Connecting transcription to messenger RNA processing

Nick Proudfoot

The production of messenger RNA by gene transcription requires at least three RNA-processing mechanisms: capping, splicing and polyadenylation. All three reactions occur in intimate association with the elongating polymerase complex through the C terminus of the largest subunit of RNA polymerase II. The processing of mRNA is therefore orchestrated to act on the nascent RNA as soon as it emerges from the polymerase complex.

Such analysis was based on the technically demanding nuclear-run-on procedure, which is only easily applied to highly active Pol II transcription units, such as globin genes or adenovirus. In essence, this technique involves the isolation of nuclei from a cell line or tissue transcribing the gene in question, although in yeast, permeabilized whole cells work equally well. These nuclei or cells are incubated with labelled ribonucleotide triphosphates, so that polymerases thus results in a huge change in the ionic nature of this domain, with the potential acquisition of 104 negatively charged phosphate groups. Initial studies in heat-shock genes indicated that this CTD-phosphorylation process switches transcription from initiation to elongation, probably through the release of a substantial part of the Pol II initiation complex.

That part of this complex associate directly with the CTD was shown by experiments in yeast. In budding yeast, the CTD is a shorter, 26-heptad-repeat domain (in mammals). Complete phosphorylation by CTD kinases thus results in a huge change in the ionic nature of this domain, with the potential acquisition of 104 negatively charged phosphate groups. Initial studies in heat-shock genes indicated that this CTD-phosphorylation process switches transcription from initiation to elongation, probably through the release of a substantial part of the Pol II initiation complex.

What has now become clear is that, once transcription elongation has started, the space created on the CTD by the loss of initiation factors might well be filled by RNA-processing factors. These effectively ‘ride along’ with the Pol II elongation complex as it makes the nascent RNA and so are ready and waiting to process the transcript as it emerges from the RNA-binding pocket of Pol II. (This subject was excellently reviewed by the DNA template model. The other model reflected the possibility that cleavage of the mRNA transcript at the poly(A) site caused the downstream RNA, still attached and actively transcribed, to be highly unstable. Because this downstream RNA lacks a protective 5’ cap or triphosphate, it would be sensitive to exonuclease attack. Degradation of this RNA might ‘catch up’ with the polymerase and then promote termination. These models have long been considered as alternatives. However, it is now becoming increasingly likely that both could be a part of the overall picture.

The Pol II C-terminal domain

A surprising structural feature of Pol II, three years ago, was the direct connection between transcription and mRNA processing. The largest subunit of Pol II contains an extraordinary C-terminal domain (CTD) comprising 52 seven-amino-acid repeats (in mammals). Most importantly, the CTD contains three serine per heptad repeat, two of which are subject to phosphorylation. Complete phosphorylation by CTD kinases thus results in a huge change in the ionic nature of this domain, with the potential acquisition of 104 negatively charged phosphate groups. Initial studies in heat-shock genes indicated that this CTD-phosphorylation process switches transcription from initiation to elongation, probably through the release of a substantial part of the Pol II initiation complex.

This experimental procedure clearly showed that an RNA-processing event can directly influence transcriptional elongation. However, because the site of polyadenylation is often several kilobases upstream from where the polymerase finally terminates, it was conceptually hard to imagine how these two events connect at the molecular level. Two models were proposed for this connection. The first model proposed that the polymerase loses a processivity factor required for efficient transcription elongation when it passes through the gene’s poly(A) signal. The polymerase then becomes prone to random release from the DNA template. The other model reflected the possibility that cleavage of the mRNA transcript at the poly(A) site caused the downstream RNA, still attached and actively transcribed, to be highly unstable. Because this downstream RNA lacks a protective 5’ cap or triphosphate, it would be sensitive to exonuclease attack. Degradation of this RNA might ‘catch up’ with the polymerase and then promote termination. These models have long been considered as alternatives. However, it is now becoming increasingly likely that both could be a part of the overall picture.

REVIEWS

N.J. Proudfoot is at the Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford, UK OX1 3RE. Email: nicholas.proudfoot@path.ox.ac.uk

TIBS 25 – JUNE 2000

0968 – 0004/00/$ – See front matter © 2000, Elsevier Science Ltd. All rights reserved. PII: S0968-0004(00)01591-7
Coupled capping

Cap addition is the first RNA-processing event to occur, soon after transcription has started (Fig. 1). 7-Methyl guanine is added to the triphosphate end of the transcript by three enzymes: a phosphatase, a guanyl transferase and a methylase. The guanyl transferase and methylase enzymes have both been shown to bind to the phosphorylated CTD and to carry out the RNA-processing reaction before the transcript has reached a size of ~30 nucleotides. It therefore seems possible that capping marks the completed switch from transcription initiation to elongation.

These initial events in mRNA production might relate to the observation that a number of genes stall transcriptional elongation soon after promoter clearance, a process called attenuation. This attenuation might be used to regulate gene expression by controlling the release of the polymerase complex from its attenuation block. This provides a sensitive way to regulate the output of these genes; capping could play a role in this attenuation process.

Coupled splicing

The next RNA-processing reaction to take place on the nascent transcript is intronic splicing. The complexity of splicing increases enormously from yeast to mammals. In budding yeast, few genes have introns and most of those that do have only one small intron near the beginning of the transcript. Whether splicing is co-transcriptional in these cases or not has yet to be determined. However, most mammalian genes are highly intronic, with well over 90% of the gene made up of noncoding introns. In fact, the relatively small exons might often be hard to find in a sea of ‘intronic rubbish’.

The co-transcriptional nature of splicing was perhaps most clearly visualized by electron-microscopic analysis of chromosomal spreads from Drosophila embryos caught in the act of transcription (Fig. 2). Most importantly, nascent RNA was also clearly visible in these images, with partially spliced introns and associated spliceosomes in evidence. Clearly, the large size of the assembled spliceosome precludes a full interaction with the CTD. What has become clear is that some splicing factors, including proteins with homology to splicing regulatory proteins (SR proteins), do associate with the CTD through a specific CTD-binding domain. This class of protein is called either a CTD-associated SR-like protein (CASP) or an SR-like CTD-associated factor (SCAF). SR proteins (including SCAFs) generally possess one or two RNA-binding domains (recognition motifs, RBMs) and an arginine-and-serine-rich region (the RS domain). RBMs often target SR proteins to exon enhancer sequences. The RS domain then appears to provide molecular ‘glue’, allowing RS-RS interactions to

---

Figure 1

The three stages of transcription by RNA polymerase II (initiation, elongation and termination) and how mRNA processing is associated with each of these stages. At initiation, transcription factors (TFs, red) associate with enhancer sequences (shown greatly simplified) and recruit RNA polymerase II (Pol II, blue) with associated general transcription factors (GTFs, green). Cleavage polyadenylation stimulatory factor (CPSF) also associates with GTFs at this stage (orange). At elongation, the Pol II C-terminal domain (CTD) becomes highly phosphorylated (purple) and many of the GTFs are replaced by only partly characterized elongation factors (Elongator, green). Also, capping enzymes (guanyl transferase (GT) and methyl transferase (MT)) associate with the CTD to promote cap addition, followed by SR-like CTD-associated factors (SCAFs) to promote spliceosome recruitment (orange). Finally, 3’ cleavage and polyadenylation are mediated by CTD-associated cleavage-stimulation factor (CstF) as well as CPSF (orange), which, in turn, attracts additional cleavage and polyadenylation factors to polyadenylate the mRNA, so releasing it from the transcription complex. This process is enhanced by splicing of the terminal intron. Termination finally occurs in part through the degradation of the remaining nascent transcript and might be enhanced by transcriptional pausing.
Figure 2

Electron micrograph directly visualizing co-transcriptional splicing, obtained by Miller chromatin spreading. The gene shown is from Drosophila and is ~6 kb long. DNA enters at the upper left of the micrograph and exits at the lower right. Transcription initiates near the DNA template and might have been caught in the process of termination and release. Bar, 200 nm. (Photograph courtesy of Ann Beyer and Yvonne Osheim, University of Virginia, VA, USA.)
RNA cleavage triggers Pol II termination and are therefore consistent with the second termination model described above. However, recent data in higher eukaryotes paint a different picture. As shown previously for co-transcriptional splicing, electron-microscopic analysis of actively transcribing minichromosomes in Xenopus oocytes visually shows that termination requires a poly(A) signal.11 Without one, long RNA transcripts still attached to elongating Pol II are visible that must have transcribed around the circular template many times. With a strong poly(A) signal at the transcript’s 3’ end, elongating Pol II might extend into the 3’ flanking region of the gene and up to the site of termination before cleavage at the poly(A) site actually occurs. As it is also known that Pol II CTD directly acts in vitro polyadenylation,29 it seems plausible that CTD-associated 3’ cleavage triggers Pol II termination. Consistent with these observations, it has been further shown that the splice-acceptor site of the terminal intron not only enhances downstream polyadenylation26 but also does so at the nascent RNA level. In effect, this splicing signal works in concert with the poly(A) signal to mediate Pol II termination.23 Thus, definition of the terminal exon appears to be the termination trigger. Once these final events in transcript maturation occur, mRNA is released from the Pol II complex, and this could leave it susceptible to 3’-end processing.14 If so, one might even consider that associated 3’-end processing coincides with transcription.32 Thus, definition of the 3’-end is also mediated by the transcription complex.14 It has been proposed that the nature of actively transcribing minichromosomes in Xenopus oocytes visually shows that termination requires a poly(A) signal.11 Without one, long RNA transcripts still attached to elongating Pol II are visible that must have transcribed around the circular template many times. With a strong poly(A) signal at the transcript’s 3’ end, elongating Pol II might extend into the 3’ flanking region of the gene and up to the site of termination before cleavage at the poly(A) site actually occurs. As it is also known that Pol II CTD directly acts in vitro polyadenylation,29 it seems plausible that CTD-associated 3’ cleavage triggers Pol II termination. Consistent with these observations, it has been further shown that the splice-acceptor site of the terminal intron not only enhances downstream polyadenylation26 but also does so at the nascent RNA level. In effect, this splicing signal works in concert with the poly(A) signal to mediate Pol II termination.23 Thus, definition of the terminal exon appears to be the termination trigger. Once these final events in transcript maturation occur, mRNA is released from the Pol II complex, and this could leave it vulnerable to 3’-end destabilization by mecha- 
nisms such as rapid degradation of the vestigial nascent RNA. The exact nature of Pol II destabilization and release from the DNA remains to be characterized. However, it seems likely that both of the early models for termination will fit into a final picture of how transcriptional termi-
nation actually occurs (Fig. 1).

Future developments

All of the above transcription–RNA processing interconnections are begin-

ning to be clarified through sophisticated nuclear-imaging technology. By using confocal microscopy on fixed nuclear preparations labelled with fluorescent antibodies against protein or modified nucleic acid probes, it is becoming possible to visualize gene expression under the microscope. These studies reveal an ordered ar-

rangement of transcription sites and spliceosomes, indicating that many RNA-processing activities coincide with these transcription sites.3-35 The use of green fluorescent protein attached to splicing factors has also allowed the visualization of co-transcriptional splicing in living cells. Such data show that splicing factors are stored in specific nuclear structures (speckles) and that these factors move out of the speckles onto actively transcribing chromatin.12 Interestingly, splicing-factor movement requires an intact CTD on the transcription complex, as CTD-deletion mutants fail to recruit splicing factors.2,13 The possibility that cell- and molecular-biological approaches are now able to target the same mechanism raises hopes that we will soon acquire a detailed picture of exactly how eukary-

otes coordinate different stages in gene expression. Three recent reviews give excellent coverage of the subject of transcription–processing interconnections for those who want to delve deeper.21,22,25

Acknowledgments

I am grateful to Ann Beyer, Walter Keller, Daniela Barilla, Saba Valabhadjan and Jim Manley for advice on the manuscript. I am indebted to Ann Beyer and Yvonne Osheim for providing the picture shown in Fig. 2.

References


TIBS 25 – JUNE 2000