

Integrating mRNA Processing with Transcription

Review

Nick J. Proudfoot,¹ Andre Furger,
and Michael J. Dye
Sir William Dunn School of Pathology
University of Oxford
South Parks Road
Oxford OX1 3RE
United Kingdom

The messenger RNA processing reactions of capping, splicing, and polyadenylation occur cotranscriptionally. They not only influence one another's efficiency and specificity, but are also coordinated by transcription. The phosphorylated CTD of RNA polymerase II provides key molecular contacts with these mRNA processing reactions throughout transcriptional elongation and termination.

Introduction

For eukaryotes, simply copying the genetic information from a DNA template into an RNA transcript does not result in the completion of messenger RNA synthesis. Before a gene transcript is ready to be transported out of the nucleus, it has to undergo three major processing events to produce the fully translatable mRNA. These comprise the acquisition of a cap structure at the 5' terminus, the splicing out of introns within the body of the pre-mRNA, and the generation of a 3' end, usually modified by the addition of a poly(A) tail. Although each of these reactions are biochemically distinct processes, they are interlinked and so influence one another's specificity and efficiency. Furthermore, 3' end formation and splicing have traditionally been viewed as posttranscriptional processing events. However, in more recent years it has emerged that these reactions are not only capable of influencing each other but are also tightly linked to transcription. It is therefore now believed that most mRNA processing reactions occur cotranscriptionally. This review describes recent progress in the field from genes in both lower and higher eukaryotes. Although the universal nature of gene expression means that most processes will occur throughout eukaryotes, it cannot be assumed that all details will be the same from yeast to man.

mRNA Processing Reactions

Capping

The first pre-mRNA processing step occurs after about 20–30 nucleotides have been synthesized. A three-step reaction adds a cap structure to the 5' end of all mRNAs (Figure 1). First, an RNA 5' triphosphatase (RTP) hydrolyzes the triphosphate of the first nucleotide to a diphosphate. Then, a guanylyltransferase (GT) catalyzes the fusion of a GMP moiety from GTP to the first nucleotide of the pre-mRNA via an unusual 5'-5' triphosphate link-

age. To finalize the cap structure, a methyltransferase (MT) methylates the N7 position of the transferred GMP. In mammals, both the triphosphatase and the guanylyltransferase activities are part of the same polypeptide at the N terminus and C terminus, respectively, while in yeast they are catalyzed by separate enzymes. This initial cap structure is then recognized by the cap binding complex (CBC), which contains two proteins, CBP20 and CBP80. Upon export through the nuclear pore complex, the nuclear cap binding proteins are replaced by the cytoplasmic translation initiation factor, eIF-4E (Shatkin and Manley, 2000, for review).

The cap structure bound to CBC is believed to play a major role in the stabilization of the mRNA, since it represents an obstacle for 5'-3' exonucleases (Beelman and Parker, 1995). In addition, in the cytoplasm, cap bound to eIF-4E and other translation initiation factors enhances translation by promoting the engagement of the ribosomal subunits with the mRNA. This is at least partly achieved by the interaction of eIF4G with poly(A)-associated PABPI, resulting in a translationally competent circular mRNA:protein complex (Sachs et al., 1997, for review).

Splicing

With a few exceptions, mammalian genes are interrupted by noncoding sequences (introns), and in order to generate a functional message from the DNA template, mRNA splicing must occur. The transcribed pre-mRNA itself contains several consensus elements in *cis*, which are essential for the splicing reaction. The 5' exon-intron junction or splice site (indicated by |) is marked by the consensus sequence AG|GURAGU (R, purine; Y, pyrimidine). The end of the intron, the 3' splice site, is defined by YAG|RNNN. About 100 nt upstream of the 3' splice site lies the branchpoint CURA^{20H}Y with a highly conserved adenosine followed by a pyrimidine-rich track.

In chemical terms splicing occurs in two steps. As shown in Figure 2, first the 2'OH of the branchpoint adenosine acts as a nucleophile to attack the 5' exon-intron border, and *trans*-esterification results in a free 5' exon and a lariat-shaped molecule consisting of the intron sequences and the 3' exon. In the second step, the 3'OH of the freed 5' exon attacks the intron-3' exon border. The subsequent *trans*-esterification results in the fusion of the two exon sequences and the release of the lariat-shaped intron (Moore and Sharp, 1993). The five small nuclear RNAs (snRNAs) U1, U2, U4, U5, and U6 exist with associated proteins in snRNPs. These snRNPs, together with many other proteins, assemble over the intron to form the spliceosome and thereby facilitate splicing (Kramer, 1996; Burge et al., 1999, for reviews).

Interaction of the U1 snRNA with the 5' splice site initiates spliceosome assembly. In higher eukaryotes, the 3' splice site and adjacent pyrimidine tract are identified through interactions with a dimeric splicing factor U2AF. This, in turn, helps recruit U2 snRNP to the branchpoint together with a branchpoint binding protein (BBP or SF1). In yeast, BBP and U2 snRNP can independently find the branchpoint (which has a more extensive

¹ Correspondence: nicholas.proudfoot@path.ox.ac.uk

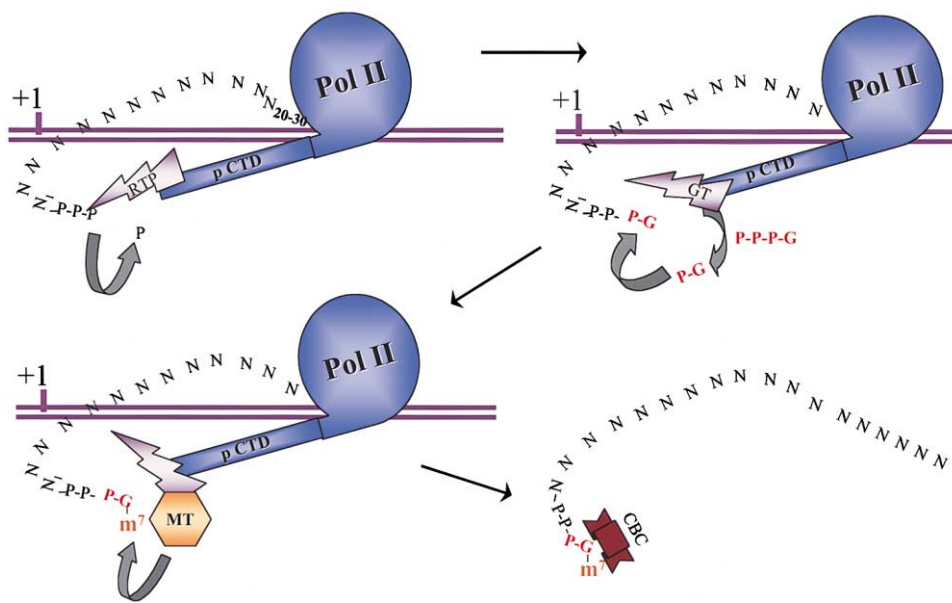


Figure 1. Capping of Pre-mRNA Occurs Early in Transcription

cis recognition sequence—UACUAA^{20H}C). U2 snRNA base pairs with the branchpoint sequences and, in so doing, bulges out the branchpoint adenosine, enhancing the first nucleophilic attack. In contrast, U4 does not directly interact with the pre-mRNA but plays an essential role in bringing U5 and U6 into the spliceosome. U6 interacts with sequences at the 5' splice site consensus after the displacement of U1. Both U2 and U6 not only interact with sites on the pre-mRNA but importantly build three essential helices by intramolecular base pairing with each other. These complex RNA-RNA interactions between U2, U6, and the pre-mRNA seem to form a catalytic core (Madhani and Guthrie, 1992; Sun and Manley, 1995) and by so doing place the 5' splice site and branchpoint in close proximity. Crosslinking experi-

ments demonstrated that nucleotides in the conserved U5 snRNA loop can interact both with 5' and 3' exon sequences (Newman, 1997, for review). The role of U5 is likely to be in positioning the exons for the second nucleophilic attack. As mentioned above, although a large number of proteins are known to be part of the spliceosome, there is now growing evidence that the catalytic component of the spliceosome may indeed be RNA based (Valadkhan and Manley, 2000). First, in vitro-transcribed U2 and U6 RNA fragments in the absence of any protein components are able to form intermolecular base pairing, similar to those obtained in the spliceosome. Second, the addition of a small branch-site like RNA oligo to this U2/U6 RNA complex was able to trigger, in the complete absence of proteins, a nucleophilic

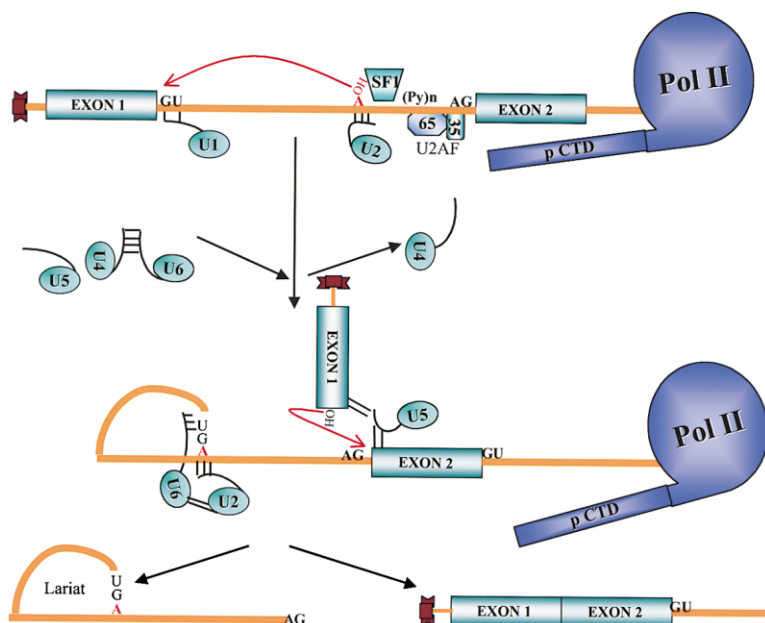


Figure 2. Simplified View of Splicing

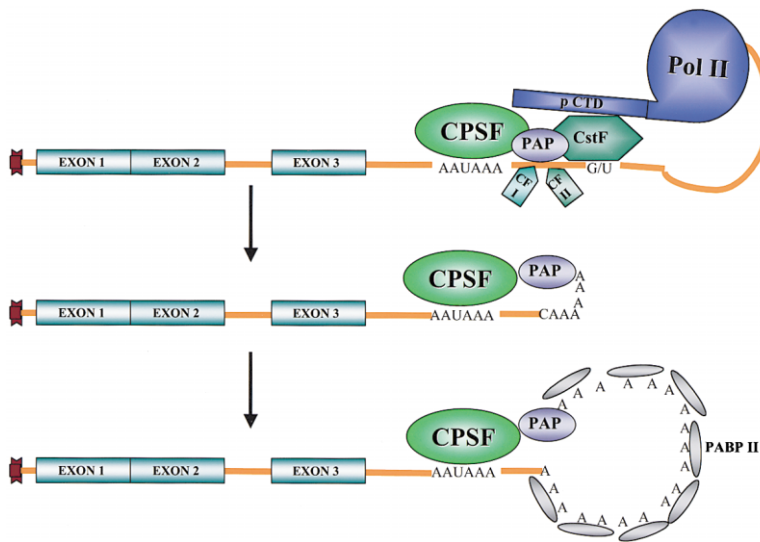


Figure 3. Polyadenylation Involves Separate RNA Cleavage and Poly(A) Synthesis

attack of the branchpoint adenosine on a distinct nucleotide within U6. The final product is a lariat-like structure between U6 and the branchpoint oligo (Valadkhan and Manley, 2001). Nevertheless, in vivo, protein components of the spliceosome play an essential part in the assembly and stabilization of a catalytically active spliceosome. There is also evidence that proteins directly interact with the catalytic core and may therefore be components of the catalytic activity in vivo (Collins and Guthrie, 2000, for review).

3' End Formation

With the exception of replication-dependent histone genes (in higher eukaryotes), all protein encoding mRNAs contain a uniform 3' end consisting of around 200 adenosine residues. The formation of this poly(A) tail is directed by sequences present on the pre-mRNA and the mammalian polyadenylation machinery, consisting of at least six multimeric protein factors (Figure 3). Prior to the addition of poly(A), the pre-mRNA must be cleaved. The site of cleavage in most pre-mRNAs lies between the highly conserved AAUAAA hexamer and a downstream sequence element (DSE), which is a U- or GU-rich motif. Cleavage itself occurs predominantly at a CA dinucleotide. In retroviruses it is also a common feature that U-rich upstream sequence elements (USE) located 5' of the AAUAAA sequence are required for full efficiency of the poly(A) signal. USEs have also been identified in two mammalian genes, C2 Complement and Lamin. These *cis* elements are recognized by two multisubunit protein complexes. The 160 kDa subunit of the cleavage and polyadenylation specificity factor (CPSF) has been shown to interact with the AAUAAA, but it is most likely that the other three subunits (CPSF-100, CPSF-73, and CPSF-30) contribute to the specificity and strength of binding. The DSE represents a platform for the interaction with the cleavage stimulatory factor (CstF) via its 64 kDa subunit. The binding of CstF and CPSF appears to be cooperative in that CstF binding to the DSE greatly enhances the affinity of CPSF to the hexamer and vice versa (Colgan and Manley, 1997; Zhao et al., 1999, for reviews). Two additional factors are essential to direct cleavage of the pre-mRNA: cleavage factor I (CF I) and cleavage factor II (CF II). CF

I consists of three subunits that are able to directly interact with a pre-mRNA substrate. CF II, only recently partially purified (de Vries et al., 2000), can be subdivided into two fractions of which only fraction CF IIAm appears to be essential for the cleavage reaction. One of the subunits of CF IIAm, hClp1, has been shown to interact with both CPSF and CF I. CF IIBm proved to be of a stimulatory nature only. Poly(A) polymerase (PAP) itself is usually required for the cleavage reaction and together with CPSF directs poly(A) addition. Poly(A) binding protein PABP II binds the emerging poly(A) tail and in turn enhances the processivity of the poly(A) polymerase (Wahle and Ruegsegger, 1999, for review).

The only known protein-encoding genes that lack a poly(A) tail are the replication-dependent histone genes. The formation of histone 3' ends differs significantly from polyadenylation, as an essential U7 snRNP complex is required. Furthermore, the mature 3' end is obtained by a single cleavage event. Two elements within the histone pre-mRNA play a crucial role in this maturation process. Sequences in the untranslated region of the histone RNA are able to form a stem-loop structure that is the site of interaction for the stem loop binding protein (SLBP). A purine-rich motif is located further downstream and provides the histone pre-mRNA with an anchor for interaction with the 5' end of U7 snRNA. The SLBP is not absolutely required for the processing reaction but proved to be indispensable in histone pre-mRNAs that have reduced complementarity to U7. RNA cleavage occurs between the two sequence elements in the histone pre-mRNA and similar to polyadenylation, the preferred nucleotide at which this occurs is an adenosine. A further factor called heat labile factor HLF (due to its sensitivity to heat treatment) is essential for the processing reaction. As with polyadenylation, the nuclease component remains to be identified (Dominsky and Marzluff, 1999, for review).

Crosstalk between mRNA Processing Events

Capping Interactions with Splicing and Polyadenylation

Although the three mRNA processing reactions were discovered independently, it emerged very early on that

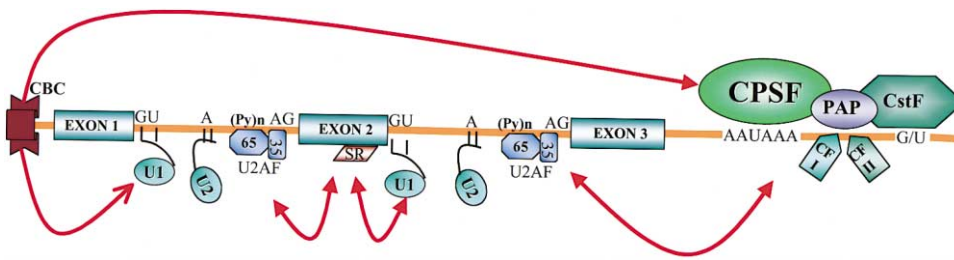


Figure 4. Molecular Interactions between mRNA Processing Reactions

these separate reactions closely influence one another (Figure 4). The development of *in vitro* systems capable of directing mRNA processing reactions using whole cell and nuclear extract also indicated that one mRNA processing reaction can enhance another. In particular, using *in vitro* systems it became clear that the cap structure played an important role in splicing. Incubation of *in vitro*-synthesized pre-mRNA in nuclear extract revealed that the addition of a cap structure significantly enhanced the formation of spliced mRNA. By incubating pre-mRNAs containing more than one intron, either in nuclear extracts or injected into *Xenopus* oocytes, the cap structure enhanced the excision of the first, cap-proximal intron, but had a lesser effect on the splicing efficiency of the second intron. The characterization of the cap binding complex (CBC) subsequently allowed a more detailed analysis of cap-splicing interconnections. Antibodies directed against the 80 kDa subunit of the CBC were used to deplete nuclear extracts, resulting in significant inhibition of splicing. Furthermore, CBC appeared to enhance the interaction of U1 snRNP with the 5' splice site (Lewis and Izaurralde, 1997, for review). Additional research suggested that the CBC positively influences the U6 snRNA 5' splice site interaction, possibly by affecting the displacement of U1 snRNP by U6 (O'Mullane and Eperon, 1998). These findings clearly underline the strong dependence of efficient splicing on the cap structure. Effects of CBC on splicing complex formation has also been demonstrated in yeast, where conditional inactivation of the yeast guanylyltransferase reduced levels of spliced mRNA *in vivo*, demonstrating that these interconnections are conserved in evolution (Shatkin and Manley, 2000). However, especially in mammalian systems, the question of how the cap and splicing components interact remains unsolved. By use of a yeast hybrid screen, a potential candidate that links of CBC with the splicing machinery has been identified as hnRNP F (Gamberi et al., 1997). hnRNP F can specifically interact with both CBC subunits and, moreover, hnRNP F-depleted extracts show a reduced splicing efficiency.

Around the same time that the cap-splicing interconnection was discovered, experiments were performed that showed a similar efficiency dependence of polyadenylation on the cap structure. As in splicing, the availability of CBC80 antibodies allowed the analysis of *in vitro* polyadenylation in CBC-depleted extracts. The presence of CBC enhanced the cleavage reaction significantly, but showed only a modest effect on the poly(A) addition reaction. Similar to splicing, CBC appeared to

have a stabilizing effect on the polyadenylation complex with its target pre-mRNA (Flaherty et al., 1997; Lewis and Izaurralde, 1997, for review).

Connections between Splicing and Polyadenylation

The fact that polyadenylation is strongly influenced by RNA splicing first became apparent through experiments that revealed the process of exon definition in RNA splicing (Niwa et al., 1992). Thus, for exons in higher eukaryotes, which are normally much shorter than their surrounding introns, the 3' and 5' splice sites flanking an exon are often recognized synergistically through interactions stretching across the exon (Figure 4). Splicing regulatory (SR) proteins are the key players in this process, often recruited to the exon through interactions between their RNA binding domains and specific exon enhancer sequences (Smith and Valcarcel, 2000, for review). They also interact through serine arginine-rich (RS) domains with other RS domain containing proteins in U1 and U2 snRNPs (Fu 1995, for review). Thus, the 70 kDa subunit of U1 snRNP and the 35 kDa subunit of U2AF both interact with these bridging SR proteins. For U2AF, the larger 65 kDa subunit also interacts with the pyrimidine tract and the 35 kDa subunit, once recruited to the 3' splice site through this molecular interaction, directly contacts the 3' splice site AG sequence (Zorio and Blumenthal, 1999; Wu et al., 1999). This exon definition mechanism also applies to the terminal exon of a gene, although in this case the end of the exon will be defined by the poly(A) signal. Both splicing and polyadenylation on either side of the terminal exon are strongly enhanced by each other (Niwa et al., 1992). The molecular connector in this case appears to be an interaction between U2AF65 and the C-terminal domain of poly(A) polymerase (Vagner et al., 2000; Figure 4). Interestingly, terminal exon definition also appears to be employed in a negative way, as a number of studies demonstrate that poly(A) site recognition is inhibited by the proximity of a 5' splice site (Proudfoot, 1996, for review). This inhibition will not normally occur in the terminal exon of a gene but may arise when a poly(A) site is positioned within the body of a gene. In this case, the interaction may be between the same poly(A) polymerase CTD and the U1 snRNP 70 kDa subunit (Gunderson et al., 1998). Also in some cases, a domain in the U1A protein of U1 snRNP, which is homologous to the PAP-interacting domain of the 70 kDa subunit, can also be demonstrated to inhibit polyadenylation at a nearby site (Gunderson et al., 1997). Examples of this type of regulation are well documented in retroviruses, where functional poly(A) signals may exist in the transcribed portion of the 5'

LTR sequence, as well as in DNA papillomaviruses, which again maximize their gene expression repertoire by employing internal poly(A) signals (Proudfoot, 1996, for review).

Another key example of splicing and polyadenylation crosstalk comes from genes where alternative polyadenylation leads to the expression of different gene products. Two systems are well characterized: IgM heavy chain genes and the Calcitonin gene (Zhao et al., 1999, for review). For the IgM gene, a relatively weak internal poly(A) site positioned within an intron is subjected to repression by the upstream 5' splice site (possibly through 70 kDa subunit of U1snRNP) and so is removed by splicing of this poly(A) site containing intron. The resulting mRNA utilizes a downstream poly(A) site, producing the membrane form of this protein, which predominates in early developing B cells. In plasma cells the weak upstream poly(A) site works more efficiently, out-competing the splicing process and so a shorter, secreted form of the protein is synthesized. Higher amounts of CstF-64 are present in plasma cells, so allowing recognition of the weaker intronic poly(A) signal (Takagaki et al., 1996; Takagaki and Manley, 1998). A similar situation exists for the Calcitonin gene. Here an internal poly(A) site is activated in thyroid tissue by a downstream positive element that appears to involve recruitment of splicing factors, including an SR protein SRp20 (Lou et al., 1998). In neural tissue this enhancer fails to act, and so the poly(A) site is spliced out and the resulting product CGRP is produced instead of Calcitonin. There are probably many more examples of alternative poly(A) site use in eukaryotes that may also rely on splicing-poly(A) factor interactions either negative or positive (for general review on polyadenylation and its regulation, see Zhao et al., 1999).

Transcription and mRNA Processing

As described in the first part of this review, it is clear that almost all steps in the production of the mature, translatable messenger RNA are interconnected. Thus, the presence of a cap modification at the 5' end of mRNA augments both splicing and polyadenylation. Also, several examples of the close integration of splicing and polyadenylation are described. In many cases it is likely that important examples of gene regulation are achieved through the interplay of these three mRNA processing mechanisms. These interconnections are now explained by the large body of evidence that demonstrates the direct association of the RNA polymerase II (Pol II) complex with mRNA processing activities at all three stages of gene transcription; initiation, elongation, and termination. Recent progress in our understanding of Pol II termination warrants a more extended account of this process (see below).

Initiation

A number of studies have revealed that there may be a direct connection between the promoter of a gene and how its transcript is processed. First, early experiments using chimeric gene constructs in which the transcription of a Pol II gene was driven by either a Pol I or Pol III promoter indicated that capping, splicing, and polyadenylation were all prevented, or at least partially repressed (Smale and Tjian, 1985; Lewis and Manley,

1986; Sisodia et al., 1987). Furthermore, exchanging the promoters of different classes of Pol II-transcribed genes by replacing an snRNA gene with one from a protein-encoding gene has been reported to perturb normal RNA processing (Hernandez, 1992, for review). Thus, for snRNAs transcribed from a globin promoter, normal snRNA 3' end formation is prevented, and instead, cryptic poly(A) signals are utilized. All of the above experiments argue that specific types of promoter recruit specific processing activities that determine exactly how the nascent transcript is processed into the mature RNA. A clear explanation for the promoter specificity of mRNA processing comes from the now-substantial weight of evidence that the Pol II transcription complex possesses a unique structural feature that acts to specifically recruit mRNA processing activities to the nascent transcript as it is generated by the transcription elongation process. The largest Pol II subunit contains a carboxyl-terminal domain (CTD) that appears to be quite separate from the main body of the enzyme, being attached to a relatively unstructured linker sequence (Cramer et al., 2001). This CTD is a heptad amino acid sequence repeated 52 times in mammals and 26 times in yeast. Two serine residues (Ser2 and Ser5) per repeat are subjected to phosphorylation so that complete phosphorylation of CTD will cause a massive increase in the acidic nature of this domain (Dahmus, 1996). Unphosphorylated CTD (Pol IIA) is required for initiation of transcription and interacts with a range of general transcription factors. However, once initiation of transcription begins, kinases, including a subunit of TFIIF and the elongation factor P-TEFb, phosphorylate the CTD to generate the elongation competent Pol IIO form (Reines et al., 1996; Price, 2000, for reviews). In part, this is accomplished by the stripping off of initiation factors and their replacement with as yet ill-defined transcription elongation factors (Wittschieben et al., 1999). Also, as a simple explanation of the above described promoter dependence of mRNA processing, recruitment of mRNA processing factors to the phosphorylated CTD occurs (for reviews see Hirose and Manley, 2000; Proudfoot, 2000). The explanation for why Pol I and III promoters are unable to generate transcripts capable of mRNA processing is thus simply explained by the fact that these polymerases lack CTDs. Similarly, snRNA and histone promoters must specify the recruitment of specific RNA processing factors that operate on their homologous RNAs and not other mRNA sequences.

It is also now clear that Pol II promoters can predetermine the balance of alternative mRNA splicing. An alternatively spliced exon (EDA) from the fibronectin gene has been investigated as part of a chimeric globin-fibronectin gene construct. The ratio of mRNAs that include or exclude this alternative exon was previously shown to be strongly influenced by the promoter used to drive transcription. Thus, the α -globin promoter resulted in a preponderance of EDA inclusion, while the fibronectin promoter gave the reverse effect (Cramer et al., 1997). Although the relative strength of the promoter did not appear to correlate with a specific splicing pattern, recent data from this system does argue that the capability of the promoter to mediate more or less processive transcription elongation can also strongly correlate with specific processing patterns. Thus, less-processive

transcription elongation promotes more efficient inclusion of EDA, while more-processive transcription has the reverse effect (Kadener et al., 2001). It is likely that a part of these processivity effects are caused by kinetic effects. The longer time taken to transcribe through an alternatively spliced region of a gene (when the Pol II complex is less processive) may allow more time for splicing factors to assemble on an exon's splice sites and so allow its inclusion in the mature mRNA sequence. Indeed, such kinetic effects have also been demonstrated for the tropomyosin gene, by placing transcription pause sites (see below) between the alternatively spliced exon 3 and a downstream intronic negative regulatory element (NRE). This slows down transcription and so allows more time to prevent the downstream NRE from inhibiting upstream splice site selection (Roberts et al., 1998). Such experiments may be of questionable physiological relevance, as they undeniably involve the analysis of unnatural gene constructions. However, they do clearly implicate the role of the promoter in determining the subsequent fate of the primary RNA transcript as it emanates from the RNA binding domain of the elongating Pol II complex.

Elongation

The critical molecular trigger that allows escape of the Pol II complex from initiation is likely to be the massive phosphorylation of its CTD, as described above. However, as with most such molecular processes, further specificity is apparent. Thus, phosphorylated serine 5 of the heptad repeat is found mainly at early stages of elongation, while serine 2 phosphorylation occurs later on during the elongation process (Komarnitsky et al., 2000). It is now apparent that a major component of the elongation factors are mRNA processing factors. All three capping enzymes are known to associate with phospho-CTD (Hirose and Manley, 2000, for review). Furthermore, in the case of the guanylyltransferase, phospho-CTD significantly enhances its enzymatic activity, principally through serine 5 phosphorylation (Ho and Shuman, 1999). Recent data has been obtained using the chromatin immunoprecipitation technique and reveals that capping enzymes associate at the beginning of the transcription unit but are lost as the elongating polymerase progresses into the body of the gene. This is in contrast to the methyltransferase, which continues to be associated with the transcription complex during elongation (Komarnitsky et al., 2000).

The role of the phospho-CTD in orchestrating mRNA splicing is currently less fully documented than for capping. This largely reflects the enormous complexity of the splicing mechanism. However, at three levels of analysis it is apparent that there is a close connection between splicing and transcription. First, electron microscopy of chromosome spreads demonstrates co-transcriptional splicing (Beyer and Osheim, 1988). Also, a number of nuclear imaging experiments relying on the colocalization of antibodies to Pol II and mRNA processing factors all point to their tight association within the nucleus (Misteli and Spector, 1998, for review). Second, it is clear that phospho-CTD enhances splicing, at least indirectly, as in several different *in vivo* experimental systems CTD truncation results in reduced splicing (Corden and Patturajan, 1997, for review). Also, *in vitro* splicing efficiency can be shown under some circum-

stances to be strongly enhanced by the addition of Pol IIO (Hirose et al., 1999). Interestingly, in another study, only splicing templates with complete exons bordered by introns were shown to be activated by phospho-CTD, suggesting that exon definition requires the presence of the elongating Pol II complex (Zeng and Berget, 2000). Finally, a direct biochemical link has been made, but only in a few specific cases (Goldstrohm et al., 2001, for recent review). Thus, a series of protein factors with some homology to SR proteins interact with Pol II phospho-CTD (Corden and Patturajan, 1997). These factors possess a well-defined CTD binding domain (in common with the yeast poly(A) factor, pCF11p, and SnRNA termination factor, Nrd1p—see below). Secondly, a component of U1 snRNP in yeast called Prp40p is known to directly interact with phospho-CTD (Morris and Greenleaf, 2000). These data have led to the attractive view that phospho-CTD may play a direct role in tethering splicing signals on the pre-mRNA to the elongating Pol II complex. When the next splice signal appears it will directly abut factors bound to the upstream splice signal and so greatly enhance splicing.

Poly(A) factors are also now well established to interact with Pol II and at least in some cases may ride with the elongation complex all the way from initiation to termination of transcription. First, CPSF has been shown to be a component of active TFIID and then transfers onto Pol IIO concomitant with elongation into the gene (Dantonel et al., 1997). Also, truncation of the CTD results in both splicing and termination-defective transcription (McCracken et al., 1997). In yeast, the cleavage/poly(A) factor pCF11 directly contacts phospho-CTD (Barilla et al., 2001), as do other factors, at least indirectly, such as Pta1p (Rodriguez et al., 2000) and Hrp1p (Komarnitsky et al., 2000). Indeed, since all cleavage/poly(A) factors exist in tight complexes, it is likely that they may all be detectable in association with Pol IIO. As well as binding to CTD, it is also clear that cleavage/poly(A) factors are more active in the presence of CTD (Hirose and Manley, 1998). Indeed, a very recent *in vitro* study supports the view that CTD is required for the cleavage activity, as in certain factor preparations careful quantitation revealed a small level contamination of Pol II, accounting for apparent CTD-independent activity (Ryan et al., 2002). Recent *in vivo* analysis has suggested that the 3' terminus of CTD, which in general has more variant heptad repeats, may be the main component of this domain and that it interacts with cleavage/polyadenylation activities, in particular CstF-50 (Fong and Bentley, 2001). However, *in vitro* analysis of CTD subdomains suggests that there is little difference between different parts of the CTD. By construction of an artificial CTD with only the consensus heptad repeat, it is clear that the critical feature for activating cleavage/polyadenylation is merely the length of the domain, which needs to be over 26 heptad repeats to function efficiently (Ryan et al., 2002).

POL II Transcriptional Termination

This final stage in the transcription process is critical for successful gene expression. First, it allows release of transcripts from the site of transcription and release of Pol II from the DNA template. This facilitates the recy-

cling of Pol II back to the promoter for further rounds of transcription. Second, it will ensure that promoters are not perturbed by read-through polymerases that have failed to terminate at upstream genes. Such transcriptional interference commonly occurs in yeast where genes are closely spaced and simultaneously expressed (Greger and Proudfoot, 1998). Compared to the large body of detailed information on transcriptional termination in bacteria (reviewed in Henkin, 2000), as well as eukaryotic Pol I and Pol III genes (reviewed in Paule and White, 2000), Pol II termination is less well understood. This is because the mature 3' ends of Pol II transcripts are formed by cotranscriptional RNA processing at positions upstream of the termination site. Transcript cleavage at the poly(A) site leaves two products which have quite different fates; the upstream product, destined to become mRNA, is stabilized by poly(A) addition, whereas the downstream product is destabilized by cleavage, separating it from the body of the transcript. These short-lived transcripts cannot be visualized by standard RNA mapping techniques.

Three different classes of Pol II genes each employ significantly different termination processes: those that encode snRNA and snoRNAs (small nucleolar RNA), the specialized histone genes, and finally, the most abundant and diverse class that encode all other proteins. This review will focus primarily on transcriptional termination of the protein-encoding genes. A comparative account of transcriptional termination in histone as well as snRNA and snoRNA genes is also presented.

Role of the Poly(A) Signal in Pol II Termination

The demonstration that transcription termination is dependent on the presence of a functional poly(A) signal (Proudfoot, 1989, for review) provided the first indication that mRNA processing and Pol II transcription are coupled events. These data were obtained by use of nascent RNA mapping techniques, so-called run-on analysis. Subsequently, run-on analysis of the *CYC1* gene in *S. cerevisiae*, using strains defective in cleavage but not in polyadenylation, showed disruption of transcriptional termination (Birse et al., 1998). Biochemical sense of these results was made when 3' end processing was shown to be dependent on the presence of Pol II CTD both in vivo (McCracken et al., 1997) and in vitro (Hirose and Manley, 1998), as described above. More recently, EM visualization of nascent transcription on the *TFIIIA* gene injected as part of a plasmid into *Xenopus* oocytes showed that poly(A) signal strength is directly correlated to termination efficiency (Osheim et al., 1999). Consistent with these data, it has also been shown that poly(A) signal strength itself is a direct measure of the speed with which the cleavage/polyadenylation apparatus assembles on the nascent transcript (Chao et al. 1999). Run-on analysis of transcription from the human β -globin gene (transiently transfected into HeLa cells) has further shown that the dependence of Pol II termination on RNA processing extends to splicing of the terminal intron (Dye and Proudfoot, 1999). This was also indicated from an RT-PCR-based study of nascent transcripts from the endogenous Balbiani Ring 1 gene in salivary glands of *C. tentans* (Bauren et al., 1998). These observations can be readily explained by the interaction of the terminal 3' splice and poly(A) sites to define the terminal exon as described above (Niwa et al., 1992).

Considering the absolute requirement of a poly(A) signal for transcription termination, it is most likely that the dependence of termination on splicing is an indirect effect.

When it was first established that 3' end processing signals are required for transcription termination, two basic models of how these two processes are coupled were put forward. The first "antiterminator" model invokes some conformational change in the transcription complex upon recognition of the poly(A) signal that makes it termination competent (Logan et al. 1987). The second "torpedo" model suggested that cleavage of the transcript at the poly(A) site initiates termination by promoting rapid degradation of the 3' product, still attached to the elongating polymerase (Connelly and Manley, 1988; Proudfoot, 1989). The results of studies on Pol II transcription termination carried out over the past decade have indicated that these two models are not mutually exclusive. Rather, a new version, combining the antiterminator and torpedo models, is likely to prevail (see below).

Kinetics of Nascent Transcript Cleavage at Poly(A) Site

Four recent studies have independently addressed the issue of when in the transcription process cleavage at the poly(A) site actually occurs. Using an RT-PCR approach it was found that uncleaved nascent transcripts of the Balbiani ring 1 gene extend up to 700 bp downstream of the poly(A) site (Bauren et al., 1998). This was also shown by a more direct method in which radio-labeled nascent transcripts of the human ϵ - and β -globin genes, from transiently transfected HeLa cells, were analyzed using a hybrid selection technique (Dye and Proudfoot, 1999). In these latter experiments, uncleaved nascent transcripts were detected that extend up to 1.5 Kb downstream of the poly(A) site. Indeed, for both globin genes, uncleaved transcripts appear to extend right up to the region where termination occurs. In the above-mentioned EM study on nascent transcripts from the *TFIIIA* gene, transcription termination appears to occur before cleavage at the poly(A) site, even though the requirement of a functional poly(A) site for termination was clearly shown (Osheim et al., 1999). Confirmatory evidence for the above results was also provided by in vitro experiments utilizing an antisense approach. Here, 3' end processing was shown to occur when Pol II is located at positions some distance downstream of the poly(A) site (Tran et al., 2001). It is clear from all the above data that 3' end processing occurs some distance downstream of the point at which the poly(A) site is transcribed. This conclusion, when combined with the fact that cleavage/poly(A) factors are bound to CTD, suggests a model for the post-poly(A) site polymerase in which nascent RNA is looped out between the CTD-associated cleavage/poly(A) factors and the Pol II RNA exit channel (Cramer et al., 2001; see Figure 5). It appears that Pol II proceeds up to the site of termination before transcript cleavage takes place. This indicates that 3' end processing is the trigger for Pol II termination.

Since it is clear that RNA processing signals affect transcription, it follows that the reverse effect of transcriptional events influencing RNA processing occurs. Thus, transcriptional pause sites located downstream of the poly(A) site are thought to cause a transient pause

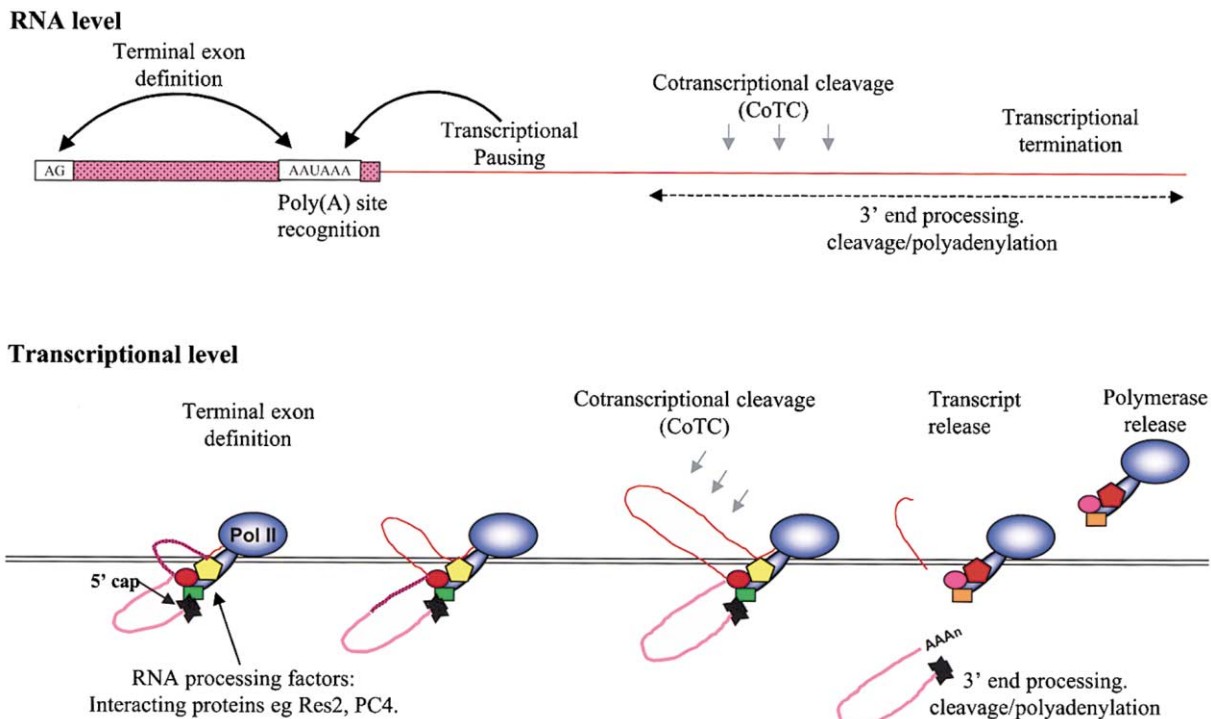


Figure 5. Transcriptional Termination in Pol II Genes Encoding mRNA

to Pol II progression and so enhance poly(A) site recognition. Studies on a number of genes have identified such pause sites. The strongest evidence for an involvement of transcriptional pausing in 3' end processing and termination comes from studies in yeast. In *S. pombe*, the poly(A) signals of the *ura4* and *nmt2* genes both possess downstream sequence elements (DSE) that are required for efficient 3' end processing and induce transcription termination. Increased Pol II density was measured over these DSEs by run-on analysis, indicating that they function as pause sites. However, they share no sequence homology, so that their specificity has not been determined (Birse et al., 1997; Aranda and Proudfoot, 1999). Studies in mammalian cells also indicated the presence of transcription pause sites located between the closely spaced α -globin genes and the C2 and factor B genes. Analysis of these pause sites in a coupled in vitro transcription-polyadenylation system showed that in some situations they can increase the efficiency of cleavage and polyadenylation and thereby promote termination (Yonaha and Proudfoot, 2000).

Further Pol II Termination Factors and Sequences

From the above account it is clear that transcript cleavage at the poly(A) site represents a key determinant in transcriptional termination. However, a number of detailed nascent transcription studies on several different genes show that mRNA 3' end processing is not the only requirement for transcription termination. Evidence has been obtained for the requirement of both specific termination factors and sequence elements located downstream of the poly(A) site.

Two recent studies in yeast have identified transcription factors as partners with the RNA processing factors

in the termination process. First, it has been demonstrated that the C-terminal domain of the *S. cerevisiae* and *S. pombe* homologs of human CstF-64 protein (Rna15p and Ctf1, respectively) is required for efficient transcriptional termination. This effect does not appear to be mediated through the 3' end processing machinery, since the remaining N-terminal portion of CstF-64 is sufficient for it to accomplish 3' end processing. Using yeast two-hybrid analysis, it was also shown that the C-terminal domain of Ctf1 interacts with Res2, a component of the transcription factor MBF. This transcription factor controls transcription initiation of genes switched on during the G1-S phase of the cell cycle. Confirmatory evidence for the role of Res2 in transcriptional termination has been provided by transcriptional run-on analysis of a Δ res2 strain. In the absence of Res2, transcription proceeds beyond the natural termination site (Aranda and Proudfoot, 2001). By a similar approach, it has been shown that the human transcriptional coactivator PC4 interacts with the CTD of CstF-64, as do their yeast homologs Rna15p and Sub1p, and so influence transcription termination (Calvo and Manley, 2001). The reported effects that these factors have on transcription termination suggests a further link between transcriptional activators acting at transcription initiation and the 3' end processing and the Pol II termination machinery (see above).

Detailed studies of nascent transcription in the 3' flanking regions of a number of genes have indicated that sequences located downstream of the poly(A) site are important players in the termination process. Early studies (Proudfoot, 1989, for review) identified a CCAAT sequence in the adenovirus major late promoter that

binds the protein CP1 and effectively terminates transcription from upstream genes. A detailed transcriptional analysis of the mouse β -major globin gene 3' flanking region also identified a number of sequence elements, including a 69 bp AT-rich sequence that effects transcription termination in the presence of a functional poly(A) site. Interestingly, the activity of this sequence is dependent on its position relative to the poly(A) site (Tantravahi et al., 1993).

To add to these data, a very recent study of transcription in the human β -globin gene 3' flanking region has identified sequence tracts located between 900 and 1600 bp downstream of the poly(A) site that are required for transcriptional termination. Interestingly, hybrid selection analysis of transcripts in these terminator tracts, using an antisense biotinylated RNA probe, spanning the poly(A) site region, shows that they are cotranscriptionally cleaved. This cleavage activity was originally referred to as pretermination cleavage (PTC) but is now renamed cotranscriptional cleavage (CoTC) to avoid confusion with the term pretermination codon. Although the activity that mediates CoTC has so far not been identified, it is clear that both 3' end processing and CoTC are required for transcription termination but occur independently of each other. Thus, the efficiency of mRNA 3' end processing appears to be unaffected by deletion of the termination region, and conversely, CoTC in the termination region transcript occurs in the absence of upstream mRNA 3' end processing (Dye and Proudfoot, 2001). In this same study, the dependence of ϵ -globin termination on long tracts of 3' flanking region was also described. However, it appears that termination elements in this region are more diffuse than in the β -globin gene. The sequence elements described in the human β -globin gene differ from those described in the mouse β -major globin gene in two ways: they are longer and they can also function at various positions relative to the poly(A) site. One characteristic that both the mouse and human termination elements share is AT richness. Hybrid selection nuclear run-on analysis of transcripts spanning the termination sequences in the β -globin 3' flanking region enables a dissection of the termination process. Current data favors a model in which Pol II termination occurs in two stages. In step 1, the termination region transcript is cotranscriptionally cleaved. Since Pol II transcription termination is dependent on cotranscriptional 3' end processing at the poly(A) signal and in vitro experiments indicate a direct involvement of Pol II CTD in this process (Hirose and Manley, 1998), it is predicted that an interaction of the transcript with the CTD may continue after CoTC has taken place. This interaction and subsequent cleavage at the poly(A) site may then have a role in mediating step 2 of the termination process, namely polymerase release. Consistent with this model, CoTC operates independently of mRNA 3' end processing, while polymerase release is entirely dependent upon it (Figure 5).

In summary, Pol II transcription termination is a complex process that is primarily regulated by mRNA 3' end processing, the efficiency of which is modulated by signals located upstream (splicing of the final intron) and downstream (pause sites) of the poly(A) site. There is also expanding evidence that dedicated signals and factors operate downstream of the poly(A) site and are

also required for termination to occur (see Figure 5). In particular, the fact that the human β - and ϵ -globin gene termination region transcripts are targets for a cotranscriptional cleavage activity is likely to be an important part of the termination process. It has yet to be determined how this cleavage event is orchestrated, but it is possible that it is carried out by a ubiquitous RNA degradation pathway or as an intrinsic transcript cleavage activity of Pol II (Dye and Proudfoot, 2001). Which-ever the case, it appears that RNA processing is involved at all stages of the Pol II transcription termination process.

Transcriptional Termination in Histone, snRNA, and snoRNA Genes

Histone pre-mRNAs like snRNAs are not spliced, and the majority are not polyadenylated. The mature 3' end of histone transcripts is formed by an endonucleolytic cleavage of the primary transcript. As with other mRNA genes, Pol II transcription does not terminate until it is some distance beyond the histone 3' processing site. An early report suggested that transcriptional termination at a position located 600 bp downstream of the mouse H2a gene was dependent on 3' end processing at the upstream site. A more recent study using an in vitro system indicates that termination of a proportion of transcripts is dependent on the presence of functional histone processing signals rather than the cleavage reaction (Dominski and Marzluff, 1999, for review).

Although most snRNAs are transcribed by Pol II, they are neither spliced nor polyadenylated. 3' end formation of mature snRNAs requires a sequence element known as the 3' box, which is located 9–19 nt downstream of the end of the nascent transcript (reviewed in Hernandez, 1992). Run-on analysis of the mouse U1 snRNA gene has shown that transcriptional termination of the majority of transcripts occurs close to the 3' box (Kunkel and Pederson, 1985). Until recently, it was unknown whether the 3' box is an RNA processing or transcription termination signal. This issue was addressed in a study employing run-on analysis and in vivo foot printing of the U1 and U2 genes (Cuello et al., 1999). Whereas U1 transcription terminates abruptly after the 3' box, nascent transcripts of the U2 gene are detected up to 250 nt downstream. The extended nature of U2 snRNA nascent transcripts indicates that the 3' box is not a transcription termination sequence. In vivo foot printing of the U1 gene showed that a protein-DNA interaction occurs downstream of the 3' box. This sequence element, in conjunction with the U2 3' box, was shown to cause transcriptional termination in transiently transfected constructs. It appears from this study that the U1 and U2 snRNA genes have a similar bipartite termination process to that of mRNA genes, requiring RNA processing and termination elements.

A recent study of 3' end formation of snRNA and snoRNA transcripts in yeast has implicated the protein Nrd1p in this process (Steinmetz et al., 2001). Note that many snoRNAs, especially in mammals, are matured from the introns of other Pol II genes and so do not possess termination signals. Nrd1p has an amino terminus that interacts with the CTD of Pol II. It also contacts an RNA binding protein termed Nab3p, which in turn is thought to recruit the CTD kinase, Ctk1p. Each of these factors appears to be involved in the *NRD1*-dependent

3' end formation pathway. Transcriptional analysis of a *nrd1* mutant strain shows that in this background, transcription on the *SNR3* gene proceeds beyond the normal termination site. These results suggest that the mechanism of transcriptional termination for many yeast snoRNA and snRNA genes has strong similarities to that of protein-encoding genes (Steinmetz et al., 2001). Furthermore, it has been demonstrated for at least some snoRNAs that factors normally associated with mRNA cleavage/polyadenylation may also mediate 3' end cleavage of these transcripts (Fatica et al., 2000).

In the final analysis, it is clear that all of these different Pol II termination processes have overlapping mechanisms. No doubt, further study will lead to a unified termination mechanism for all Pol II transcribed genes.

Acknowledgments

Research in our laboratory is funded by the Wellcome Trust and Medical Research Council of Great Britain. We are grateful to members of the lab for advice and encouragement. We also are indebted to Jim Manley for critical reading of the manuscript.

References

- Aranda, A., and Proudfoot, N.J. (1999). Definition of transcriptional pause elements in fission yeast. *Mol. Cell. Biol.* 19, 1251–1261.
- Aranda, A., and Proudfoot, N. (2001). Transcriptional termination factors for RNA polymerase II in yeast. *Mol. Cell* 7, 1003–1011.
- Barilla, D., Lee, B.A., and Proudfoot, N.J. (2001). Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* 98, 445–450.
- Bauren, G., Belikov, S., and Wieslander, L. (1998). Transcriptional termination in the Balbiani ring 1 gene is closely coupled to 3'-end formation and excision of the 3'-terminal intron. *Genes Dev.* 12, 2759–2769.
- Beelman, C.A., and Parker, R. (1995). Degradation of mRNA in eukaryotes. *Cell* 81, 179–183.
- Beyer, A.L., and Osheim, Y.N. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev.* 2, 754–765.
- Birse, C.E., Lee, B.A., Hansen, K., and Proudfoot, N.J. (1997). Transcriptional termination signals for RNA polymerase II in fission yeast. *EMBO J.* 16, 3633–3643.
- Birse, C.E., Minvielle-Sebastia, L., Lee, B.A., Keller, W., and Proudfoot, N.J. (1998). Coupling termination of transcription to messenger RNA maturation in yeast. *Science* 280, 298–301.
- Burge, C.B., Tuschl, T., and Sharp, P.A. (1999). Splicing of precursors to mRNAs by the spliceosomes. In *The RNA World*, R.F. Gesteland, T.R. Cech, and J.F. Atkins, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 525–560.
- Calvo, O., and Manley, J.L. (2001). Evolutionarily conserved interaction between CstF-64 and PC4 links transcription, polyadenylation, and termination. *Mol. Cell* 7, 1013–1023.
- Chao, L.C., Jamil, A., Kim, S.J., Huang, L., and Martinson, H.G. (1999). Assembly of the cleavage and polyadenylation apparatus requires about 10 seconds in vivo and is faster for strong than for weak poly(A) sites. *Mol. Cell. Biol.* 19, 5588–5600.
- Colgan, D.F., and Manley, J.L. (1997). Mechanism and regulation of mRNA polyadenylation. *Genes Dev.* 11, 2755–2766.
- Collins, C.A., and Guthrie, C. (2000). The question remains: is the spliceosome a ribozyme? *Nat. Struct. Biol.* 7, 850–854.
- Connelly, S., and Manley, J.L. (1988). A functional mRNA polyadenylation signal is required for transcriptional termination by RNA polymerase II. *Genes Dev.* 4, 440–452.
- Corden, J.L., and Patturajan, M. (1997). A CTD function linking transcription to splicing. *Trends Biochem. Sci.* 22, 413–416.
- Cramer, P., Pesce, C.G., Baralle, F.E., and Kornblihtt, A.R. (1997). Functional association between promoter structure and transcript alternative splicing. *Proc. Natl. Acad. Sci. USA* 94, 11456–11460.
- Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* 292, 1863–1876.
- Cuello, P., Boyd, D.C., Dye, M.J., Proudfoot, N.J., and Murphy, S. (1999). Transcription of the human U2 snRNA genes continues beyond the 3' box in vivo. *EMBO J.* 18, 2867–2877.
- Dominsky, Z., and Marzluff, W.F. (1999). Formation of the 3' end of histone mRNA. *Gene* 239, 1–14.
- Dahmus, M.E. (1996). Reversible phosphorylation of the C-terminal domain of RNA polymerase II. *J. Biol. Chem.* 271, 19009–19012.
- Dantonel, J.C., Murthy, K.G., Manley, J.L., and Tora, L. (1997). Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA. *Nature* 389, 399–402.
- de Vries, H., Ruegsegger, U., Hubner, W., Friedlein, A., Langen, H., and Keller, W. (2000). Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors. *EMBO J.* 19, 5895–5904.
- Dominski, Z., and Marzluff, W.F. (1999). Formation of the 3' end of histone mRNA. *Gene* 239, 1–14.
- Dye, M.J., and Proudfoot, N.J. (1999). Terminal exon definition occurs cotranscriptionally and promotes termination of RNA polymerase II. *Mol. Cell* 3, 371–378.
- Dye, M.J., and Proudfoot, N.J. (2001). Multiple transcript cleavage precedes polymerase release in termination by RNA polymerase II. *Cell* 105, 669–681.
- Fatica, A., Morlando, M., and Bozzoni, I. (2000). Yeast snoRNA accumulation relies on a cleavage-dependent/polyadenylation-independent 3'-processing apparatus. *EMBO J.* 19, 6218–6229.
- Flaherty, S.M., Fortes, P., Izauralde, E., Mattaj, I.W., and Gilmartin, G.M. (1997). Participation of the nuclear cap binding complex in pre-mRNA 3' processing. *Proc. Natl. Acad. Sci. USA* 94, 11893–11898.
- Fong, N., and Bentley, D.L. (2001). Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. *Genes Dev.* 15, 1783–1795.
- Fu, X.D. (1995). The superfamily of arginine/serine-rich splicing factors. *RNA* 1, 663–680.
- Gamberi, C., Izauralde, E., Beisel, C., and Mattaj, I.W. (1997). Interaction between the human nuclear cap-binding protein complex and hnRNP F. *Mol. Cell. Biol.* 17, 2587–2597.
- Goldstrohm, A.C., Greenleaf, A.L., and Garcia-Blanco, M.A. (2001). Co-transcriptional splicing of pre-messenger RNAs: considerations for the mechanism of alternative splicing. *Gene* 277, 31–47.
- Greger, I.H., and Proudfoot, N.J. (1998). Poly(A) signals control both transcriptional termination and initiation between the tandem GAL10 and GAL7 genes of *Saccharomyces cerevisiae*. *EMBO J.* 17, 4771–4779.
- Gunderson, S.I., Vagner, S., Polycarpou-Schwarz, M., and Mattaj, I.W. (1997). Involvement of the carboxyl terminus of vertebrate poly(A) polymerase in U1A autoregulation and in the coupling of splicing and polyadenylation. *Genes Dev.* 11, 761–773.
- Gunderson, S.I., Polycarpou-Schwarz, M., and Mattaj, I.W. (1998). U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(A) polymerase. *Mol. Cell* 1, 255–264.
- Henkin, T.M. (2000). Transcription termination control in bacteria. *Curr. Opin. Microbiol.* 3, 149–153.
- Hernandez, N. (1992). Transcription of vertebrate snRNA genes and related genes. In *Transcriptional Regulation*, S. McKnight and K.R. Yamamoto, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 281–313.
- Hirose, Y., and Manley, J.L. (1998). RNA polymerase II is an essential mRNA polyadenylation factor. *Nature* 395, 93–96.
- Hirose, Y., and Manley, J.L. (2000). RNA polymerase II and the integration of nuclear events. *Genes Dev.* 14, 1415–1429.

- Hirose, Y., Tacke, R., and Manley, J.L. (1999). Phosphorylated RNA polymerase II stimulates pre-mRNA splicing. *Genes Dev.* 13, 1234–1239.
- Ho, C.K., and Shuman, S. (1999). Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. *Mol. Cell* 3, 405–411.
- Kadener, S., Cramer, P., Nogues, G., Cazalla, D., de la Mata, M., Fededa, J.P., Werbach, S.E., Srebrow, A., and Kornblihtt, A.R. (2001). Antagonistic effects of T-Ag and VP16 reveal a role for RNA pol II elongation on alternative splicing. *EMBO J.* 20, 5759–5768.
- Komarnitsky, P., Cho, E.J., and Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev.* 14, 2452–2460.
- Kramer, A. (1996). The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu. Rev. Biochem.* 65, 367–409.
- Kunkel, G.R., and Pederson, T. (1985). Transcription boundaries of U1 small nuclear RNA. *Mol. Cell. Biol.* 5, 2332–2340.
- Lewis, J.D., and Izaurralde, E. (1997). The role of the cap structure in RNA processing and nuclear export. *Eur. J. Biochem.* 247, 461–469.
- Lewis, E.D., and Manley, J.L. (1986). Polyadenylation of an mRNA precursor occurs independently of transcription by RNA polymerase II in vivo. *Proc. Natl. Acad. Sci. USA* 22, 8555–8559.
- Logan, J., Falck-Pedersen, E., Darnell, J.E., Jr., and Shenk, T. (1987). A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. *Proc. Natl. Acad. Sci. USA* 84, 8306–8310.
- Lou, H., Neugebauer, K.M., Gagel, R.F., and Berget, S.M. (1998). Regulation of alternative polyadenylation by U1 snRNPs and SRP20. *Mol. Cell. Biol.* 18, 4977–4985.
- Madhani, H.D., and Guthrie, C. (1992). A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell* 71, 803–817.
- McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., and Bentley, D.L. (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* 385, 357–361.
- Misteli, T., and Spector, D.L. (1998). The cellular organization of gene expression. *Curr. Opin. Cell Biol.* 10, 323–331.
- Moore, M.J., and Sharp, P.A. (1993). Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. *Nature* 365, 364–368.
- Morris, D.P., and Greenleaf, A.L. (2000). The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J. Biol. Chem.* 275, 39935–39943.
- Newman, A.J. (1997). The role of U5 snRNP in pre-mRNA splicing. *EMBO J.* 16, 5797–5800.
- Niwa, M., MacDonald, C.C., and Berget, S.M. (1992). Are vertebrate exons scanned during splice-site selection? *Nature* 360, 277–280.
- O'Mullane, L., and Eperon, I.C. (1998). The pre-mRNA 5' cap determines whether U6 small nuclear RNA succeeds U1 small nuclear ribonucleoprotein particle at 5' splice sites. *Mol. Cell. Biol.* 18, 7510–7520.
- Osheim, Y.N., Proudfoot, N.J., and Beyer, A.L. (1999). EM visualization of transcription by RNA polymerase II: downstream termination requires a poly(A) signal but not transcript cleavage. *Mol. Cell* 3, 379–387.
- Paule, M.R., and White, R.J. (2000). Survey and summary: transcription by RNA polymerases I and III. *Nucleic Acids Res.* 28, 1283–1298.
- Price, D.H. (2000). P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II. *Mol. Cell. Biol.* 20, 2629–2634.
- Proudfoot, N.J. (1989). How RNA polymerase II terminates transcription in higher eukaryotes. *Trends Biochem. Sci.* 14, 105–110.
- Proudfoot, N. (1996). Ending the message is not so simple. *Cell* 87, 779–781.
- Proudfoot, N. (2000). Connecting transcription to messenger RNA processing. *Trends Biochem. Sci.* 25, 290–293.
- Reines, D., Conaway, J.W., and Conaway, R.C. (1996). The RNA polymerase II general elongation factors. *Trends Biochem. Sci.* 21, 351–355.
- Roberts, G.C., Gooding, C., Mak, H.Y., Proudfoot, N.J., and Smith, C.W. (1998). Co-transcriptional commitment to alternative splice site selection. *Nucleic Acids Res.* 26, 5568–5572.
- Rodriguez, C.R., Cho, E.J., Keogh, M.C., Moore, C.L., Greenleaf, A.L., and Buratowski, S. (2000). Kin28, the TFIH-associated carboxy-terminal domain kinase, facilitates the recruitment of mRNA processing machinery to RNA polymerase II. *Mol. Cell. Biol.* 20, 104–112.
- Ryan, K., Murthy, K.G.K., Kaneko, S., and Manley, J.L. (2002). Requirements of the RNA polymerase II C-terminal domain for reconstituting pre-mRNA 3' cleavage. *Mol. Cell. Biol.*, in press.
- Sachs, A.B., Sarnow, P., and Hentze, M.W. (1997). Starting at the beginning, middle, and end: translation initiation in eukaryotes. *Cell* 89, 831–838.
- Shatkin, A.J., and Manley, J.L. (2000). The ends of the affair: capping and polyadenylation. *Nat. Struct. Biol.* 7, 838–842.
- Sisodia, S.S., Sollner-Webb, B., and Cleveland, D.W. (1987). Specificity of RNA maturation pathways: RNAs transcribed by RNA polymerase III are not substrates for splicing or polyadenylation. *Mol. Cell. Biol.* 7, 3602–3612.
- Smale, S.T., and Tjian, R. (1985). Transcription of herpes simplex virus tk sequences under the control of wild-type and mutant human RNA polymerase I promoters. *Mol. Cell. Biol.* 5, 352–362.
- Smith, C.W., and Valcarcel, J. (2000). Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem. Sci.* 25, 381–388.
- Steinmetz, E.J., Conrad, N.K., Brow, D.A., and Corden, J.L. (2001). RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. *Nature* 413, 327–331.
- Sun, J.S., and Manley, J.L. (1995). A novel U2–U6 snRNA structure is necessary for mammalian mRNA splicing. *Genes Dev.* 9, 843–854.
- Takagaki, Y., and Manley, J.L. (1998). Levels of polyadenylation factor CstF-64 control IgM heavy chain mRNA accumulation and other events associated with B cell differentiation. *Mol. Cell* 2, 761–771.
- Takagaki, Y., Seipelt, R.L., Peterson, M.L., and Manley, J.L. (1996). The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation. *Cell* 87, 941–952.
- Tantravahi, J., Alvira, M., and Falck-Pedersen, E. (1993). Characterization of the mouse beta maj globin transcription termination region: a spacing sequence is required between the poly(A) signal sequence and multiple downstream termination elements. *Mol. Cell. Biol.* 13, 578–587.
- Tran, D.P., Kim, S.J., Park, N.J., Jew, T.M., and Martinson, H.G. (2001). Mechanism of poly(A) signal transduction to RNA polymerase II in vitro. *Mol. Cell. Biol.* 21, 7495–7508.
- Vagner, S., Vagner, C., and Mattaj, I.W. (2000). The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF 65 to couple 3'-end processing and splicing. *Genes Dev.* 14, 403–413.
- Valadkhan, S., and Manley, J.L. (2000). A tertiary interaction detected in a human U2–U6 snRNA complex assembled in vitro resembles a genetically proven interaction in yeast. *RNA* 6, 206–219.
- Valadkhan, S., and Manley, J.L. (2001). Splicing-related catalysis by protein-free snRNAs. *Nature* 413, 701–707.
- Wahle, E., and Rueggsegger, U. (1999). 3'-end processing of pre-mRNA in eukaryotes. *FEMS Microbiol. Rev.* 23, 277–295.
- Wittschieben, B.O., Otero, G., de Bizemont, T., Fellows, J., Erdjument-Bromage, H., Ohba, R., Li, Y., Allis, C.D., Tempst, P., and Svejstrup, J.Q. (1999). A novel histone acetyltransferase is an integral subunit of elongating RNA polymerase II holoenzyme. *Mol. Cell* 4, 123–128.
- Wu, S., Romfo, C.M., Nilsen, T.W., and Green, M.R. (1999). Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature* 16, 832–835.
- Yonaha, M., and Proudfoot, N.J. (2000). Transcriptional termination and coupled polyadenylation in vitro. *EMBO J.* 19, 3770–3777.

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

Zhao, J., Hyman, L., and Moore, C. (1999). Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol. Mol. Biol. Rev.* 63, 405–445.

Zorio, D.A., and Blumenthal, T. (1999). Both subunits of U2AF recognize the 3' splice site in *Caenorhabditis elegans*. *Nature* 398, 835–838.