

---

# A journey to the end of the message

---

**JAMES L. MANLEY**

Department of Biological Sciences, Columbia University, New York, New York 10027, USA

As do many good things, my interest in how mRNA 3' ends are made came about serendipitously. Ever since I was an undergrad here at Columbia in the early 1970s, I had been interested in figuring out how RNA polymerase (RNAP) begins transcription. This fascination began while I worked in the lab of the late Geoffrey Zubay, who had just developed the “Zubay system,” which for the first time allowed accurate transcription, and then translation, of exogenous DNA (a *lacZ* gene carried in a transducing  $\lambda$  phage; this was in the days before recombinant DNA) in an extract of *Escherichia coli*. I continued studies on translation using the Zubay system as a graduate student with Ray Gesteland at Cold Spring Harbor (this involved a detour into something called “protein splicing,” but that’s another story...). However, I remained fascinated with how RNAP initiates transcription, especially in mammalian cells. This was at the time a challenging problem, and it was even thought in some circles that RNAP II might initiate transcription randomly, with mature mRNAs made entirely by RNA processing.

I chose for my postdoc to work with Malcom Gefter at MIT. This was because his group had just provided evidence that RNAP II in nuclei isolated from mammalian cells could indeed initiate transcription. When I arrived we decided to collaborate with Phil Sharp to extend these studies, using HeLa cell nuclei isolated from adenovirus-infected cells, which allowed us to take advantage of the highly expressed major late transcription unit. Indeed, we were able to show that transcription initiated in these isolated nuclei at a site precisely corresponding to the 5' end of the late RNA. We used the adenovirus system for several things, including subsequent development of the “Manley system,” which allowed accurate transcription of exogenously added DNA in HeLa cell extract. But we continued to analyze transcription in the isolated nuclei. A major goal was to determine if the newly synthesized transcripts were spliced in the nuclei. (These experiments were done in the late 1970s, a couple of years after the discovery of splicing, also of course in the adenovirus system, but well before the development of conditions that would allow splicing in cell extracts.) Alas, splicing did not occur under our conditions (although we were able to pro-

vide evidence that RNAs that had undergone splicing towards their 5' ends in vivo could be elongated in vitro, in retrospect providing perhaps the earliest evidence of transcription-coupled splicing). What we did find, however, was evidence for robust 3' cleavage and polyadenylation (PA), at several of the multiple sites in the late transcription unit. By using pulse-chase experiments, we were able to show that these 3' ends were indeed created by endonucleolytic cleavage of a longer precursor, and that they were polyadenylated essentially immediately upon cleavage; i.e., that cleavage and poly(A) synthesis are coupled.

I set up my lab at Columbia in 1980, and decided to pursue several of the projects related to mRNA synthesis I began at MIT. But I was especially excited about figuring out how polyadenylation works. One reason was that it seemed like it should be a relatively straightforward undertaking; after all, all we'd need to do is purify an endonuclease and a poly(A) polymerase (PAP) and we'd be done. Boy, was I wrong! My colleagues and I, and many other labs, have been working on this problem for 35 years now, and while we've learned a good deal there's still a lot we don't know. Much of this reflects the remarkable complexity of the “polyadenylation machinery,” which is of course much more than just the nuclease and PAP.

Our first goal was to develop conditions to purify the required factors. We initially developed conditions where a precursor RNA synthesized in HeLa cell extract could be polyadenylated in an AAUAAA-dependent manner, although without cleavage, i.e., at its end created by “run-off” transcription. (An interesting aside is that the PA machinery in the extract could use the TATAAA box in the adenovirus major late promoter, when transcribed into RNA, as a signal for PA. Although this is unlikely to be of physiological significance, this finding suggested a similarity between core RNAP II promoters and “core” PA signals, a view that has persisted to this day.) Shortly after this, Claire Moore and Phil Sharp succeeded in reproducing the complete PA reaction in similar extracts. But while establishing that the entire reaction can occur in vitro, this left the problem that these reactions were coupled with transcription, preventing fractionation of the extracts. This problem was overcome with the advent

---

**Corresponding author:** [jlm2@columbia.edu](mailto:jlm2@columbia.edu)

Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.050724.115>. Freely available online through the RNA Open Access option.

© 2015 Manley This article, published in *RNA*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

of systems employing phage RNAPs to produce RNAs that could be added exogenously to processing reactions. With these methods, Yoshio Takagaki and Lisa Ryner in my lab soon showed that a PAP and a separable “cleavage specificity factor” were necessary and sufficient for PA *in vitro*.

This set the stage for purification and characterization of the factors necessary for polyadenylation. A number of labs played significant roles, including those of Joe Nevins and Walter Keller, in addition to my own. Analogous to what was happening at about the same time in the transcription field, multiple distinct multisubunit protein factors were uncovered that are necessary, together with PAP, for 3′ cleavage and polyadenylation. And in both cases, these factors were found to be responsible for recognizing specific sequences in the DNA/RNA, and for recruiting RNAP II or PAP, both of which are sequence-non-specific on their own. In the case of PA, four factors were eventually identified, Cleavage Factors I and II (CFI and CFII), Cleavage-Polyadenylation Specificity Factor (CPSF, which we know now recognizes AAUAAA and contains the endonuclease, denoted CPSF-73), and Cleavage Stimulation Factor (CstF; we initially thought this factor was stimulatory but not essential. However, this was because CstF was present in one of our other fractions, and we now know that CstF is, like the other factors, essential).

Identification of the four factors was finished by the end of the 1980s. One interesting diversion along the way was the belief held by some that one of these factors must be a snRNP. This was entirely reasonable, as it was known of course that splicing requires snRNPs, and perhaps more relevantly, histone mRNA 3′ end formation had just been found to also require a snRNP (U7). Indeed, more than one paper was published in one of everyone’s favorite journals suggesting the involvement of a snRNP in polyadenylation, with attention focused on the then mysterious U11 snRNP. This background demanded extra vigilance in characterizing our purified fractions, to make sure a snRNP wasn’t lurking in one of them, which indeed none were.

During the course of the 1990s, and beyond, the four factors and PAP were all characterized further, cDNAs encoding subunits cloned, and even more essential factors identified. I won’t go into the details of all that here, as a number of excellent reviews have been published over the years. What I will do instead is describe what led us to the discovery of one unexpected PA factor: RNAP II. The first evidence that RNAP II, and specifically the C-terminal domain of its largest subunit (CTD), is important for PA came from studies of David Bentley and colleagues, who showed that mRNAs transcribed by a CTD-lacking RNAP II in transfected cells were inefficiently spliced or polyadenylated. An explanation for this was that RNAP II, via the CTD, helps to recruit RNA processing factors to their sites of action. While this is indeed part of the story, we showed the next year that RNAP II, and specifically the CTD, does more than that and is indeed required for 3′ cleavage *in vitro*, in the absence of transcription.

How we stumbled upon the role of the CTD in PA, however, initially had nothing to do with transcription. Rather, it stemmed from our investigation of an apparent energy (ATP) requirement of the cleavage reaction. ATP, and creatine phosphate (CP) to regenerate ATP, were routinely added to *in vitro* reactions, and ATP was long thought to be essential. However, this seemed odd, as the cleavage reaction could proceed in the presence of EDTA. So Yutaka Hirose decided to investigate this and found that CP, but not ATP, was required. Significantly, hydrolysis of the high energy phosphate bond did not occur, and arginine phosphate functioned as well as CP. These results were puzzling, suggesting that CP or a related compound functioned as a cofactor for cleavage. This seemed unlikely, and we were indeed pointed in the right direction by a conversation with Jim Dahlberg at an RNA Society meeting, who suggested that these compounds, which were required in high concentrations, might function as molecular mimics for an unknown phosphoprotein. As phosphoarginine is rare in eukaryotic proteins, we first tested whether phosphoserine could activate cleavage, which it efficiently did. Yutaka then tested the ability of several phosphoproteins involved in RNA processing (PAP and SR protein splicing factors) to substitute for CP, but these were inactive. What worked though was—you guessed it—purified RNAP II. And the CTD, known to be extensively phosphorylated on serines in the heptad repeats, is necessary and sufficient to activate cleavage. Precisely how the CTD works, and how CP can substitute for it, remains unknown. One interesting possibility is that the CTD may function analogously to the way it does in facilitating 5′ capping, which, as shown by Stewart Shuman and colleagues, is both to recruit capping enzyme and then to allosterically activate it. The complexity of the PA machinery has so far not allowed the kind of analysis that would be necessary to show this.

Our cloning of PAP also merits a brief mention. A PAP activity was first partially purified by Mary Edmonds and colleagues way back in 1960, over ten years before the discovery of mRNA poly(A) tails. Several groups subsequently characterized PAP activities with somewhat different properties, and it was unclear if any of these corresponded to the PAP that functions in mRNA polyadenylation. However, AQHead preparations prepared in the lab of Fred Bollum in Bethesda looked especially promising, so I contacted him when we first set out to clone PAP cDNAs. Unfortunately, or so it seemed, they no longer worked on PAP, and only had a preparation that had been in a freezer for 15 years. But undeterred, Tobias Raabe proceeded with this sample, obtained peptide sequence, and isolated PAP cDNAs, proving that nothing ages like a good PAP preparation! Since then, PAP has been extensively characterized by many labs. For example, over the years we showed that PAP, although a single subunit enzyme, is highly regulated, by alternative splicing of its pre-mRNA and by multiple post-translational modifications, including phosphorylation, sumoylation and, most recently, PARylation.

After 35 years of polyadenylation, what remains for the future? Well, quite a bit actually, but I'll conclude by saying a bit about just one thing, alternative polyadenylation (APA), which is currently the hottest topic at the end of the message. APA is the selection of one of several possible cleavage sites in an mRNA precursor, and is quite common, with over 70% of human primary transcripts subject to APA. APA can alter coding potential of transcripts, although more frequently it affects the length of the 3' UTR, and hence the constellation of miRNA and protein binding sites that populate these regions. A growing number of excellent studies over the last several years have documented widespread changes in APA that occur during, for example, cell differentiation and disease, especially cancer. Important questions concern the actual physiological significance of these changes, and also

the nature of the underlying mechanisms. With respect to the latter, while multiple mechanisms may exist, I favor the idea that changes in the make-up of the core PA machinery are a major contributor. I confess some bias here, as my lab provided evidence for this back in the 1990s, by showing that changes in levels of one PA factor, CstF64, during B cell differentiation were sufficient to switch APA in the IgM-heavy chain transcripts, helping determine whether membrane-bound or secreted antibodies are produced. Indeed, much more recently changes in a variety of PA factors have been observed in different circumstances, and to affect APA in different ways. Thus an intriguing possibility is the existence of a "PA factor code," in which different combinations of PA factors lead to distinct choices of APA in different cell types. Time will tell whether this model has any validity.