Pre-mRNA Processing and the CTD of RNA Polymerase II: The Tail That Wags the Dog?

Minireview

Eric J. Steinmetz
Department of Biomolecular Chemistry
University of Wisconsin Medical School
1300 University Avenue
Madison, Wisconsin 53706-1532

Eukaryotic cells must execute a complex program to generate mature functional messenger RNA (mRNA). After the transcription of a nuclear gene by RNA polymerase II (RNA Pol II) is initiated, introns must be removed from the pre-mRNA and the mature 3′ end of the transcript defined by specific cleavage and polyadenylation. Although the splicing and 3′-end formation reactions can occur in isolation from transcription in vitro, there have been periodic reminders that transcription and pre-mRNA processing may be coupled in vivo. Electron microscopy and microdissection studies illustrate that splicing can occur cotranscriptionally (Beyer and Osheim, 1988; Wetterberg et al., 1996), and available evidence indicates that 3′-end processing not only precedes but also may be a prerequisite for termination of transcription (Connelly and Manley, 1988; Russo and Sherman, 1989). However, factors or mechanisms that might couple mRNA processing to transcription have been elusive.

Recent reports from a number of sources are beginning to fill in the missing pieces of the puzzle. The striking revelation linking these studies is the finding that splicing factors and the 3′-end formation apparatus are connected to the transcriptional machinery through the carboxy-terminal domain, or CTD, of the largest subunit of RNA polymerase II. This physical association may ensure that the nascent transcript is presented directly to the pre-mRNA processing machinery even as its synthesis continues, and furthermore may enable coordinate regulation of transcription and processing. In order to consider these implications, it is necessary first to touch briefly on the role of the CTD and its reversible phosphorylation in the transcription cycle, a subject of recent reviews (Koleske and Young, 1995; Dahmus, 1996).

The CTD is a conserved and essential structure consisting of a repeated heptapeptide with the consensus sequence, YSPTSPS. Mouse has 52 tandem repeats, while the yeast S. cerevisiae has 26 or 27. Two forms of RNA Pol II differing in the extent of phosphorylation on the CTD can be distinguished and are thought to have distinct functions in the transcription cycle. RNA Pol IIa, with a hypophosphorylated CTD, is the form that assembles into transcription initiation complexes. Interactions between the CTD of RNA Pol IIa and factors that may facilitate initiation complex assembly have been examined as a possible mechanism for transcriptional activation. Coincident with or shortly after transcription initiation, the CTD becomes hyperphosphorylated to generate the form of the enzyme known as RNA Pol IIo. The precise role of CTD phosphorylation in the transcription cycle remains unclear, but it may convert the polymerase from a form prone to pausing or termination into a form capable of processive elongation. After transcription terminates, initiation competence presumably is restored by dephosphorylation.

Additional functions of the CTD in the processing of pre-mRNA have been suggested previously (Corden, 1990; Greenleaf, 1993). The CTD is present only on RNA Pol II, which is uniquely responsible for synthesis of spliced and polyadenylated mRNA, and pre-mRNA transcripts engineered to be synthesized by RNA polymerase III in mammalian cells fail to be spliced (Sisodia et al., 1987). Greenleaf (1993) proposed that the negative charge of the hyperphosphorylated CTD on RNA Pol IIo might facilitate direct electrostatic interactions with highly basic arginine–serine (RS) dipeptide repeat sequences characteristic of many splicing factors in multicellular eukaryotes, including members of the SR protein family (reviewed by Fu, 1995). This prediction has proven to be very close to the mark.

Association of Pre-mRNA Splicing Factors with the CTD

Several groups have now established that SR proteins and other spliceosome components are indeed physically associated with RNA polymerase II. Antibodies specific for the RNA Pol II CTD have been found to coimmunoprecipitate from mammalian cell extracts a set of proteins reactive with anti-SR protein monoclonal antibodies (Yuryev et al., 1996; Kim et al., 1997) as well as spliceosomal snRNPs (Chabot et al., 1995; Vincent et al., 1996). Reciprocal coprecipitation of RNA Pol II by antibodies directed against SR proteins or against the Sm antigens common to spliceosomal snRNPs further supports the existence of a bona fide complex (Mortillaro et al., 1996; Kim et al., 1997). Components of this complex are not simply tethered to each other by nascent RNA, since the association of Pol II with SR and Sm proteins is resistant to exhaustive ribonuclease treatment and is maintained during stages of the cell cycle when transcription does not occur (Kim et al., 1997). Intermediates in the splicing reaction generated in vitro from exogenous pre-mRNA can be precipitated by anti-CTD antibodies, implying that functional spliceosomes associate with RNA Pol II even when it is not transcriptionally engaged (Chabot et al., 1995; Mortillaro et al., 1996).

RNA Pol IIo, the hyperphosphorylated form of the enzyme implicated in processive transcriptional elongation, is preferentially associated with splicing complexes. This is evident in the coprecipitation of Pol IIo, but not Pol IIa, by anti-SR and anti-Sm antibodies (Mortillaro et al., 1996; Kim et al., 1997). Several lines of evidence indicate that association with the RNA Pol IIo CTD is important for spliceosome function in vitro and in vivo. Remarkably, in vitro splicing of exogenous pre-mRNA substrate can be inhibited by anti-CTD antibody or by CTD peptides (Yuryev et al., 1996). Excess peptides or antibody may displace multiple splicing factors from the CTD, preventing their colocalization and thus inhibiting spliceosome assembly. Interestingly, the second catalytic step of splicing appears to be inhibited at low peptide concentrations, while the first step is inhibited only at higher peptide concentrations, consistent with differential titration of factors required for the

first and second steps. Thus, the CTD may provide a platform for the recruitment and stepwise assembly of spliceosome components.

Splicing in vivo likewise is inhibited by expression of polypeptides containing CTD repeats but lacking most of the rest of the RNA Pol II large subunit (Du and Warren, 1997). The recombinant CTD polypeptides are subject to phosphorylation and localize to the nucleus, where their presence results in dispersal of splicing factors from discrete loci of concentration. Splicing is also inhibited when cells transfected with an α-amanitin-resistant, CTD-truncated allele of the RNA Pol II large subunit are treated with the antibiotic, confining transcriptional activity to the truncated form of the polymerase (McCracken et al., 1997). The simplest interpretation of these results is that spliceosomes must be associated with transcriptionally engaged polymerase in order to catalyze efficient splicing in vivo.

Association of Cleavage/Polyadenylation Factors with the CTD

In vivo experiments employing α-amanitin-resistant forms of RNA Pol II revealed that 3′-end processing, like pre-mRNA splicing, is inhibited when the transcriptionally active polymerase lacks an intact CTD (McCracken et al., 1997). Furthermore, cleavage stimulatory factor (CstF) and cleavage/polyadenylation specificity factor (CPSF), but not poly(A) polymerase, are retained when HeLa cell extracts are chromatographed on CTD-affinity columns. A similar profile of polyadenylation factors copurifies extensively with RNA Pol II through affinity and gel-filtration steps. Binding experiments with in vitro–translated CstF subunits revealed that two of its three subunits are able to bind directly to immobilized CTD peptides. It is not yet known if binding of CPSF is direct or mediated through its association with CstF.

The finding that polyadenylation factors associate with the CTD suggests an explanation for the dependence of transcription termination on functional 3′-end formation signals (Connelly and Manley, 1988) and for the observation from the current study that termination in response to these signals also depends on an intact CTD. McCracken et al. (1997) propose that the recognition of polyadenylation signals may alter the association of processing factors with the CTD, which may cause the polymerase to switch to a termination-competent form. However, it remains equally possible that the act of cleavage itself, generating an uncapped 5′ end, is the primary event resulting in subsequent termination (Connelly and Manley, 1988). It should be possible now to design experiments to distinguish between these possibilities.

Novel CTD-Binding Proteins

What molecules may connect spliceosomes to the CTD? Yuryev et al. (1996) used the yeast two-hybrid interaction screen to identify mammalian factors that bind to repeats 36–52 of the mouse CTD, composed entirely of nonconsensus heptapeptides. Proteins of two novel classes were recovered. Members of both classes contain RS and RE dipeptide repeat domains resembling those found in SR proteins and other splicing factors, in apparent support of the prediction by Greenleaf (1993). However, interaction with the CTD in each case is mediated by a short domain distinct from the alternating

arginine repeats. Representatives of one class, typified by the rat rA1 and rA9 proteins, share an 80-residue CTD-binding domain at their carboxyl termini. Direct association of this domain with hyper- or minimally phosphorylated yeast RNA Pol II has been demonstrated in vitro. A second class is defined by the rA4 and rA8 proteins, which interact with the CTD in the two-hybrid assay via an amino-terminal domain of ~120 residues. The latter proteins further resemble SR proteins and SR-related splicing factors in having an RNP-consensus RNA recognition motif (RRM).

The CTD-binding proteins identified by Yuryev et al. (1996) are good candidates for some of the anti-SR reactive factors identified in RNA Pol II immunoprecipitates and may connect spliceosomes to the CTD through protein–protein or RNA–protein interactions. However, while they bear strong resemblance to known splicing factors, a role for these proteins in splicing has not yet been demonstrated. Thus, although these studies are tantalizingly close to the identification of a connection between spliceosomes and the CTD, the circle has not been completely closed.

What About Yeast?

Although the splicing reaction mechanism and most of the spliceosome's major components are well-conserved, pre-mRNA splicing in yeast and mammalian cells differs in several respects. Only about 4% of protein-coding genes in S. cerevisiae have introns, and alternative splicing is virtually nonexistent. Splicing proteins with RS domains, which play such pivotal roles in both constitutive and alternative splicing in multicellular organisms, are very poorly represented in S. cerevisiae and S. pombe. Genetic studies of CTD function in S. cerevisiae suggest that its major cellular role is related to transcriptional activation, since growth defects due to CTD truncation can be suppressed by mutations in transcription factors (Koleske and Young, 1995). Finally, the rule that transcripts synthesized by RNA Pol III are not spliced is violated by several different yeast species that have mRNA-type introns within their RNA Pol III–transcribed genes encoding the spliceosomal RNA U6 (Tani and Ohshima, 1991).

It is remarkable, therefore, that S. cerevisiae harbors a relative of the rA4/rA8 class of mammalian CTD-binding, SR-like proteins (Steinmetz and Brow, 1996). The yeast Nrd1 protein has an amino-terminal domain with sequence similarity to the CTD-binding domains of rA4 and rA8, and this region of Nrd1p was also found to interact weakly with nonconsensus repeats of the mouse CTD in the two-hybrid assay (Yuryev et al., 1996). Binding to the yeast CTD, which consists mostly of perfect matches to the heptapeptide consensus, is inferred but has not yet been demonstrated. Nrd1p also has a single RRM as well as a segment rich in RE and RS dipeptides, both situated in relative positions within the protein's primary structure comparable to the corresponding domains in the rA8 protein. Interestingly, homology searches have identified an additional class of proteins conserved at least among S. cerevisiae, S. pombe, and C. elegans that share a similar N-terminal CTD-binding domain but have zinc-finger repeats in place of the RRM and RE/RS domains found in the Nrd1, rA4, and rA8 proteins.
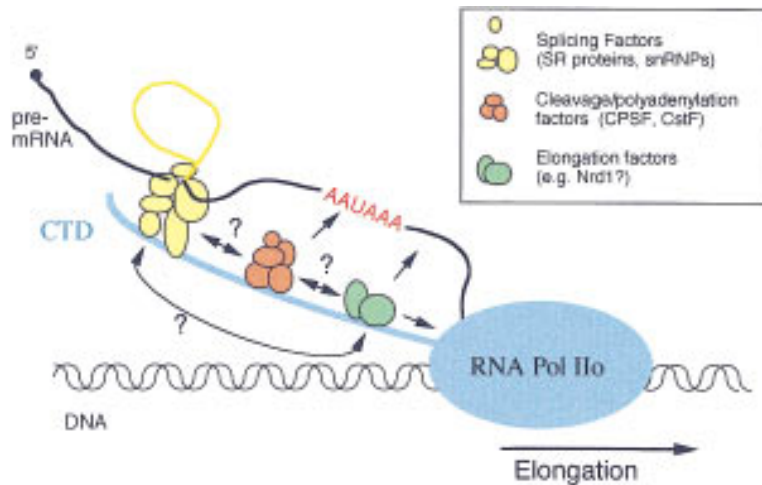
Figure 1. Schematic Representation of the Association of Pre-mRNA Processing Factors with the Carboxy-Terminal Domain (CTD) of Transcriptionally Engaged RNA Polymerase II

Potential physical and functional interactions between splicing, polyadenylation, and elongation factors are indicated by double-headed arrows.

Nrd1p mediates the severe decrease in synthesis of pre-mRNA that is caused by insertion of a complex sequence element into the intron of a reporter gene (Steinmetz and Brow, 1996). This effect can be overcome by substitution of a single amino acid in the Nrd1p RRM or by point mutations clustered within the sequence element, suggesting that binding of Nrd1p to the sequence element is required to decrease pre-mRNA synthesis. A likely mechanism for Nrd1p action is through control of transcript elongation, since truncated transcripts having 3′ ends just downstream of the sequence element accumulate in place of full-length pre-mRNA. Disruption of the *NRD1* gene is lethal, but the amino-terminal portion of the protein containing the CTD-binding domain is sufficient for viability, suggesting that interactions with the CTD may serve an essential function. Modulation of this function through interactions at Nrd1p's RNA-binding domain may lead to transcriptional pausing or termination. The apparent conservation of this CTD-binding domain is intriguing, and it will be of interest to learn whether any of the related CTD-binding factors might have a similar capacity to mediate control of pre-mRNA synthesis.

Implications

The studies described herein foster an expanded view of cellular functions of the CTD of RNA polymerase II. The biogenesis of mRNA in mammalian cells can now be seen as a set of coordinated processes (see Figure 1), with the CTD participating not only in the activation of transcription, but also as the foundation of an "RNA factory" (McCracken et al., 1997). At the level of resolution afforded by currently available data, it cannot be discerned whether this factory comprises an ordered assembly line or simply provides the necessary processing machinery in proximity to the site of RNA synthesis. In either case, the localization of this machinery may have significant ramifications for regulation of gene expression.

Although splicing in vivo can occur cotranscriptionally, this is not obligatory since some introns are not removed until after the RNA has been released from the chromosome (Wetterberg et al., 1996). Yet in vitro studies suggest that association with the CTD is maintained during the catalytic steps of splicing even when polymerase is not engaged in the synthesis of the splicing substrate. At what point are spliceosomes released from the CTD in vivo? Do processing factors dissociate from the CTD as they associate with nascent pre-mRNA, or might a polymerase molecule be committed to its newly synthesized pre-mRNA transcript until processing is complete, even after transcription has terminated? Can multiple spliceosomes assemble on a single CTD simultaneously, or must each intron be spliced in succession? What might be the significance for alternative splicing? Might microheterogeneous control of CTD phosphorylation and, thereby, factor binding play a role in splicing regulation? The answers to these questions will require detailed characterization of the molecular connections between the spliceosome and the CTD and dissection of the precise role of the CTD in spliceosome assembly and function.

It is attractive to consider the mRNA factory as a highly integrated machine. For example, intercommunication between splicing and polyadenylation factors (e.g., Lutz et al., 1996), facilitated by their colocalization, may allow the coordinated selection of splice junctions and cleavage/polyadenylation sites. The apparent role of the CTD in transcriptional activation raises the possibility that the specificity or efficiency of processing may be determined in part by promoter-specific factors, or that association with processing factors in turn may influence initiation. Promoter-dependent selection of snRNA versus mRNA 3′-end formation pathways (Dahlberg and Lund, 1988), for example, conceivably could be mediated by CTD-bound factors. Accumulating evidence implicates the CTD as a target for control of transcriptional elongation, whether through modulation of its phosphorylation state or its association with specific factors. An exciting corollary is that signals from the processing machinery may be propagated through the CTD to coordinate the continued elongation of a transcript with its processing, providing a form of mRNA surveillance or "quality control." Such a mechanism could lie at the heart of the observed coupling between 3′-end processing and transcript termination, as proposed by McCracken et al. (1997). A similar coupling mechanism could signal pausing or termination of transcription in response to delays or errors in spliceosome assembly or

function. These signals might be relayed through factors like the yeast Nrd1 protein or related mammalian SR-like proteins. Clearly, researchers investigating RNA Pol II transcription and those studying pre-mRNA processing will both contribute to this unfolding picture. With their combined efforts, a ''grand unified theory'' of mRNA biogenesis may soon be at hand.

Selected Reading

Beyer, A.L., and Osheim, Y.N. (1988). Genes Dev. *2*, 754–765.

Chabot, B., Bisotto, S., and Vincent, M. (1995). Nucleic Acids Res. *23*, 3206–3213.

Connelly, S., and Manley, J.L. (1988). Genes Dev. *2*, 440–452.

Corden, J.L. (1990). Trends Biochem. Sci. *15*, 383–387.

Dahlberg, J.E., and Lund, E. (1988). In Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles, M.L. Birnsteil, ed. (Berlin: Springer-Verlag), pp. 38–70.

Dahmus, M.E. (1996). J. Biol. Chem. *271*, 19009–19012.

Du, L., and Warren, S.L. (1997). J. Cell Biol. *136*, 5–18.

Fu, X.-D. (1995). RNA *1*, 663–680.

Greenleaf, A. (1993). Trends Biochem. Sci. *18*, 117–119.

Kim, E., Du, L., Bregman, D.B., and Warren, S.L. (1997). J. Cell Biol. *136*, 19–28.

Koleske, A.J., and Young, R.A. (1995). Trends Biochem. Sci. *20*, 113–116.

Lutz, C.S., Murthi, K.G.K., Schek, N., O'Connor, P.J., Manley, J., and Alwine, J.C. (1996). Genes Dev. *10*, 325–337.

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., and Bentley, D.L. (1997). Nature *385*, 357–361.

Mortillaro, M.J., Blencowe, B.J., Wei, X., Nakayasu, H., Du, L., Warren, S.L., Sharp, P.A., and Berezney, R. (1996). Proc. Natl. Acad. Sci. USA *93*, 8253–8257.

Russo, P., and Sherman, F. (1989). Proc. Natl. Acad. Sci. USA *86*, 8348–8352.

Sisodia, S.S., Sollner-Webb, B., and Cleveland, D.W. (1987). Mol. Cell. Biol. *7*, 3602–3612.

Steinmetz, E.J., and Brow, D.A. (1996). Mol. Cell. Biol. *16*, 6993–7003.

Tani, T., and Ohshima, Y. (1991). Genes Dev. *5*, 1022–1031.

Vincent, M., Lauriault, P., Dubois, M.-F., Lavoie, S., Bensaude, O., and Chabot, B. (1996). Nucleic Acids Res. *24*, 4649–4652.

Wetterberg, I., Baurén, G., and Wieslander, L. (1996). RNA *2*, 641–651.

Yuryev, A., Patturajan, M., Litingtung, Y., Joshi, R.V., Gentile, C., Gebara, M., and Corden, J.L. (1996). Proc. Natl. Acad. Sci. USA *93*, 6975–6980.