CTD. Previous studies showed that removal of more than half of the CTD is lethal in yeast (Nonet et al., 1987), *Drosophila* (Zehring et al., 1988), and mammalian cells, indicating that the upstream half of the CTD is essential for cell viability. In addition, a positive and incremental effect on gene expression and cell growth is achieved as heptapeptides are added to the upstream half of the CTD (Nonet et al., 1987; Scafe et al., 1990). These genetic studies indicated that partial truncation of the CTD leads to partial functional deficits in gene expression, but the molecular basis of these effects is poorly understood. Consistent with a transcriptional role for Pol II's CTD, mouse Pol II molecules containing five or fewer CTD heptapeptide repeats cannot respond to enhancer-driven activators in vivo (Gerber et al., 1995).

Multiple groups have reported that the unphosphorylated CTD binds to transcription factors, such as TATA-binding protein (TBP), TFIIF and TFIIE (Kim et al., 1997). The ability of the unphosphorylated CTD to interact with general transcription factors and the suppresser of RNA polymerase B (SRB) mediator complex suggests a transcriptional role for the CTD, and it is consistent with the idea that phosphorylation of the CTD releases the polymerase from the promoter-bound transcription factors (discussed by Koleske and Young, 1995; Dahmus, 1996). Although there is increasing evidence indicating a transcriptional role for the CTD, it remains unclear whether CTD phosphorylation regulates transcription, or whether it merely coincides with transcriptional initiation (see Kim et al., 1997 and references therein). Indeed, it is possible that CTD is a multifunctional domain with roles in transcription as well as other processes, which may not be revealed by genetic selection (viability) or in vitro transcription assays.

Please address all correspondence to S.L. Warren, Brady Memorial Laboratories, Room B117, Department of Pathology, Yale University School of Medicine, P.O. Box 208023, New Haven, CT 06520-8023. Tel.: (203) 737-2247. Fax: (203) 785-7303. E-mail: stephen.warren@yale.edu

The current address of L. Du is Genome Therapeutics Corporation, Waltham, MA 02154.

1. *Abbreviations used in this paper*: CBs, coiled bodies; CTD, carboxy-terminal domain of RNA polymerase II; MIG, mitotic interchromatin granule cluster; nt, nucleotide; Pol IIa, unphosphorylated largest RNA polymerase II subunit; Pol II LS, largest subunit of RNA polymerase II; Pol IIo, hyperphosphorylated largest RNA polymerase II subunit; RT-PCR, reverse transcription-polymerase chain reaction; Sm sn RNP, Smith antigen-containing small nuclear ribonucleoprotein; SR, serine-arginine dipeptide repeat motif.

Little attention has been paid to the phosphorylation state of the CTD after the polymerase disengages from chromatin in vivo. Recently, a fraction of Pol IIo was immunolocalized in 20–50 discrete nuclear domains ("speckles"), which are enriched with serine/arginine dipeptide repeat motif (SR) splicing proteins and Sm snRNPs (Bregman et al., 1995; Blencowe et al., 1996; Zeng, 1997). In addition, Pol IIo, SR proteins and Smith antigen-containing small nuclear ribonucleo proteins (Sm snRNPs) become sequestered in dot-like nonchromosomal domains during mitosis, when transcription is inactive (Warren et al., 1992; Bregman et al., 1994). These immunolocalization experiments revealed Pol IIo molecules in the same nonchromosomal location as certain splicing factors, but it was the preceding study which showed for the first time that splicing factors are associated with Pol IIo in the absence of pre-mRNA, and at times when the polymerase is not engaged in transcription (Kim et al., 1997). The latter findings, together with the observation that anti-CTD phosphoepitope-specific mAbs H5 and H14 can release Pol IIo from the splicing factors in vitro (Kim et al., 1997), strongly imply that Pol IIo's association with the splicing factors is mediated by the hyperphosphorylated CTD. Indeed, the results of the latter study prompted us to ask whether the CTD interacts with the pre-mRNA splicing process in vivo.

Below, we show that overexpression of CTD-derived proteins results in the dispersal of Sm snRNPs and SR splicing factors from a speckled pattern to a diffuse nucleoplasmic distribution. This property is selective, since other types of nuclear domains remain intact. Next, we show that CTD-derived proteins block the accumulation of spliced, but not unspliced, human β-globin transcripts in vivo. Interestingly, the stepwise addition of heptapeptide repeats to a fusion protein potentiates its ability to disrupt the splicing factor domains, and to inhibit splicing in vivo. These results, in conjunction with the preceding study, strongly suggest that the highly conserved and repetitive CTD links splicing components to a key subunit of RNA polymerase II, thereby helping to coordinate the processes of transcription and splicing.

Materials and Methods

Plasmids Expressing Flag-tagged CTD-derived Proteins

Epitope-tagged CTD expression plasmids were created using standard techniques (Sambrook et al., 1989). Full-length CTD coding sequences were obtained from a human Pol II LS cDNA isolated and sequenced by (Du, L., unpublished results). The Pol II LS cDNA was authenticated by comparison to EMBL sequence X63564 (Wintzerith et al., 1992). A 2.1-kb BamHI fragment containing the COOH-terminal domain plus 146 bp of 3′-untranslated mRNA was subcloned into the BamHI site of pcDNA3AB, an expression vector derived from pcDNA3 (Invitrogen, San Diego, CA). pcDNA3AB has a Flag® epitope (AspTyrLysAspAsp AspAspLys; Kodak) immediately upstream of the multiple cloning site (Morrow, J., personal communication). The full-length Flag-tagged CTD expression plasmid is termed "pF-CTD52" to indicate the presence of 52 heptapeptide repeats. pF-CTD52 is predicted to express a fusion protein comprised of an NH_2-terminal Flag peptide® attached to 636 amino acids derived from the COOH terminus of human Pol II LS. The latter segment includes residues 1335-1588 (immediately upstream of the CTD) and residues 1589-1970, which contain 52 tandemly repeated heptapeptides. pF-CTD32 (analogous to pF-CTD52, but lacking heptapeptides 33-52) was derived from a BamHI/EcoRI cDNA clone isolated from a human fetal liver library (Stratagene, La Jolla, CA); sequence analysis revealed this fragment to be truncated within the 32nd repeat of the CTD coding sequence (Du, L., and S. Warren, unpublished results). pF-CTD26 (identical to pF-CTD52, but lacking heptapeptides 27-52) was made as follows: the 2.1-kb BamHI fragment of human Pol II LS cDNA (from nt 4001 to nt 6059 of coding region) was digested with SpeI. The resulting 1.3-kb BamHI–SpeI fragment was subcloned into the BamHI–XbaI sites of pcDNA3AB. pF-CTD13, pF-CTD6, pF-CTD3, and pF-CTD1 were generated by PCR mutagenesis. The forward primer for each of these reactions was p4204U (5′AAGAGGTGGTGGACAAGATGGATG-3′), an oligonucleotide that hybridizes to a 24–nucleotide sequence 183-160 base pairs upstream of the BamHI site located at nucleotides 4001-4006 within the coding portion of the human Pol II LS cDNA. Reverse primers included: p5394 (5′-GCGAATTCGCTGGGAGAGGTGGGCGAATAGCT-3′) for pF-CTD13; p5264 (5′-GCGAATTCGGACTGGTTGGAGAATAGGATGGA-3′) for pF-CTD6; p5205 (5′-GCGAATTCAAGAGGGACTCTGGGGTGTGTAGCC-3′) for pF-CTD3, and p1CTD (5′-GCGAATTCAGCTTGGACTAGTGGGTGAGTAGCTGGGAGACATGGCGCCACCTGGTGA-3′) for pF-CTD1. The PCR products were digested with EcoRI (encoded in downstream primers) and BamHI (present in Pol II cDNA 183-160 nucleotides downstream of the upstream primer). The PCR products were subcloned into the BamHI/EcoRI sites of pcDNA3AB.

One control plasmid, pF-CTDless.3, expresses the Flag-tagged NH_2-terminal 282 amino acids of Pol II LS. This segment was generated by PCR amplification, using human Pol II LS cDNA as the template. The oligos for this reaction were p337U (5′-GCGAATTCGGCTTTTTGTAGTGAGGTTTG-3′) and p1209L (5′GCGAATTCGTCAGCC-AGTTTGTGAGTCAGGTC-3′). The amplified segment of DNA was digested with EcoRI and subcloned into the EcoRI site of pcDNA2AB. Another control plasmid, pF-CTDless.1, expresses a Flag-tagged ~25-kD segment of Pol II immediately upstream of the CTD. This control sequence corresponds to a 714-bp BamHI–SmaI fragment derived from the Pol II LS cDNA, which was subcloned into the BamHI–EcoRV sites of pcDNA3AB. A third control, pF-CTDless.2, expresses a Flag-tagged ~22-kD segment of Pol II LS, which contains a 500-bp fragment immediately downstream from the BamHI site. This fragment was generated by PCR amplification using oligos p4204U (see above) and p4869L (5′-GCGAATTCAGCCGGTGGGTCCAGCAGC-3′). The PCR product was digested with BamHI and EcoRI and subcloned into pcDNA3AB. This F-CTDless.2 protein is similar to F-CTDless.1, but lacks a heptapeptide-like sequence (MFFGSAPSPMGGISPAMTPWNQGATPAYGAWSPSVGSGMTPGAAGFSPSAASDASGFSPGYSPAWSPTPGSPGSPGPSSPYIPSPGGA), which precedes the CTD.

Plasmids Expressing βGalactosidase-linked CTD Proteins

The βGal-CTD fusion constructs were made as follows: First, the stop codon at the end of the βGalactosidase gene was replaced with restriction sites EcoNI, BamHI, and SalI, which were recombinantly added to the COOH terminus. This PCR reaction used pSVβ (Promega, Madison, WI) as a template, and two primers: a downstream adapter oligo which contains a SalI site (pMCS, 5′-GCGTCGACTCTAGAATTCGCGGATCCTCCTGAAGGTTTTTGACACCAGACCAACTGG-3′) and an internal oligo p3047 (5′-GGATTGGTGGCGACGACTCCTGGA-3′). The 130-bp PCR product was digested with EspI and SalI and inserted back into pSVβ that had been cut with EspI and SalI to make pSVβMCS. Next, the βGal coding sequence was excised from pSVβMCS with SmaI and BamHI, and subcloned into pcDNA3 that has been cut with HindIII, filled in with Klenow and cut with BamHI. The resulting vector, pcDNAβGal, preserves all the cloning sites downstream of BamHI from pcDNA3. Finally, CTD-26, CTD-32, and CTD-52 fragments with BamHI–EcoRI ends were subcloned into the corresponding sites of pcDNAβGal to generate the βGal-CTD series (Fig. 1 shows only pβGal-CTD52, pβGal-CTDless, and βGal).

Plasmids Expressing Human β-Globin Genes and Recombinant CTD-derived Proteins

Plasmids that co-express Flag-tagged CTD-derived proteins and human β-globin genes are generically termed "pF-CTD$_x$β-globin [+/−]," where "F" refers to the Flag peptide coding sequence, "CTD" refers to the sequence of the CTD-derived protein, "*x*" refers to the number of heptapeptide repeats, "β-globin" refers to the β-globin gene, and the "[+]" and "[−]" signs designate the relative orientation of the two transcription

units. The plasmids were constructed as follows: A 2.7-kb HindIII–FspI fragment containing the 2.3-kb human β-globin gene plus an SV40 enhancer element was excised from pUCβ128SV (Caceres et al., 1994), filled in with Klenow fragment of DNA polymerase I, and subcloned into CTD expression plasmids pF-CTD1, pF-CTD6, pF-CTD13, and pF-CTD52, each of which had been digested with EcoRV. For controls, β-globin genes were subcloned into pF-CTDless.1, pF-CTDless.3, and pβgal as illustrated in Fig. 7.

Antibodies

For a description of mAbs H5, H14, Y12, and B1C8 see preceding paper (Kim et al., 1997). mAb M2 (Kodak) is an IgG that binds to the Flag® peptide, AspTyrLysAspAspAspAspLys. mAb anti-βGal is an IgG that binds to βGalactosidase (Promega). pAb anti-βGal is a polyclonal antibody that binds to βGalactosidase (Cappel, Malvern PA). mAb 138 is an IgG directed against ND55, a protein in N10 domains (Ascoli and Maul, 1991). Anti-coilin is a rabbit antiserum directed against p80 coilin (Andrade et al., 1993).

Cell Culture and Transient Plasmid Transfections

See preceding paper (Kim et al., 1997). For in vivo splicing assays, 10^6 HeLa cells were seeded in a 60-mm petri dish and transfected with each of the plasmids (5 μg) using 45 μl of Lipofectamine®.

SDS-PAGE and Immunoblotting

See preceding paper (Kim et al., 1997).

Immunofluorescence Microscopy and Image Analysis

See preceding paper (Kim et al., 1997).

Quantitative RT-PCR and RNase Protection Assays

Quantitative RT-PCR was carried out as follows: Total RNA was prepared from HeLa cells 1–2 d after transfection using UltraSpec RNA® (Biotecx, Houston, TX), and digested with RNase-free RQ1 DNase (Promega) to remove contaminating DNA. The RNA was phenol extracted, ethanol precipitated, and dissolved in water. A reverse primer (449 nucleotides downstream from the HindIII site) that hybridizes to the second exon of the β-globin gene, 5′-CAGGAGTGGACAGATCCC-3′, was used for reverse transcription, followed by 10, 12, and 14 cycles of PCR amplification using a forward oligo 5′-TCAAACAGACACCATGGTGCACCTGACT-3′ which hybridizes to exon 1 of β-globin (167 nucleotides downstream from the HindIII site).

The RNase protection assay was carried out using the RPA II Ribonuclease Protection Kit (Ambion, Austin, TX) according to the manufacturer's procedures. Briefly, a HindIII–BamHI fragment containing exon 1, intron 1, and most of exon 2 of human β-globin was subcloned into the corresponding sites of BlueScript-SKII (Strategene) and digested with BbvII located within intron 1 to yield a linearized template. A complementary RNA probe was synthesized in vitro with T3 RNA polymerase (New England Biolabs, Boston, MA) in the presence of 10 U of RNasin (Promega), 0.5 mM of ATP, GTP, UTP, 3 μM of CTP, and 100 μCi [α-^{32}P]CTP, which yielded an internally radiolabeled 343-nt fragment covering exon 2 and the 3′ half of intron 1. The probe was purified on a 4.5% denaturing polyacrylamide gel and hybridized to total RNA prepared from transfected HeLa cells at 45°C overnight. Hybridization mixtures were digested with RNase A/T1, precipitated, solubilized with gel loading buffer, and separated on a 4% denaturing polyacrylamide gel. The dried gel was exposed to hyperfilm (Amersham Corp., Arlington Heights, IL) overnight and scanned into a digital image using ScanJet (Hewlett Packard) and analyzed using NIH Image® software.

Results

The results of the preceding study (Kim et al., 1997) led us to hypothesize that Pol IIo's association with pre-mRNA splicing factors is mediated by the CTD. To test this hypothesis we first asked whether CTD-derived sequences, which lack the catalytic and DNA-binding regions of the Pol II LS, can target indicator proteins to speckle domains. For this purpose, the Flag peptide (Flag) or βGalactosidase (βGal) sequences were recombinantly added to the NH_2 terminus of the CTD-containing proteins. The resulting fusion proteins were transiently expressed and immunolocalized in CV1 cells. A similar approach has been used to show that certain SR domains can target βGal to speckle domains (Li and Bingham, 1991). Next, we asked whether the CTD-derived fusion proteins interfere with splicing in vivo. For this purpose, we co-expressed human β-globin pre-mRNAs and CTD-derived proteins in HeLa cells, and quantitated the efficiency of β-globin splicing in vivo. A similar approach has been used to assess the in vivo properties of splicing factors (Romac and Keane, 1995; Caceres, 1994).

Plasmids Expressing Unidirectionally Truncated CTD Sequences

We constructed plasmid vectors that express a variety of CTD-containing fusion proteins shown in Fig. 1 (Materials and Methods). The expression and intracellular distribution of each fusion protein has been documented by immunoblotting, immunoprecipitation, and immunostaining with antibodies directed at the Flag epitope or βGal (see below). During our investigations, we sought to determine the minimum number of heptapeptide repeats required to achieve certain biological effects (see below). Therefore, the CTD-containing fusion proteins were unidirectionally truncated from the COOH terminus, giving rise to a nested set of proteins containing heptapeptides 1-52 (F-CTD52 and βGal-CTD52); 1-32 (F-CTD32 and βGal-CTD32); 1-26 (F-CTD26 and βGal-CTD26); 1-13 (F-CTD13); 1-6 (F-CTD6); 1-3 (F-CTD3), or only the first heptapeptide (F-CTD1) (Fig. 1). Several control proteins were used: (*a*) F-CTDless.1; (*b*) F-CTDless.2; (*c*) F-CTDless.3; (*d*) βGal-CTDless and (*e*) βGal (Materials and Methods).

CTD-derived Proteins Accumulate in the Nucleus

The CTD-derived fusion proteins must gain access to the nucleus to interact with splicing factors. At the beginning of the study we immunolocalized each fusion protein in CV1 or HeLa cells to confirm that our experimental approach meets this requirement. Plasmids expressing each of the 13 proteins illustrated in Fig. 1 were transfected into cells (Materials and Methods). 2 d later, the cells were fixed and double immunostained with: (*a*) an antibody directed at the indicator portion of the fusion protein (anti-Flag or anti-βGal), and (*b*) an antibody directed at the CTD portion of the fusion protein (mAb H5 or mAb H14).

A representative experiment is shown in Fig. 2 *C*. In this case, CV1 cells were transfected with pF-CTD52, fixed and double stained with anti-Flag mAb M2 and anti-CTD mAb H14. A cell expressing the F-CTD52 protein is shown at the top of each panel, and an untransfected cell is shown at the bottom. Interestingly, F-CTD52 is distributed almost exclusively in the nucleus, even though it lacks a conventional nuclear localization signal (Fig. 2 *C*, *left panel*). The F-CTD52 protein is present in the diffuse nucleoplasm, but it is most concentrated in ~50 discrete, nonnucleolar sites (Fig. 2 *C*, *left panel*, *arrows*). In addi-

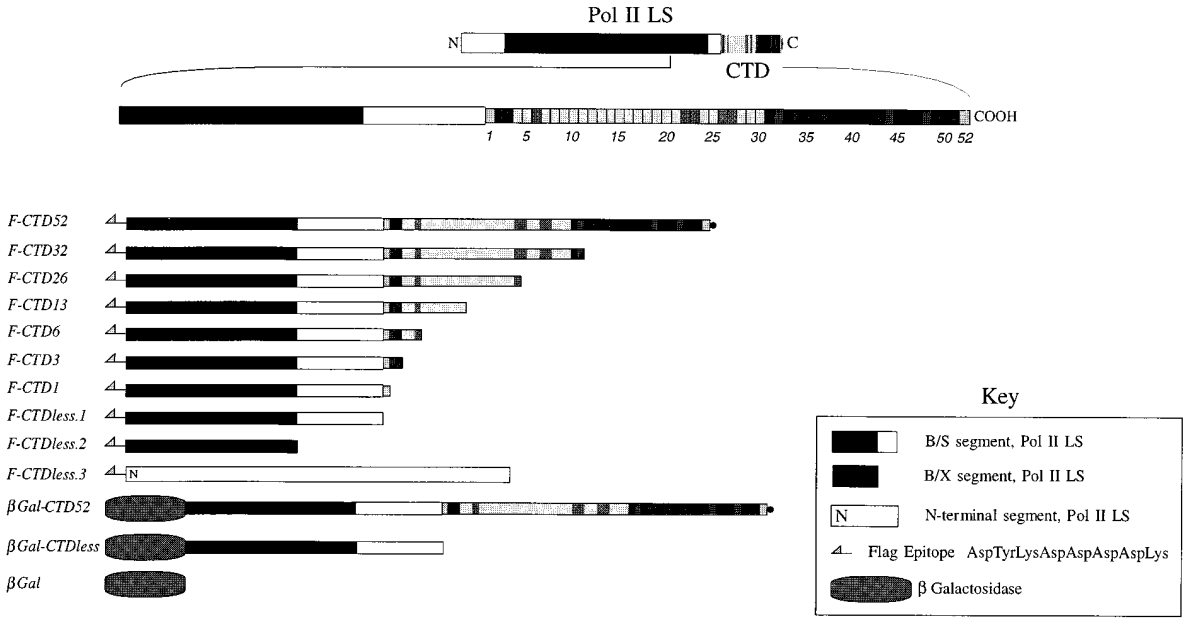

Figure 1. Fusion proteins derived from Pol II's CTD. The largest subunit of RNA Polymerase II (Pol II LS) is illustrated schematically (*top*). An expanded view of the CTD shows 52 heptapeptide repeats represented by variably shaded boxes. Lightly shaded boxes represent consensus repeats (YSPTSPS) and more darkly shaded boxes represent variant repeats (see Corden et al., 1985; Wintzerith et al., 1992). The CTD coding sequence was unidirectionally truncated from the COOH terminus and recombinantly fused to the Flag® peptide (*Flag symbol*) or βGalactosidase (*oval symbol*). The resulting fusion proteins are described by nomenclature that begins with the NH_2 terminus and ends with the COOH terminus, including the number of heptapeptide repeats. Symbols are summarized in the key (for details see Materials and Methods).

tion, the transfected cell nucleus is much more intensely stained by mAb H14 than the untransfected cell nucleus (Fig. 2 *C*, *right panel*). The nuclear "dots" are also intensely stained by mAb H14 (Fig. 2 *C*, *right panel*, *arrows*), and mAb H5 (Du, L., and S.L. Warren, unpublished results) antibodies, both of which recognize CTD phosphoepitopes (Kim et al., 1997). The above results indicate that F-CTD52 accumulates in the nucleus, and suggest that CTD heptapeptides on the F-CTD52 protein are phosphorylated in vivo. All of the CTD-derived and control proteins illustrated in Fig. 1 are expressed and enter the nucleus (see below).

The CTD-derived Proteins Are Phosphorylated In Vivo

All observations indicating an association between Pol II LS and splicing factors suggest a mechanism involving a hyperphosphorylated CTD (Bregman et al., 1995; Kim et al., 1997; Blencowe et al., 1996). Therefore, if the CTD-derived proteins are expected to interact with splicing factors in the nucleus, they probably need to be phosphorylated in vivo. The immunolocalization studies described above suggest strongly that the CTD-derived fusion proteins are phosphorylated in vivo. To confirm this impression, and to establish the electrophoretic mobility of each CTD-derived protein, whole cell extracts were prepared from cells transfected with each plasmid in Fig. 1. The samples were subjected to 5–15% gradient SDS-PAGE and immunoblotted with: (*a*) mAbs directed against CTD-specific phosphoepitopes (H5 or H14); or (*b*) mAbs directed at the indicator part of the protein (Flag or βGal) (Fig. 2, *A* and *B*).

An analysis of the Flag-tagged proteins is presented in Fig. 2 *A*. mAbs H14 and H5 blot a ~240-kD protein corresponding to endogenous Pol IIo in all of the extracts (Fig. 2 *A*, *right panels*, *IIo*). In cells transfected with the pFCTD series of plasmids, mAbs H5 and H14 14 blot a nested set of fusion proteins. In this experiment, mAb H5 immunoblots F-CTD26, F-CTD32, and F-CTD52 (Fig. 2 *A*, lanes *15–17*), and mAb H14 immunoblots pF-CTD6, pF-CTD13, pF-CTD26, pF-CTD32, and pF-CTD52 (Fig. 2 *A*, lanes *21–25*). As expected, the stepwise removal of heptapeptide repeats incrementally increases the electrophoretic mobility of the proteins. However, the apparent mol wt of each fusion protein significantly exceeds its predicted size. For example, F-CTD52 migrates as a 120/130-kD doublet, even though it has a predicted mol wt of ~66 kD (Fig. 2 *A*, lanes *17* and *25*). Repeated immunoblotting experiments reveal that many of the CTD-derived proteins migrate as closely spaced doublets (Du, L., and S.L. Warren, unpublished results).

The anomalous SDS-PAGE mobilities of the CTD-derived proteins, and our observation that alkaline phosphatase treatment of the filters abolishes mAb H14 and H5 immunoreactivity (Du, L. and S.L. Warren, unpublished results), indicate that the CTD-derived proteins are phosphorylated. Together with our previous studies showing that mAbs H5 and H14 recognize distinct phosphoepitopes on the CTD of native Pol II (Kim et al., 1997), these data indicate that the phosphorylation sites are within the CTD portion of the fusion proteins.

Some of the Flag-tagged CTD proteins are immunoblotted weakly, or not at all, by anti-Flag mAb M2 (Fig. 2 *A*, lanes *2–8*). However, all of the Flag-tagged CTD-derived proteins are expressed in HeLa or CV1 cells, since anti-

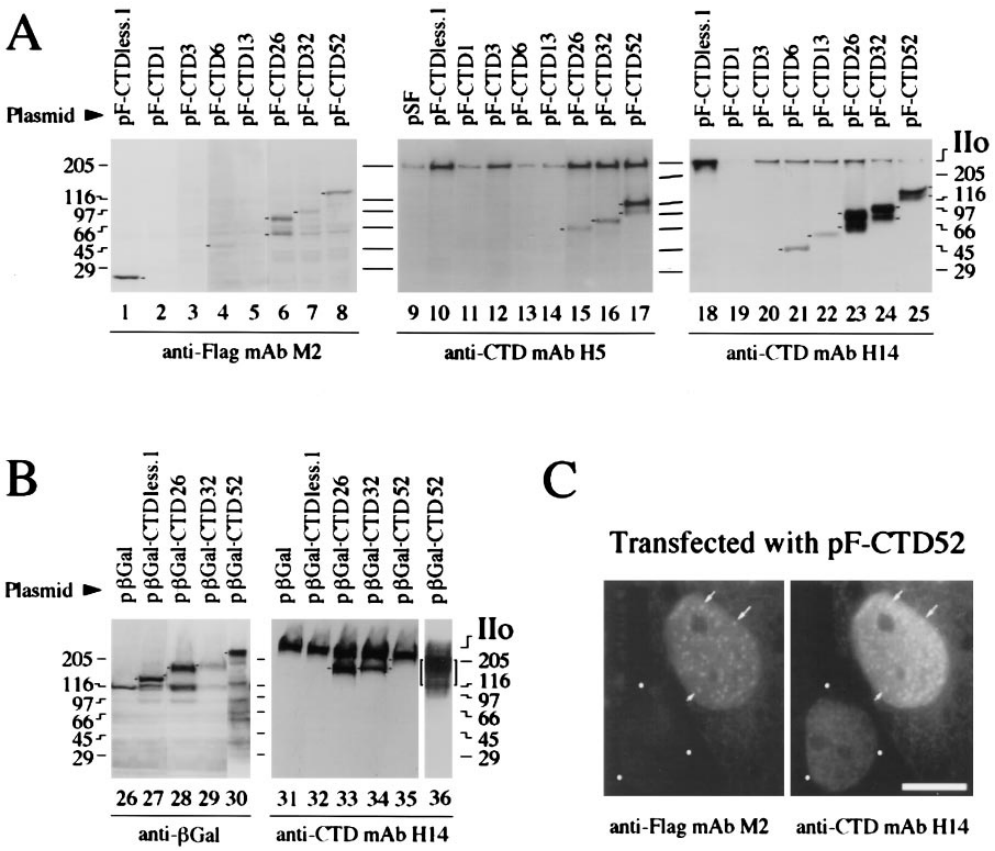

Figure 2. Expression, in vivo phosphorylation, and nuclear localization of CTD-derived fusion proteins. (*A* and *B*) Immunoblotting. CV1 cells were transfected with each of the plasmids listed above the panels (see Materials and Methods). 2 d later, the cells were lysed in SDS sample buffer, subjected to 5–15% gradient SDS-PAGE, and immunoblotted with the antibodies listed below each panel. mAb M2 is directed against the Flag® epitope, anti-βGal is directed against βGalactosidase, and mAbs H5 and H14 are directed against CTD phosphoepitopes (Kim et al., 1997). Numbers at the margins indicate apparent molecular weights in kilodaltons. *pSF*, control plasmid expressing a Flag® tagged ~30-kD segment of human β-spectrin. *IIo*, hyperphosphorylated largest subunit of Pol II. (*C*) Immunofluorescence microscopy. CV1 cells were transfected with pF-CTD52. 2 d later the cells were fixed and double immunostained with anti-Flag® mAb M2 and anti-CTD mAb H14 (see Materials and Methods). Anti-Flag staining is visualized by rhodamine (*left panel*) and mAb H14 staining is visualized by FITC (*right panel*). The cell at the top expresses the F-CTD52 protein, and the cell at the bottom is an untransfected control. Note that mAb M2 labeling is almost exclusively intranuclear. In addition, the untransfected cell nucleus is weakly immunostained by mAb H14, whereas the transfected cell nucleus is intensely labeled. mAbs M2 and H14 stain the diffuse nucleoplasm, but they also stain ~50 discrete "dots." Bar, 10 μM.

Flag mAb M2 stains the nucleus in cells transfected by pF-CTD1 (Fig. 4 *A*), pF-CTD3 (Fig. 4 *D*), pF-CTD6 (Fig. 4 *G*), pF-CTD13 (Fig. 4 *J*), pF-CTD26 (Fig. 4 *M*), pF-CTD32 (data not shown), and pF-CTD52 (Fig. 3 *A* and *G*). We have repeatedly immunostained the Flag-tagged CTD-derived proteins with mAb M2, but it has not been easy to reproducibly detect many of the CTD-derived proteins by immunoblotting with the same antibody. The basis for this discrepancy is not understood. One factor may be low transfection efficiencies; expression of CTD-derived proteins in a small fraction of cells is difficult to detect by immunoblotting, but easy to detect by in situ methods such as immunostaining.

Some short CTD-derived proteins are not immunoblotted by mAbs H5 and H14 (Fig. 2 *A*). Nevertheless, we believe that all of the FCTD proteins are phosphorylated in the cell, as indicated by enhanced mAb H14 immunostaining of transfected cell nuclei (see Fig. 4). The inability of mAb H5 to immunoblot F-CTD1, F-CTD3, F-CTD6, and F-CTD13, and the inability of mAb H14 to immunoblot F-CTD1 and F-CTD3, may be explained by three factors: (*1*) Transfection efficiencies vary widely from experiment to experiment and from plasmid to plasmid. (*2*) Fusion proteins with only a few heptapeptides have fewer potential phosphorylation sites, and hence fewer mAb H5- and H14-binding sites, than proteins with long CTD segments (e.g., F-CTD52 has ~50-fold more phosphorylation sites than F-CTD1). (*3*) Finally, it is possible that downstream heptapeptides are better kinase substrates than upstream heptapeptides. In this regard, it is interesting to note that repeats 1-3 diverge from the YSPTSPS consensus sequence more than other repeats in the CTD.

An immunoblot of selected βGal-linked CTD proteins is presented in Fig. 2 *B*. mAb H14 immunoblots a ~240-kD protein corresponding to endogenous Pol IIo (Fig. 2 *B*, lanes *31–35*). mAb H14 also immunoblots βGalCTD fusion proteins in cells transfected with pβGal-CTD26, pβGal-CTD32, and pβGal-CTD52 (Fig. 2 *B*, lanes *33–36*). Hyperphosphorylated βGal-CTD52 comigrates with Pol IIo at ~240 kD; however, more rapidly migrating species are observed in some experiments (Fig. 2 *B*, lane *36*). Finally, immunoblotting with an antibody directed at βGal reveals the expected stepwise increase in the PAGE mobility of these proteins (Fig. 2 *B*, lanes *26–30*).

Expression of F-CTD52 or βGal-CTD52 Induces the SR-related Splicing Factor B1C8 to Redistribute from Discrete Domains to a Diffuse Nucleoplasmic Pattern

The F-CTD52 protein is phosphorylated on CTD epitopes

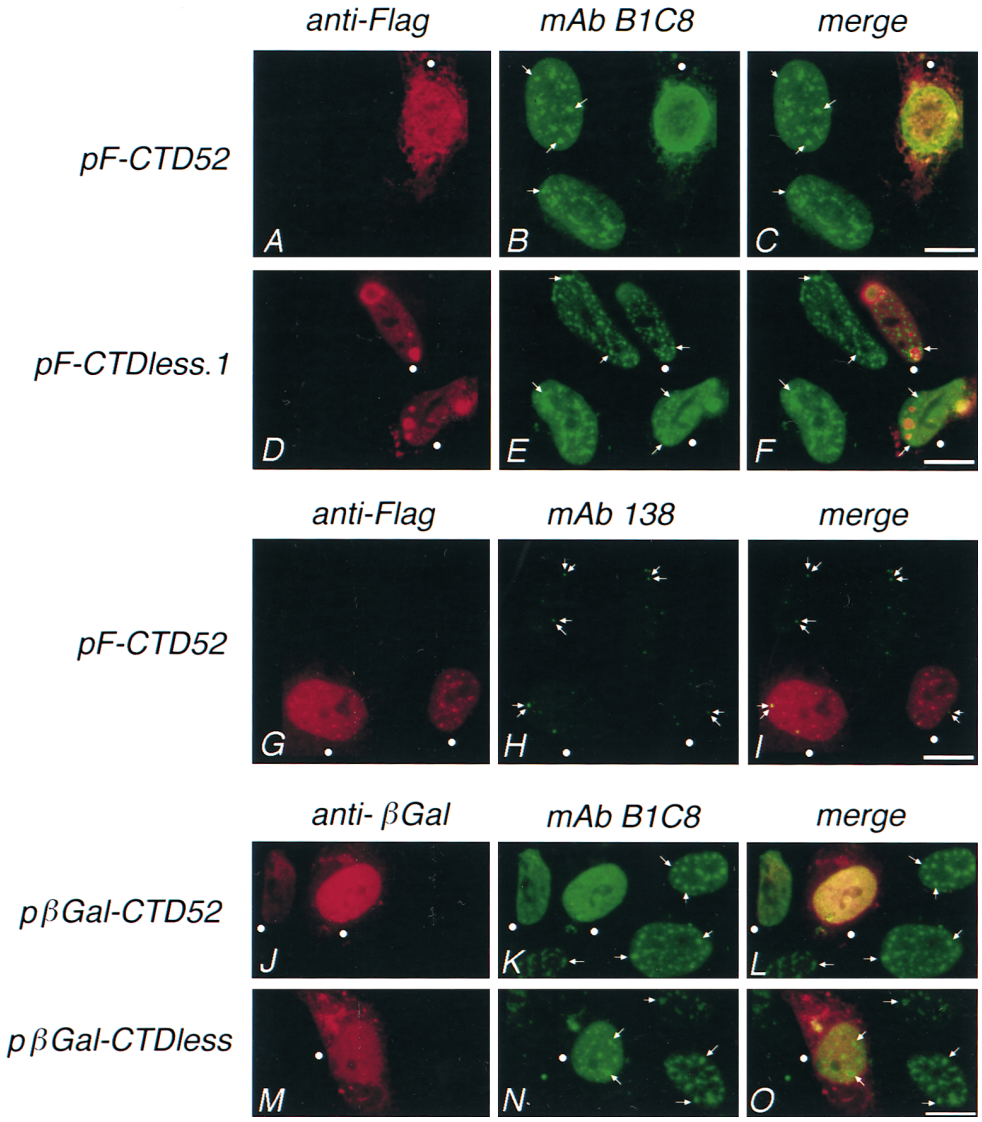

Figure 3. CTD52 disrupts the speckled distribution of SR splicing factor B1C8. CV1 cells were transfected with plasmids encoding fusion proteins listed at the left margin. 2 d later, the cells were fixed and double immunostained (see Materials and Methods). CTD-derived fusion proteins and control proteins were immunolocalized with anti-Flag® mAb M2 (*A*, *D*, and *G*) or anti-βGal (*J* and *M*). A 160-kD SR-related family splicing factor (Blencowe et al., 1995) was immunolocalized with mAb B1C8 (*B*, *E*, *K*, and *N*). ND55 was immunolocalized with mAb 138 (Ascoli and Maul, 1991). Red pseudocolor indicates distribution of Flag-tagged or βgal-linked fusion proteins. Green pseudocolor indicates distribution of endogenous nuclear proteins B1C8 or ND55. Red and green digital images were merged (*C*, *F*, *I*, *L*, and *O*), and areas of overlap between the distributions of transiently expressed fusion proteins and endogenous proteins are pseudocolored yellow. *White dots*, nuclei expressing fusion proteins; *single arrows*, B1C8 speckles; *double arrows*, ND55 in N10/PML domains (*H*). Bars, 10 μm.

(Fig. 2 *A*, lanes *17* and *25*), and it enters the nucleus where it is frequently, but not always, observed in discrete nuclear dots (Fig. 2 *C*, *arrows*). One possible explanation for this distribution is that CTD52 targets the Flag peptide to splicing factor domains, perhaps reflecting its ability to associate with Sm snRNPs and SR family splicing proteins, which are most concentrated in the speckles. To further explore this idea, we sought evidence that the F-CTD52 containing dots overlap or colocalize with speckle domains. Thus, CV1 cells were transfected with pF-CTD52, and the cells were double immunostained with anti-Flag mAb M2 (IgG) and mAb B1C8 (IgM), which recognizes an SR-related splicing protein in the speckle domains (Blencowe et al., 1994, 1995).

This experiment yielded a striking (and unexpected) result: the B1C8 splicing factor is distributed in a speckled pattern in untransfected cell nuclei (Fig. 3 *B*, *left side of panel*, *arrows*), but it is distributed in a nearly uniform, diffuse nuclear pattern in every cell expressing the F-CTD52 protein (Fig. 3 *B*, *right side of panel*). Control proteins such as F-CTDless.1 accumulate in the nucleus, but they have little effect on the distribution of B1C8 (Fig. 3 *E*, *right side of panel*, *arrows*).

Next we sought to confirm that the CTD is responsible for the redistribution of B1C8. For this purpose, CV1 cells were transfected with pβGal-CTD52, and the cells were double immunostained with mAb B1C8 and anti-βGal (Fig. *3*, *J–L*). Again, B1C8 has a speckled distribution in control cells (Fig. 3 *K*, *arrows*), but it has a diffuse nuclear distribution in cells expressing βGal-CTD52 (Fig. 3 *K*). B1C8 remains in a speckled distribution in nuclei expressing similar levels of a control protein, βGal-CTDless (Fig. 3 *N*, *arrows*).

We asked whether F-CTD52 alters the distribution of proteins located in other types of nuclear domains. ND55 (55 kD) is one of several proteins localized in ~10 highly circumscribed nuclear dots, referred to as "N10 domains" or "PML bodies" (Ascoli and Maul, 1991). N10 domains are dynamic structures. For example, the number of N10 domains increases after growth factor stimulation, and they disassemble following virus infection (Maul and Everett, 1994; Terris et al., 1995). Several proteins in N10

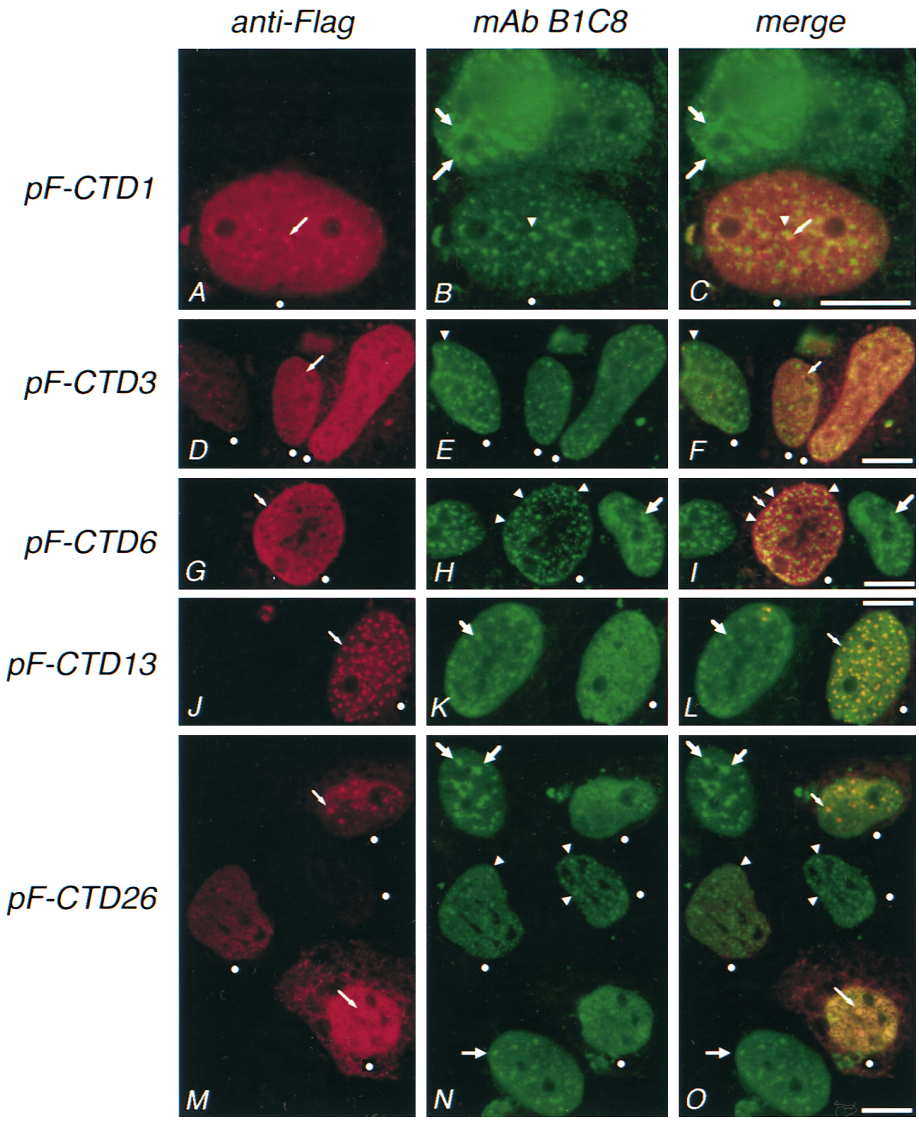

Figure 4. Addition of heptapeptide repeats to Flag-tagged fusion proteins potentiates their disruptive effect on B1C8 speckles. CV1 cells were transfected with plasmids encoding fusion proteins listed at the left margin. 2 d later, the cells were fixed and double immunostained as described in Fig. 3. The CTD-derived fusion proteins were immunolocalized with anti-Flag® mAb M2 (*A*, *D*, *G*, *J*, and *M*). A 160-kD SR-related family splicing factor was immunolocalized with mAb B1C8 (*B*, *E*, *H*, *K*, and *N*). Digital images were pseudocolored and merged as described in Fig. 3 (*C*, *F*, *I*, *L*, and *O*). Red pseudocolor indicates distribution of CTD-derived fusion proteins. Green pseudocolor indicates distribution of SR splicing factor B1C8. Yellow pseudocolor indicates overlap between red and green. *White dots*, nuclei expressing fusion proteins; *thick arrows*, intact B1C8 speckles; *thin arrows*, Flag-tagged CTD-derived protein in discrete nuclear sites; *arrowhead*, B1C8 speckles immediately adjacent to F-CTD1 (*B* and *C*) and FCTD6 (*H* and *I*). Bar, 10 μm.

domains have been identified, but none appear to have a role in pre-mRNA splicing. Cells were transfected with pF-CTD52 and double immunostained with anti-ND55 mAb 138 (IgM) and anti-Flag mAb M2 (IgG). Our results indicate that F-CTD52 does not alter the distribution of ND55, which remains exclusively in the N10 domains (Fig. 3, *G–I*).

We had originally predicted that CTD heptapeptides would behave like SR domains, which target indicator proteins to the splicing factor domains (Li and Bingham, 1991). However, our experiment was complicated by the fact that the B1C8 speckles, our intended landmarks, disperse in the presence of F-CTD52. Nevertheless, this outcome was gratifying, because CTD52 alters the distribution of an SR-related splicing factor that is colocalized with native Pol IIo molecules in the speckles. These results are consistent with the idea that the CTD interacts with splicing factors in the speckles.

Addition of Heptapeptide Repeats to the Fusion Protein Leads to an Incremental Disruption of B1C8 Speckles

Our next goal was to determine how many heptapeptide repeats are required to induce the redistribution of B1C8. Therefore, we performed the following "heptapeptide titration" experiment: CV1 cells were transfected with a nested set of Flag-tagged CTD-derived proteins: pF-CTD26, pF-CTD13, pF-CTD6, pF-CTD3, and pF-CTD1. 2 d later, the cells were fixed and double immunostained with anti-Flag mAb M2 (IgG) and mAb B1C8 (IgM) (Fig. 4).

First, consider the results obtained with F-CTD26. Immunostaining with mAb M2 reveals four transfected cell nuclei (Fig. 4 *M*). Note that mAb M2 staining is almost exclusively intranuclear, and the level of FCTD26 expression varies widely among the four cells (Fig. 4 *M*, transfected nuclei marked by white dots). Diffuse mAb M2-immunoreactivity is observed in all four nuclei, but two nuclei also contain discrete dots harboring the F-CTD26 protein (Fig. 4 *M*, *arrows*). The nucleus expressing the highest level of F-CTD26 has a completely dispersed pattern of B1C8 staining (Fig. 4, *M* and *N*, *lower right corner*). The nucleus expressing the second highest level of F-CTD26 has a nearly complete dispersal of B1C8 staining (Fig. 4, *M* and *N*, *upper right corner*). The two nuclei expressing low levels of F-CTD6 have a partial dispersal of the B1C8

staining pattern as indicated by the multiple diminutive B1C8 speckles (Fig. 4, *M* and *N*, *middle of panel*). Finally, the two untransfected nuclei each contain ~20 prominent B1C8-speckles (Fig. 4 *N*, *thick arrows*). These results indicate that the upstream half of the CTD retains the ability to disrupt the distribution of B1C8, and the degree of B1C8 disruption is proportional to the level of CTD-derived protein in the nucleus. Similar results were obtained with F-CTD32 (Du, L., and S.L. Warren, unpublished results).

Next, consider the results obtained with F-CTD13. Immunostaining with anti-Flag mAb M2 reveals a transfected cell nucleus (Fig. 4, *J–L*, *right*) and an untransfected cell nucleus (Fig. 4, *J–L*, *left*). mAb M2 staining is almost exclusively intranuclear; the F-CTD13 protein is distributed in ~75 discrete dots, as well as the diffuse nucleoplasm (Fig. 4 *J*, *right*). The nucleus expressing F-CTD13 has a dispersed pattern of B1C8 staining (Fig. 4 *K*, *right*) and the control nucleus has a typical speckled pattern (Fig. 4 *K*, *left*). Thus, removal of 75% of the heptapeptides from the CTD does not abolish the B1C8-disrupting property of the fusion protein.

Consider the results obtained with F-CTD6 and F-CTD3. Three representative nuclei expressing low, medium, and high levels of the F-CTD3 protein are presented (Fig. 4 *D*). Again, mAb M2 staining is almost exclusively intranuclear, and the distribution of F-CTD3 is diffuse with a few discrete dots (Fig. 4 *D*, *arrow*). The nucleus expressing a low level of F-CTD3 retains a prominent speckled pattern of B1C8 staining (Fig. 4 *E*, *left*, *arrowhead*). Nuclei expressing higher levels of F-CTD3 protein have a partial disruption of B1C8 staining, as indicated by diminutive speckles (Fig. 4 *E*, *center* and *right*). Partial disruption of B1C8 stained speckles is observed in a nucleus expressing F-CTD6. Note that the transfected nucleus has diminutive B1C8 speckles (Fig. 4, *G–I*, *center of panel*).

Finally, consider the results obtained with F-CTD1. A representative transfected cell nucleus reveals mAb M2 staining in a diffuse and punctate distribution (Fig. 4 *A*). Most nuclei expressing the F-CTD1 protein have prominent B1C8 containing speckles, as shown here (Fig. 4 *B*, *arrowhead*). When the anti-B1C8 and anti-Flag images are merged, one observes a close spatial relationship between the B1C8-speckles and F-CTD1 dots (Fig. 4 *C*, *arrowhead* and *thin arrow*). Close examination of a nucleus expressing F-CTD6 reveals a similar phenomenon (Fig. 4 *I*). Many of the overexpressed CTD proteins form discrete dots, and in nuclei containing intact B1C8 speckles the CTD-rich dots do not coincide with the speckles. Quantitative image analysis is needed to determine whether the CTD-rich dots are organized randomly with respect to the B1C8 speckles, or whether they reproducibly form at the periphery of the speckles. It is possible that the CTD-rich dots revealed by mAb M2 staining might be aggregated, Flag-tagged CTD proteins, which are randomly distributed in the nucleus.

The effect of CTD length (i.e., number of heptapeptide repeats) on B1C8-speckles was quantitated as follows: CV1 cells were transfected with each of the Flag-tagged CTD-derived plasmids in Fig. 1. The cells were fixed and double stained with anti-Flag mAb M2 and a mAb directed against B1C8 as described in Fig. 4. The pattern of B1C8 staining in each transfected cell nucleus was scored as "intact" (20–50 prominent speckles) or "disrupted" (diffuse pattern or diminutive speckles). Multiple sets of experiments were conducted, and 150–250 nuclei were scored for each plasmid (see Materials and Methods).

The scoring results are presented in Fig. 5. Intact B1C8 speckles were observed in >90% of control (untransfected) nuclei (Fig. 5, *light gray bar*). Intact speckles were observed in ~76% of nuclei expressing a control protein, F-CTDless.1. The significance of this reduction is uncertain, but it is interesting to note that F-CTDless.1 contains a heptapeptide-like sequence on its COOH terminus, which was derived from the region upstream of the CTD (Materials and Methods). These heptapeptide-like sequences are deleted in F-CTDless.2, and interestingly, intact speckles were observed in ~86% of nuclei expressing this control protein. Expression of F-CTDless.3, which has no heptapeptide-like sequences, does not reduce the frequency of intact B1C8 speckles (Du, L., and S.L. Warren, unpublished results). Intact B1C8 speckles were observed in ~70% of cell nuclei expressing F-CTD1, and significantly, the addition of 2–4 heptapeptides markedly increases the B1C8 disrupting activity: only ~30% of nuclei expressing F-CTD3 or F-CTD6 have intact B1C8 speckles. The addition of 7, 20, or 26 heptapeptides to F-CTD6 does not further reduce the frequency of nuclei with intact B1C8 speckles, but the longer CTD segments (e.g., F-CTD13, F-CTD26, and F-CTD32) induce a more severe disruption of the B1C8 speckles than short CTD segments (not reflected by the histogram in Fig. 5). Significantly, F-CTD52 induces a complete disruption of the B1C8 speckled pattern in nearly 100% of the transfected nuclei. A similar trend was observed with a nested set of CTD sequences linked to βGal (Du, L., and S.L. Warren, unpublished results). These data indicate that the speckled distribution of an SR splicing protein (B1C8) is incrementally disrupted by the stepwise addition of heptapeptide repeats to the fusion protein.

Multiple SR Splicing Factors and Sm snRNPs Redistribute from a Speckled to a Diffuse Pattern in Nuclei Expressing CTD-derived Proteins

B1C8 is one of many SR family splicing proteins in speckle domains (reviewed by Fu, 1995). Other anti-SR mAbs, such as 3C5 104 (Roth et al., 1991), NM22, and NM4 (Blencowe et al., 1995) recognize multiple overlapping sets of SR family proteins. To ascertain whether CTD-derived proteins alter the distribution of the SR proteins recognized by these reagents, we repeated the experiment described in Fig. 3, *A–C*, except mAb 3C5, mAb 104, mAb NM4, or NM22 was substituted for mAb B1C8. Our results indicate that F-CTD52 disrupts the speckled staining pattern of all four antibodies (Du, L., and S.L. Warren, unpublished results).

Speckle domains are also enriched with other classes of splicing factors, such as Sm snRNPs and U-rich snRNAs (reviewed by Fu, 1995; Sharp, 1994). The preceding study showed that Pol IIo can be co-immunoprecipitated with antibodies directed at Sm snRNPs (Kim et al., 1997), so we asked whether CTD-derived proteins induced Sm snRNP antigens to become dispersed. The Sm snRNPs were local-

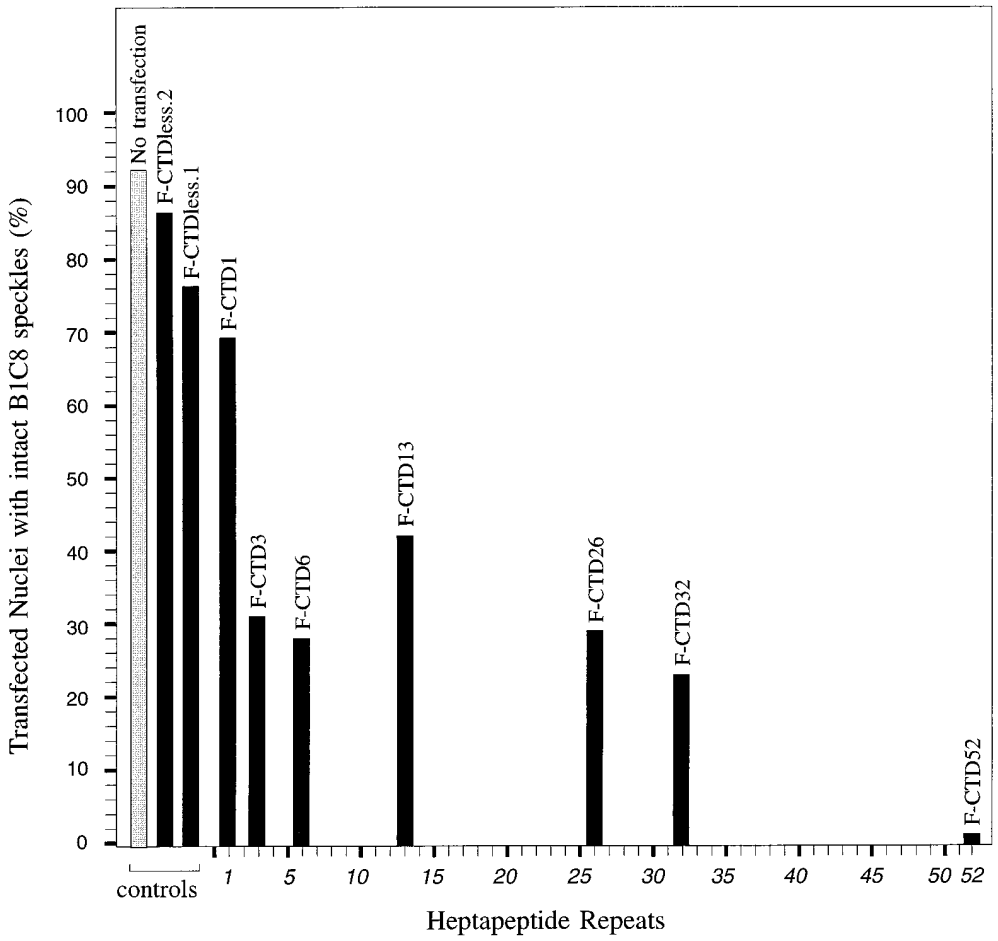

Figure 5. Relationship between CTD length and disruptive effect on B1C8 speckles. CV1 cells were transfected with plasmids encoding the fusion proteins listed above the histogram bars. 2 d later, the cells were fixed and double stained with antibodies directed at the Flag® epitope and B1C8 as described in Fig. 4. The pattern of B1C8 staining in each transfected cell nucleus was scored as "intact" (20–50 prominent speckles) or "disrupted" (diffuse pattern or diminutive speckles). Data were pooled from multiple experiments performed on different days. 150–250 nuclei were scored for each plasmid.

ized with mAb Y12 (an IgG), so the anti-Flag mAb M2 could not be used for double staining. We addressed this problem in two ways. In the first experiment, transfected cell nuclei were distinguished from untransfected nuclei by immunostaining with mAb H5 (IgM). This antibody recognizes phosphoepitopes on the CTD, and it stains nuclei expressing phosphorylated CTD-derived proteins much more intensely than control nuclei (Du, L., and S.L. Warren, unpublished results). In a second experiment, CV1 cells were transfected with pβGal-CTD52, and double stained with anti-βGal (rabbit IgG) and mAb Y12.

The results are presented in Fig. 6, *A–F.* Three untransfected cell nuclei are immunostained relatively weakly with mAb H5 (Fig. 6 *A*). In contrast, one nucleus expressing F-CTD52 is intensely immunostained (Fig. 6 *A*, *upper right*). Sm antigens are observed in speckle domains of the untransfected nuclei, but they are diffusely distributed in the transfected cell nucleus (Fig. 6 *B*, *upper right*). The βGal-linked CTD-52 protein has a more striking effect on the Sm antigens. Immunostaining with anti-βGal reveals a brightly stained nucleus expressing the βGal-CTD52 protein, and three faintly stained control cell nuclei (Fig. 6 *D*). Examination of the same cells stained with mAb Y12 reveals that the Sm antigens are distributed more diffusely in the transfected cell nucleus than in the untransfected cell nuclei (Fig. 6 *E*, *upper right*).

Expression of CTD-derived Fusion Proteins Disrupts Speckle Domains, but Not Coiled Bodies

Coiled bodies (CBs) are dot-like nuclear domains that contain certain snRNPs and snRNAs that are also present in the speckle domains (Lamond and Carmo-Fonseca, 1993). Most cultured mammalian cells have 2–5 CBs, which are easily visualized by immunostaining with antibodies directed at the p80 coilin autoantigen (Andrade et al., 1993). Speckles and CBs both contain certain splicing components, but their composition is otherwise very different: Pol IIo and SR splicing factors are present in speckle domains, but they have not been reported in CBs. Similarly, CBs contain p80 coilin, fibrillarin, and Nopp140, which have not been reported in speckle domains. Finally, transcriptional inhibitors and heat shock cause CBs to shrink and speckle domains to enlarge, suggesting distinct physiological roles for these two types of domains (Lamond and Carmo-Fonseca, 1993).

To ascertain whether the CTD-derived proteins disrupt the organization of CBs, each Flag-tagged CTD-derived protein was expressed transiently in CV1 cells, which were fixed and double immunostained with anti-p80 coilin and anti-Flag mAb M2. Our results indicate that the distribution of p80-coilin is unaffected by CTD-derived proteins F-CTD52, F-CTD32, F-CTD26, and F-CTD13. In the example presented here, CBs are observed in a control cell

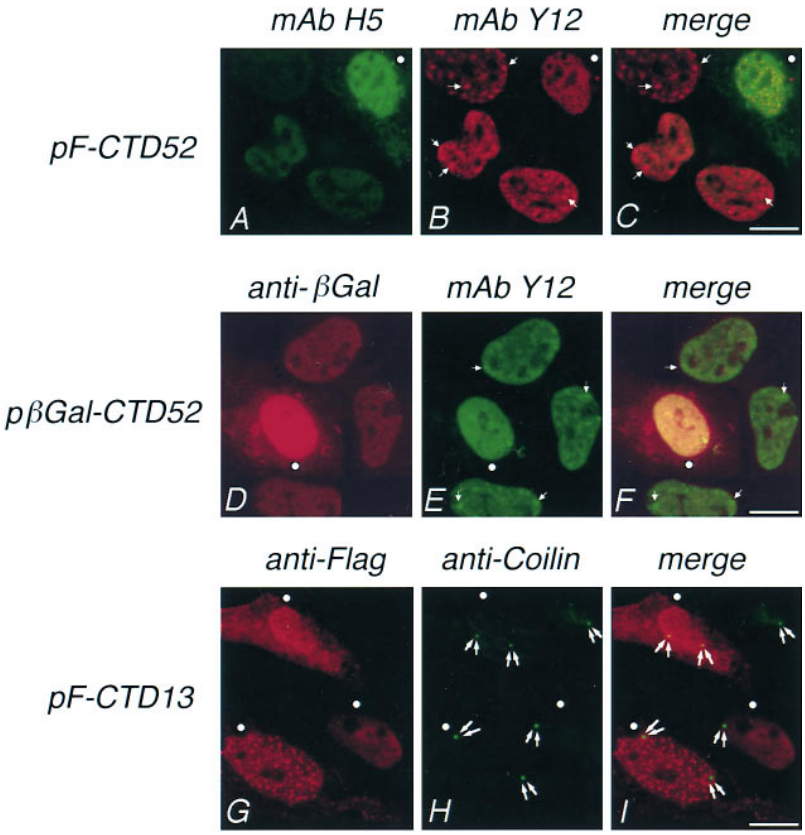

Figure 6. CTD52 disrupts the speckled distribution of Sm snRNPs without altering the distribution of p80 coilin. Transfection and immunostaining were performed as described above. (*A–C*) CV1 cells were transfected with pF-CTD52, fixed and double immunostained with mY12 (directed at Sm snRNPs; Lerner et al., 1981), and mAb H5 (directed at CTD phosphoepitopes; Bregman et al., 1995; Kim et al., 1997). A nucleus expressing F-CTD52 is identified by the intense immunostaining with mAb H5 (*upper right corner*). Three untransfected cell nuclei are identified by weaker mAb H5 immunostaining. (*D–F*) CV1 cells were transfected with pβGal-CTD52. A nucleus expressing βGal-CTD52 is identified by intense immunostaining with mAb anti-βGal, and three untransfected cell nuclei are identified by faint immunostaining with mAb anti-βGal. (*G–I*) CV1 cells were transfected with pF-CTD13. The distribution of F-CTD13 is revealed by red pseudocolor in three transfected cell nuclei (*G* and *I*). The distribution of p80 coilin is revealed by green pseudocolor in the three cells expressing F-CTD13, and in one untransfected cell (*H* and *I*). *White dots*, nuclei expressing CTD-derived fusion proteins; *single arrows*, speckles stained by mAb Y12; *double arrows*, p80 coilin in coiled bodies (*H* and *I*). Bars, 10 μm.

nucleus as well as three nuclei expressing F-CTD13 (Fig. 6, *G–I*, *double arrows*).

Expression of F-CTD52 Blocks the Accumulation of Spliced, but Not Unspliced, β-Globin Transcripts In Vivo

The transfection experiments described above, in conjunction with the accompanying study (Kim et al., 1997), support the hypothesis that Pol IIo associates with SR splicing factors and Sm snRNPs and via its CTD, but they do not provide evidence indicating a functional relationship between the CTD and pre-mRNA splicing. If the processes of transcription and splicing are linked via a CTD-mediated mechanism, then one might expect the CTD-derived proteins to interfere with transcription, splicing or both processes. We approached this question by co-expressing CTD-derived proteins and β-globin transcripts in the same nucleus. A series of double expression plasmids was created by inserting intact β-globin genes into pF-CTD52, pF-CTD13, pF-CTD6, and pF-CTD1 (Fig. 7 *A*). As controls, intact β-globin genes were inserted into pF-CTDless.1, pF-CTDless.3, and pβGal (Fig. 7 *A*). Each Flag-tagged CTD coding sequence (or control sequence) is under the control of a CMV promoter, whereas the β-globin gene is driven by its own promoter. The β-globin genes were also inserted in the opposite orientation relative to the Flag-tagged CTD coding sequences to control for possible *cis* effects (Fig. 7 *B*).

First, HeLa cells were transfected with pF-CTDless.1β-globin [+], pF-CTD13β-globin [+], or pF-CTD52β-globin [+] (Fig. 8 *A*). 24 h later, spliced and unspliced β-globin transcripts were quantitated by RT-PCR. PCR primers (P1 and P2) hybridize with sequences within exons 1 and 2, and therefore amplify a segment that includes intron 1 (Fig. 7 *A*). The PCR products corresponding to spliced and unspliced β-globin transcripts are 170 and 300 nucleotides, respectively. The results of this experiment indicate that co-expression of F-CTD52 reduces the amount of spliced β-globin transcript compared to the control, F-CTDless.1 (Fig. 8 *A*, *S*, lanes *2* and *4*). In contrast, slightly more unspliced β-globin transcript accumulates in cells co-expressing F-CTD52 than in the control cells (Fig. 8 *A*, *U*, lanes *2* and *4*). An intermediate effect is achieved by co-expressing F-CTD13 (Fig. 8 *A*, lane *3*, *U*).

This result was confirmed using an RNase protection assay (Fig. 8 *B*, lanes *2–4*). Here, a 343-nt protecting RNA probe (Fig. 8 *B*, lane *8*) was designed to hybridize with 203 nucleotides of the second β-globin exon and 73 nucleotides at the 3′ end of intron 1. Thus, unspliced β-globin transcripts protect 276 nucleotides, and spliced transcripts protect 203 nucleotides of the radiolabeled probe (Fig. 7 *A*). Similar amounts of spliced and unspliced β-globin transcripts are present in cells expressing the control protein (Fig. 8 *B*, lane *2*); however, one observes no spliced β-globin RNA in cells co-expressing the FCTD52 protein (Fig. 8 *B*, lane *4*, *S*). Significantly, this reduction is accompanied by an increase of unspliced β-globin transcript (Fig. 8 *B*, lane *4*, *U*). Splicing is inhibited to a lesser degree by F-CTD13 than F-CTD52 (Fig. 8 *B*, lane *3*, *U*).

To control for possible *cis* effects between the β-globin gene and CMV-Flag-CTD transcription unit, we reversed their relative orientation on the plasmids. The resulting plasmid constructs (pF-CTDless.1β-globin [−], pF-CTD13 β-globin

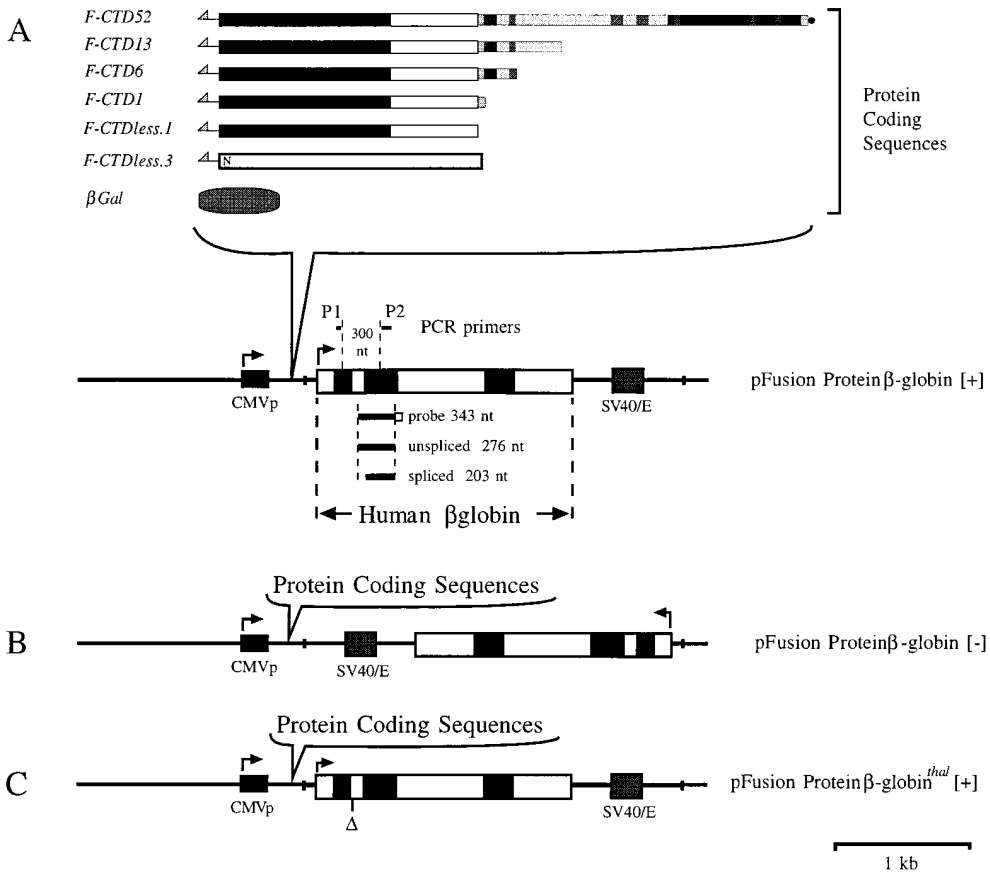

Figure 7. Plasmids expressing human β-globin transcripts and CTD-derived fusion proteins. (*A*) A wild-type human β-globin gene with a downstream SV40 enhancer (SV40E) was inserted into an EcoRV site in multiple plasmids that express Flag-tagged proteins or βGal (Fig. 1). For brevity the illustration depicts the insertion of various protein-encoding sequences into a site upstream of the β-globin gene. The Flag-tagged proteins and βGal coding sequences are under the control of the CMV promoter (CMVp) (for details see Materials and Methods). β-Globin introns are represented by open boxes, exons by black boxes, and noncoding flanking sequences open boxes at the ends of the gene. β-Globin and CMV promoters are indicated by bent arrows. The resulting constructs are generically termed "FusionProteinβ-globin [+]." The plus sign indicates that the two genes are oriented in the same direction. The primers (P1 and P2) hybridize with complementary (cDNA) sequences within exons 1 and 2, respectively. PCR amplification with P1 and P2 yields 170-nt and 300-nt DNA fragments corresponding to spliced and unspliced transcripts, respectively. The 343-nt RNA probe used for RNase protection is shown below the β-globin gene. The open box on this probe represents a nonhybridizing portion derived from pBluescript, and the black bar hybridizes with a 276-nt segment of the unspliced β-globin transcript. The 276-nt segment spans an intron-exon boundary including 203 nucleotides of exon 2 and 73 nucleotides of intron 1. Therefore, the spliced and unspliced β-globin transcripts protect 203 and 276 nucleotide segments of the probe, respectively. (*B*) A wild-type human β-globin gene with a downstream SV40 enhancer (SV40E) was also inserted in the opposite orientation of the EcoRV site in the plasmids expressing Flag-tagged proteins or βGal. The resulting constructs are generically termed "FusionProteinβ-globin [−]." The minus sign indicates that the two genes are oriented in the opposite direction. For convenience, the protein-encoding sequences are not shown. (*C*) A thalassemic human β-globin gene with a downstream SV40 enhancer (SV40E) was inserted in the positive orientation into the EcoRV site in the plasmids expressing Flag-tagged proteins or βGal. The resulting constructs are generically termed "FusionProteinβ-globin^{thal} [+]." The thalassemic allele is mutated at first residue of intron 1 (G to A transition) (*delta symbol*). Splicing of exons 1 and 2 is achieved by using three cryptic 5′ splice sites and the normal 3′ splice site (see Caceres et al., 1995). The oligonucleotide used for RNase protection spans the 3′ splice site, but it is downstream of the cryptic 5′ splice sites. Therefore, all three variably spliced transcripts register as 203 nucleotide RNAs in the RNase protection assay. For convenience, the protein-encoding sequences are not shown.

[−], or pF-CTD52β-globin [−]) were transfected into HeLa cells, and the RNase protection assay was performed. Again, one observes a reduction of spliced β-globin RNA in cells co-expressing the F-CTD52 protein (Fig. 8 *A*, lane *7*, *U*). A less severe inhibitory effect is achieved by co-expressing F-CTD13 (Fig. 8 *A*, lane *6*, *U*). More unspliced β-globin transcript accumulates in cells expressing CTD-derived proteins than in control cells (Fig. 8, *A* and *B*). This result indicates that CTD-derived proteins do not block in transcription by RNA polymerase II. Indeed, CTD-derived proteins selectively interfere with splicing.

The last series of experiments used a thalassemic β-globin gene that has a G to A transition at position 1 in intron 1 (see Caceres et al., 1994). The thalassemic pre-mRNAs are spliced at three cryptic 5′ splice sites, each located upstream of the 343-nt RNA probe used in the RNase protection assay. All three cryptically spliced products should protect 203 nucleotides of the radiolabeled probe, because they all use the same 3′ splice site. The thalassemic gene was substituted for wild-type β-globin in pF-CTDless.1β-globin [+], pF-CTD13β-globin [+], and pF-CTD52β-globin [+] (Fig. 7 *C*), the resulting plasmids were transfected into HeLa cells, and RNase protection experiments were performed as before. We have repeatedly found that splicing of this thalassemic transcript is particularly sensitive to the inhibitory effects of the CTD-derived proteins (Fig. 8 *C*).

Finally, we asked whether the removal of heptapeptide repeats from F-CTD52 progressively decreases the inhibitory effect on in vivo splicing. To test this idea, HeLa cells were transfected with plasmids that co-express the β-globin^{thal} transcript and one of a nested set of CTD-derived proteins.

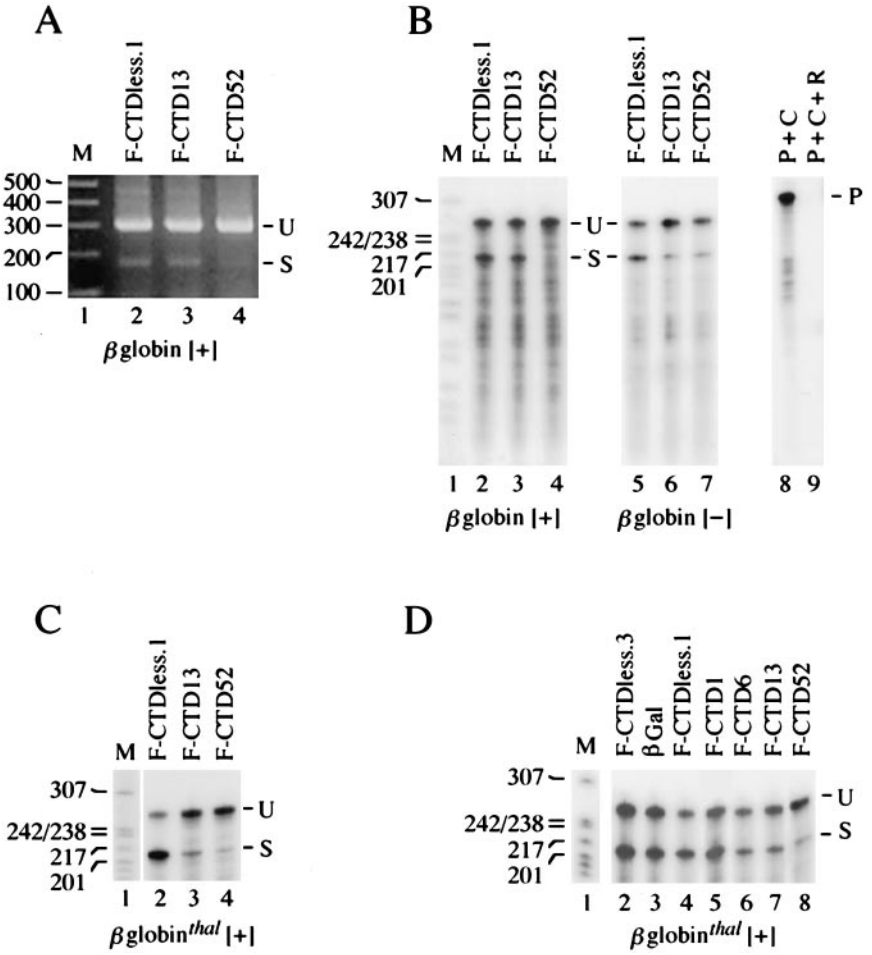

Figure 8. CTD-derived fusion proteins block the accumulation of spliced, but not unspliced, β-globin transcripts in vivo. (*A*) HeLa cells were transfected with pF-CTDless.1β-globin [+], pF-CTD_{13} β-globin [+], or pF-CTD_{52} β-globin [+]. 1 d later, total RNA was prepared from the cells as described in the Materials and Methods. The RNA was reverse transcribed and amplified by PCR using primers P1 and P2 (see Fig. 7 *A*). The PCR products were separated by agarose gel electrophoresis, stained with ethidium bromide, and photographed. (*B*) HeLa cells were transfected with pF-CTDless.1β-globin [+], pF-CTD_{13} β-globin [+], pF-CTD_{52} β-globin [+], pF-CTDless.1 β-globin [−], pF-CTD_{13} β-globin [−], or pF-CTD_{52} β-globin [−]. 1 d later, RNA was prepared and subjected to the RNase protection assay described in Fig. 7 and Materials and Methods. Protected RNAs were separated by electrophoresis and processed for autoradiography. (*C*) HeLa cells were transfected with pF-CTDless.1 β-$globin^{thal}$ [+], pF-CTD_{13} β-$globin^{thal}$ [+], or pF-CTD_{52} β-$globin^{thal}$ [+]. 1 d later, RNase protection assays were performed as described in *B*. (*D*) HeLa cells were transfected with pF-CTDless.3β-$globin^{thal}$ [+], pβGalβ-$globin^{thal}$ [+], pF-CTDless.1β-$globin^{thal}$ [+], pF-CTD_1 β-$globin^{thal}$ [+], pF-CTD_6 β-$globin^{thal}$ [+], pF-CTD_{13} β-$globin^{thal}$ [+], and pF-CTD_{52} β-$globin^{thal}$ [+]. 1 d later, RNase protection assays were performed as described in *B*. In *A–D*, the expressed Flag-tagged proteins are indicated at the top of the panel. The β-globin transcripts, and their orientation relative to the CMV driven transcription unit, are indicated below the panels. *U*, unspliced; *S*, spliced; *P*, intact probe; *C*, control yeast RNA; *R*, RNase added; *M*, molecular weight markers. MWs are indicated in base pairs at the left hand margin.

An RNase protection assay was performed as before (Fig. 8 *D*, lanes *2–8*). The ratio of unspliced to spliced β-$globin^{thal}$ transcripts is not significantly different in cells expressing F-CTD-less.3, βGal, F-CTD-less.1, or F-CTD1 (Fig. 8 *D*, lanes *2–5*); however, this ratio increases progressively as one adds 6, 13, and 52 heptapeptide repeats to the fusion protein (Fig. 8 *D*, lanes *6–8*). Indeed, a comparison of spliced β-$globin^{thal}$ transcripts in cells expressing F-CTD1, F-CTD6, F-CTD13, and F-CTD52 reveals a graded inhibition of splicing, which correlates with the number of heptapeptides added to the fusion protein.

Discussion

The experiments in the preceding paper, including the key observation that anti-CTD phosphoepitope-specific mAbs H5 and H14 release Pol IIo from immunoprecipitates prepared with splicing factor antibodies, suggested that Pol II's CTD may interact with certain splicing components. As an initial test of this hypothesis, we transfected cells with CTD-containing proteins that lack DNA-binding and catalytic domains, and asked whether they would localize specifically in the splicing factor domains. Originally, we had predicted that the CTD might behave like SR domains, which target indicator proteins to speckle domains (Li and Bingham, 1991). However, we found that the CTD-derived fusion proteins induce a striking disruption of the speckle domains. The disruptive effect is global, since multiple SR family proteins and Sm snRNPs redistribute from the speckles to a diffuse nucleoplasmic distribution. Moreover, the effect is specific for speckle domains, since the distribution of proteins in other types of domains is unaffected.

The ability of CTD-derived proteins to disrupt splicing factor domains led us to ask whether these proteins can specifically affect pre-mRNA splicing in vivo. Overexpression of the F-CTD52 protein blocks the accumulation of spliced β-globin transcripts, but it does not block Pol II–mediated transcription as indicated by the abundance of unspliced β-globin transcripts (Fig. 8). The selective effect of F-CTD52 on in vivo splicing provides the strongest evidence that one function of the CTD is related to pre-mRNA splicing.

Removal of heptapeptides diminishes two in vivo properties of the Flag-tagged fusion proteins: their ability to inhibit splicing (Fig. 8) and their ability to disrupt speckle domains (Figs. *4* and *5*). The correlation between the speckle

disruption and inhibition of splicing leads to the question of whether pre-mRNA splicing takes place in the speckles. Studies from other investigators indicate that certain Pol II transcripts are produced and spliced in nucleoplasmic sites outside of the SR protein-rich speckle domains (Zhang et al., 1994), while other transcripts are produced and spliced within, or at the periphery of, the speckle domains (Xing et al., 1993, 1995). The observation that CTD-derived proteins disrupt speckle domains and interfere with splicing argues that the CTD selectively affects (or interacts with) splicing components, but it does not help define where Pol II transcription and pre-mRNA splicing take place relative to the SR protein-rich speckle domains.

Two recent studies provide independent evidence that Pol IIo is associated with splicing factors. Yuryev and colleagues used a yeast two-hybrid screen to identify CTD interacting proteins in rat cells (Yuryev et al., 1996). Four proteins were identified, each containing repetitive SerArg dipeptide (SR) motifs characteristic of the SR superfamily of proteins; however, the SR domains in these proteins do not bind to the CTD. One of these proteins, rA1, was shown to bind yeast Pol II in overlay assays. These investigators also reported that wild-type, but not mutant CTD peptides inhibit in vitro splicing reactions. Significantly, rA1 is the only putative CTD-binding protein reported to interact with Pol II molecules containing a hyperphosphorylated CTD. Independently, Pol IIo was also detected in pre-mRNA splicing complexes assembled in vitro (Blencowe et al., 1996).

Pol II transcription and pre-mRNA splicing are known to be closely associated processes in evolutionarily diverse eukaryotic species (reviewed by Beyer and Osheim, 1991). Pol II transcripts are decorated with spliceosomal molecules in the lampbrush chromosomes (Gall, 1991; and references therein) and in polytene chromosomes (Matunis et al., 1993; Barén and Wieslander, 1994). Visualization of transcriptonally active chromatin by electron microscopy strongly suggests that introns are excised cotranscriptionally (Beyer and Osheim, 1988). Recently, co-transcriptional splicing has been directly demonstrated in polytene chromosomes of *C. tentans* (Barén and Wieslander, 1994). Transcription and splicing are also closely associated in mammalian cell nuclei. Fluorescent in situ hybridization experiments reveal that synthesis and splicing of specific Pol II transcripts takes place in coincident foci in mammalian cell nuclei (Xing et al., 1993; Zhang et al., 1994; Xing et al., 1995). Additional evidence indicating that Pol II transcription and splicing may be coordinated processes comes from plasmid transfection and viral infection studies. All of the above studies indicate a close connection between Pol II transcription and splicing, but the mechanism by which spliceosomes are recruited to Pol II transcripts remains poorly understood.

Nascent pre-mRNAs may contain all of the information required to recruit splicing factors. According to this model, transcription and splicing machinery would be linked exclusively by the pre-mRNA that is synthesized by the polymerase. Alternatively, the basal Pol II transcription machinery may participate directly in the recruitment and assembly of splicing factors on the nascent pre-mRNAs, as proposed in a speculative, but prescient model (Greenleaf, 1993). According to this model, the phosphorylated CTD helps recruit SR splicing factors to nascent Pol II transcripts. Thus, in vivo spliceosome assembly would take place processively on pre-mRNAs as they emerge from the polymerase.

The results presented here provide compelling experimental support for the model proposed by Greenleaf (1993) a few years ago. The preceding study showed that splicing factors associate with Pol IIo without the direct involvement of RNA, and surprisingly, the association is maintained at times when the polymerase is not engaged in transcription (Kim et al., 1997). The present study shows that CTD-derived proteins are phosphorylated in vivo and accumulate in the nucleus, where they disrupt splicing factor domains and interfere with pre-mRNA splicing. In agreement with these in vivo results, CTD heptapeptides were shown to specifically inhibit in vitro splicing reactions (Yuryev et al., 1996). Taken together, these studies provide evidence for a functional interaction between Pol II's CTD and the splicing process, and they strongly imply that transcription and pre-mRNA splicing are coordinated by a mechanism involving a phosphorylated form of the CTD.

We thank Euikyung Kim and David Bregman for helpful comments and for critically reading the manuscript. We thank Michael Dahmus, Jeffery Nickerson, Phil Cohen, Ed Chan, Gerd Maul, Xiang-Dong Fu, and Joan Steitz for providing antibodies (see Materials and Methods). We thank Adrian Krainer for providing β-globin genes, and Jon S. Morrow for pcDNA3AB and pSF. We thank David Ward for the use of his imaging facilities, and we are particularly grateful for the excellent technical assistance of Xun Sun.

The work was supported by the Council for Tobacco Research (#3881) and National Institutes of Health (K08-CA01339) to S.L. Warren.

Received for publication 1 June 1996 and in revised form 1 November 1996.

References

Andrade, L.E., E.M. Tan, and E.K.L. Chan. 1993. Immunocytochemical analysis of the coiled body in the cell cycle and during cell proliferation. *Proc. Natl. Acad. Sci. USA*. 90:1947–1951.

Ascoli, C.A., and G.G. Maul. 1991. Identification of a novel nuclear domain. *J. Cell Biol.* 112:785–795.

Barén, G., and L. Wieslander. 1994. Splicing of Balbiani Ring 1 gene pre-mRNA occurs simultaneously with transcription. *Cell.* 76:183–192.

Beyer, A.L., and Y.N. Osheim. 1988. Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev.* 2:754–765.

Beyer, A.L., and Y.M. Osheim. 1991. Visualization of RNA transcription and processing. *Semin. Cell Biol.* 2:131–140.

Blencowe, B.J., J.A. Nickerson, R. Issner, S. Penman, and P.A. Sharp. 1994. Association of nuclear matrix antigens with exon-containing splicing complexes. *J. Cell Biol.* 127:593–607.

Blencowe, B.J., R. Issner, J. Kim, P. McCaw, and P.A. Sharp. 1995. New proteins related to the SerArg family of splicing factors. *RNA.* 1:852–865.

Blencowe, B.J., M.J. Mortillaro, X. Wei, H. Nakayasu, L. Du, S.L. Warren, P.A. Sharp, and R. Berezney. 1996. A hyperphosphorylated form of the largest subunit of RNA polymerase II is associated with splicing complexes. *Proc. Natl. Acad. Sci. USA*. 93:8253–8257.

Bregman, D.B., L. Du, S. Ribisi, and S.L. Warren. 1994. Cytostellin distributes to nuclear regions enriched with splicing factors. *J. Cell Sci.* 10:387–396.

Bregman, D.B., L. Du, S. van der Zee, and S.L. Warren. 1995. Transcription-dependent redistribution of the large subunit of RNA polymerase II to discrete nuclear domains. *J. Cell Biol.* 129:287–298.

Caceres, J.F., S. Stamm, D.M. Helfman, and A.R. Krainer. 1994. Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. *Science (Wash. DC)*. 265:1706–1709.

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

Fu, X-D. 1995. The superfamily of arginine/serine-rich spliting factors. *RNA.* 1: 663–680.

Gall, J.G. 1991. Spliceosomes and snurposomes. *Science. (Wash. DC)*. 252: 1499–1500.

Gerber, H.P., M. Hagmann, K. Seipel, O. Georgiev, L.A. West, Y. Litingtung,

W. Schaffner, and J.L. Corden. 1995. RNA polymerase II C-terminal domain required for enhancer-driven transcription. *Nature (Lond.).* 374:660–662.
Greenleaf, A.L. 1993. Positive patches and negative noodles: linking RNA processing to transcription? *Trends Biochem. Sci.* 18:117–119.
Kim, E., L. Du, D. Bregman, and S. Warren. 1997. Splicing factors associate with hyperphosphorylated RNA polymerase II in the absence of pre-mRNA. *J. Cell Biol.* In press.
Koleske, A.J., and R.A. Young. 1995. The RNA polymerase II holoenzyme and its implications for gene regulation. *Trends Biochem. Sci.* 20:113–116.
Lamond, A.I., and M. Carmo-Fonseca. 1993. The coiled body. *Trends Cell Biol.* 3:198–204.
Lerner, E.A., M.R. Lerner, C. Janeway, and J.A. Steitz. 1981. Monoclonal antibodies to nucleic acid containing cellular constituents: probes for molecular biology and autoimmune disease. *Proc. Natl. Acad. Sci. USA.* 78:2737–2741.
Li, H., and P.M. Bingham. 1991. Arginine/Serine-rich domains of the *su(w^a)* and *tra* RNA processing regulators target proteins to a subnuclear compartment implicated in splicing. *Cell.* 67:335–342.
Matunis, E., M.J. Matunis, and G. Dreyfuss. 1993. Association of individual hnRNP proteins and snRNPs with nascent transcripts. *J. Cell Biol.* 121:219–228.
Maul, G.G., and R.D. Everett. 1994. The nuclear location of PML, a cellular member of the C3HC4 zinc-binding domain protein family, is rearranged during herpes simplex virus infection by the C3HC4 viral protein ICP0. *J. Gen. Virol.* 75:1223–1233.
Nonet, M., D. Sweetser, and R.A. Young. 1987. Functional redundancy and structural polymorphism in the large subunit of RNA polymerase II. *Cell.* 50:909–915.
Romac, J.M., and J.D. Keene. 1995. Overexpression of the arginine-rich carboxy-terminal region of U1 snRNP 70K inhibits both splicing and nucleocytoplasmic transport of mRNA. *Genes Dev.* 9:1400–1410.
Roth, M.B., A.M. Zahler, and J.A. Stolk. 1991. A conserved family of nuclear phosphoproteins localized to sites of polymerase II transcription. *J. Cell Biol.* 115:587–596.
Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. Molecular Cloning: A Laboratory Manual. 2nd Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Scafe, C., D. Chao, J. Lopes, J.P. Hirsch, S. Henry, and R.A. Young. 1990. RNA polymerase II C-terminal repeat influences response to transcriptional enhancer signals. *Nature (Lond.).* 347:491–494.
Sharp, P.A. 1994. Split genes and RNA splicing. *Cell.* 77:805–815.
Terris, B., V. Baldin, S. Dubois, C. Degott, J.-F., Flejou, and D. Henin. 1995. PML nuclear bodies are general targets for inflammation and cell proliferation. *Cancer Res.* 55:1590–1597.
Warren, S.L., A.S. Landolfi, C. Curtis, and J.S. Morrow. 1992. Cytostellin: a novel, highly conserved protein that undergoes continuous redistribution during the cell cycle. *J. Cell Sci.* 103:381–388.
Wintzerith, M., J. Acker, S. Vicaire, M. Vigneron, and C. Kedinger. 1992. Complete sequence of the human RNA polymerase II largest subunit. *Nucleic Acids Res.* 20:910.
Xing, Y., C.V. Johnson, P.R. Dobner, and J.B. Lawrence. 1993. Higher level organization of individual gene transcription and RNA splicing. *Science (Wash. DC).* 259:1326–1330.
Xing, Y., C.V. Johnson, P.T. Moen, J.A. McNeil, and J.B. Lawrence. 1995. Nonrandom gene organization: structural arrangements of specific pre-mRNA transcription and splicing with SC–35 domains. *J. Cell Biol.* 131:1635–1647.
Yuryev, A., M. Patturajan, Y. Litingtung, R.V. Joshi, C. Gentile, M. Gebara, and J.L. Corden. 1996. The CTD of RNA polymerase II interacts with a novel set of SR-like proteins. *Proc. Natl. Acad. Sci. USA.* 93:6975–6980.
Zeng, C., E. Kim, S.L. Warren, and S.M. Berget. 1997. Dynamic relocalization of transcription and splicing factors dependent on transcriptional activity *EMBO (Eur. Mol. Biol. Organ.) J.* In press.
Zehring, W.A., J.M. Lee, J.R. Weeks, R.S. Jokerst, and A.L. Greenleaf. 1988. The C-terminal repeat domain of RNA polymerase II largest subunit is essential *in vivo* but is not required for accurate transcription initiation *in vitro. Proc. Natl. Acad. Sci. USA.* 85:3698–3702.
Zhang, G., K.L. Taneja, R.H. Singer, and M.R. Green. 1994. Localization of pre-mRNA splicing in mammalian nuclei. *Nature (Lond.).* 372:809–812.