Minireview

Coordination between transcription and pre-mRNA processing

P. Cramer[b], A. Srebrow[a], S. Kadener[a], S. Werbajh[a], M. de la Mata[a], G. Melen[a], G. Nogués[a], A.R. Kornblihtt[a,*]

[a] *Laboratorio de Fisiología y Biología Molecular, Departamento de Ciencias Biológicas, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pabellón 2, C1428EHA Buenos Aires, Argentina*
[b] *Department of Molecular and Cellular Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA 02138, USA*

Received 14 April 2001; accepted 27 April 2001

First published online 17 May 2001

Edited by Gunnar von Heijne

Abstract A large body of work has proved that transcription by RNA polymerase II and pre-mRNA processing are coordinated events within the cell nucleus. Capping, splicing and polyadenylation occur while transcription proceeds, suggesting that RNA polymerase II plays a role in the regulation of these events. The presence and degree of phosphorylation of the carboxy-terminal domain of RNA polymerase II large subunit is important for functioning of the capping enzymes, the assembly of spliceosomes and the binding of the cleavage/polyadenylation complex. Nuclear architecture and gene promoter structure have also been shown to play key roles in coupling between transcription and splicing. © 2001 Published by Elsevier Science B.V. on behalf of the Federation of European Biochemical Societies.

Key words: RNA polymerase II; mRNA processing; Carboxy-terminal domain; Coupling

1. Introduction

Eukaryotic gene expression involves several steps that start with transcription. Three RNA polymerases (RNA pol) are in charge of transcribing the genes in the nucleus: RNA pol I transcribes most of the ribosomal RNAs (28S, 18S and 5.8S rRNAs); RNA pol II transcribes messenger RNAs (mRNAs), as well as some small nuclear RNAs (snRNAs); and RNA pol III synthesizes 5S rRNA, transfer RNAs (tRNAs) and some other snRNAs.

Many of these RNAs undergo modifications or *processing* that determine specific features, such as their location, stability, or interaction with macromolecular complexes. In the past few years, transcription and processing of RNAs have been studied thoroughly so we now have a large body of information about both the protein machineries involved and the sequences required on the DNA template and on the transcripts. However, most of this knowledge comes from in vitro experiments in which each step of RNA synthesis was studied in isolation in the test tube, thus leading to the view that these steps were independent of each other and took place subsequently. Evidence accumulated in recent years has proved this concept wrong as it shows that several of the steps involved in gene expression are coordinated within the cell nucleus. In fact, this coordination in some cases requires the interaction of genes and transcripts with protein complexes of such large size that they can be distinguished by light microscopy.

In this review we will focus on the coupling of transcription and RNA processing of mRNAs by RNA pol II. We encourage the reading of other recent reviews in the field [1–3].

2. RNA pol II plays a key role in coordination between transcription and processing

Transcripts synthesized by RNA pol II undergo specific and extensive processing before being transported to the cytoplasm. Capping at the 5′ end minimizes mRNA degradation and most importantly permits its interaction with the ribosomes in the cytoplasm. The 3′ end is completed by the addition of a polyadenosine monophosphate tail, resulting in increased mRNA stability. Transcribed intervening sequences, called introns, are removed. Far from being unrelated events that take place sequentially, transcription and processing are coordinated in both time and space. In many cases these modifications take place as transcription proceeds. This led to the idea that RNA pol II itself might be responsible for coordinating all the steps required for mature mRNA biogenesis.

RNA pol II is a multimeric protein. The enzymatic activity is located in its largest subunit, characterized by a carboxy-terminal domain (CTD) composed of 52 repeats of the heptad consensus peptide YSPTSPS in mammals. The fact that this domain is a unique feature of this protein, together with the presence of serines and threonines in the repeat that could be candidates for phosphorylation, suggested that the CTD and its putative phosphorylated residues could be the signals used by the protein machineries involved in mRNA synthesis and processing to contact the right polymerase [4]. In fact, RNA pol II is phosphorylated in vivo and can be isolated from mammals and flies in a hyperphosphorylated form (RNA pol IIO) or a hypophosphorylated form (RNA pol IIA) distinguishable by specific antibodies [5]. Immunoprecipitation of RNA pol II transcribing the *Drosophila melanogaster* hsp70 inducible gene showed that whereas elongating complexes are composed of RNA pol IIO, complexes that paused 20–40 bases downstream of the transcription start site are rich in RNA pol IIA [6]. The cdk7-cyclin H component of TFIIH in mammals and Kin28 in yeast [7] phosphorylate the CTD in

*Corresponding author. Fax: (54)-11-576 3321.
E-mail: ark@bg.fcen.uba.ar

PII: S0014-5793(01)02485-1

the preinitiation complex stimulating transcription by RNA pol II. P-TEFb in mammals [8], and in yeast SRB10, SRB11 [9] and CTK1 [10] also function as CTD kinases. The role of each kinase in transcription remains unclear. It has been proposed that different promoters could recruit different kinases [5]. This could result in varying elongation efficiencies that could in turn affect splicing or polyadenylation. A phosphatase called FCP1 has been described to act as an inhibitor of early elongating RNA pol II [11,12], supporting the view that cycles of phosphorylation/dephosphorylation are involved in transcription regulation at the level of elongation.

3. Capping and transcription are tightly coupled

Capping is the best described example of a pre-mRNA processing reaction coupled to transcription. After RNA pol II transcribes the first 25–30 nucleotides, the capping enzymes proceed to remove the γ-phosphate from the first nucleotide of the RNA, transfer GMP from GTP to this latter to form the GpppN, which is later methylated at the N7 position of the guanine. These activities are performed by two proteins in mammals: a bifunctional polypeptide with a triphosphatase domain in its amino-terminus and a guanylyltransferase domain in its carboxy-terminus, and a separate methyltransferase. In yeast, the triphosphatase (Cet1) and the guanylyltransferase (Ceg1) are separate subunits while a third polypeptide, called Abd1, methylates the guanine. The mammalian capping enzyme binds RNA pol IIO but not RNA pol IIA [13,14]. This interaction is sensitive to the CTD phosphorylation status as demonstrated by in vitro binding of the guanylyltransferase domain to CTD peptides phosphorylated at Ser5 (but not at Ser2) [15]. As a consequence of this interaction the affinity of guanylyltransferase for GTP increases two-fold. Moreover, in vivo functional studies performed in HeLa cells have shown that upon inhibition of the endogenous RNA pol II by the addition of α-amanitin and simultaneous transfection with a plasmid expressing an α-amanitin-resistant, CTD-less version of the enzyme, the cells are unable to cap an mRNA encoded in a reporter plasmid [16].

In yeast, direct contacts of Ceg1, Cet1 and Abd1 with the CTD were demonstrated in vivo. Surprisingly, Ceg1 is released early in elongation while Abd1 travels with the polymerase as far as the 3′ end of the gene [17]. Moreover, these contacts are restricted to genes that are being actively transcribed and require a functional Kin28 (but not CTK1 or SRB10) that phosphorylates Ser5 in the CTD. It seems then that CTD phosphorylation at this position is critical for mRNA capping. These findings are consistent with the previous finding that a viable mutant carrying a shortened CTD combined with a viable mutation of the capping enzyme leads to death [18].

4. Nuclear structure and the coordination between transcription and splicing

Most metazoan RNA pol II genes include introns that are eliminated from the pre-mRNA through splicing. The spliceosome, composed of small nuclear ribonucleoproteins (snRNP) and non-snRNP proteins, interacts with sequences in the RNA molecule helped by members of a family of arginine/serine-rich proteins (SR proteins). It was initially assumed that mRNA molecules were spliced only after being completely synthesized and released from the DNA template. However, examination of expression of the rat fibronectin (FN) gene employing fluorescent probes showed that the transcripts do not diffuse freely in the nucleoplasm from the site of transcription but form one elongated 'track' per allele. Interestingly, tracks are constituted of unspliced mRNA in the proximity of the gene and of mature mRNA towards the more distal parts [19]. Cytological examination of insect genes caught in active transcription showed that some introns are excised as the mRNA is being synthesized [20,21]. The CTD seems to play a central role in linking mRNA synthesis and splicing machineries. Antibodies directed against RNA pol II [22] or the CTD co-immunoprecipitate SR proteins and inhibit splicing both in vitro and in vivo [23,24]. Moreover, the phosphorylated form of CTD stimulates the early steps of spliceosome assembly thereby stimulating splicing while the hypophosphorylated form inhibits the formation of these complexes [25]. The CTD seems to accomplish this by bridging the splicing factors bound to the 3′ and 5′ splice sites at the ends of each exon [26], consistent with the mechanism of exon definition of splicing [27]. The role of CTD in splicing is so significant that transcripts generated by truncated versions of RNA pol II lacking the CTD cannot be spliced in vivo [16].

Studies on nuclear architecture also reveal coupling of transcription and splicing (for recent reviews see [28,29]). The nucleus presents several domains [30]. Among them, splicing factor compartments (SFCs, formerly called speckles [31,32]), which are present at 20–50 per cell, have been shown to associate with transcribing genes by means of live cell microscopy [33]. While this dynamic interaction is inhibited by α-amanitin it remains controversial whether it is intron-dependent [34] or -independent [35].

Coupling between transcription and splicing is not limited to just a temporal or spatial coincidence of the corresponding machineries. Our laboratory has shown that the structure of the promoter regulating the expression of a gene can affect splicing. Human cell lines were transiently transfected with a series of human α-globin/FN minigenes that include the alternatively spliced EDI exon of FN but differ in the promoter driving their expression. When the FN promoter included point mutations that had been previously shown to affect the recruiting of specific transcription factors [36] the ratio between the two alternatively spliced forms (EDI+/EDI−) was much higher than the one obtained from expression from the wild type FN promoter [37]. Furthermore, a variant of the plasmid carrying the human α-globin promoter expressed a dramatically lower proportion of EDI. Moreover, the promoter affected the responsiveness of this alternative splicing to the activation by the SR proteins SF2/ASF and 9G8 [38]. As these splicing factors require the presence of an intact splicing enhancer on the alternative exon in order to stimulate EDI inclusion, it was suggested that the transcription machinery modulates their recruitment to the splicing enhancer. A possible mechanism that would fit in these results is that the promoter itself is somehow responsible for recruiting these splicing factors to the site of transcription, possibly through transcription factors that bind the promoter or the transcriptional enhancers. The finding that p52, a transcriptional coactivator, directly interacts with SF2/ASF stimulating pre-mRNA splicing [39] is consistent with this model. Furthermore, some proteins could display a dual function, acting in both processes as is the case of a transcriptional activator

of human papilloma virus [40] or the thermogenic coactivator PGC-1 [41]. Alternatively, the RNA polymerase could be responsible for recruiting these proteins, perhaps through its CTD. The role of the promoter in this case could be to affect the extent of CTD phosphorylation which could in turn modify the ability of the CTD to interact with SR proteins [38].

5. Polyadenylation and transcription are interdependent events

With the exception of histone mRNAs, a polyadenosine monophosphate tail is added at the 3′ end of all eukaryotic messengers, resulting in increased RNA stability and also in improved translation. In mammals the cleavage specificity polyadenylation factor (CSPF) recognizes the consensus sequence AAUAAA (poly(A) signal), and the cleavage stimulation factor some U- or GU-rich elements located downstream. Although the cleavage reaction itself takes place some 10–30 nucleotides downstream of the poly(A) signal and requires the presence of two other proteins (CF I_m and CF II_m), RNA pol II does not stop transcription at this point but continues elongating the RNA for up to 1000 nucleotides. The poly(A) tail is then added by poly(A) polymerase with the assistance of poly(A) binding protein. It remains to be established which of the four proteins involved in cleavage cuts the mRNA, and more importantly, which are the signals for transcription stop in eukaryotes.

RNA pol II plays a key role in polyadenylation. Three of the four subunits that compose CSPF are actually part of TFIID (a basal transcription factor) and are transferred to the RNA pol II after initiation [42]. Immunoprecipitation of these factors with antibodies directed against the CTD confirms that this machinery contacts the CTD, as in the case of the capping reaction. Again, polyadenylation is impaired in HeLa cells in which the only active RNA pol II has a truncated CTD [16]. Then, we have to imagine that the transcribing RNA pol II carries at least part of the cleavage/polyadenylation machinery along the transcribing gene. It seems that its role is more than just transporting these proteins to the poly(A) signal, as both RNA pol II and the CTD alone were shown to stimulate polyadenylation in vitro, in a transcription-free assay [43]. Coupling between transcription and polyadenylation appears to be conserved from yeast to human as illustrated by the recent finding that phosphorylated CTD binds to a cleavage complex called CF1A in yeast [44].

6. Transcription termination and 3′ end mRNA processing cannot be resolved

Termination of transcription requires a functional polyadenylation site [45,46]. Electron micrographs of DNA:protein:RNA complexes isolated from *Xenopus laevis* oocytes injected with plasmids expressing pre-mRNAs that differ in the strength of their poly(A) signals revealed that the length of the nascent transcripts is constant when the poly(A) signal is strong, whereas a mutated signal correlates with much longer mRNAs that result from several rounds of transcription of the circular template [47].

Extensive mutation analysis made it possible to conclude that, besides a functional poly(A) signal, termination requires a functional acceptor site in the last intron [48]. The timing of the events at the 3′ end of the transcript also seems to be important as demonstrated by coupled polyadenylation–transcription studies in vitro. The presence of physiological pausing elements downstream of the poly(A) signal, followed by artificial 'roadblocks', activates polyadenylation [49]. This pausing of the enzyme might help its interactions with the

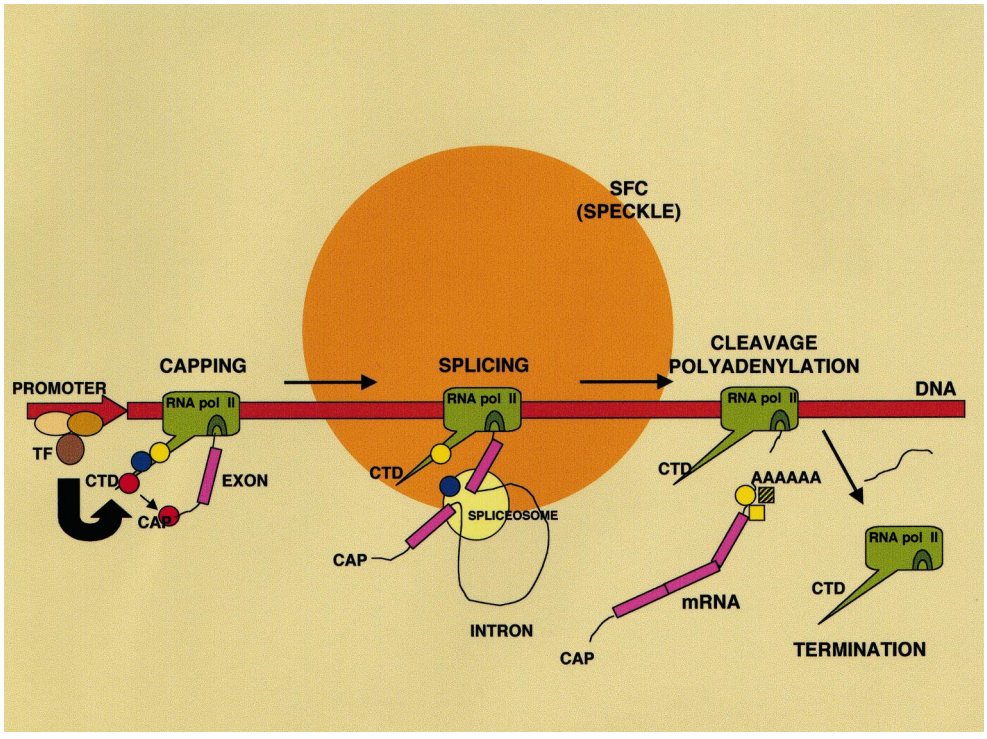

Fig. 1. Involvement of RNA pol II in different steps of pre-mRNA processing. TF: transcription factors. Small circles represent proteins involved in pre-mRNA processing that interact with RNA pol II, participating in capping (red), splicing (blue) or cleavage/polyadenylation (yellow) mechanisms. Yellow squares represent factors involved in cleavage/polyadenylation which do not interact with RNA pol II.

factors engaged in 3′ processing, including splicing of the last intron (Fig. 1).

7. Perspectives

The study of the molecular bases of pre-mRNA processing and, in particular, of alternative splicing regulation becomes especially relevant in the post-genome era. The identification of all human genes will be insufficient to understand cell-specific patterns of gene expression since it is estimated that the 30 000 genes in the human genome could produce several hundred thousand different proteins. The existence of a supramolecular compartmentalization in the mammalian cell nucleus and the discovery of functional and structural coordination between transcription and pre-mRNA processing establish a new conceptual frame to investigate the control of gene expression. The evidence discussed here implies that coupling between transcription and processing can be quite complex, comprising contacts between RNA pol II, cleavage/polyadenylation and splicing machineries. It has not been studied so far how many of the repeats in the CTD and which of them need to be phosphorylated for the coupling between transcription and processing. Perhaps only the overall quantity of phosphorylated residues (namely the extent of negative charge of the CTD) is what counts, or perhaps the phosphorylation of only some of the repeats is crucial. It may well be the case that different processing machineries require different 'formats' of the CTD. On the other hand, the physiological roles of the modulation of splicing by promoter structure and changes in nuclear architecture remain to be determined.

Acknowledgements: This review was written with the support of the Fundación Antorchas, the Agencia Nacional de Promoción Científica y Tecnológica, the International Center for Genetic Engineering and Biotechnology and the University of Buenos Aires.

References

[1] Bentley, D. (1999) Curr. Opin. Cell Biol. 11, 347–351.
[2] Hirose, Y. and Manley, J.L. (2000) Genes Dev. 14, 1415–1429.
[3] Proudfoot, N. (2000) Trends Biochem. Sci. 25, 290–293.
[4] Greenleaf, A.L. (1993) Trends Biochem. Sci. 18, 117–119.
[5] Dahmus, M.E. (1996) J. Biol. Chem. 271, 19009–19012.
[6] O'Brien, T., Hardin, S., Greenleaf, A.L. and Lis, T.J. (1994) Nature 370, 75–77.
[7] Svejstrup, J.Q., Feaver, W.J. and Kornberg, R.D. (1996) J. Biol. Chem. 271, 643–645.
[8] Kim, J.B. and Sharp, P.A. (2001) J. Biol. Chem., 5 Jan (electronic publication ahead of print).
[9] Liao, S.M. et al. (1995) Nature 374, 193–196.
[10] Sterner, D.E., Lee, J.M., Hardin, S.E. and Greenleaf, A.L. (1995) Mol. Cell. Biol. 15, 5716–5724.
[11] Cho, H., Kim, T.H., Mancebo, H., Lane, W.S., Flores, O. and Reinberg, D. (1999) Genes Dev. 13, 1540–1552.
[12] Marshall, N.F. and Dahmus, M.E. (2000) J. Biol. Chem. 275, 32430–32437.
[13] Yue, Z., Maldonado, E., Pillutla, R., cho, H., Reinberg, D. and Shatkin, A.J. (1997) Proc. Natl. Acad. Sci. USA 94, 12898–12903.
[14] McCracken, S. et al. (1997) Genes Dev. 11, 3309–3318.
[15] Ho, C.K. and Shuman, S. (1999) Mol. Cell 3, 405–411.
[16] McCracken, S. et al. (1997) Nature 385, 357–361.
[17] Schroeder, S.C., Schwer, B., Shuman, S. and Bentley, D. (2000) Genes Dev. 14, 2435–2440.
[18] Cho, E.J., Takagi, T., Moore, C.R. and Buratowski, S. (1997) Genes Dev. 11, 3319–3326.
[19] Xing, Y., Johnson, C.V., Dobner, P.R. and Lawrence, J.B. (1993) Science 259, 1326–1330.
[20] Bauren, G. and Wieslander, L. (1994) Cell 76, 183–192.
[21] Beyer, A.N. and Osheim, Y.N. (1988) Genes Dev. 2, 754–765.
[22] 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.
[23] Yuryev, A., Patturajan, M., Litingtung, Y., Joshi, R., Gentile, C., Gebara, M. and Corden, J. (1996) Proc. Natl. Acad. Sci. USA 93, 6975–6980.
[24] Du, L. and Warren, S.L. (1997) J. Cell Biol. 136, 5–18.
[25] Hirose, Y., Tacke, R. and Manley, J.L. (1999) Genes Dev. 13, 1234–1239.
[26] Zeng, C. and Berget, S.M. (2000) Mol. Cell. Biol. 20, 8290–8301.
[27] Berget, S.M. (1995) J. Biol. Chem. 270, 2411–2414.
[28] Misteli, T. (2000) J. Cell Sci. 113, 1841–1849.
[29] Lewis, J.D. and Tollervey, D. (2000) Science 288, 1385–1389.
[30] Lamond, A.I. and Earnshaw, W.C. (1998) Science 280, 547–553.
[31] Spector, D.L. (1993) Annu. Rev. Cell Biol. 9, 265–315.
[32] Misteli, T. and Spector, D.L. (1998) Curr. Opin. Cell Biol. 10, 322–331.
[33] Misteli, T. and Spector, D.L. (1997) Nature 387, 523–527.
[34] Huang, S. and Spector, D.L. (1996) J. Cell Biol. 131, 719–732.
[35] Jolly, C., Vourc'h, C., Robert-Nicoud, M. and Morimoto, R.I. (1999) J. Cell Biol. 145, 1133–1143.
[36] Alonso, C.R., Pesce, C.G. and Kornblihtt, A.R. (1996) J. Biol. Chem. 271, 22271–22279.
[37] Cramer, P., Pesce, G.C., Baralle, F.E. and Kornblihtt, F.E. (1997) Proc. Natl. Acad. Sci. USA 94, 11456–11460.
[38] Cramer, P., Cáceres, J.F., Cazalla, D., Kadener, S., Muro, A.F., Baralle, F.E. and Kornblihtt, A.R. (1999) Mol. Cell 4, 251–258.
[39] Ge, H., Si, Y. and Wolffe, A.P. (1998) Mol. Cell 2, 751–759.
[40] Lai, M., Teh, B. and Tarn, W. (2000) J. Biol. Chem. 274, 11832–11841.
[41] Monsalve, M., Wu, Z., Adelmant, G., Puigserver, P., Fan, M. and Spiegelman, B.M. (2000) Mol. Cell 6, 307–316.
[42] Dantonel, J.C., Murthy, K.G., Manley, J.L. and Tora, L. (1997) Nature 389, 399–402.
[43] Hirose, Y. and Manley, J.L. (1998) Nature 395, 93–96.
[44] Barilla, D., Lee, B.A. and Proudfoot, N.J. (2001) Proc. Natl. Acad. Sci. USA 98, 445–450.
[45] Proudfoot, N.J. (1989) Trends Biochem. Sci. 14, 105–110.
[46] Niwa, M., Rose, S.D. and Berget, S.M. (1990) Genes Dev. 4, 1552–1559.
[47] Osheim, Y.N., Proudfoot, N.J. and Beyer, A.L. (1999) Mol. Cell 3, 379–387.
[48] Dye, M.J. and Proudfoot, N.J. (1999) Mol. Cell 3, 371–378.
[49] Yonaha, M. and Proudfoot, N.J. (2000) EMBO J. 19, 3770–3777.