The union of transcription and mRNA processing: 20 years of coupling

DAVID L. BENTLEY

Department of Biochemistry and Molecular Genetics, University of Colorado School of Medicine, Aurora, Colorado 80045, USA

Twenty years ago the communities studying eukaryotic transcription and mRNA processing were pretty segregated and seemed happy for the most part to attend their respective Cold Spring Harbor meetings in blissful ignorance of one another. Transcription people, myself included, were content not to concern ourselves with how a spliceosome worked, and equally, I imagine, RNA processing people were not too concerned about how an RNA polymerase II (pol II) transcription complex was put together. My blinkered perspective began to open up on the day of our lab's Christmas lunch in 1995. That morning Nova Fong showed me a freshly developed RNAse protection assay (RPA) to look at processing of the poly(A) site on transcripts made by a mutant form of pol II that lacked the C-terminal domain (CTD). This enigmatic, highly conserved domain on the polymerase large subunit comprises 52 tandem heptad repeats (consensus YSPTSPS). It was co-discovered by Jeff Corden a decade earlier, and he had given us a cleverly designed toxin-resistant truncation mutant. We thought the CTD might be required to overcome the transcription elongation block that occurs near the start sites of genes like c-myc but in reality early elongation was enhanced, not inhibited, by removing the CTD. As we reviewed these disappointing results and tried to pick up the pieces, it became clear that a "mystery" band on the gels was always associated with the CTD-truncated polymerase, and this band could be explained if there were RNAs that read all the way around the plasmid template without terminating. Seminal work by the Manley, Proudfoot, Darnell, and Shenk labs showed that transcription termination required a functional poly(A) site, in the first clear case of communication between transcription and mRNA processing. So we asked whether there was a problem with 3′ end processing when pol II lacks its CTD and Nova's RPA showed that those transcripts ran right through the poly(A) site without getting cleaved and polyadenylated. Soon afterwards Nova and Susan McCracken found that splicing and capping also depend on an intact pol II CTD, and my lab learned to embrace the world of mRNA processing. Over the past 20 years the ties between the transcription and mRNA maturation have multiplied and strengthened enormously through insights made in many labs and I submit that this union has greatly enriched both fields.

The big question 20 years ago was how could a mutation at the heart of the transcription machinery, in the polymerase itself, possibly cause all three major mRNA processing steps to fail? There seemed to be something special about RNA pol II itself because the Tjian and Cleveland labs had found in the late '80s that RNA pol I and pol III were not up to the job of making mature mRNA effectively. Somehow the "right" RNA polymerase seemed necessary to determine the proper fate of pre-mRNAs. Truncation of the CTD made pol II lose its special "mRNA processing mojo"; without it, transcripts lost their identity as mRNA precursors and were treated as if they had been made by the "wrong" RNA polymerase. The CTD therefore appeared to provide a molecular lynchpin that coupled mRNA processing with transcription by pol II.

Such coupling is only possible of course if transcription and processing actually occur at the same time and place and this was graphically displayed in electron micrographs of fly genes from the Beyer, Wieslander, and Daneholt labs. These images revealed the removal of introns from growing transcripts still tethered to the template by RNA polymerases caught in the act of transcription. The early EM demonstrations that splicing could occur co-transcriptionally were validated decades later by Rosbash, Neugebauer, and others using genome-wide approaches that show co-transcriptional splicing is the norm rather than the exception. Twenty years ago mRNA maturation events were usually described as "post-transcriptional" and now they are more accurately described as "co-transcriptional." This shift is more than just a semantic one; it reflects a realization that the natural substrate for mRNA processing and packaging factors is a nascent RNA chain that is being expelled through an exit channel in the polymerase as the transcription elongation complex (TEC) makes its way along a chromatin template at rates of ∼0.5–4.0 kb per minute.

The physical intimacy between the mRNA processing and transcription machines is attested by direct binding of capping, splicing, and 3′ end processing factors to the CTD and their co-purification with pol II in "mRNA factory" complexes. Even more compelling are the structures of CTD-bound forms of capping enzymes and 3′ end processing factors revealed at atomic resolution by the Lima and Cramer labs. A major ongoing challenge is for structural and mechanistic studies to discover how functional coupling of synthesis and maturation of the mRNP really works. In its simplest form, coupling serves a localization role to help processing factors and pre-mRNAs find one another in a timely fashion within the labyrinthine nuclear landscape. Coupling mechanisms probably first evolved for localization and later acquired more complex regulatory functions that influence the activities of interacting partners as shown for CTD interactions with 3′ processing, splicing, and capping factors by the Manley and Shuman labs.

Thinking about transcription in the context of simultaneous mRNA processing has helped us to understand why the TEC undergoes dynamic modifications. The Dahmus lab discovered that the CTD becomes phosphorylated shortly after transcription initiates and Steve Buratowski showed that phosphorylations on different residues of the heptad repeats change in intricate ways that are synched with the cycle of transcript initiation, elongation, and termination. Thus phosphorylation on Ser5 of the YSPTSPS repeats is high early in the cycle when polymerase is at the 5′ end of a gene, and low at later times when polymerase is at the 3′ end. Ser2 phosphorylation, on the other hand, follows a distinct set of rules that results in the opposite profile: low 5′ and high 3′ modification. How can one make sense of the dizzying dynamics of CTD phosphorylation? Only by considering what is happening co-transcriptionally to the nascent RNA does all this complexity start to make sense. Capping enzyme is bound and activated specifically by heptads phosphorylated on Ser5, while the 3′ end processing factor Pcf11 is recruited specifically to Ser2 phosphorylated heptads. One can therefore rationalize CTD Ser5 and Ser2 phosphorylation dynamics in a satisfying way by taking into account the need to bring capping and 3′ end processing factors on board at the right times to modify the 5′ and 3′ ends of the nascent transcript. The CTD plays an instructional role in directing recruitment of processing factors to the TEC, and Buratowski suggested that the instructions are written in code of phosphates on the heptad repeats. There is good reason to believe that the ancestral CTD code word was a combination of two Ser5 phosphorylated heptads that instructs capping. Indeed Schwer and Shuman showed this is the only essential CTD code word in fission yeast. How elaborate the code has become in multicellular organisms is an important unanswered question.

The cross fertilization of transcription and mRNA processing research has worked in both directions. Consideration of mRNA processing helped to make sense of dynamic CTD phosphorylation and thinking about transcription elongation has inspired new models for how coupled mRNA processing works. Timing is a critical component of coupling mechanisms as first suggested by Weissman's "first come first served" model for why 5′ introns tend to be removed before 3′ introns. Alberto Kornblihtt first realized that the rate of transcript growth could influence whether or not an alternative exon is removed and it now seems clear that the speed of transcription really matters for many splicing decisions.

Almost two decades ago Kornblihtt showed that promoters, the core regulators of transcription, can influence the outcome of alternative splicing decisions and recent results from Jean Beggs suggest that splicing factors regulate transcription elongation and CTD phosphorylation. The fields of transcription and mRNA processing seem to have come full circle on one another over the last 20 years and I am confident their paths will continue to intertwine in fascinating ways for years to come. Among upcoming highlights, we might expect:

1. Deciphering of the CTD code and its full impact on mRNA maturation in multicellular organisms. The CTD is not a unique structure but I hope we will resolve the general shapes that this flexible domain assumes and how they change in different phosphorylation states.
2. A set of rules for how alternative splicing and polyadenylation decisions are made that takes into account not only association with RNA binding factors but also local transcription elongation rates, CTD phosphorylation, and chromatin modification.
3. An understanding of co-transcriptional RNA folding and how it affects mRNA maturation.
4. A mechanism for how mRNPs are assembled, a process that starts co-transcriptionally.
5. Who knows, perhaps in another 20 years, the biennial Cold Spring Harbor meetings on transcription and eukaryotic mRNA processing will merge into an annual meeting on coupled mRNP biogenesis.

Acknowledgments