CHAPTER 15

RNA Processing II: Capping and Polyadenylation

Caps in an Egyptian market. © *Iconotec.com.*

Besides splicing, eukaryotic cells perform several other kinds of processing on their RNAs. Messenger RNAs are subject to two kinds of processing, known as capping and polyadenylation. In capping, a special blocking nucleotide (a cap) is added to the 5′-end of a pre-mRNA. In polyadenylation, a string of AMPs (poly[A]) is added to the 3′-end of the pre-mRNA. These steps are essential for the proper function of mRNAs and will be our topics in this chapter.

15.1 Capping

By 1974, several investigators had discovered that mRNA from a variety of eukaryotic species and viruses was methylated. Moreover, a significant amount of this methylation was clustered at the 5′-end of mRNAs, in structures we call **caps.** In this section we will examine the structure and synthesis of these caps.

Cap Structure

Before gene cloning became routine, viral mRNAs were much easier to purify and investigate than cellular mRNAs. Thus, the first caps to be characterized came from viral RNAs. Bernard Moss and his colleagues produced vaccinia virus mRNAs in vitro and isolated their caps as follows: They labeled the methyl groups in the RNA with [^{3}H] S-adenosylmethionine (AdoMet, a methyl donor), or with ^{32}P-nucleotides, then subjected the labeled RNA to base hydrolysis. The major products of this hydrolysis were mononucleotides, but the cap could be separated from these by DEAE-cellulose chromatography. Figure 15.1 shows

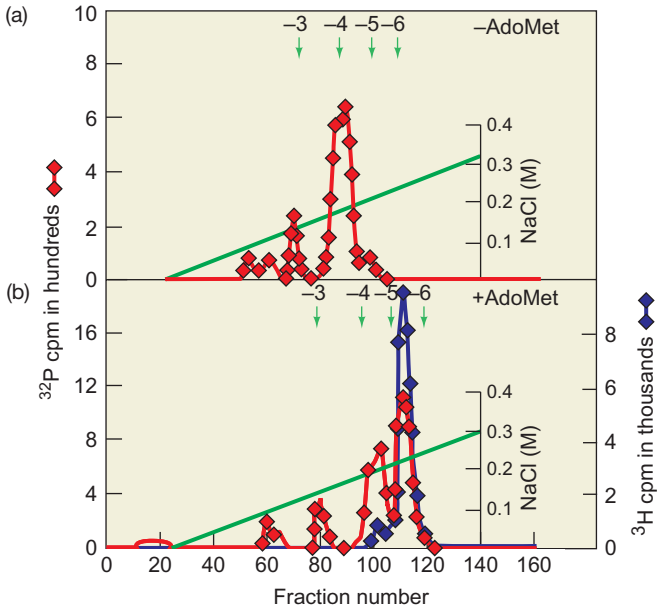

Figure 15.1 DEAE-cellulose chromatographic purification of vaccinia virus caps. Wei and Moss allowed vaccinia virus particles to synthesize caps in the presence of [β, γ-^{32}P]GTP and in the **(a)** absence and **(b)** presence of *S*-adenosyl[methyl-^{3}H]methionine. Then they digested the labeled, capped RNAs with KOH and separated the products by DEAE-cellulose column chromatography. ^{3}H (blue) and ^{32}P (red) radioactivities (in counts per minute) are plotted versus column fraction number. Salt concentrations (green) of each fraction are also plotted. The positions and net charges of markers are shown at the top of each panel. (*Source:* Adapted from Wei, C.M. and B. Moss, Methylated nucleotides block 5′-terminus of vaccinia virus messenger RNA, *Proceedings of the National Academy of Sciences USA* 72(1):318–322, January 1975.)

the chromatographic behavior of the vaccinia virus caps. They behaved as a substance with a net charge near −5. Furthermore, the red and blue curves in Figure 15.1b show that the ^{3}H(methyl) and ^{32}P labels essentially coincided, demonstrating that the caps were methylated. Aaron Shatkin and his coworkers obtained very similar results with reovirus caps.

To determine the exact structure of the reovirus cap, Yasuhiro Furuichi and Kin-Ichiro Miura performed the following series of experiments. They found that they could label the cap with [β,γ-^{32}P]ATP (but not with [γ-^{32}P]ATP). This result indicated that the β-phosphate, but not the γ-phosphate, was retained in the cap. Because the β-phosphate of a nucleoside triphosphate remains only in the first nucleotide in an RNA, this finding reinforced the notion that the cap was at the 5′-terminus of the RNA. But the β-phosphate must be protected, or blocked, by some substance (X), because it cannot be removed with alkaline phosphatase.

This raised the next question: What is X? The blocking agent could be removed with phosphodiesterase, which cuts both phosphodiester and phosphoanhydride bonds (e.g., the bond between the α- and β-phosphates in a nucleotide). This enzyme released a charged substance likely to be Xp. Next, Furuichi and Miura removed the phosphate from Xp with phosphomonoesterase, leaving just X, and subjected this substance to paper electrophoresis, followed by paper chromatography. Figure 15.2 shows that X coelectrophoresed with **7-methylguanosine (m^{7}G).** Thus, the capping substance is m^{7}G.

Another product of phosphodiesterase cleavage of the cap was pAm (2′-O-methyl-AMP). Thus, m^{7}G is linked to pAm in the cap. What is the nature of the linkage? The following two considerations tell us that it is a triphosphate: (1) The α-phosphate, but not the β- or γ-phosphate, of GTP was retained in the cap. (2) The β- and α-phosphates of ATP are retained in the cap. Thus, because one phosphate comes from the capping GTP, and two come from the nucleotide (ATP) that initiated RNA synthesis, there are three phosphates (a triphosphate linkage) between the capping nucleotide (m^{7}G) and the next nucleotide. Furthermore, because both ATP and GTP have their phosphates in the 5′-position, the linkage is very likely to be 5′ to 5′.

How do we explain the charge of the reovirus cap, about −5? Figure 15.3 provides a rationale. Three negative charges come from the triphosphate linkage between the m^{7}G and the penultimate (next-to-end) nucleotide. One negative charge comes from the phosphodiester bond between the penultimate nucleotide and the next nucleotide. (This bond is not broken by alkali because the 2′-hydroxyl group, which is needed for cleavage, is methylated.) Two more negative charges come from the terminal phosphate in the cap. This makes a total of six negative charges, but the m^{7}G provides a positive charge, which gives the purified reovirus cap a charge of about −5.

Other viral and cellular mRNAs have similar caps, although the extent of 2′-O-methylation can vary to produce

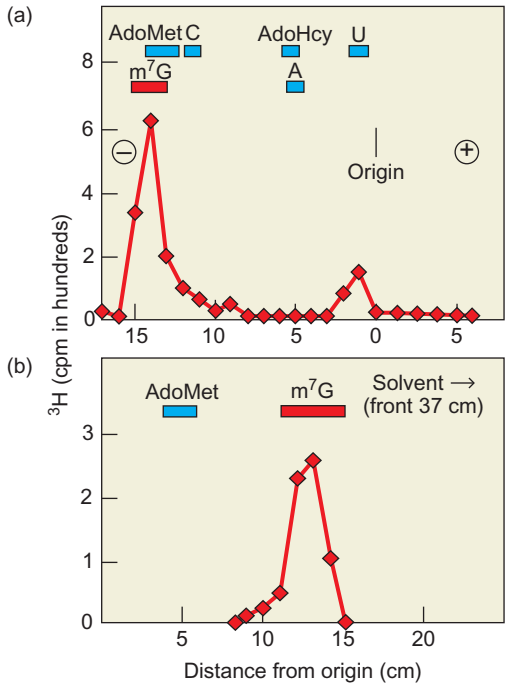

Figure 15.2 Identification of the capping substance (X) as 7-methylguanosine. Miura and Furuichi used phosphomonoesterase to digest the ^{3}H-labeled capping substance (Xp) to yield X. They electrophoresed this digest **(a)** along with a series of markers (*S*-adenosylmethionine, AdoMet; m^{7}G; *S*-adenosylhomocysteine, AdoHcy; adenosine, A; and uridine, U). Because electrophoresis did not resolve AdoMet and m^{7}G, these workers subjected the digest to paper chromatography **(b)** along with markers for AdoMet and m^{7}G. The radioactivity in X cochromatographed with the m^{7}G marker. (*Source:* Data from Furuichi, Y. and K. -I. Miura, A blocked structure at the 5′ terminus of mRNA from cytoplasmic polyhedrosis virus. *Nature* 253:375, 1975.)

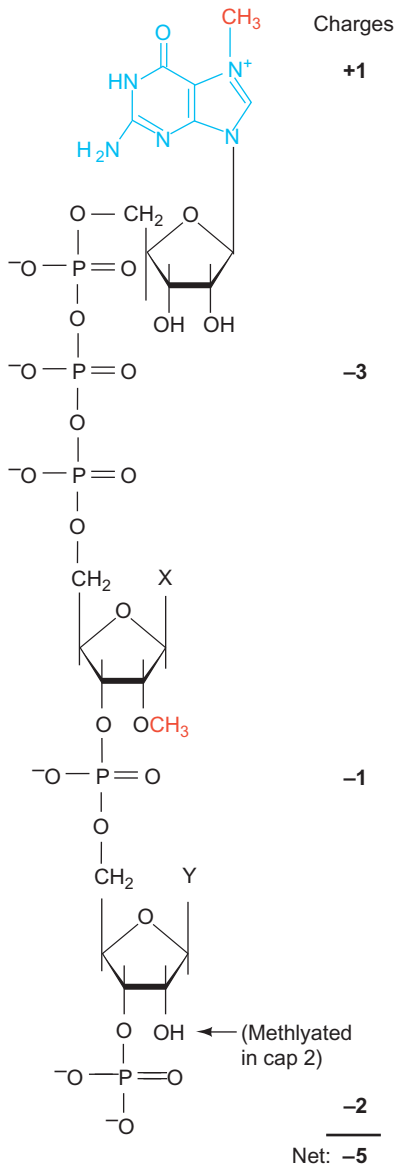

Figure 15.3 Reovirus cap structure (cap 1), highlighting the charges. The m^{7}G (blue guanine with red methyl group) contributes a positive charge, the triphosphate linkage contributes three negative charges, the phosphodiester bond contributes one negative charge, and the terminal phosphate contributes two negative charges. The net charge is therefore about −5. The 2′-hydroxyl group on the ribose attached to the Y base would be methylated in cap 2.

three forms of cap. **Cap 1** is the same as the cap shown in Figure 15.3. **Cap 2** has another 2′-O-methylated nucleotide (two in a row). And **cap 0** has no 2′-O-methylated nucleotides. Cap 2 is found only in eukaryotic cells, cap 1 is found in both cellular and viral RNAs, and cap 0 is found only in certain viral RNAs. Most of the snRNAs (Chapter 14) have another kind of cap, which contains a trimethylated guanosine. We will discuss these caps later in this chapter.

Cap Synthesis

To determine how caps are made, Moss and his colleagues, and Furuichi and Shatkin and their colleagues, studied capping of model substrates in vitro. These investigators used cores from vaccinia virus and reovirus, respectively, to provide the capping enzymes. Both these human viruses replicate in the cytoplasm of their host cells, so they do not have access to the host nuclear machinery. Therefore, they must carry their own transcription and capping systems right in their virus cores. In both viruses, we observe the same sequence of events, as illustrated in Figure 15.4. (a) A nucleotide phosphohydrolase (also called **RNA triphosphatase**) clips the γ-phosphate off the triphosphate at the 5′-end of the growing RNA (or model substrate), leaving a diphosphate. (b) A guanylyl transferase attaches GMP from GTP to the diphosphate at the end of the RNA, forming the 5′–5′-triphosphate linkage. (c) A methyltransferase transfers the methyl group from *S*-adenosylmethionine (AdoMet) to the 7-nitrogen of the capping guanine. (d) Another methyltransferase uses another molecule of AdoMet to methylate the 2′-hydroxyl of the penultimate nucleotide.

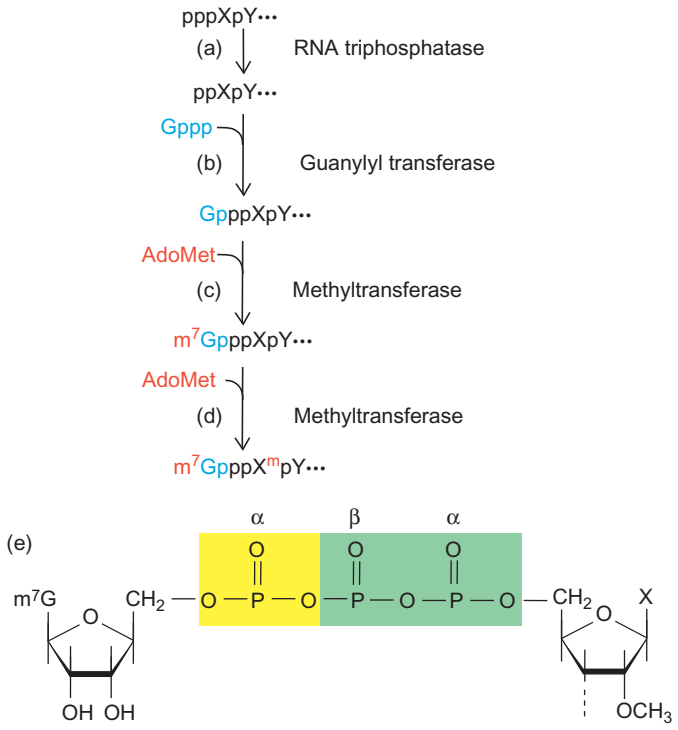

Figure 15.4 Sequence of events in capping. **(a)** RNA triphosphatase cleaves the γ-phosphate from the 5′-end of the growing RNA. **(b)** Guanylyl transferase adds the GMP part of GTP (blue) to form a triphosphate linkage, blocking the 5′-end of the RNA. **(c)** A methyltransferase adds a methyl group (red) from AdoMet to the N^7 of the blocking guanine. **(d)** Another methyltransferase adds a methyl group (red) from AdoMet to the 2′-hydroxyl group of the penultimate nucleotide. The product is cap 1. To form a cap 2, the next nucleotide (Y) would be methylated in a repeat of step (d). **(e)** The origin of the phosphates in the triphosphate linkage. The α- and β-phosphates from the initiating nucleotide (XTP) are highlighted in green, and the α-phosphate from the capping GTP is highlighted in yellow.

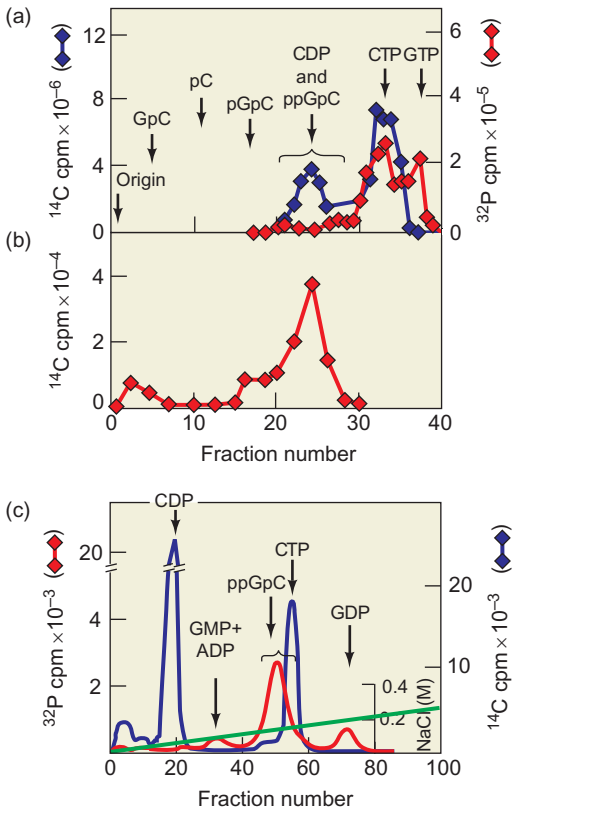

Figure 15.5 Identification of ppGpC as an intermediate in reovirus cap synthesis. **(a)** First purification step. Furuichi and colleagues added [^{14}C]CTP and [^{32}P]GTP to reovirus cores to label caps and capping intermediates. Then they analyzed the mixture by paper electrophoresis with the markers listed at top. One radioactive intermediate (bracket) coelectrophoresed with the ppGpC and CDP markers. **(b)** Conversion of ppGpC to GpC. Furuichi and colleagues treated the bracketed radioactive material from panel (a) with alkaline phosphatase, which should convert ppGpC to GpC, then electrophoresed the products. This time, a significant peak (though not the main peak) coelectrophoresed with the GpC marker. **(c)** Positive identification of ppGpC. Furuichi and colleagues subjected the bracketed material in (a) to ion-exchange chromatography on Dowex resin with the markers indicated at top. The major ^{32}P peak (red) coincided with the ppGpC marker. (*Source:* Adapted from Furuichi Y., S. Muthukrishnan, J. Tomasz, and A.J. Shatkin, Mechanism of formation of reovirus mRNA 5′-terminal blocked and methylated sequence m^7GpppG^mpC. *Journal of Biological Chemistry* 251:5051, 1976.)

To verify that this really is the correct pathway, the investigators isolated each of the enzymes we have listed and all of the intermediates. For example, Furuichi and colleagues started with the labeled model substrate pppGpC, which resembles the 5′-end of a newly initiated reovirus mRNA. How do we know that the virus cores can remove a terminal phosphate and convert this starting material to ppGpC? These workers blocked the guanylyl transferase reaction with an excess of by-product (PP_i), which should cause ppGpC to build up, if it exists. They looked directly for this intermediate by the scheme in Figure 15.5. First, they performed paper electrophoresis with markers and showed that a significant labeled product coelectrophoresed with the ppGpC marker. Unfortunately, CDP also electrophoresed to this position, so the product could not be clearly identified. Next, they treated the product with alkaline phosphatase to convert any ppGpC to GpC and reelectrophoresed it. Now a peak of radioactivity appeared in the GpC position. This was encouraging, but to positively identify ppGpC, these workers subjected the putative ppGpC peak from panel (a) to ion-exchange chromatography on a Dowex resin and obtained a radioactive peak that comigrated uniquely with the ppGpC marker. Thus, ppGpC is a real intermediate in the capping scheme. Relatively little ^{14}C radioactivity appeared in the ppGpC peak because of the lower radioactivity of the ^{14}C label.

When is the cap added? In some viruses, such as cytoplasmic polyhedrosis virus (CPV), lack of AdoMet completely inhibits transcription, suggesting that transcription depends on capping. This implies that capping in this virus is a very early event and presumably occurs soon after the first phosphodiester bond forms in the pre-mRNA. In other

viruses, such as vaccinia virus, transcription occurs normally in the absence of AdoMet, so transcription and capping may not be so tightly coupled in that virus.

Unlike CPV and vaccinia virus, adenovirus replicates in the nucleus and therefore presumably takes advantage of the host cell's capping system. Adenovirus should therefore tell us more about when capping of eukaryotic pre-mRNAs occurs. James Darnell and colleagues performed an experiment that showed that adenovirus capping occurs early in the transcription process. These workers measured the incorporation of [^{3}H]adenosine into the cap and the first dozen or so adenylate residues of the adenovirus major late transcripts (pre-mRNAs). First, they added [^{3}H]adenosine to label the cap (the bold A in m^{7}GpppA) and other adenosines in adenovirus pre-mRNAs during the late phase of infection. Then they separated large from small mRNA precursors by gradient centrifugation. Then they hybridized the small RNAs to a small restriction fragment that included the major late transcription start site. Any short RNAs that hybridized to this fragment were likely to be newly initiated RNAs, not just degradation products of mature RNAs. They eluted these nascent fragments from the hybrids and looked to see whether they were capped. Indeed they were, and no pppA, which would have been present on uncapped RNA, could be detected. This experiment demonstrated that caps are added to adenovirus major late pre-mRNA before the chain length reaches about 70 nt. It is now generally accepted that capping in eukaryotic cells occurs even earlier than that: before the pre-mRNA chain length reaches 30 nt.

SUMMARY Caps are made in steps: First, an RNA triphosphatase removes the terminal phosphate from a pre-mRNA; next, a guanylyl transferase adds the capping GMP (from GTP). Next, two methyltransferases methylate the N^7 of the capping guanosine and the 2′-O-methyl group of the penultimate nucleotide. These events occur early in the transcription process, before the chain length reaches 30 nt.

Functions of Caps

Caps appear to serve at least four functions. (1) They protect mRNAs from degradation. (2) They enhance the translatability of mRNAs. (3) They enhance the transport of mRNAs from the nucleus into the cytoplasm. (4) They enhance the efficiency of splicing of mRNAs. In this section we will discuss the first three of these functions, then deal with the fourth later in the chapter.

Protection The cap is joined to the rest of the mRNA through a triphosphate linkage found nowhere else in the RNA. The cap might therefore be expected to protect the mRNA from attack by RNases that begin at the 5′-end of their substrates and that cannot cleave triphosphate linkages. In fact, good evidence supports the notion that caps protect mRNAs from degradation.

Furuichi, Shatkin, and colleagues showed in 1977 that capped reovirus RNAs are much more stable than uncapped RNAs. They synthesized newly labeled reovirus RNA that was either capped with m^{7}GpppG, "blocked" with GpppG, or uncapped. Then they injected each of the three kinds of RNA into *Xenopus* oocytes, left them there for 8 h, then purified them and analyzed them by glycerol gradient ultracentrifugation. Reovirus RNAs exist in three size classes, termed large (l), medium (m), and small (s). Figure 15.6a shows a glycerol gradient ultracentrifugation separation of these three RNA classes. Furuichi and colleagues included RNAs with all three kinds of 5′-ends in this experiment, and no significant differences could be seen. All three size classes are clearly visible. Figure 15.6b shows what happened to these RNAs after 8 h in *Xenopus* oocytes. RNAs with all three kinds of 5′-ends had suffered degradation, but this degradation was much more pronounced for the uncapped RNAs. Thus, the *Xenopus* oocytes contain nucleases that

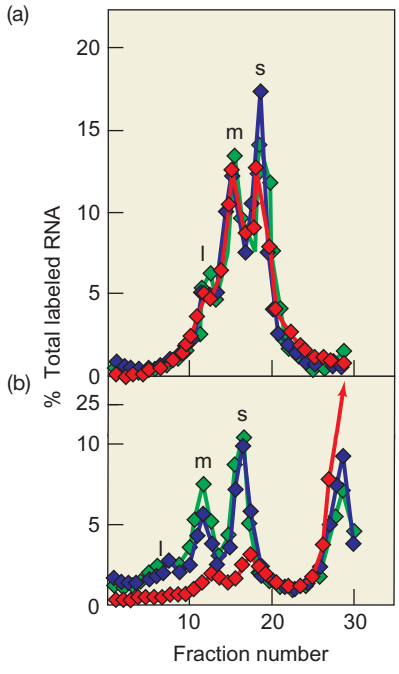

Figure 15.6 Effect of cap on reovirus RNA stability. (a) Appearance of newly synthesized RNAs. Furuichi and colleagues made labeled reovirus RNAs with capped (green), blocked (blue), or uncapped (red) 5′-ends, then subjected these RNAs to glycerol gradient ultracentrifugation. The three size classes of RNA are labeled l, m, and s. **(b)** Effect of incubation in *Xenopus* oocytes. Furuichi and colleagues injected the RNAs with the three different 5′-ends into *Xenopus* oocytes. After 8 h they purified the RNAs and performed the same sedimentation analysis as in panel (a). Colors have the same meaning as in panel (a). (*Source:* Adapted from Furuichi, Y., A. LaFiandra, and A.J. Shatkin, 5′-terminal structure and mRNA stability. *Nature* 266:236, 1977.)

Table 15.1 Synergism Between Poly(A) and Cap during Translation of Luciferase mRNA in Tobacco Protoplasts

mRNA	Luciferase mRNA Half-Life (min)	Luciferase Activity (light units/mg protein)	Relative Effect of Poly(A) on Activity	Relative Effect of Cap on Activity
Uncapped				
Poly(A)$^-$	31	2941	1	1
Poly(A)$^+$	44	4480	1.5	1
Capped				
Poly(A)$^-$	53	62,595	1	21
Poly(A)$^+$	100	1,331,917	21	297

Source: Gallie, D.R., The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency, *Genes & Development* 5:2108–2116, 1991. Copyright © Cold Spring Harbor, NY. Reprinted by permission.

degrade the viral RNAs, but the caps appear to provide some protection from these nucleases.

Translatability Another important function of the cap is to provide translatability. We will see in Chapter 17 that a eukaryotic mRNA gains access to the ribosome for translation via a cap-binding protein that recognizes the cap. If there is no cap, the cap-binding protein cannot bind and the mRNA is very poorly translated. Using an in vivo assay, Daniel Gallie documented the stimulatory effect of the cap on translation. In this procedure, Gallie introduced the firefly luciferase mRNA, with and without a cap, and with and without poly(A), into tobacco cells. Luciferase is an easy product to detect because of the light it generates in the presence of luciferin and ATP. Table 15.1 illustrates that the poly(A) at the 3′-end and the cap at the 5′-end act synergistically to stabilize and, especially, to enhance the translation of luciferase mRNA. Poly(A) provided a 21-fold boost in translation of a capped mRNA, but that was a minor effect compared with the 297-fold stimulation of translation that the cap conferred on a polyadenylated mRNA. Of course, mRNA stability also figured into these numbers, but its effect was not great.

Transport of RNA The cap also appears to facilitate the transport of at least some mature RNAs out of the nucleus. Jörg Hamm and Iain Mattaj studied the behavior of U1 snRNA to reach this conclusion. Most of the snRNA genes, including the U1 snRNA gene, are normally transcribed by RNA polymerase II, and the transcripts receive monomethylated (m^7G) caps in the nucleus. They migrate briefly to the cytoplasm, where they bind to proteins to form snRNPs, and their caps are modified to trimethylated ($m^{2,2,7}G$) structures. Then they reenter the nucleus, where they participate in splicing and other activities. The U6 snRNA is exceptional. It is made by polymerase III and is not capped. It retains its terminal triphosphate and remains in the nucleus. Hamm and Mattaj wondered what would happen if they arranged for the U1 snRNA gene to be transcribed by polymerase III instead of polymerase II. If it failed to be capped and remained in the nucleus, that would suggest that capping is important for transporting an RNA out of the nucleus.

Thus, Hamm and Mattaj placed the *Xenopus* U1 snRNA gene under the control of the human U6 snRNA promoter, so it would be transcribed by polymerase III. Then they injected this construct into *Xenopus* oocyte nuclei, along with a labeled nucleotide and a *Xenopus* 5S rRNA gene, which acted as an internal control. They also included 1 μg/mL of α-amanitin to inhibit RNA polymerase II and therefore ensure that no transcripts of the U1 gene would be made by polymerase II. In addition to the wild-type U1 gene, these workers also used several mutant U1 genes, with lesions in the regions coding for protein-binding sites. Loss of ability to associate with the proper proteins in the cytoplasm rendered the products of these mutant genes unable to return to the nucleus once they had been transported to the cytoplasm. Twelve hours after injection, Hamm and Mattaj dissected the oocytes into nuclear and cytoplasmic fractions and electrophoresed the labeled products in each. They compared the cellular locations of capped U1 snRNAs made by RNA polymerase II and uncapped U1 snRNA made by polymerase III.

Virtually all the uncapped U1 snRNA made by polymerase III remained in the nucleus. On the other hand, the U1 snRNAs made by polymerase II were transported to the cytoplasm. These results are consistent with the hypothesis that capping is required for U1 snRNA to be transported out of the nucleus.

Finally, as we will see later in this chapter, the cap is essential for proper splicing of a pre-mRNA.

SUMMARY The cap provides: (1) protection of the mRNA from degradation; (2) enhancement of the mRNA's translatability; (3) transport of at least some RNAs out of the nucleus; and (4) proper splicing of the pre-mRNA.

15.2 Polyadenylation

We have already seen that hnRNA is a precursor to mRNA. One finding that suggested such a relationship between these two types of RNA was that they shared a unique structure at their 3′-ends: a long chain of AMP residues called **poly(A).** Neither rRNA nor tRNA has a poly(A) tail. The process of adding poly(A) to RNA is called **polyadenylation.** Let us examine first the nature of poly(A) and then the polyadenylation process.

Poly(A)

James Darnell and his coworkers performed much of the early work on poly(A) and polyadenylation. To purify HeLa cell poly(A) from the rest of the mRNA molecule, Diana Sheiness and Darnell released it with two enzymes: RNase A, which cuts after the pyrimidine nucleotides C and U, and RNase T1, which cuts after G nucleotides. In other words, they cut the RNA after every nucleotide except the A's, preserving only pure runs of A's. Next, Sheiness and Darnell electrophoresed the poly(A)s from nuclei and from cytoplasm to determine their sizes. Figure 15.7 shows the results, which demonstrate that both poly(A)s have major peaks that electrophoresed more slowly than 5S rRNA, at about 7S. Sheiness and Darnell estimated that this corresponded to about 150–200 nt. The poly(A) species observed in this experiment were labeled for only 12 min, so they were newly synthesized. Little difference in size between these fresh nuclear and cytoplasmic poly(A)s is noticeable. However, cytoplasmic poly(A) is subject to shortening, as we will see later in this chapter. Now that poly(A)s from many different organisms have been analyzed, we see an average size of fresh poly(A) of about 250 nt.

It is apparent that the poly(A) goes on the 3′-end of the mRNA or hnRNA because it can be released very quickly with an enzyme that degrades RNAs from the 3′-end inward. Furthermore, complete RNase digestion of poly(A) yielded one molecule of adenosine and about 200 molecules of AMP. Figure 15.8 demonstrates that this requires poly(A) to be at the 3′-end of the molecule. This experiment also reinforced the conclusion that poly(A) is about 200 nt long.

We also know that poly(A) is not made by transcribing DNA because genomes contain no runs of T's long enough to encode it. In particular, we find no runs of T's at the ends of any of the thousands of eukaryotic genes that have been sequenced. Furthermore, actinomycin D, which inhibits DNA-directed transcription, does not inhibit polyadenylation. Thus, poly(A) must be added posttranscriptionally. In fact, there is an enzyme in nuclei called **poly(A) polymerase (PAP)** that adds AMP residues one at a time to mRNA precursors.

We know that poly(A) is added to mRNA precursors because it is found on hnRNA. Even specific unspliced mRNA precursors (the 15S mouse globin mRNA precursor, for example) contain poly(A). However, as we will see later in this chapter, splicing of some introns in a pre-mRNA can occur before polyadenylation. Once an mRNA enters the cytoplasm, its poly(A) turns over; in other words, it is constantly being broken down by RNases and rebuilt by a cytoplasmic poly(A) polymerase.

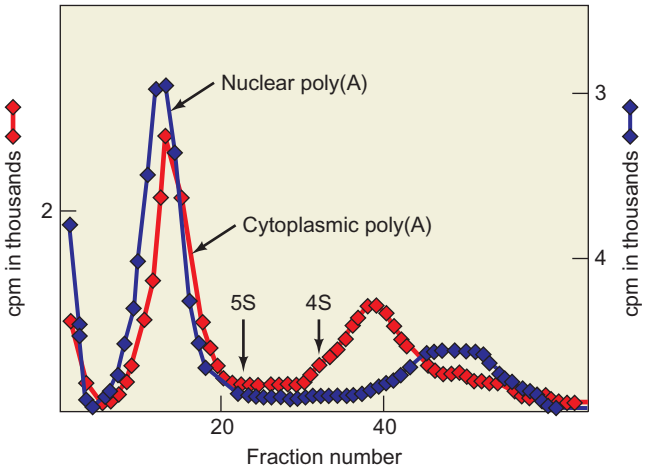

Figure 15.7 Size of poly(A). Sheiness and Darnell isolated radioactively labeled hnRNA from the nuclei (blue), and mRNA from the cytoplasm (red) of HeLa cells, then released poly(A) from these RNAs by RNase A and RNase T1 treatment. They electrophoresed the poly(A)s, collected fractions, and determined their radioactivities by scintillation counting (Chapter 5). They included 4S tRNA and 5S rRNA as size markers. Both poly(A)s electrophoresed more slowly than the 5S marker, corresponding to molecules about 200 nt long.
(*Source:* Adapted from Sheiness, D. and J.E. Darnell, Polyadenylic acid segment in mRNA becomes shorter with age. *Nature New Biology* 241:267, 1973.)

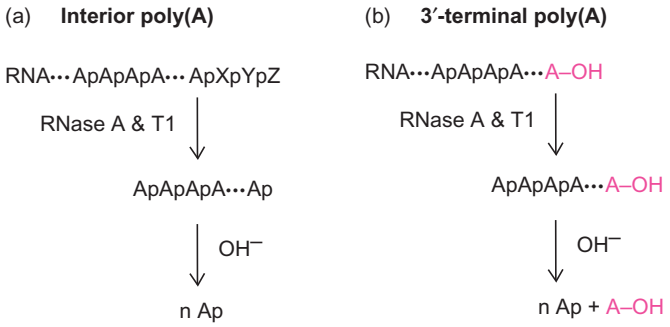

Figure 15.8 Finding poly(A) at the 3′-end of hnRNA and mRNA. (**a**) Interior poly(A). If poly(A) were located in the interior of an RNA molecule, RNase A and RNase T1 digestion would yield poly(A) with a phosphate at the 3′-end, then base hydrolysis would give only AMP. (**b**) Poly(A) at the 3′-end of hnRNA and mRNA. Because poly(A) is located at the 3′-end of these RNA molecules, RNase A and T1 digestion yields poly(A) with an unphosphorylated adenosine at the 3′-end. Base hydrolysis gives AMP plus one molecule of adenosine. In fact, the ratio of AMP to adenosine was 200, suggesting a poly(A) length of about 200 nt.

SUMMARY Most eukaryotic mRNAs and their precursors have a chain of AMP residues about 250 nt long at their 3′-ends. This poly(A) is added post-transcriptionally by poly(A) polymerase.

Functions of Poly(A)

Most mRNAs contain poly(A). One noteworthy exception is the histone mRNAs, which manage to perform their functions without detectable poly(A) tails. This exception notwithstanding, the near universality of poly(A) in eukaryotes raises the question: What is the purpose of poly(A)? One line of evidence suggests that it helps protect mRNAs from degradation. Another indicates that it stimulates translation of mRNAs to which it is attached. Still others show that poly(A) plays a role in splicing and transport of mRNA out of the nucleus. Here we will consider evidence for the effect of poly(A) on mRNA stability and translatability. We will return to the themes of splicing and transport at the end of this chapter.

Protection of mRNA To examine the stabilizing effect of poly(A), Michel Revel and colleagues injected globin mRNA, with and without poly(A) attached, into *Xenopus* oocytes and measured the rate of globin synthesis at various intervals over a 2-day period. They found that there was little difference at first. However, after only 6 h, the mRNA without poly(A) [**poly(A)$^-$ RNA**] could no longer support translation, while the mRNA with poly(A) [**poly(A)$^+$ RNA**] was still quite actively translated (Figure 15.9). The simplest explanation for this behavior is that the poly(A)$^+$ RNA has a longer lifetime than the poly(A)$^-$ RNA, and that poly(A) is therefore the protective agent. However, as we will see, other experiments have shown no protective effect of poly(A) on certain other mRNAs. Regardless, it is clear that poly(A) plays an even bigger role in efficiency of translation of mRNA.

Translatability of mRNA Several lines of evidence indicate that poly(A) also enhances the translatability of an mRNA. One of the proteins that binds to a eukaryotic mRNA during translation is **poly(A)-binding protein I, (PAB I).** Binding to this protein seems to boost the efficiency with which an mRNA is translated. One line of evidence in favor of this hypothesis is that excess poly(A) added to an in vitro reaction inhibited translation of a capped, polyadenylated mRNA. This finding suggested that the excess poly(A) was competing with the poly(A) on the mRNA for an essential factor, presumably for PAB I. Without this factor, the mRNA could not be translated well. Carrying this argument one step further leads to the conclusion that poly(A)$^-$ RNA, because it cannot bind PAB I, cannot be translated efficiently.

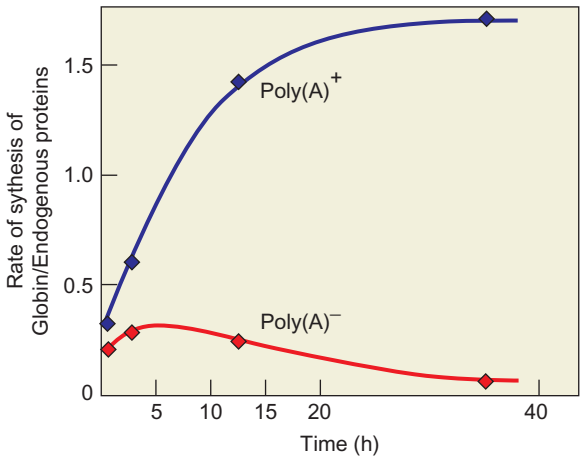

Figure 15.9 Time course of translation of poly(A)$^+$ (blue) and poly(A)$^-$ (red) globin mRNA. Revel and colleagues plotted the ratio of radioactivity incorporated into globin and endogenous protein versus the midpoint of the labeling time. (*Source:* Adapted from Huez, G., G. Marbaix, E. Hubert, M. Leclereq, U. Nudel, H. Soreq, R. Solomon, B. Lebleu, M. Revel, and U.Z. Littauer, Role of the polyadenylate segment in the translation of globin messenger RNA in *Xenopus* oocytes. *Proceedings of the National Academy of Sciences USA* 71(8):3143–3146, August 1974.)

To test the hypothesis that poly(A)$^-$ RNA is not translated efficiently, David Munroe and Allan Jacobson compared the rates of translation of two synthetic mRNAs, with and without poly(A), in rabbit reticulocyte extracts. They made the mRNAs (rabbit β-globin [RBG] mRNA and vesicular stomatitis virus N gene [VSV.N] mRNA) by cloning their respective genes into plasmids under the control of the phage SP6 promoter, then transcribing these genes in vitro with SP6 RNA polymerase. They endowed the synthetic mRNAs with various length poly(A) tails by adding poly(T) to their respective genes with terminal transferase and dTTP for varying lengths of time before cloning and transcription.

Munroe and Jacobson tested the poly(A)$^+$ and poly(A)$^-$ mRNAs for both translatability and stability in the reticulocyte extract. Figure 15.10 shows the effects of both capping and polyadenylation on translatability of the VSV.N mRNA. Both capped and uncapped mRNAs were translated better with poly(A) than without. Further experiments showed that polyadenylation made no difference in the stability of either mRNA. Munroe and Jacobson interpreted these results to mean that the extra translatability conferred by poly(A) was not due to stabilization of the mRNAs, but to enhanced translation per se. If so, what aspect of translation is enhanced by poly(A)? These studies suggested that it is a step at the very beginning of the translation process: association between mRNA and ribosomes. We will see in Chapter 17 that many ribosomes bind sequentially at the beginning of eukaryotic mRNAs and read the message in tandem. An mRNA with more than one ribosome translating it at once is called a polysome. Munroe

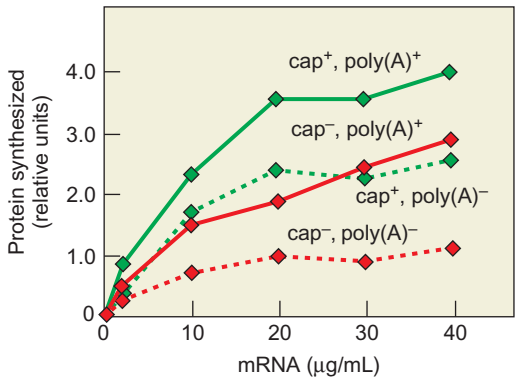

Figure 15.10 Effect of polyadenylation on translatability of mRNAs. Munroe and Jacobson incubated VSV.N mRNAs with [^{35}S]methionine in rabbit reticulocyte extracts. The mRNAs were capped (green) or uncapped (red), and poly(A)$^+$ (68 As; solid lines) or poly(A)$^-$ (dashed lines). After allowing 30 min for protein synthesis, these workers electrophoresed the labeled products and measured the radioactivity of the newly made protein by quantitative fluorography. Poly(A) enhanced the translatability of both capped and uncapped mRNAs. (*Source:* Adapted from Munroe, D. and A. Jacobson, mRNA poly(A) tail, a 3′ enhancer of a translational initiation. *Molecular and Cellular Biology* 10:3445, 1990.)

and Jacobson contended that poly(A)$^+$ mRNA forms polysomes more successfully than poly(A)$^-$ mRNA.

These workers measured the incorporation of labeled mRNAs into polysomes as follows: They labeled poly(A)$^+$ mRNA with ^{32}P and poly(A)$^-$ mRNA with ^{3}H, then incubated these RNAs together in a reticulocyte extract. Then they separated polysomes from monosomes by sucrose gradient ultracentrifugation. Figure 15.11a indicates that the poly(A)$^+$ VSV.N mRNA was significantly more associated with polysomes than was poly(A)$^-$ mRNA. In parallel experiments, the RBG mRNA exhibited the same behavior. Figure 15.11b shows the effect of *length* of poly(A) attached to RBG mRNA on the extent of polysome formation. We see the greatest increase as the poly(A) grows from 5 to 30 nt, and a more gradual increase as more A residues are added.

Munroe and Jacobson's finding that poly(A) did not affect the stability of mRNAs seems to contradict the earlier work by Revel and colleagues. Perhaps the discrepancy arises from the fact that the early work was done in intact frog eggs, whereas the later work used a cell-free system. Earlier in this chapter, Table 15.1 showed that poly(A) stimulated transcription of luciferase mRNA. The stabilizing effect of poly(A) on this mRNA was twofold at most, whereas the overall increase in luciferase production caused by poly(A) was up to 20-fold. Thus, this system also suggested that enhancement of translatability by poly(A) seems to be more important than mRNA stabilization.

In Chapter 17, we will see how poly(A) can both protect and stimulate the translation of an mRNA. Briefly, poly(A) can bind to cytoplasmic poly(A)-binding proteins. These in turn can bind to a translation initiation factor (eIF4G), which binds to a cap-binding protein, bound to the cap. In this way, the poly(A) at the 3′-end, and the cap at the 5′-end of the mRNA are brought together, effectively circularizing the mRNA. The mRNA in this closed loop

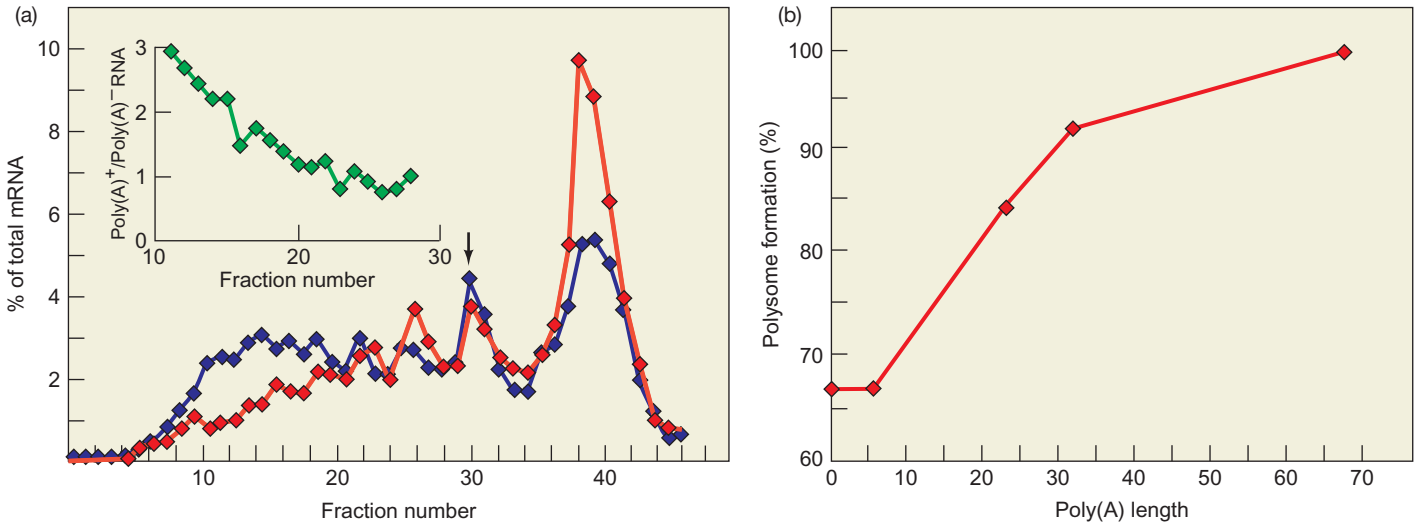

Figure 15.11 Effect of polyadenylation on recruitment of mRNA to polysomes. (a) Polysome profiles. Munroe and Jacobson mixed ^{32}P-labeled poly(A)$^+$ (blue) and ^{3}H-labeled poly(A)$^-$ (red) mRNA with a rabbit reticulocyte extract, then separated polysomes from monosomes by sucrose gradient ultracentrifugation. The arrow denotes the monosome peak; fractions to the left of this peak are polysomes, and one can see the disome, trisome, and even higher polysome peaks. The poly(A)$^+$ mRNA is clearly better at associating with polysomes, especially the higher polysomes. The inset shows the ratio of poly(A)$^+$ to poly(A)$^-$ RNA in fractions 11–28. Again, this demonstrates a preferential association of poly(A)$^+$ mRNA with polysomes (the lower fraction numbers). **(b)** Efficiency of polysome formation as a function of poly(A) length on VSV.N mRNA. The efficiency at a tail length of 68 is taken as 100%. (*Source:* Adapted from Munroe, D. and A. Jacobson, mRNA poly(A) tail, a 3′ enhancer of a translational initiation. *Molecular and Cellular Biology* 10:3447–8, 1990.)

form, with proteins binding to both its ends, is more stable than linear, naked mRNA would be. The mRNA is also more readily translated in this loop form, partly because the eIF4G, which ties the loop together, can help recruit the ribosomes to the mRNA.

SUMMARY Poly(A) enhances both the lifetime and translatability of mRNA. The relative importance of these two effects seems to vary from one system to another. At least in rabbit reticulocyte extracts, poly(A) seems to enhance translatability by helping to recruit mRNA to polysomes.

Basic Mechanism of Polyadenylation

It would be logical to assume that poly(A) polymerase simply waits for a transcript to be finished, then adds poly(A) to the 3′-end of the RNA. However, this is not what ordinarily happens. Instead, the mechanism of polyadenylation usually involves clipping an mRNA precursor, even before transcription has terminated, and then adding poly(A) to the newly exposed 3′-end (Figure 15.12). Thus, contrary to expectations, RNA polymerase can still be elongating an RNA chain, while the polyadenylation apparatus has already located a polyadenylation signal somewhere upstream, cut the growing RNA, and polyadenylated it.

Joseph Nevins and James Darnell provided some of the first evidence for this model of polyadenylation. They chose to study the adenovirus major late transcription unit because it serves as the template for several different overlapping mRNAs, each of which is polyadenylated at one of five separate sites. Recall from Chapter 14 that each of these mature mRNAs has the same three leader exons spliced to a different coding region. The poly(A) of each is attached to the 3′-end of the coding region. There are two alternative hypotheses for the relationship between transcription termination and polyadenylation in this system. (1) Transcription terminates immediately downstream of a polyadenylation site, and then polyadenylation occurs. For example, if gene A is being expressed, transcription will proceed only to the end of coding region A, then terminate, and then polyadenylation will occur at the 3′-end left by that termination event. (2) Transcription goes at least to the end of the last coding exon, and polyadenylation can occur at any polyadenylation site, presumably even before transcription of the whole major late region is complete.

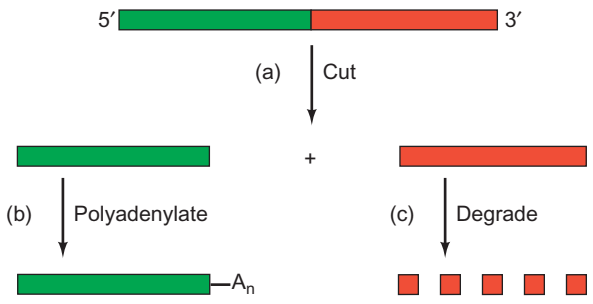

Figure 15.12 Overview of the polyadenylation process. (a) Cutting. The first step is cleaving the transcript, which may actually still be in the process of being made. The cut occurs at the end of the RNA region (green) that will be included in the mature mRNA. **(b)** Polyadenylation. The poly(A) polymerase adds poly(A) to the 3′-end of the mRNA. **(c)** Degradation of the extra RNA. All RNA (red) lying beyond the polyadenylation site is superfluous and is destroyed.

The first hypothesis, that transcription does not always go clear to the end, was easy to eliminate. Nevins and Darnell hybridized radioactive RNA made in cells late in infection to DNA fragments from various positions throughout the major late region. If primary transcripts of the first gene stopped after the first polyadenylation site, and only transcripts of the last gene made it all the way to the end, then much more RNA would hybridize to fragments near the 5′-end of the major late region than to fragments near the 3′-end. But RNA hybridized to all the fragments equally well—to fragments near the 3′-end of the region just as well as to fragments near the 5′-end. Therefore, once a transcript of the major late region is begun, it is elongated all the way to the end of the region before it terminates. In other words, the major late region contains only one transcription terminator, and it lies at the end of the region. Thus, this whole region can be called a **transcription unit** to denote the fact that it is transcribed as a whole, even though it contains multiple genes. Nevins and Darnell went on to show that clipping and polyadenylation usually occurred before transcription had terminated.

This behavior of transcribing far past a polyadenylation site before clipping and polyadenylating the transcript seems wasteful because all the RNA past the polyadenylation site will be destroyed without being used. So the question naturally arises: Is this method of polyadenylation unique to viruses, or does it also occur in ordinary cellular transcripts? To find out, Erhard Hofer and James Darnell isolated labeled RNA from Friend mouse erythroleukemia cells that had been induced with dimethyl sulfoxide (DMSO) to synthesize large quantities of globin, and therefore to transcribe the globin genes at a high rate. They hybridized the labeled transcripts to cloned fragments representing various parts of the mouse β-globin gene, and regions downstream of the gene (Figure 15.13). They observed just as much hybridization to fragments lying over 500 bp downstream of the polyadenylation site as to fragments within the globin gene. This demonstrated that transcription continues at least 500 bp downstream of this polyadenylation site. In further studies, these workers found that transcription finally terminated in regions lying even farther downstream. Thus, transcription significantly beyond the polyadenylation site occurs in cellular, as well as viral, transcripts.

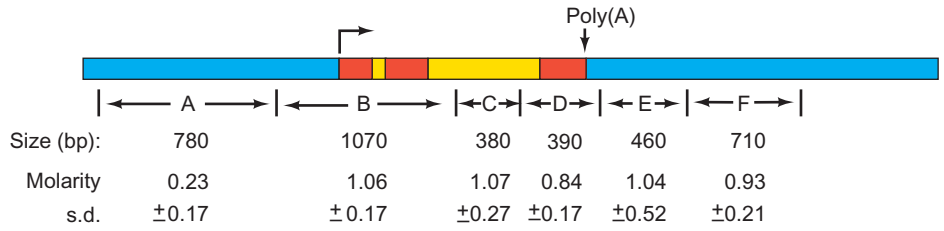

Figure 15.13 Transcription beyond the polyadenylation site. Hofer and Darnell isolated nuclei from DMSO-stimulated Friend erythroleukemia cells and incubated them with [^{32}P]UTP to label run-on RNA—mostly globin pre-mRNA. Then they hybridized this labeled RNA to DNA fragments A–F, whose locations and sizes are given in the diagram at top. The molarities of RNA hybridization to each fragment are given beneath each, with their standard deviations (s.d.). In the physical map at top, the exons are in red and the introns are in yellow. (*Source:* Adapted from E. Hofer and J.E. Darnell, The primary transcription unit of the mouse β-major globin gene. *Cell* 23:586, 1981.)

SUMMARY Transcription of eukaryotic genes extends beyond the polyadenylation site. Then the transcript is cleaved and polyadenylated at the 3′-end created by the cleavage.

Polyadenylation Signals

If the polyadenylation apparatus does not recognize the ends of transcripts, but binds somewhere in the middle to cleave and polyadenylate, what is it about a polyadenylation site that attracts this apparatus? The answer to this question depends on what kind of eukaryote or virus we are discussing. Let us first consider mammalian **polyadenylation signals.** By 1981, molecular biologists had examined the sequences of dozens of mammalian genes and had found that the most obvious common feature they had was the sequence AATAAA about 20 bp before the polyadenylation site. At the RNA level, the sequence AAUAAA occurs in most mammalian mRNAs about 20 nt upstream of their poly(A). Molly Fitzgerald and Thomas Shenk tested the importance of the **AAUAAA** sequence in two ways. First, they deleted nucleotides between this sequence and the polyadenylation site and sequenced the 3′-ends of the resulting RNAs. They found that the deletions simply shifted the polyadenylation site downstream by roughly the number of nucleotides deleted.

This result suggested that the AAUAAA sequence is at least part of a signal that causes polyadenylation approximately 20 nt downstream. If so, then deleting this sequence should abolish polyadenylation altogether. These workers used an S1 assay as follows to show that it did. They created a recombinant SV40 virus (mutant 1471) with duplicate polyadenylation signals 240 bp apart, at the end of the late region. S1 analysis of the 3′-ends of the late transcripts (Chapter 5) revealed two signals 240 bp apart (Figure 15.14). [We can ignore the poly(A) in this kind of experiment because it does not hybridize to the probe.] Thus, both polyadenylation sites worked, implying some readthrough of the first site. Then Fitzgerald and Shenk deleted either the first AATAAA (mutant 1474) or the second AATAAA (mutant 1475) and reran the S1 assay. This time, the polyadenylation site just downstream of the deleted AATAAA did not function, demonstrating that AAUAAA in the pre-mRNA is necessary for polyadenylation. We shall see shortly, however, that this is only part of the mammalian polyadenylation signal.

Is the AAUAAA invariant, or is some variation tolerated? Early experiments with manipulated signals (AAUACA, AAUUAA, AACAAA, and AAUGAA) suggested that no deviation from AAUAAA could occur without destroying polyadenylation. But by 1990, a compilation of polyadenylation signals from 269 vertebrate cDNAs showed some variation in these natural signals, especially in the second nucleotide. Marvin Wickens compiled these data, which defined a consensus sequence (Figure 15.15). The most common sequence, at the RNA level, is AAUAAA, and it is the most efficient in promoting polyadenylation. The most common variant is AUUAAA, and it is about 80% as efficient as AAUAAA. The other variants are much less common, and also much less efficient.

By now it has also become clear that AAUAAA by itself is not sufficient for polyadenylation. If it were, then polyadenylation would occur downstream of the many AAUAAA sequences found in introns, but it does not. Several investigators found that polyadenylation can be disrupted by deleting sequences immediately downstream of the polyadenylation site. This raised the suspicion that the region just downstream of the polyadenylation site contains another element of the polyadenylation signal. The problem was that that region is not highly conserved among vertebrates. Instead, there is simply a tendency for it to be GU- or U-rich.

These considerations suggested that the minimum efficient polyadenylation signal is the sequence AAUAAA followed about 20 bp later by a GU- or U-rich sequence. Anna Gil and Nicholas Proudfoot tested this hypothesis by examining the very efficient rabbit β-globin polyadenylation signal, which contains an AAUAAA, followed 24 bp later by a GU-rich region, immediately followed by a U-rich region. Throughout this discussion, we will refer to the sequences of the RNA (e.g., AAUAAA), even though the

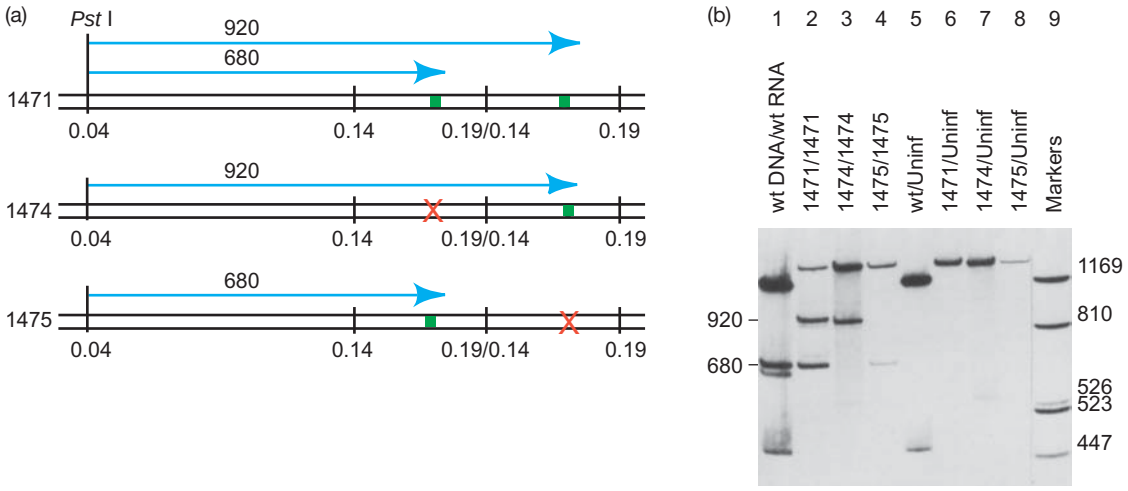

Figure 15.14 Importance of the AAUAAA sequence to polyadenylation. Fitzgerald and Shenk created recombinant SV40 viruses with the following characteristics **(a)** Mutant 1471 contained duplicate late polyadenylation sites (green) 240 bp apart within the duplicated region, which extends from 0.14 to 0.19 map units. Mutant 1474 contained a 16-bp deletion (red) at the AAUAAA in the upstream site, and mutant 1475 contained the same kind of 16-bp deletion (red) in the downstream site, resulting in the loss of the corresponding AAUAAA sequences in the pre-mRNAs produced by these mutant genes. Then they performed S1 analysis with a probe that should yield a 680-nt signal if the upstream polyadenylation signal works, and a 920-nt signal if the downstream polyadenylation signal works (blue arrows). **(b)** Experimental results. The lanes are marked at the top with the probe designation, followed by the RNA (or template) designation. Lane 1, using only wild-type probe and template, showed the wild-type signal at 680 nt, as well as an artifactual signal not usually seen. Lanes 5–8 are uninfected negative controls. The top band in each lane represents reannealed S1 probe and can be ignored. The results, also diagrammed in panel (a), show that deletion of an AAUAAA prevents polyadenylation at that site. (*Source:* Adapted from Fitzgerald, M. and T. Shenk, The sequence 5′-AAUAAA-3′ forms part of the recognition site for polyadenylation of late SV40 mRNAs. *Cell* 24 (April 1981) p. 257, f. 7.)

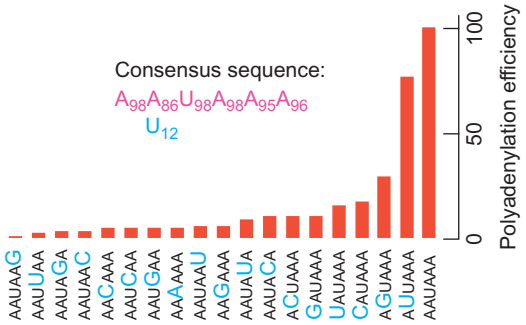

Figure 15.15 Summary of data on 369 vertebrate polyadenylation signals. The consensus sequence (in RNA form) appears at top, with the frequency of appearance of each base. The substitution of U for A in the second position is frequent enough (12%) that it is listed separately, below the main consensus sequence. Below, the polyadenylation efficiency is plotted for each variant polyadenylation signal. The base that deviates from normal is printed larger than the others in blue. The standard AAUAAA is given at the bottom, with the next most frequent (and active) variant (AUUAAA) just above it. (*Source:* Adapted from Wickens, M., How the messenger got its tail: addition of poly(A) in the nucleus. *Trends in Biochemical Sciences,* 15:278, 1990.)

mutations were of course made in the DNA. They began by inserting an extra copy of the whole polyadenylation signal upstream of the natural one, then testing for polyadenylation at the two sites of this mutant clone (clone 3) by S1 analysis. This DNA supported polyadenylation at the new site at a rate 90% as high as at the original site. Thus, the inserted polyadenylation site is active. Next, they created a new mutant clone [clone 2(v)] by deleting a 35-bp fragment containing the GU- and U-rich region (GU/U) in the new (upstream) polyadenylation signal. This abolished polyadenylation at the new site, reaffirming that this 35-bp fragment is a vital part of the polyadenylation signal.

To define the minimum efficient polyadenylation site, these workers added back various sequences to clone 2(v) and tested for polyadenylation. They showed that neither the GU-rich nor the U-rich sequence by itself could reconstitute an efficient polyadenylation signal: Clone GT had the GU-rich region, but was only 30% as active as the wild-type signal; clone A–T had the U-rich region, but had only 30% of the normal activity. Furthermore, the position of the GU/U region was important. In clone C–GT/T it was shifted 16 bp further downstream of the AAUAAA element, and this clone had less than 10% of normal activity. Moreover, the spacing between the GU-rich and U-rich sequences was important. Clone GT–T had both, but they were separated by an extra 5 bp, and this mutant signal had only 30% of the normal activity. Thus, an efficient polyadenylation signal has an AAUAAA motif followed 23–24 bp later by a GU-rich motif, followed immediately by a U-rich motif.

Plants and yeast mRNAs are also polyadenylated, but their polyadenylation signals are different from those of mammals. Yeast genes usually lack an AAUAAA sequence near their polyadenylation sites. In fact, it is difficult to discern a pattern in the yeast polyadenylation signals, other

than a general AU-richness upstream of the polyadenylation site. Plant genes may have an AAUAAA in the appropriate position, and deletion of this sequence prevents polyadenylation. But plant and animal polyadenylation signals are not the same: Single-base substitutions within the AAUAAA of the cauliflower mosaic virus do not have near the negative effect they have in vertebrate polyadenylation signals. Furthermore, animal signals do not function when placed at the ends of plant genes in plant cells.

SUMMARY An efficient mammalian polyadenylation signal consists of an AAUAAA motif about 20 nt upstream of a polyadenylation site in a pre-mRNA, followed 23 or 24 bp later by a GU-rich motif, followed immediately by a U-rich motif. Many variations on this theme occur in nature, which results in variations in efficiency of polyadenylation. Plant polyadenylation signals also usually contain an AAUAAA motif, but more variation is allowed in this region than in an animal AAUAAA. Yeast polyadenylation signals differ even more, and rarely contain an AAUAAA motif.

Cleavage and Polyadenylation of a Pre-mRNA

The process commonly known as polyadenylation really involves both RNA cleavage and polyadenylation. In this section we will briefly discuss the factors involved in the cleavage reaction, then discuss the polyadenylation reaction in more detail.

Pre-mRNA Cleavage Several proteins are necessary for cleavage of mammalian pre-mRNAs prior to polyadenylation. One of these proteins is also required for polyadenylation, so it was initially called "cleavage and polyadenylation factor," or "CPF," but it is now known as **cleavage and polyadenylation specificity factor (CPSF).** Cross-linking experiments have demonstrated that this protein binds to the AAUAAA signal. Shenk and colleagues reported in 1994 that another factor participates in recognizing the polyadenylation site. This is the **cleavage stimulation factor (CstF),** which, according to cross-linking data, binds to the G/U-rich region. Thus, CPSF and CstF bind to sites flanking the cleavage and polyadenylation site. Binding of either CPSF or CstF alone is unstable, but together the two factors bind cooperatively and stably.

Still another pair of RNA-binding proteins required for cleavage are the **cleavage factors I** and **II** (**CF I** and **CF II**). It is also likely that poly(A) polymerase itself is required for cleavage because cleavage is followed immediately by polyadenylation. In fact, the coupling between cleavage and polyadenylation is so strong that no cleaved, unpolyadenylated RNAs can be detected.

Another protein that is intimately involved in cleavage is RNA polymerase II. The first hint of this involvement was the discovery that RNAs made in vitro by RNA polymerase II were capped, spliced, and polyadenylated properly, but those made by polymerases I and III were not. In fact, even RNAs made by RNA polymerase II lacking the carboxyl-terminal domain (CTD) of the largest subunit were not efficiently spliced and polyadenylated. These data suggested that the CTD was involved somehow in splicing and polyadenylation.

In light of these data, Yutaka Hirose and James Manley performed experiments to test the role of the CTD, including its phosphorylation status, in polyadenylation. In 1998 they reported that the CTD stimulates the cleavage reaction, and this stimulation is not dependent on transcription. First, these workers tested the phosphorylated and unphosphorylated forms of polymerase II (IIO and IIA, Chapter 10) for ability to stimulate cleavage in the presence of all the other cleavage and polyadenylation factors. They incubated ^{32}P-labeled adenovirus L3 pre-mRNA with CPSF, CstF, CF I, CF II, poly(A) polymerase, and either RNA polymerase IIA or IIO. After the incubation period, they electrophoresed the products and autoradiographed the gel to see if the pre-mRNA had been cleaved in the right place. Figure 15.16 depicts the results. Both polymerases IIA and IIO stimulated correct cleavage

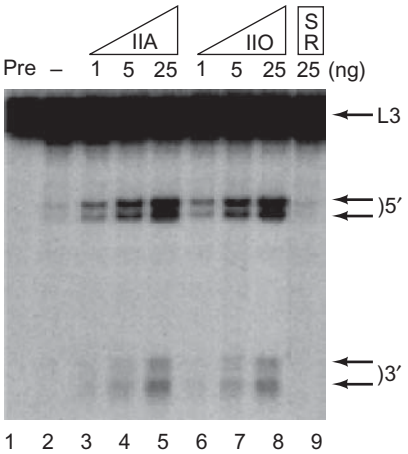

Figure 15.16 Effect of RNA polymerases IIA and IIO on prepolyadenylation mRNA cleavage in vitro. Hirose and Manley prepared a ^{32}P-labeled adenovirus L3 pre-mRNA and incubated it with all the cleavage and polyadenylation factors [CPSF, CstF, CF I, CF II, and poly(A) polymerase] plus polymerase IIA, IIO, no protein (–), or purified HeLa cell SR proteins, as indicated at top. (The amounts of the various proteins are given in nanograms.) Then the investigators electrophoresed the RNA products and detected them by autoradiography. The positions of the 5′- and 3′-cleavage fragments, and the pre-mRNA are indicated at right. Lane 1 contained precursor alone. Both IIA and IIO stimulated cleavage of the pre-mRNA to the appropriate 5′- and 3′-fragments. (*Source:* Hirose, Y. and Manley, J. RNA polymerase II is an essential mRNA polyadenylation factor. *Nature* 395 (3 Sep 1998) f. 2, p. 94. Copyright © Macmillan Magazines Ltd.)

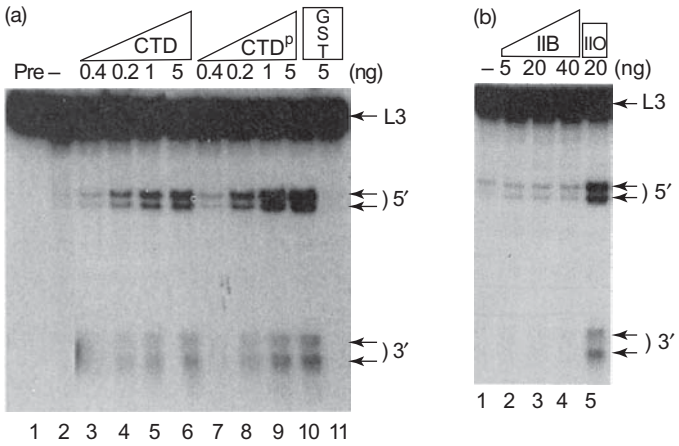

Figure 15.17 Effect of the Rpb1 CTD on prepolyadenylation mRNA cleavage in vitro. Hirose and Manley incubated a labeled pre-mRNA with cleavage and polyadenylation factors and assayed for cleavage as in Figure 15.16. **(a)** They included phosphorylated or unphosphorylated GST–CTD fusion proteins or GST alone, as indicated at top, in the cleavage reaction. **(b)** They included RNA polymerase IIB or IIO, as indicated at top, in the cleavage reaction. The phosphorylated CTD stimulated cleavage more than the unphosphorylated CTD; polymerase IIB, which lacks the CTD, did not stimulate cleavage at all. (*Source:* Hirose, Y. and Manley, J. RNA polymerase II is an essential mRNA polyadenylation factor. *Nature* 395 (3 Sep 1998) f. 3, p. 94. Copyright © Macmillan Magazines Ltd.)

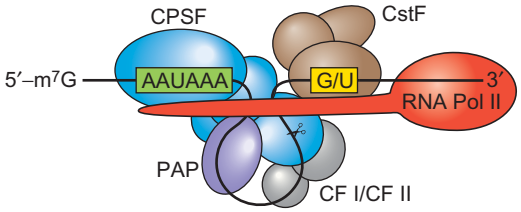

Figure 15.18 A model for the precleavage complex. This partly hypothetical model shows the apparent positions of all the proteins presumed to be involved in cleavage, with respect to the two parts of the polyadenylation signal (green and yellow). The scissors symbol denotes the active site of CPSF-73. (*Source:* Adapted from Wahle, E. and W. Keller, The biochemistry of polyadenylation, *Trends in Biochemical Sciences* 21 [1996] pp. 247–250, 1996.)

of the pre-mRNA, yielding 5′- and 3′-fragments of the expected sizes.

To verify that the CTD is the important part of polymerase II in stimulating cleavage, Hirose and Manley expressed the CTD as a fusion protein with glutathione-*S*-transferase (Chapter 4), then purified the fusion protein by glutathione affinity chromatography. They phosphorylated part of the fusion protein preparation on its CTD component and tested the phosphorylated and unphosphorylated fusion proteins in the cleavage assay with the adenovirus L3 pre-mRNA. Figure 15.17a shows that both the phosphorylated and unphosphorylated CTDs stimulated cleavage, but the phosphorylated form worked about five times better than the unphosphorylated one. That makes sense because the CTD is phosphorylated in polymerase IIO, which is the form that carries out transcription. It is unclear why phosphorylation made no difference when whole polymerase II was used in Figure 15.16.

If the CTD is the key to stimulating cleavage of the pre-mRNA, then polymerase IIB, the proteolytic product of polymerase IIA that lacks the CTD, should not stimulate, and Figure 15.17b shows that it does not. Thus, RNA polymerase II, and the CTD in particular, appears to be required for efficient cleavage of a pre-mRNA prior to polyadenylation. Figure 15.18 summarizes our knowledge about the complex of proteins that assembles on a pre-mRNA just before cleavage.

We have seen that an array of multisubunit complexes are required for cleavage at the polyadenylation site, but what protein carries out the cleavage itself? That question remained open until 2003, when Masayuki Nashimoto and colleagues discovered that one of the subunits of CPSF (**CPSF-73**) is related to the enzyme (ELAC2) that cleaves pre-tRNAs to generate their 3′-ends (Chapter 16). This finding led to the suggestion that CPSF-73 is the cleavage enzyme. This is an attractive notion because of the symmetry between ELAC2, which cleaves off the 3′-ends of pre-tRNAs prior to the untemplated addition of CCA, and CPSF-73, which cleaves off the 3′-ends of pre-mRNAs prior to the untemplated addition of poly(A). Both ELAC2 and CPSF-73 are unusual RNases that contain two zinc ions at their active sites. They belong to a family of hydrolases (enzymes that carry out hydrolytic reactions, such as hydrolyzing RNA phosphodiester bonds) known as the β-lactamase superfamily of zinc-dependent hydrolases.

Now James Manley and Liang Tong have provided strong evidence that CPSF-73 really is the enzyme that cleaves pre-mRNAs prior to polyadenylation. First, they obtained the crystal structure of human CPSF-73 (amino acids 1–460) in complex with a sulfate group, which mimics the scissile phosphodiester group (the one where the break will occur) in the pre-mRNA at the active site of the enzyme. They found that CPSF-73 contains a Zn-binding motif that coordinates two zinc ions that are essential for its RNase activity. These two zinc ions coordinate a hydroxide ion that is in perfect position to attack the scissile phosphodiester bond (represented by the sulfate) in the active site of the enzyme.

To demonstrate that CPSF-73 has endonuclease activity, Manley and Tong expressed the human CPSF-73 gene in bacteria and tested the product for the ability to cleave an SV40 late pre-mRNA. It did have weak endonuclease activity, producing a variety of cleavage products. By contrast, a mutant CPSF-73, which was missing two of the ligands for the zinc ions, was inactive. Although these data were not as clean as one might hope, taken together with the structural

studies on the enzyme, they strongly suggest that CPSF-73 is indeed the endonuclease that cleaves the pre-mRNA prior to polyadenylation.

SUMMARY Polyadenylation requires both cleavage of the pre-mRNA and polyadenylation at the cleavage site. Cleavage in mammals requires several proteins: CPSF, CstF, CF I, CF II, poly(A) polymerase, and RNA polymerase II (in particular, the CTD of Rpb1). One of the subunits of CPSF (CPSF-73) appears to cleave the pre-mRNA prior to polyadenylation.

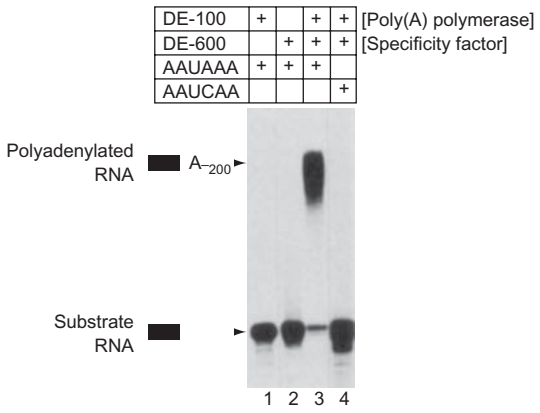

Figure 15.19 Separation of poly(A) polymerase and specificity factor activities. Wickens and colleagues separated HeLa cell poly(A) polymerase and specificity factor activities by DEAE-Sepharose chromatography. The polymerase eluted at 100 mM salt, so it is called the DE-100 fraction; the specificity factor eluted at 600 mM salt, so it is designated the DE-600 fraction. These workers tested the separated activities on a labeled synthetic substrate consisting of nucleotides −58 to +1 of SV40 late mRNA, whose 3′-end is at the normal polyadenylation site. After they incubated the two fractions, separately or together, with the substrate and ATP, they electrophoresed the labeled RNA and autoradiographed the gel. The components in the reactions in each lane are listed at top. The positions of substrate and polyadenylated product are listed at left. (*Source:* Bardwell, V.J., D. Zarkower, M. Edmonds, and M. Wickens, The enzyme that adds poly(A) to mRNAs is a classical poly(A) polymerase. *Molecular and Cellular Biology* 10 (Feb 1990) p. 847, f. 1. American Society for Microbiology.)

Initiation of Polyadenylation Once a pre-mRNA has been cleaved downstream of its AAUAAA motif, it is ready to be polyadenylated. The polyadenylation of a cleaved RNA occurs in two phases. The first, initiation, depends on the AAUAAA signal and involves slow addition of at least 10 A's to the pre-mRNA. The second phase, elongation, is independent of the AAUAAA motif, but depends on the oligo(A) added in the first phase. This second phase involves the rapid addition of 200 or more A's to the RNA. Let us begin with the initiation phase.

Strictly speaking, the entity we have been calling "the polyadenylation signal" is really the cleavage signal. It is what attracts the cleavage enzyme to cut the RNA about 20 nt downstream of the AAUAAA motif. Polyadenylation itself, that is, the addition of poly(A) to the 3′-end created by the cleavage enzyme, cannot use the same signal. This must be true because the cleavage enzyme has already removed the downstream part of the signal (the GU-rich and U-rich elements).

What is the signal that causes polyadenylation itself? It seems to be AAUAAA, followed by at least 8 nt at the end of the RNA. We know this because short synthetic oligonucleotides (as short as 11 nt) containing AAUAAA can be polyadenylated in vitro. The optimal length between the AAUAAA and the end of the RNA is 8 nt.

To study the process of polyadenylation by itself in vitro, it is necessary to divorce it from the cleavage reaction. Molecular biologists accomplish this by using labeled, short RNAs that have an AAUAAA sequence at least 8 nt from the 3′-end. These substrates mimic pre-mRNAs that have just been cleaved and are ready to be polyadenylated. The assay for polyadenylation is electrophoresis of the labeled RNA. If poly(A) has been added, the RNA will be much bigger and will therefore electrophorese much more slowly. It will also be less discrete in size, because the poly(A) tail varies somewhat in length from molecule to molecule. In this section, we will use the term *polyadenylation* to refer to the addition of poly(A) to the 3′-end of such a model RNA substrate.

Figure 15.19 shows how Marvin Wickens and his colleagues used this assay to demonstrate that two fractions are needed for polyadenylation: poly(A) polymerase and a specificity factor. We now know that this specificity factor is CPSF. At high substrate concentrations, the poly(A) polymerase can catalyze the addition of poly(A) to the 3′-end of any RNA, but at low substrate concentrations it cannot polyadenylate by itself (lane 1). Neither can CPSF, which recognizes the AAUAAA signal (lane 2). But together, these two substances can polyadenylate the synthetic substrate (lane 3). Lane 4 demonstrates that both fractions together will not polyadenylate a substrate with an aberrant signal (AAUCAA).

Michael Sheets and Wickens questioned whether polyadenylation is carried out in phases, and they used several different model RNA substrates to answer this question. The first substrate is simply the same terminal 58 nt of the SV40 late mRNA, including the AAUAAA, used in Figure 15.19. The second is the same RNA with 40 A's [a short poly(A)] at the 3′-end. The third is the same RNA with 40 nt from the vector instead of a short poly(A) at the 3′-end. They also used an analogous set of three substrates that had an AAGAAA signal instead of AAUAAA.

Sheets and Wickens used each of these substrates in standard polyadenylation reactions with HeLa cell nuclear extracts. Figure 15.20, lanes 1–4, shows that the extract could polyadenylate the usual model substrate with an AAUAAA signal. Lanes 5–8 show that polyadenylation also occurred with the model substrate that already had 40 A's at

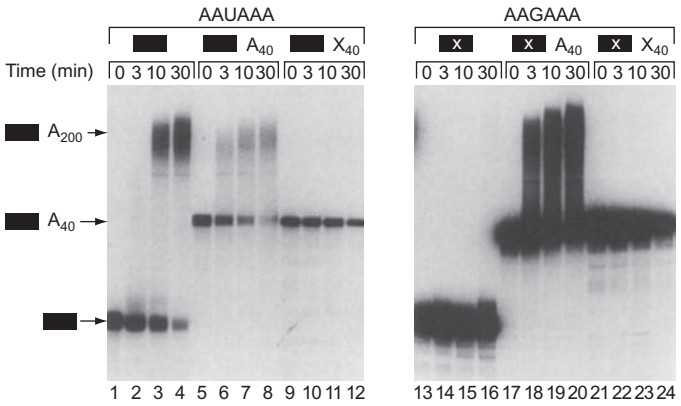

Figure 15.20 Demonstration of two phases in polyadenylation. Sheets and Wickens performed polyadenylation reactions in HeLa nuclear extracts with the following labeled substrates: 1. The standard 58-nt substrate containing the 3′-end of an SV40 late mRNA, represented by a black box; 2. The same RNA with a 40-nt poly(A), represented by a black box followed by A_{40}; 3. The same RNA with a 40-nt 3′-tag containing vector sequence, represented by a black box followed by X_{40}; substrates 1–3 containing an aberrant AAGAAA instead of AAUAAA are represented with white X's within the black boxes. Sheets and Wickens used four different reaction times with each substrate, and the substrate in each set of lanes is indicated by its symbol at top. The electrophoretic mobility of substrates and products are indicated at left. (*Source:* Sheets and Wickens, Two phases in the addition of a poly(A) tail. *Genes & Development* 3 (1989) p. 1402, f. 1. Cold Spring Harbor Laboratory Press.)

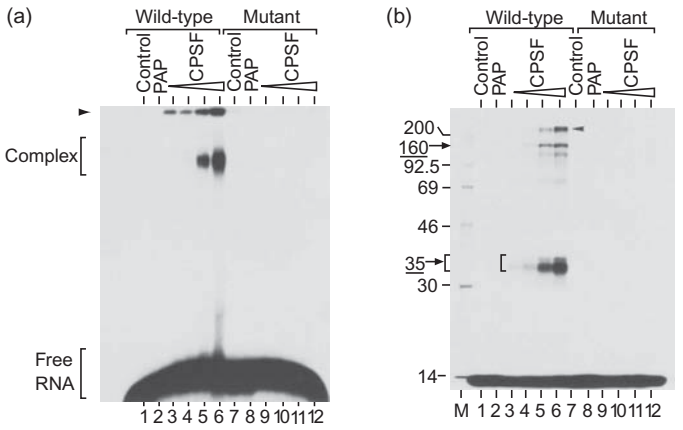

Figure 15.21 CPSF binds to the AAUAAA motif. (a) Gel mobility shift assay. Keller and colleagues mixed a labeled oligoribonucleotide with poly(A) polymerase (PAP), or CPSF in various concentrations, then electrophoresed the mixture. The wild-type oligo contained the AAUAAA motif, and the mutant oligo contained an AAGAAA motif. The controls contained no added proteins. CPSF could form a complex with the wild-type but not the mutant oligo. The band at the top in both panels (arrowheads) is material that remained at the top of the gel, rather than a specific band. **(b)** SDS-PAGE of proteins cross-linked to oligoribonucleotides. Keller and colleagues illuminated each of the mixtures from panel (a) with ultraviolet light to cross-link proteins to the oligo. Then they electrophoresed the complexes on an SDS polyacrylamide gel. Major bands appeared at about 35 and 160 kD (arrows). (*Source:* Keller, W., S. Bienroth, K.M. Lang, and G. Christofori, Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3′ processing signal AAUAAA. *EMBO Journal* 10 (1991) p. 4243, f. 2.)

its end (A_{40}). The polyadenylated signal was weaker in this case, but the radioactivity of the substrate was also lower. On the other hand, the extract could not polyadenylate the model substrate with 40 non-poly(A) nucleotides at its end (X_{40}). Lanes 13–16 demonstrate that the extract could not polyadenylate the substrate with an aberrant AAGAAA signal and no poly(A) pre-added. However, lanes 17–20 make the most telling point: The extract is able to polyadenylate the substrate with an aberrant AAGAAA signal and 40 A's already added to the end. Thus, by the time 40 A's have been added, polyadenylation is independent of the AAUAAA signal. But these extra nucleotides must be A's; the X_{40} substrate with an aberrant AAGAAA signal could not be polyadenylated (lanes 21–24).

Sheets and Wickens went on to show that the shortest poly(A) that could override the effect of a mutation in AAUAAA is 9 A's, but 10 A's work even better. These findings suggest the following hypothesis: After cleavage of the pre-mRNA, the first phase of polyadenylation, initiation, begins. It depends on the AAUAAA signal and CPSF until the poly(A) reaches about 10 A's in length. At that point, polyadenylation enters the elongation phase and is independent of the AAUAAA and CPSF, but dependent on the poly(A) at the 3′-end of the RNA.

If CPSF recognizes the poladenylation signal AAUAAA, we would predict that CPSF binds to this signal in the pre-mRNA. Walter Keller and colleagues have demonstrated this directly, using gel mobility shift and RNA–protein cross-linking procedures. Figure 15.21 illustrates the results of both kinds of experiments. Panel (a) shows that CPSF binds to a labeled RNA containing an AAUAAA signal, but not to the same RNA with a U→G mutation in the AAUAAA motif. Panel (b) demonstrates that an oligonucleotide bearing an AAUAAA motif, but not an AAGAAA motif, can be cross-linked to two polypeptides (about 35 and 160 kD) in a CPSF preparation. Furthermore, these complexes will not form in the presence of unlabeled competitor RNAs containing AAUAAA; competitor RNAs containing AAGAAA cannot compete. All of these findings bolster the conclusion that CPSF binds directly to the AAUAAA motif.

SUMMARY Short RNAs that mimic a newly created mRNA 3′-end can be polyadenylated. The optimal signal for initiation of such polyadenylation of a cleaved substrate is AAUAAA followed by at least 8 nt. Once the poly(A) reaches about 10 nt in length, further polyadenylation becomes independent of the AAUAAA signal and depends on the poly(A) itself. Two proteins participate in the initiation process: poly(A) polymerase and CPSF, which binds to the AAUAAA motif.

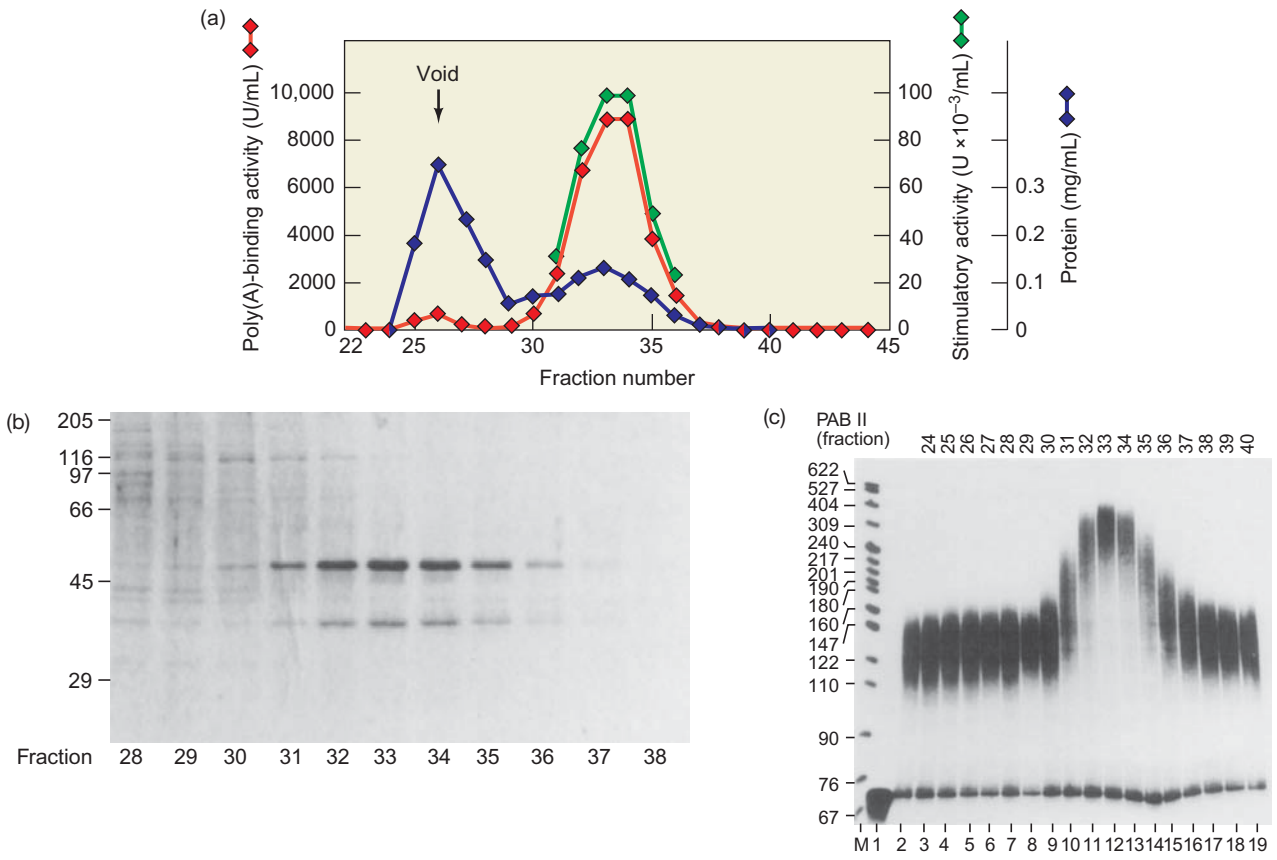

Figure 15.22 Purification of a poly(A)-binding protein. (a) Summary of results. Wahle subjected the poly(A)-binding protein to a final gel filtration chromatographic purification step on Sephadex G-100. In this panel, he plotted three parameters against fraction number from the G-100 column. Red, poly(A)-binding activity determined by a filter binding assay; green, polyadenylation-stimulating activity [see panel (c)]; blue, protein concentration. "Void" indicates proteins that eluted in the void volume. These large proteins were not included in the gel spaces on the column. **(b)** SDS-PAGE analysis. Wahle subjected aliquots of fractions from the G-100 column in panel (a) to SDS-PAGE and stained the proteins in the gel with Coomassie Blue. Sizes of marker polypeptides are given at left. A 49-kD polypeptide reached maximum concentration in the fractions (32–35) that had peak poly(A)-binding activity and polyadenylation-stimulatory activity. **(c)** Assay for polyadenylation stimulatory activity. Wahle added aliquots of each fraction from the G-100 column to standard polyadenylation reactions containing labeled L3pre RNA substrate. Lane 1 contained only substrate, with no poly(A) polymerase. The increase in size of poly(A) indicates stimulatory activity, which peaked in fractions 32–35. (*Source:* Wahle, E., A novel poly(A)-binding protein acts as a specificity factor in the second phase of messenger RNA polyadenylation. *Cell* 66 (23 Aug 1991) p. 761, f. 1. Reprinted by permission of Elsevier Science.)

Elongation of Poly(A) We have seen that elongation of an initiated poly(A) chain 10 nt or more in length is independent of CPSF. However, purified poly(A) polymerase binds to and elongates poly(A) only very poorly by itself. This implies that another specificity factor can recognize an initiated poly(A) and direct poly(A) polymerase to elongate it. Elmar Wahle has purified a poly(A)-binding protein that has these characteristics.

Figure 15.22b shows the results of PAGE on fractions from the last step in purification of the poly(A)-binding protein. A major 49-kD polypeptide is visible, as well as a minor polypeptide with a lower molecular mass. Because the latter band varied in abundance, and was even invisible in some preparations, Wahle concluded that it was not related to the poly(A)-binding protein. Wahle tested the fractions containing the 49-kD protein for poly(A) binding by a nitrocellulose filter binding assay [panel (a)], and found that the peak of poly(A)-binding activity coincided with the peak of abundance of the 49-kD polypeptide. Next, he tested the same fractions for ability to stimulate polyadenylation of a model RNA substrate in the presence of poly(A) polymerase and CPSF [panel (c)]. Again, he found that the peak of activity coincided with the abundance of the 49-kD polypeptide. Thus, the 49-kD polypeptide is a poly(A)-binding protein, but differs from the major, 70-kD poly(A)-binding protein, (**PAB I**) found earlier in the cytoplasm, so Wahle named it **poly(A)-binding protein II** (**PAB II**).

PAB II can stimulate polyadenylation of a model substrate, just as CPSF can, but it binds to poly(A) rather than to the AAUAAA motif. This suggests that PAB II is active in elongation, rather than initiation, of polyadenylation. If so,

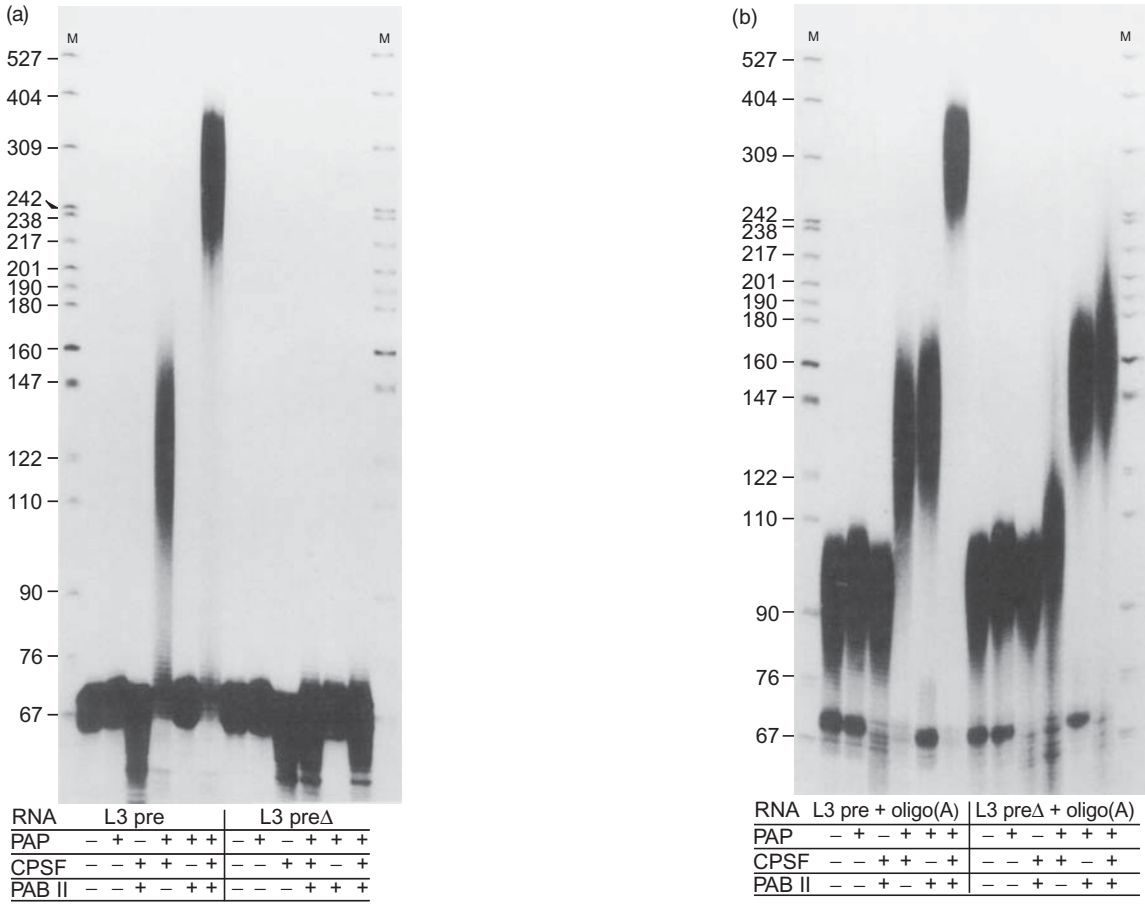

Figure 15.23 Effect of CPSF and PAB II on polyadenylation of model substrates. (a) Polyadenylation of RNAs lacking oligo(A). Wahle carried out polyadenylation reactions in the presence of the RNAs and proteins listed at bottom. L3 pre was the standard substrate RNA with an AAUAAA motif; L3 preΔ was the same, except that AAUAAA was mutated to AAGAAA. PAB II could not direct polyadenylation of L3 pre without help from CPSF. **(b)** Polyadenylation of RNAs containing oligo(A). All conditions were the same as in panel (a) except that the substrates contained oligo(A) at their 3′-ends. This allowed PAB II to work in the absence of CPSF and to work on the substrate with a mutant AAUAAA motif. The first and last lanes in both panels contained markers. (*Source:* Wahle, E., A novel poly(A)-binding protein acts as a specificity factor in the second phase of messenger RNA polyadenylation. *Cell* 66 (23 Aug 1991) p. 764, f. 5. Reprinted by permission of Elsevier Science.)

then its substrate preference should be different from that of CPSF. In particular, it should stimulate polyadenylation of RNAs that already have an oligo(A) attached, but not RNAs with no oligo(A). The results in Figure 15.23 confirm this prediction. Panel (a) shows that an RNA lacking oligo(A) (L3 pre) could be polyadenylated by poly(A) polymerase (PAP) plus CPSF, but not by PAP plus PAB II. However, PAP plus CPSF plus PAB II polyadenylated this substrate best of all. Presumably, CPSF serves as the initiation factor, then PAB II directs the polyadenylation of the substrate once an oligo(A) has been added, and does this better than CPSF can. Predictably, an L3 pre substrate with a mutant AAUAAA signal (AAGAAA) could not be polyadenylated by any combination of factors, because it depends on CPSF for initiation, and CPSF depends on an AAUAAA signal.

Figure 15.23b shows that the same RNA with an oligo(A) at the end behaved differently. It could be polyadenylated by PAP in conjunction with *either* CPSF or PAB II. This makes sense because this substrate has an oligo(A) that PAB II can recognize. It is interesting that both factors together produced even better polyadenylation of this substrate. This suggests that PAP might interact with both factors, directly or indirectly, during the elongation phase. Finally, panel (b) demonstrates that PAB II, in the absence of CPSF, could direct efficient polyadenylation of the mutant RNA with an AAGAAA motif, as long as the RNA had an oligo(A) to begin with. Again, this makes sense because the oligo(A) provides a recognition site for PAB II and therefore makes it independent of CPSF and the AAUAAA motif.

Figure 15.24 presents a model of initiation and elongation of polyadenylation. Optimal activity during the initiation phase requires PAP, CPSF, CstF, CF I, CF II and the two-part polyadenylation signal (the AAUAAA and G/U motifs flanking the polyadenylation site). The elongation phase requires PAP, PAB II, and an oligo(A) at least 10 nt long.

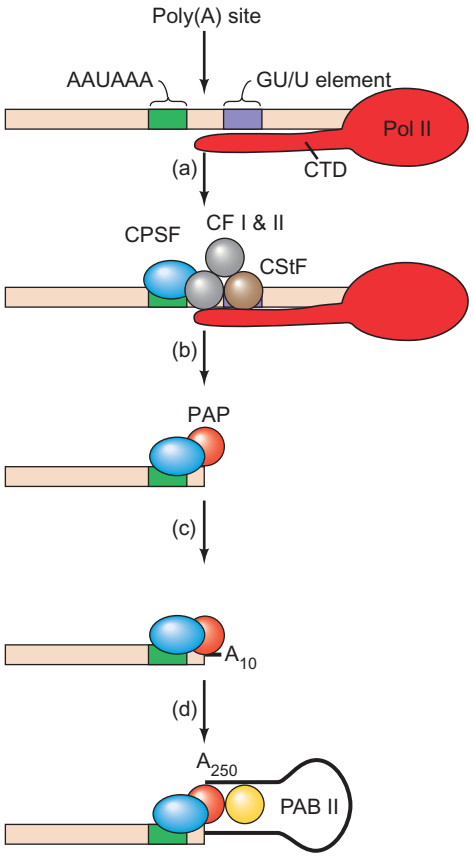

Figure 15.24 Model for polyadenylation. **(a)** CPSF (blue), CstF (brown), and CF I and II (gray) assemble on the pre-mRNA, guided by the AAUAAA and GU/U motifs. **(b)** Cleavage occurs, stimulated by the CTD of RNA polymerase II; CstF and CF I and II leave the complex; and poly(A) polymerase (PAP, purple) enters. **(c)** poly(A) polymerase, aided by CPSF, initiates poly(A) synthesis, yielding an oligo(A) at least 10 nt long. **(d)** PAB II (yellow) enters the complex and allows the rapid extension of the oligo(A) to a full-length poly(A). At this point, the complex presumably dissociates.

It is enhanced by CPSF. Table 15.2 lists all these protein factors, their structures, and their roles.

SUMMARY Elongation of poly(A) in mammals requires a specificity factor called poly(A)-binding protein II (PAB II). This protein binds to a preinitiated oligo(A) and aids poly(A) polymerase in elongating the poly(A) to 250 nt or more. PAB II acts independently of the AAUAAA motif. It depends only on poly(A), but its activity is enhanced by CPSF.

Poly(A) Polymerase

In 1991, James Manley and colleagues cloned cDNAs encoding bovine poly(A) polymerase (PAP). Sequencing of these clones revealed two different cDNAs that differed at their 3′-ends, apparently because of two alternative splicing schemes. This in turn should give rise to two different PAPs (*PAP I* and *PAP II*) that differ in their carboxyl termini. PAP II has several regions whose sequences match (more or less) the consensus sequences of known functional domains of other proteins. These are, in order from N-terminus to C-terminus: an RNA-binding domain (RBD); a polymerase module (PM); two nuclear localization signals (NLS 1 and 2); and several serine/threonine-rich regions (S/T). By 1996, four additional PAP cDNAs had been discovered. Two of these were short and could arise from polyadenylation within the pre-mRNA. Another was long and could come from a pseudogene (Chapter 23). The most important PAP in most tissues is probably PAP II.

Because the polymerase module, which presumably catalyzes the polyadenylation reaction, lies near the amino terminus of the protein, it would be interesting to know how much of the carboxyl end of the protein is required for activity. To examine the importance of the carboxyl end, Manley and colleagues expressed full-length and 3′-deleted versions of the PAP I cDNA by transcribing them in vitro with SP6 RNA polymerase, then translating these transcripts in cell-free reticulocyte extracts. This generated a full-length protein of 689 amino acids, and truncated proteins of 538, 379, and 308 amino acids. Then they tested each of these proteins for specific polyadenylation activity in the presence of calf thymus CPSF. The full-length and 538-amino-acid proteins had activity, but the smaller proteins did not. Thus, the S/T domain is not necessary for activity, but sequences extending at least 150 amino acids toward the carboxyl terminus from the polymerase module are essential, at least in vitro.

SUMMARY Cloning and sequencing cDNAs encoding calf thymus poly(A) polymerase reveal a mixture of 5 cDNAs derived from alternative splicing and alternative polyadenylation. The structures of the enzymes predicted from the longest of these sequences include an RNA-binding domain, a polymerase module, two nuclear localization signals, and a serine/threonine-rich region. The latter region, but none of the rest, is dispensable for activity in vitro.

Turnover of Poly(A)

Figure 15.7 showed some evidence of a slight difference in size between nuclear and cytoplasmic poly(A). However, that experiment involved newly labeled RNA, so the poly(A) had not had much time to break down. Sheiness and Darnell performed another study on RNA from cells that were continuously labeled with RNA precursors for 48 h. This procedure gave a population of poly(A)s at their "steady-state" sizes; that is, the natural sizes one would

Table 15.2 Mammalian Factors Required for 3′-Cleavage and Polyadenylation

Factor	Polypeptides (kD)	Properties
Poly(A) polymerase (PAP)	82	Required for cleavage and polyadenylation; catalyzes poly(A) synthesis
Cleavage and polyadenylation specificity factor (CPSF)	160 100 73 30	Required for cleavage and polyadenylation; binds AAUAAA and interacts with PAP and CstF; CPSF-73 cleaves RNA
Cleavage stimulation factor (CstF)	77 64 50	Required only for cleavage; binds the downstream element and interacts with CPSF
Cleavage factor I (CF I)	68 59 25	Required only for cleavage; binds RNA
Cleavage factor II (CF II)	Unknown	Required only for cleavage
RNA polymerase II (especially CTD)	Many	Required only for cleavage
Poly(A)-binding protein II (PAB II)	49	Stimulates poly(A) elongation; binds growing poly(A) tail; essential for poly(A) tail length control

Source: Adapted from Wahle, E. and W. Keller, The biochemistry of polyadenylation, *Trends in Biochemical Sciences* 21: 247–250. Copyright © 1996 with permission of Elseiver Science.

observe by peeking into a cell at any given time. Figure 15.25 shows an apparent difference in the sizes of nuclear and cytoplasmic poly(A)s. The major peak of nuclear poly(A) was 210 ± 20 nt, whereas the major peak of cytoplasmic poly(A) was 190 ± 20 nt. Furthermore, the cytoplasmic poly(A) peak showed a much broader skew toward smaller species than the nuclear poly(A) peak. This broad peak encompassed RNAs at least as small as 50 nt. Thus, poly(A) seems to undergo considerable shortening in the cytoplasm.

In 1970, Maurice Sussman proposed a "ticketing" hypothesis that held that each mRNA has a "ticket" that allows it entry to the ribosome for translation. Each time it is translated, the mRNA gets its "ticket punched." When it accumulates enough "punches," it can no longer be translated. Poly(A) would make an ideal ticket; the punches would then be progressive shortening of the poly(A) every time it is translated. To test this idea, Sheiness and Darnell tested the rate of shortening of poly(A) in the cytoplasm under normal conditions, and in the presence of emetine, which inhibits translation. They observed no difference in the size of cytoplasmic poly(A), whether or not translation was occurring. Thus, the shortening of poly(A) does not depend on translation, and the ticket, if it exists at all, seems not to be poly(A).

Poly(A) is not just shortened in the cytoplasm; it turns over. That is, it is constantly being shortened by RNases and lengthened by a cytoplasmic poly(A) polymerase. The general trend, however, is toward shortening, and ultimately an mRNA will lose all or almost all of its poly(A). By that time, its demise is near.

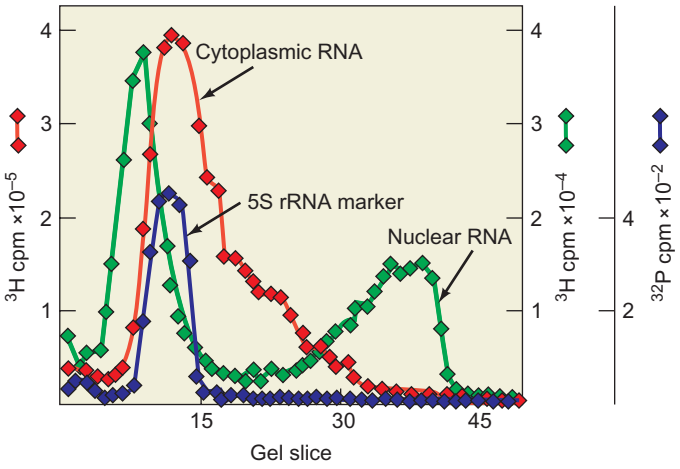

Figure 15.25 Shortening of cytoplasmic poly(A). Sheiness and Damell labeled HeLa cells with ^{3}H-adenine for 48 h, then isolated nuclear (green) and cytoplasmic (red) poly(A)$^+$ RNA and analyzed it by gel electrophoresis. They also included a [^{32}P]5S rRNA as a marker (blue). (*Source:* Adapted from Sheiness, D. and J.E. Darnell, Polyadenylic acid segment in mRNA becomes shorter with age. *Nature New Biology* 241:266, 1973.)

SUMMARY Poly(A) turns over in the cytoplasm. RNases tear it down, and poly(A) polymerase builds it back up. When the poly(A) is gone, the mRNA is slated for destruction.

Cytoplasmic Polyadenylation The best studied cases of cytoplasmic polyadenylation are those that occur during oocyte maturation. Maturation of *Xenopus* oocytes, for example, occurs in vitro on stimulation by progesterone. The immature oocyte cytoplasm contains a large store of mRNAs called **maternal messages,** or **maternal mRNAs,** many of which are almost fully deadenylated and are

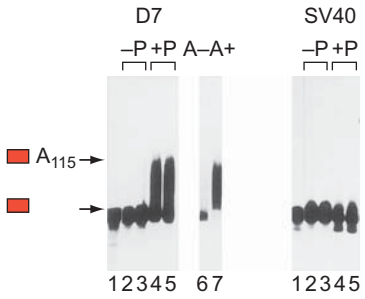

Figure 15.26 Maturation-specific polyadenylation of two RNAs. Wickens and colleagues injected labeled RNAs into *Xenopus* oocyte cytoplasm and stimulated maturation-specific polyadenylation with progesterone. After a 12-h incubation, they isolated the labeled RNA products, electrophoresed them, and visualized them by autoradiography. The two RNAs, as indicated at top, were synthetic 3′-fragments of either the *Xenopus* mRNA (D7), which normally undergoes maturation-specific polyadenylation, or an SV40 mRNA, which does not. The mobilities of unpolyadenylated RNA and RNA with a 115-nt poly(A) are indicated by the red boxes at left. The presence or absence of progesterone during the incubation is indicated at top by +P and −P, respectively. Lanes 6 and 7 contained RNA that was fractionated by oligo(dT)-cellulose chromatography. RNA that did not bind to the resin is designated A−, and RNA that did bind is designated A+. (*Source:* Fox et al., Poly(A) addition during maturation of frog oocytes: Distinct nuclear and cytoplasmic activities and regulation by the sequence UUUUUAU. *Genes & Development* 3 (1989) p. 2154, f. 3. Cold Spring Harbor Laboratory Press.)

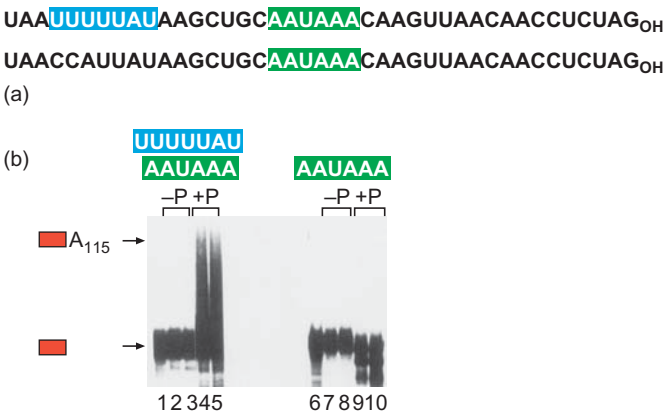

Figure 15.27 Demonstration that UUUUUAU confers maturation-specific polyadenylation. Wickens and colleagues performed the same experiment as described in Figure 15.26, using the same SV40 3′-mRNA fragment with and without an added UUUUUAU motif upstream of the AAUAAA motif. **(a)** Sequences of the two injected RNAs, with the UUUUUAU and AAUAAA motifs highlighted. **(b)** Results. Lanes 2–5 contained RNA from oocytes injected with the RNA having both a UUUUUAU and an AAUAAA sequence, as shown at top. Lanes 7–10 contained RNA from oocytes injected with the RNA having only an AAUAAA sequence. Presence or absence of progesterone during the incubation is indicated at top as in Figure 15.26. Lanes 1 and 6 had uninjected RNA. Markers at left as in Figure 15.26. The UUUUUAU motif was essential for polyadenylation. (*Source:* Fox et al., *Genes & Development* 3 (1989) p. 2155, f. 5. Cold Spring Harbor Laboratory Press.)

not translated. During maturation, some maternal mRNAs are polyadenylated, and others are deadenylated.

To find out what controls this maturation-specific cytoplasmic polyadenylation, Wickens and colleagues injected two mRNAs into *Xenopus* oocyte cytoplasm. The first was a synthetic 3′-fragment of D7 mRNA, a *Xenopus* mRNA known to undergo maturation-specific polyadenylation. The second was a synthetic 3′-fragment of an SV40 mRNA. As Figure 15.26 shows, the D7 RNA was polyadenylated, but the SV40 RNA was not. This implied that the D7 RNA contained a sequence or sequences that are required for maturation-specific polyadenylation, and that these are lacking in the SV40 RNA.

Wickens and colleagues noted that *Xenopus* RNAs that were known to undergo polyadenylation during oocyte maturation all contained the sequence UUUUUAU, or a close relative, upstream of the AAUAAA signal. Is this the key? To find out, these workers inserted this sequence upstream of the AAUAAA in the SV40 RNA and retested it. Figure 15.27 demonstrates that addition of this sequence caused polyadenylation of the SV40 RNA. In light of this character, the UUUUUAU sequence has been dubbed the **cytoplasmic polyadenylation element (CPE).**

Is the AAUAAA also required for cytoplasmic polyadenylation? To answer this question, Wickens and colleagues made point mutations in the AAUAAA motif and injected the mutated RNAs into oocyte cytoplasm. They found that alteration of AAUAAA to either AAUAUA or AAGAAA completely abolished polyadenylation. Thus, this motif is required for both nuclear and cytoplasmic polyadenylation.

SUMMARY Maturation-specific polyadenylation of *Xenopus* maternal mRNAs in the cytoplasm depends on two sequence motifs: the AAUAAA motif near the end of the mRNA and an upstream motif called the cytoplasmic polyadenylation element (CPE), which is UUUUUAU or a closely related sequence.

15.3 Coordination of mRNA Processing Events

Now that we have studied capping, polyadenylation, and splicing, we can appreciate that these processes are related. In particular, the cap can be essential for splicing, but only for splicing out the first intron. Similarly, the poly(A) can be essential for splicing out the last intron. Let us first consider the role of the CTD of the Rpb1 subunit of RNA polymerase II in coordinating capping, splicing, and polyadenylation. Then we will discuss the mechanism of termination of transcription of class II genes and its relationship to polyadenylation.

Binding of the CTD of Rpb1 to mRNA-Processing Proteins

In this chapter and in Chapter 14, we have seen evidence that all three of the mRNA-processing events—splicing, capping, and polyadenylation—take place during transcription. Capping occurs when the nascent mRNA is less than 30 nt long, when the 5′-end of the RNA first emerges from the polymerase. Polyadenylation occurs when the still-growing mRNA is cut at the polyadenylation site. And splicing at least begins when transcription is still underway. We have also just learned that capping and polyadenylation both stimulate splicing, at least of the first and last introns, respectively.

The unifying element for all these processing activities is the CTD of the Rpb1 subunit of RNA polymerase II. We have seen evidence in this chapter for the involvement of the CTD in polyadenylation, but it also plays a part in splicing and capping. In fact, direct evidence shows that the capping, polyadenylating, and splicing enzymes bind directly to the CTD, which provides a platform for all three activities.

For example, consider the evidence for interaction between the capping enzymes and the CTD, presented in 1997 by David Bentley and colleagues. They made affinity columns containing glutathione-*S*-transferase (GST) coupled to: wild-type CTD; wild-type phosphorylated CTD; mutant CTD; or just GST with no CTD attached. Then they subjected HeLa cell extracts to affinity chromatography on each of these columns and tested the eluates for guanylyl transferase activity. The guanylyl transferase assay was done by mixing an eluate with [^{32}P]GTP and observing the transfer of [^{32}P]GMP to form a covalent adduct with the enzyme. This labeled enzyme was then detected by SDS-PAGE and autoradiography. Figure 15.28 shows that the guanylyl transferase bound to the CTD, but only to its phosphorylated form.

Using a very similar experimental approach, Nick Proudfoot and colleagues demonstrated in 2001 that several subunits of the yeast cleavage/polyadenylation factor 1A (CF 1A) bind to the CTD in its phosphorylated form. Other components of the cleavage and polyadenylation complex appeared not to bind directly to the CTD, but they are tightly bound in the complex with other proteins that do bind to the CTD. Other, more indirect evidence also points to the association between the polyadenylation complex and the CTD: Polyadenylation does not function very well when RNA polymerase is lacking its CTD; and the CTD, particularly in its phosphorylated form, stimulates polyadenylation in vitro.

Strong evidence also exists for interactions between the CTD and proteins involved in splicing pre-mRNAs. For example, Daniel Morris and Arno Greenleaf showed in 2000 that a yeast splicing factor, Prp40 (a component of U1 snRNP) binds to the phosphorylated CTD. Morris and Greenleaf used a "**Far Western blot**" to demonstrate binding

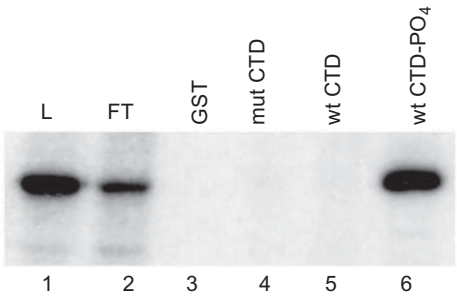

Figure 15.28 A mammalian capping guanylyl transferase binds to the phosphorylated CTD. Bentley and colleagues subjected HeLa cell nuclear extracts to affinity chromatography on resins containing the substances indicated at top, then tested the eluates for guanylyl transferase by observing the formation of a [^{32}P]GMP adduct with the enzyme, which could be identified by SDS-PAGE and autoradiography. L (lane 1) refers to the whole extract loaded onto the column; FT (lane 2) refers to the material that flowed through the column. Lanes 3–6 contain the results of guanylyl transferase assays on material subjected to affinity chromatography on resins containing GST (lane 3), and GST coupled to mutated CTD (lane 4); wild-type CTD (lane 5); and phosphorylated wild-type CTD (lane 6). The guanylyl transferase bound only to the phosphorylated CTD. (*Source:* McCracken et al., *Genes and Development* v. 11, p. 3310.)

between Prp40 and the CTD. A Far Western blot is similar to a Western blot in that it begins with electrophoresis of a protein or proteins by SDS-PAGE and blotting of the electrophoresed proteins to a membrane such as nitrocellulose. However, whereas a Western blot would be probed with an antibody, a Far Western blot is probed with another protein suspected of binding to a protein on the blot. In this case, Prp40 (and other so-called WW proteins) were electrophoresed and blotted, then probed with [^{32}P]β-galactosidase-CTD. (The CTD was expressed as a fusion protein with β-galactosidase, for ease of purification, then labeled by phosphorylation in vitro.) WW proteins are characterized by a domain including two tryptophan (W) residues and are frequently involved in RNA synthesis and processing.

Figure 15.29 shows the results of this analysis. Panel (a) depicts a gel stained with Coomassie Blue, a dye that binds to all proteins; so this panel shows the spectrum of polypeptides contained in all the protein preparations, including Prp40, loaded on the gel. The largest polypeptide in each lane is the parent; the smaller polypeptides are likely to be degradation products of the parent. Panel (b) depicts the same gel subjected to Far Western blotting and probed with [^{32}P]β-galactosidase-CTD. Clearly, Ess1, Prp40, and Rsp5 bind to the CTD. However, simply having a WW domain does not guarantee CTD-binding activity, as the other two WW proteins failed to bind the CTD probe.

SUMMARY Capping, polyadenylation, and splicing proteins all associate with the CTD during transcription.

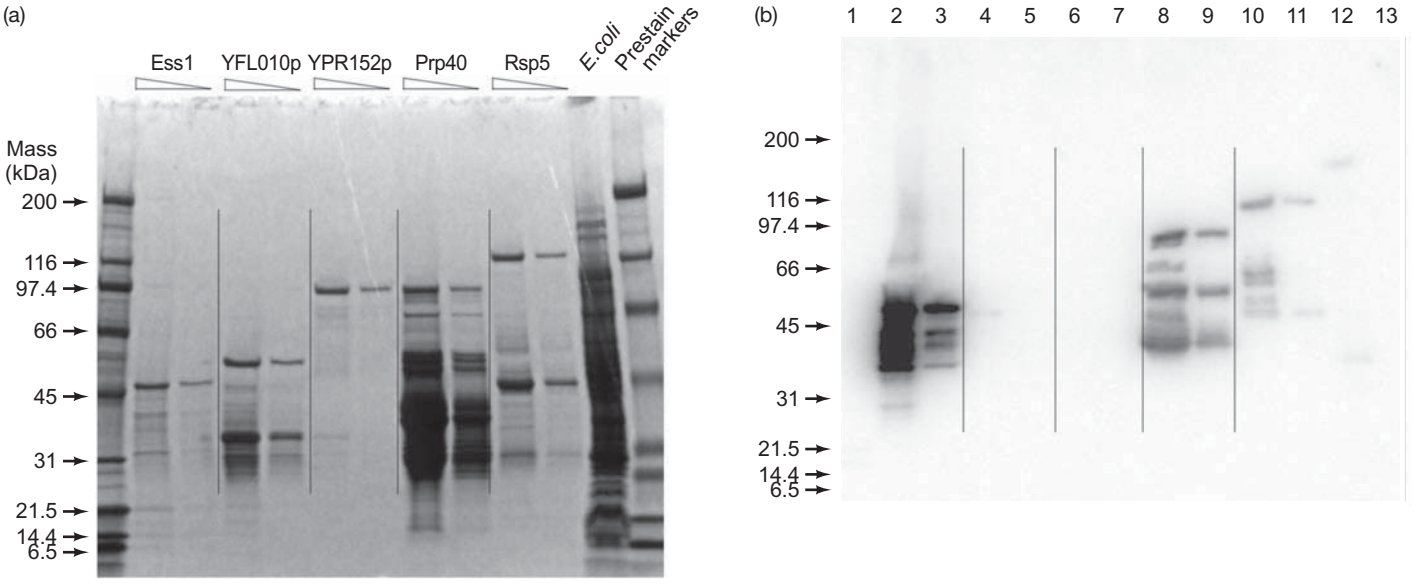

Figure 15.29 Interactions between Prp40 (and other proteins) and the CTD of Rpb1. (a) Gel electrophoresis. Morris and Greenleaf subjected five proteins known to have WW domains to SDS-PAGE and then stained the gel with Coomassie blue. The even-numbered lanes (2, 4, 6, 8, and 10) contained 500 ng of the proteins indicated at top, and the odd-numbered lanes (3, 5, 7, 9, and 11) contained 50 ng of the same proteins. The top band in each lane contains the whole, parent protein. Lanes 1 and 13 contained standard protein markers. Lane 12 contained *E. coli* proteins. **(b)** Far Western blot analysis. A gel electrophoresed in duplicate with the stained gel in panel (a) was blotted to a nitrocellulose membrane, and probed with [^{32}P] β-galactosidase-CTD, then subjected to phosphorimaging. (*Source: Journal of Biological Chemistry* by Morris and Greenleaf. Copyright 2000 by Am. Soc. For Biochemistry & Molecular Biol. Reproduced with permission of Am. Soc. For Biochemistry & Molecular Biol. in the format Textbook via Copyright Clearance Center.)

Changes in Association of RNA-Processing Proteins with the CTD Correlate with Changes in CTD Phosphorylation

The fact that all three classes of major mRNA-processing proteins bind to the CTD raises a question: We know that the CTD is long and could bind to many proteins at once, but does it associate simultaneously with all the proteins and RNAs involved in all three processing events?

The answer is that proteins come to and go from the CTD as they are needed for the task at hand. Moreover, these comings and goings are correlated with changes in CTD phosphorylation during transcription. Steven Buratowski and coworkers investigated the association of capping and polyadenylation enzymes with yeast polymerase II near the promoter (shortly after initiation) and remote from the promoter (during elongation, long after initiation). They also examined the state of phosphorylation of the CTD near promoters or remote from promoters.

They discovered that the capping enzyme (the guanylyl transferase) associates with the CTD near the promoter (shortly after initiation), but not in the interior of the gene. By contrast, the cap methyl transferase and the polyadenylation factor Hrp1/CFIB associate with the CTD both near and remote from the promoter. Thus, these factors are present on the transcription complex during both initiation and elongation. Moreover, these workers discovered that serine 5 of the CTD heptads is phosphorylated when the complex is near promoters, but not later during elongation, while serine 2 of the CTD heptads has a complementary pattern of phosphorylation: It is phosphorylated during elongation (remote from promoters) but not earlier, when the polymerase is still near the promoter.

To reach these conclusions, Buratowski and coworkers exploited the chromatin immunoprecipitation (ChIP) technique described in Chapter 5. They immunoprecipitated chromatin with antibodies against the capping and polyadenylation proteins to catch chromatin being transcribed by polymerase that is interacting with these proteins. Then they probed the precipitated chromatin by PCR with primers that would amplify DNA regions close to promoters or remote from promoters of several different genes.

What can we learn from such an assay? One possible outcome is the following: Chromatin immunoprecipitated with an antibody directed against a particular protein gives a strong PCR signal with primers that hybridize near a promoter, but only a weak signal with primers that hybridize to the interior of a gene. This would indicate that this protein is associated with the transcribing complex at or shortly after initiation of transcription, but not later during the elongation phase.

Figure 15.30 shows the results of the ChIP assay with antibodies against: the yeast capping enzyme guanylyl transferase (α-Ceg1); yeast polyadenylation factor (α-Hrp 1); and the Rpb3 subunit of yeast RNA polymerase II (α-HA-Rpb3).

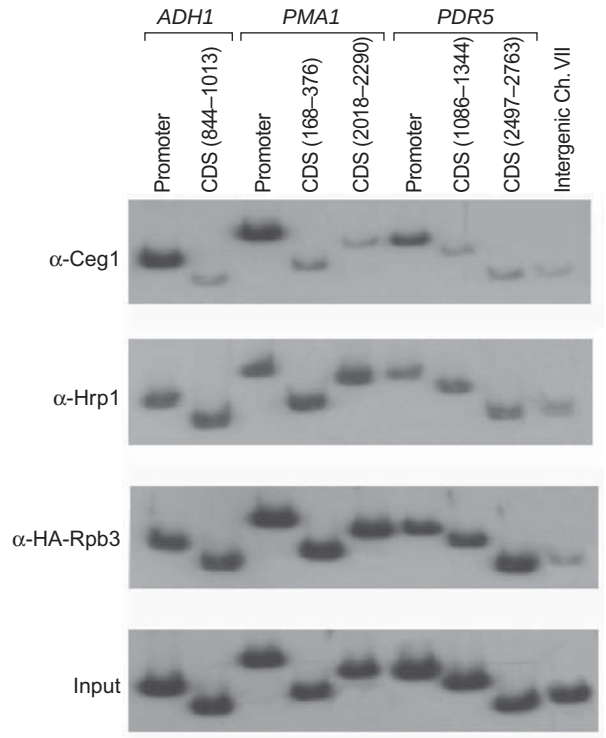

Figure 15.30 ChIP analysis of proteins associated with the transcription complex on three yeast genes. Buratowski and coworkers performed ChIP analysis of the association of three proteins (the capping guanylyl transferase, a polyadenylation factor, and the Rpb3 subunit of RNA polymerase II) with the transcription complex when it is near the promoter or remote from the promoter of three different genes (*ADH1, PMA1,* and *PDR5*). They used the following antibodies to immunoprecipitate chromatin: an antibody against the capping guanylyl transferase (α-Cegl); an antibody against a polyadenylation factor (α-Hrp1); and an antibody against the Rpb3 subunit of RNA polymerase II (α-HA-Rpb3). The antibodies used in each experiment are listed at left. Then they performed PCR on the precipitated chromatin with primers specific for promoter regions or coding sequences (CDS) of the three genes to determine whether the transcription complex was near the promoters of the genes or not. Strong signals, with abundant PCR product, indicate that the corresponding DNA, near or remote from the promoter, was present in the precipitated chromatin. The bottom panel contains PCR results on the input chromatin, showing that all areas of the genes were equally represented before immunoprecipitation. The last lane in each panel is a negative control, with the results of PCR with primers specific for an intergenic, untranscribed region of chromosome VII. This region was present in the input chromatin, but not immunoprecipitated by any of the antibodies. (*Source:* Reprinted by permission of S. Buratowski from "Komarnitsky, Cho, and Buratowski (2000) *Genes and Development* v. 14, pp. 2452–2460" © Cold Spring Harbor Laboratory Press.)

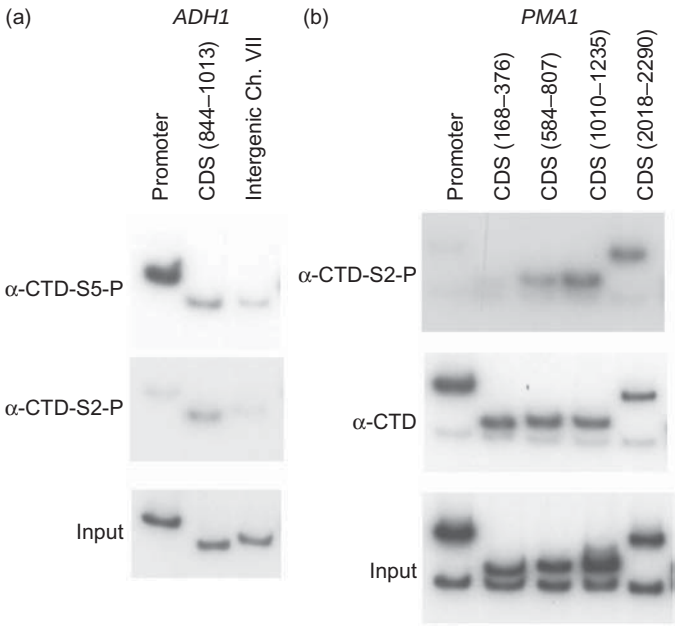

Figure 15.31 ChIP analysis of the phosphorylation state of the CTD of RNA polymerase II at various stages of transcription. Buratowski and coworkers performed ChIP analysis of the association of two phosphorylated forms of the CTD of the Rpb1 subunit of RNA polymerase II with chromatin near or remote from the promoters of two genes. **(a)** Transcription of the *ADH1* gene. Chromatin was immunoprecipitated with antibodies against the CTD phosphorylated on either the serine 2 or serine 5 of the heptad, as indicated at left (α-CTD-S2-P and α-CTD-S5-P, respectively). Then the precipitated chromatin was subjected to PCR with primers specific for regions near the promoter, or remote from the promoter, or an intergenic region, as indicated at top. **(b)** Transcription of the *PMA1* gene. Chromatin was immunoprecipitated with antibodies against the CTD phosphorylated on serine 2 or the unphosphorylated CTD, as indicated at left. PCR primers, indicated at top, were specific for the promoter, or regions progressively more remote from the promoter (CDS = coding sequences). Input chromatin controls are at bottom in both panels. (*Source:* Reprinted by permission of S. Buratowski from "Komarnitsky, Cho, and Buratowski (2000) *Genes and Development* v. 14, pp. 2452–2460" © Cold Spring Harbor Laboratory Press.)

The chromatin immunoprecipitated with each of these antibodies was subjected to PCR with primers specific for promoter regions and interiors of three yeast genes: alcohol dehydrogenase (*ADH1*); cytoplasmic H^+ ATPase (*PMA1*); and a multidrug resistance factor (*PDR5*). The results with all three genes were consistent and demonstrated that: (1) the guanylyl transferase (capping enzyme) associates with the transcription complex only when it is near the promoter; (2) the polyadenylation factor associates with the transcription complex both near and remote from the promoter; and, as expected, the Rpb3 subunit of RNA polymerase is present in the transcription complex both near and remote from the promoter.

Thus, there is a dynamic shift of proteins associating with the transcription complex through the CTD of Rpb1. Some are present only early during the transcription process; others are present for much longer. What causes these changes in the spectrum of proteins associated with the CTD? It is known that the phosphorylation state of the CTD changes during transcription, so perhaps this plays a role.

To investigate this possibility, Buratowski and coworkers performed ChIP assays using antibodies directed against specific phosphorylated amino acids (serine 2 and serine 5) within the heptad repeats of the CTD. The ChIP assays in Figure 15.31 reveal that serine 5 phosphorylation is found primarily in transcription complexes close to the promoter, while serine 2

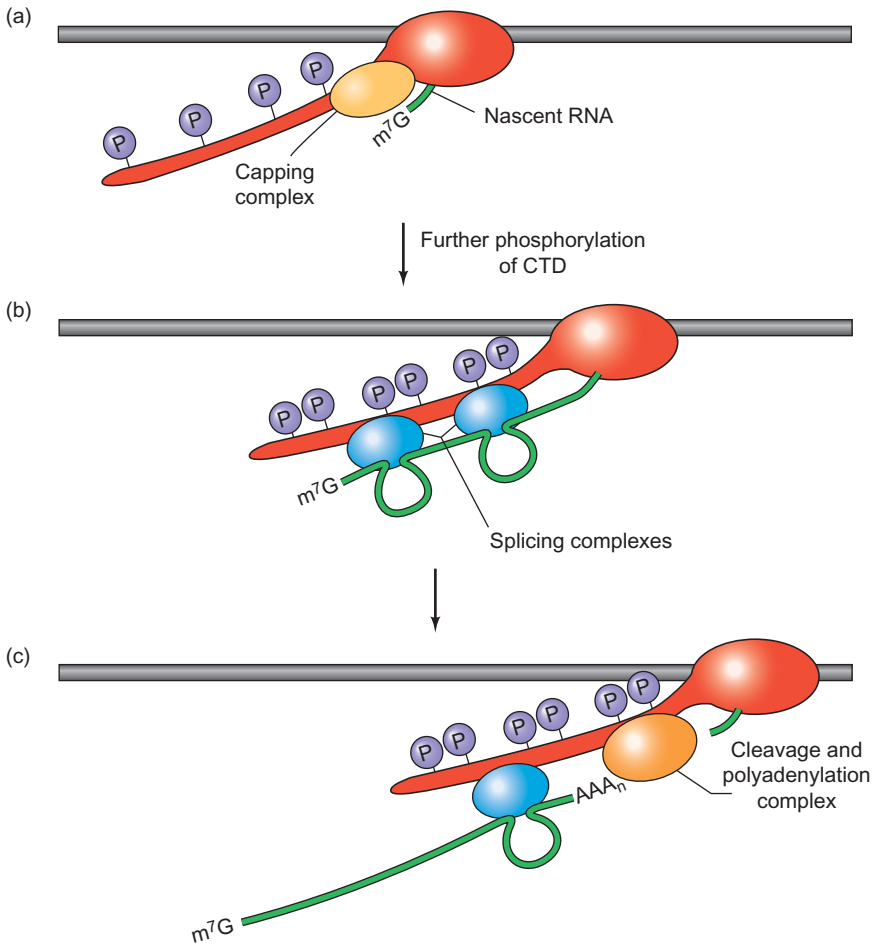

Figure 15.32 Hypothesis of RNA processing organized by CTD. **(a)** RNA polymerase (red) has begun synthesizing a nascent RNA (green). The partially phosphorylated CTD has attracted the capping complex (yellow), which adds a cap to the new RNA as soon as it is available. **(b)** The CTD has become further phosphorylated (presumably including a shift from serine 5 to serine 2 phosphorylation) and has attracted the splicing complex (blue), which defines exons as they are transcribed and splices out the introns in between. **(c)** The CTD is associated with the cleavage and polyadenylation complex (orange), which may have been present since initiation, and this complex has cleaved and begun polyadenylating the transcript. (*Source:* Adapted from Orphanides, G. and D. Reinberg, A unified theory of gene expression. *Cell* 108 [2000] p. 446, f. 3.)

phosphorylation occurs chiefly in transcription complexes remote from the promoter. Thus, it is not surprising that phosphorylation of serine 5 of the CTD helps recruit the capping complex, which needs to operate shortly after elongation begins. It is also quite possible that the shift in CTD phosphorylation from serine 5 to serine 2, as the transcription complex moves away from the promoter, causes some RNA-processing proteins (e.g., the capping complex) to leave the transcription complex and may even attract a new class of proteins. Figure 15.32 summarizes this hypothesis.

SUMMARY The phosphorylation state of the CTD of Rpb1 in transcription complexes in yeast changes as transcription progresses. Transcription complexes close to the promoter contain phosphorylated serine 5, while complexes farther from the promoter contain phosphorylated serine 2. The spectrum of proteins associated with the CTD also changes. For example, the capping guanylyl transferase is present early in the transcription process, when the complex is close to the promoter, but not later. And this enzyme, along with the rest of the capping complex, is recruited by phosphorylation of serine 5 of the heptad in the polymerase II CTD. By contrast, the polyadenylation factor Hrp1 is present in transcription complexes both near and remote from the promoter.

A CTD Code?

In 2007, Shona Murphy and colleagues showed that serine 7 of the CTD can also be phosphorylated. This raises the number of different phosphorylation states in a given repeat

within the CTD to eight (ranging from no phosphates to three phosphates per repeat). It is also possible that the phosphorylation varies from repeat to repeat, opening up many more variations in CTD phosphorylation state.

Even the potential for eight different states in a given repeat raises the possibility of a "CTD code" that signals for transcription of different gene sets and for different RNA modifications. Indeed, there is evidence for such a CTD code. Murphy and colleagues showed in 2007 that phosphorylation of serine 7 is required for expression of the U2 snRNA gene in human cells. On the other hand, Dirk Eick and colleagues demonstrated that phosphorylation of serine 7 is not required for expression of protein-encoding genes.

Human snRNAs synthesized by polymerase II, including U1 and U2 snRNAs, are not polyadenylated. Instead, their genes contain a conserved **3′ box** element that is essential for proper 3′-end processing. Transcription termination occurs downstream of the 3′ box, and this 3′ box is required for the subsequent clipping that yields the primary 3′-ends that can then be processed in the cytoplasm to mature 3′-ends.

Murphy and colleagues started with an α-amanitin-resistant human polymerase II with an Rpb1 CTD containing only the first 25 heptads. These are the ones with canonical sequences ending in serine 7; most of the last 27 heptads have lysine or threonine instead of serine in the seventh position. The α-amanitin-resistance of this polymerase allowed it to be assayed in cells that also carried an endogenous wild-type polymerase II. Next, Murphy and colleagues mutated the α-amanitin-resistant polymerase to change all 25 serine 7's to alanines, and assayed for proper 3′-end processing by RNase protection analysis. They found that the mutant polymerase was deficient in U2 snRNA processing, but was normal in processing a protein-encoding pre-mRNA.

Note that this transcription control does not occur at the initiation level; the mutant polymerase still initiates at a normal level. Instead, control occurs at the termination or 3′-end processing level. Murphy and colleagues investigated the binding of the **Integrator complex,** a group of 12 polypeptides that are required for U1 and U2 snRNA 3′-end processing, to the mutant polymerase with all its serine 7's changed to alanines. They tagged one of the subunits of the Integrator complex with a TAP epitope and used ChIP to detect binding of the Integrator complex to the mutant RNA polymerase II. Whereas the Integrator complex binds well to the CTD of normal polymerase II, Murphy and colleagues found that it does not bind to the mutant polymerase lacking serine 7 in its CTD. This suggested that serine 7 phosphorylation is required for Integrator complex binding, and thus for proper 3′-end processing of U1 and U2 snRNA transcripts. This is the best evidence to date for a CTD code that affects gene expression.

SUMMARY In addition to serines 2 and 5, serine 7 of the heptad repeat in the Rpb1 CTD is phosphorylated during transcription. This raises the number of combinations of phosphorylated and unphosphorylated serines in each repeat to eight, and raises the possibility of a CTD code that governs which genes are expressed. One piece of evidence for such a code is the fact that loss of serine 7 from the repeats prevents 3′-end processing of U2 snRNA transcripts, and therefore prevents expression of the U2 snRNA gene.

Coupling Transcription Termination with mRNA 3′-end Processing

Termination of transcription of class II genes has been notoriously difficult to study, largely because the mature 3′-end of the mRNA is not the same as the termination site. Instead, as we have already learned, a longer, pre-mRNA must be cleaved at the polyadenylation site and then polyadenylated. This leaves a relatively stable mRNA and an unstable 3′-fragment that is rapidly degraded. It is the 3′-end of this unstable part of the RNA that is the true termination site. Despite this difficulty, several investigators have successfully studied termination in class II genes and have discovered that termination is coupled to cleavage at the polyadenylation site, in that each process depends on the other. Indeed, cleavage of the nascent RNA at the termination site may even precede cleavage at the polyadenylation site.

First of all, how do we know that termination is coupled to mRNA processing? Proudfoot and colleagues made this connection in their studies of yeast class II transcription termination. In particular, they examined the *CYC1* gene of the yeast *Saccharomyces cerevisiae* and found that mutations in proteins involved in cleavage at the polyadenylation site inhibited termination, whereas mutations in proteins involved in polyadenylation per se had little effect on termination.

Proudfoot and colleagues cloned the yeast *CYC1* gene into a plasmid (p*GCYC1*) in which it would be expressed under the control of the strong *GAL1/10* promoter. They made a similar construct (p*Gcyc1-512*), which lacked the normal polyadenylation signal at the end of the *CYC1* gene. Next, they transfected yeast cells with these plasmids and assayed first for the expression level of the gene by Northern blotting. Figure 15.33a shows the results: The loss of the polyadenylation site greatly reduced expression from the gene. The control showed that expression of another gene (*ACT1*) was not affected, so the loss of the *CYC1* signal was not due to differences in loading or blotting of the two lanes.

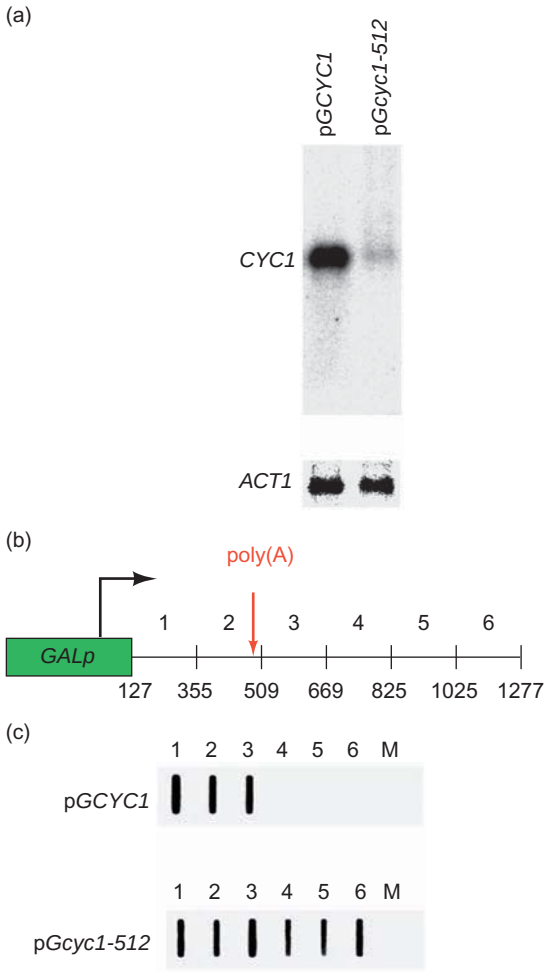

Figure 15.33 Linkage between polyadenylation and termination of transcription. (a) Northern blot analysis. Proudfoot and colleagues Northern blotted transcripts from cells bearing the wild-type gene (p*GCYC1*) or a gene lacking the *CYC1* polyadenylation site (p*Gcyc1-512*). Then they hybridized the blot with a labeled *CYC1* probe. After the first hybridization, they stripped the blot and reprobed with an actin gene probe (*ACT1*) as a control for blotting efficiency. **(b)** Map of the region used in nuclear run-on transcription analysis. Proudfoot and colleagues cloned the yeast *CYC1* gene under the control of the strong *GAL1/10* promoter (*GALp,* green) into a plasmid and placed this construct into yeast cells for analysis. For nuclear run-on analysis, they dot-blotted fragments 1–6, whose relative positions are given. The location of the polyadenylation site (red) in fragment 2 is indicated. (**c**) Results of run-on analysis. Proudfoot and colleagues hybridized dot blots of fragments 1–6, (panel b) to labeled nuclear run-on transcripts from cells carrying the wild-type or mutant *CYC1* gene, as indicated at left. M designates a negative control with M13 DNA on the dot blot. (*Source:* Birse et al *Science* 280: p. 299. © 1988 by the AAAS.)

One reason for the poor expression could be failure to terminate transcription properly. To see if termination really did fail, Proudfoot and colleagues performed a nuclear run-on analysis as follows: They dot-blotted fragments of the *CYC1* gene, including fragments encompassing about 800 bp downstream of the polyadenylation site, as illustrated in Figure 15.33b. Then they hybridized labeled nuclear run-on RNA from cells transfected with either the wild-type *CYC1* gene or the mutant gene lacking the polyadenylation site. Figure 15.33c shows the results. Transcription of the wild-type gene terminated in fragment 3, just downstream of the polyadenylation site. We know that termination occurred in fragment 3 because no transcripts hybridized to fragment 4. But transcription of the mutant gene extended far past the normal termination site, at least into fragment 6, showing that normal termination had failed.

As we have learned, polyadenylation really consists of two steps: RNA cleavage and then polyadenylation. In principle, one of these steps, and not the other, could be coupled to termination. To explore this issue, Proudfoot and colleagues performed a new run-on transcription assay with yeast strains bearing temperature-sensitive mutations in the genes encoding cleavage and polyadenylation factors. Again, they did Northern blots first and discovered that all of the mutants showed depressed levels of *CYC1* mRNA at the nonpermissive temperature. Again, failure to polyadenylate the transcript and failure to terminate the transcript could both have led to its instability.

The run-on transcription assay gave a more complete answer. Some of the mutations caused a failure of termination, but others did not. Is there a pattern here? Indeed, there is. The former set of genes encode proteins involved in cleavage prior to polyadenylation, while the latter set encode proteins involved in polyadenylation after cleavage. Thus, it appears that cleavage at the polyadenylation site, not polyadenylation per se, is coupled to termination of transcription.

We know that the cleavage and polyadenylation factors associate with the CTD of the Rpb1 subunit of RNA polymerase II. The fact that active cleavage factors are required for termination implicates the CTD in termination as well as in other aspects of mRNA maturation. We will return to this theme in the next section.

SUMMARY Transcription termination and mRNA 3′-end processing are coupled in the following way: An intact polyadenylation site and active factors that cleave at the polyadenylation site are required for transcription termination, at least in yeast. Active factors that polyadenylate a cleaved pre-mRNA are not required for termination.

Mechanism of Termination

Michael Dye and Proudfoot performed a detailed analysis of termination in the human β- and ε-globin genes in 2001. They made the following discoveries: (1) The region downstream of the polyadenylation site is essential for termination. (2) Cleavage of the nascent transcript at multiple sites

downstream of the polyadenylation site is required for termination. (3) This transcript cleavage occurs cotranscriptionally and, presumably, precedes cleavage at the polyadenylation site. Then, in 2004, they discovered that the cleavage of the nascent transcript is an autocatalytic event: The RNA cleaves itself.

In their 2001 study, Dye and Proudfoot put the human β-globin gene, including 1.7 kb of its 3′-flanking region, into a plasmid under control of a strong enhancer–promoter combination from the human immunodeficiency virus (HIV). Then they placed this construct into HeLa cells where the β-globin gene could be expressed. The HIV enhancer–promoter has the advantage that the transcription it directs depends on a viral transactivating factor called Tat, so transcription can be turned on and off easily by adding or removing Tat.

Next, these workers performed nuclear run-on analysis of the cloned gene and compared the results to those from the β-globin gene in its natural chromosomal context, under control of its own promoter. Figure 15.34a shows a map of the β-globin gene, including the downstream region, with its own promoter, and the results of the nuclear run-on experiment. Transcription continued through region 10, which lies 1.7 kb downstream of the polyadenylation site. Figure 15.34b shows a map of the cloned β-globin gene under control of the HIV enhancer–promoter, and the results of the nuclear run-on experiment. Again, transcription continued through region 10, but fell off significantly after region 10. The DNA beyond region 10 encompassed regions A and B of the vector, and region U3 of the HIV enhancer–promoter. Thus, termination had occurred at least by region 10, and transcription and termination appeared to be working normally in this cloned construct.

Next, Dye and Proudfoot narrowed down the part of the 3′-flanking region that was important for termination of transcription. They did this by deleting parts of the region and testing by nuclear run-on analysis to see whether termination still occurred. They discovered that deleting regions 8–10 prevented termination. Thus, regions 4–7 were not sufficient for termination. On the other hand, they discovered that deleting regions 5–8, but retaining 9 and 10, or even deleting regions 5–9, but retaining 10, maintained termination. Most strikingly, deleting all regions downstream of 4, except region 8, maintained termination. Thus, regions 8, 9, and 10, individually or together, all could direct termination.

Because region 8 (as well as 9 and 10) appeared to have a termination sequence that operated by causing cleavage of the growing transcript during transcription, Proudfoot and colleagues named it the cotranscriptional cleavage element (**CoTC element**). Then, in 2004, Proudfoot and Alexander Akoulitchev and their colleagues discovered an important secret of the CoTC element: It encodes an autocatalytic domain that can cleave the growing RNA. When they incubated a transcript containing the full-length CoTC

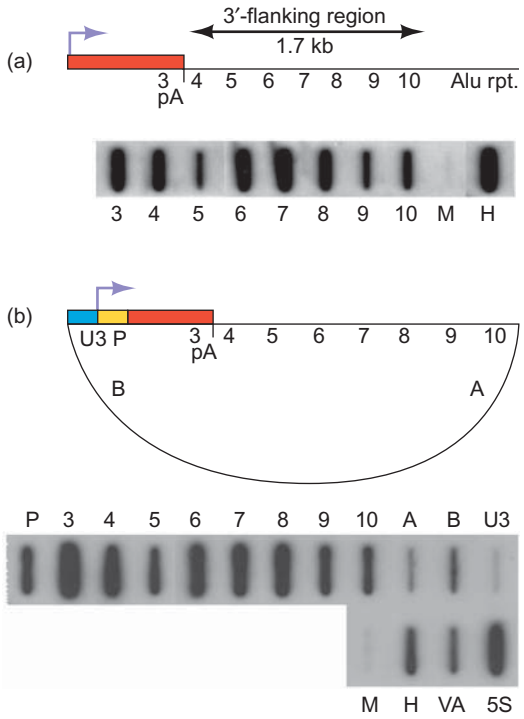

Figure 15.34 Nuclear run-on analysis of natural and cloned β-globin genes. (a) Gene in its chromosomal context. A map of the human gene is shown, including the promoter (purple arrow denotes transcription start site), the coding region (red), the polyadenylation site (pA), and 1.7 kb of downstream sequence (regions 4–10). The results of nuclear run-on analysis are shown below the map, including regions 3–10 and two controls, M and H. M is a negative control containing phage M13 DNA. H is a positive control containing human histone DNA. The histone gene will be transcribed by RNA polymerase II in the cell. **(b)** Gene under control of the HIV enhancer/promoter. The map shows the HIV enhancer region (blue), the HIV promoter region (yellow), the start of transcription (purple arrow), and the coding region (red). Regions A and B lie within the plasmid cloning vector. The results of nuclear run-on analysis are shown below the map. M and H have the same meaning as in panel (a). VA represents an adenovirus VA1 gene, cotransfected along with the β-globin plasmid. This gene is transcribed by RNA polymerase III. 5S denotes hybridization to a 5S rRNA probe, which detects in vivo transcription of the human 5S rRNA gene by RNA polymerase III. (*Source:* Reprinted from *Cell* v. 105, Dye and Proudfoot, p. 670 © 2001, with permission from Elsevier Science.)

element with Mg^{2+} and GTP, but no proteins, the RNA decayed much faster than a control RNA, with a half-life of just 38 min. By making deletions within the CoTC element, these workers were able to narrow the autocatalytic site's location down to a 200-nt sequence [**CoTC(r)**] at the 5′-end of the CoTC element (Figure 15.35). This 200-nt sequence decayed with a half-life of just 15 min in vitro. By contrast, the mutant sequence (*mut*Δ) containing nucleotides 50–150 had no autocatalytic activity.

Is the CoTC element important in transcription termination? To find out, the investigators inserted the β-globin gene into a plasmid and placed the plasmid into HeLa cells. They also replaced the CoTC element at the end of the β-globin

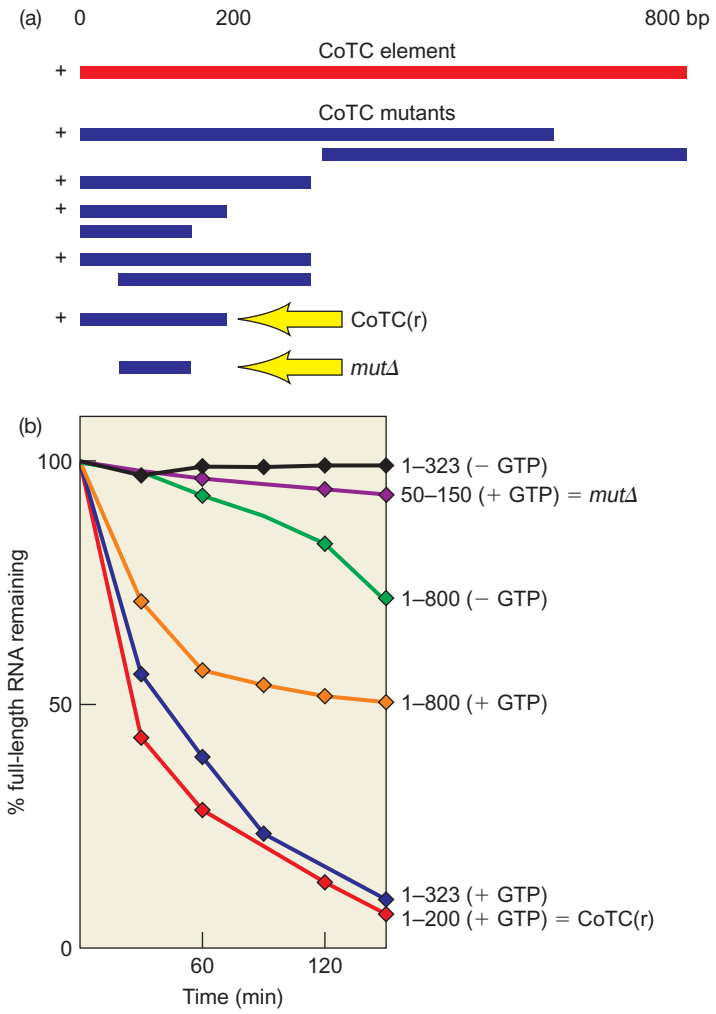

Figure 15.35 Finding the catalytic site in the CoTC element. **(a)** The mutants. Proudfoot, Akoulitchev, and colleagues started with the 800-bp CoTC element at top (red bar) and made deletion mutants that were transcribed to yield the RNAs illustrated below (blue bars). Deletions are denoted by gaps in the bars. Mutant RNAs that retained catalytic activity are marked with plus signs at left. The arrows point to: CoTC(r), the RNA containing nucleotides 1–200, which retained activity; and *mut*Δ, the RNA containing nucleotides 50–150, which lacked activity. **(b)** Experimental results. The fraction of full-length RNA remaining is plotted versus reaction time. We see that the reaction depends on GTP, and that the CoTC(r) RNA that includes nucleotides 1–200 retains full catalytic activity. (*Source:* Adapted from A. Teixeira et al., Autocatalytic RNA cleavage in the human beta-globin pre-mRNA promotes transcription termination. *Nature* 432:526, 2006.)

gene with its mutant forms, including CoTC(r) (the minimal autocatalytic element) and *mut*Δ (the element lacking autocatalytic activity). Then they performed nuclear run-on analysis to see whether transcription termination occurred normally. They found that the gene with the CoTC(r) element at its end terminated transcription almost as well as wild-type, while the gene with the *mut*Δ element at its end allowed transcription to continue past the normal termination site. In experiments with other mutant CoTC elements, they found that the autocatalytic activity of CoTC correlated very well with termination activity. Thus, the autocatalytic activity appears to be required for proper termination.

Is an autocatalytic CoTC-like element a general requirement for transcription termination in eukaryotes? The β-globin genes of primates do contain a conserved CoTC element, with the highest level of conservation in the catalytic core. Such elements are not detected in less related organisms, presumably because of greater sequence divergence. However, the CoTC element itself could not have been identified as a self-cleaving ribozyme on the basis of sequence alone, so there may be CoTC-like elements downstream of the poly(A) sites of many more eukaryotic genes.

Is simple cleavage of a growing RNA at a CoTC or other site sufficient to cause termination? Perhaps not, as we now have evidence for another phenomenon that operates on RNA polymerases that are extending transcripts beyond their poly(A) sites: The polymerases are "torpedoed." Figure 15.36 illustrates this torpedo mechanism, which resembles the rho-dependent mechanism of termination we studied in Chapter 6. First the RNA is cleaved

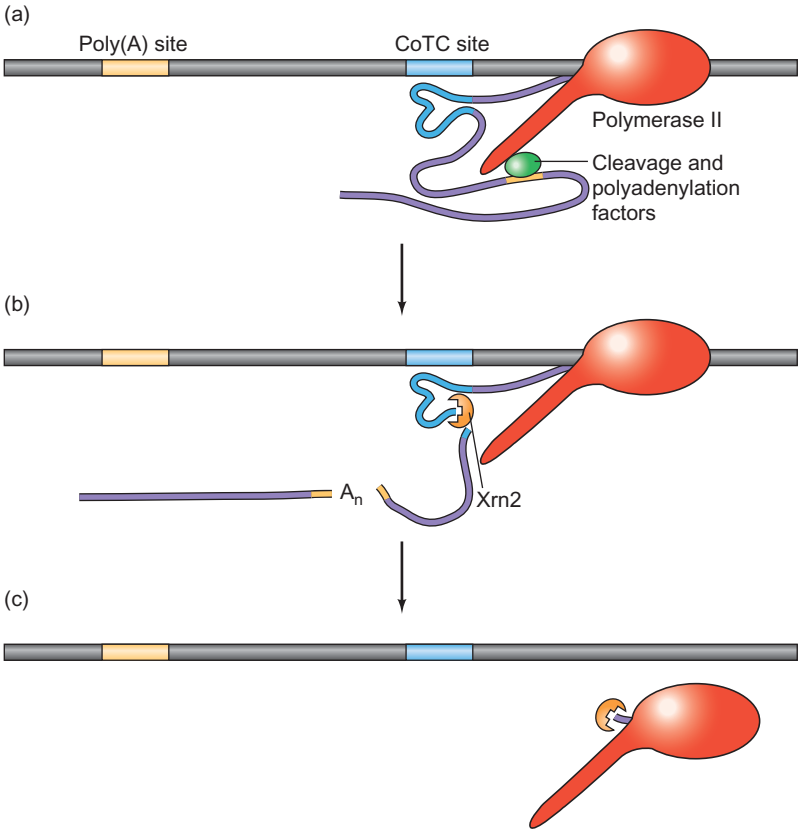

Figure 15.36 A torpedo model for transcription termination in the human β-globin gene. (a) The RNA polymerase (red) has transcribed both the poly(A) site (yellow) and the CoTC site (blue). Cleavage and polyadenylation factors (green) have assembled at the poly(A) site and are also attached to the CTD of the polymerase. **(b)** The cleavage and polyadenylation process is complete, and the mRNA has its poly(A) tail. Also, the CoTC sequence in the transcript has undergone self-cleavage, and the Xrn2 exonuclease (orange) has loaded onto the newly-created RNA 5′-end. **(c)** Xrn2 has degraded the growing RNA nucleotide by nucleotide, has caught the RNA polymerase, and has somehow torpedoed it, causing the polymerase to dissociate from the template and terminate transcription.

downstream of the poly(A) site at a CoTC or other site, then an exonuclease binds to the newly generated RNA free end and begins degrading the RNA, "chasing" the polymerase that is elongating the RNA. When the exonuclease catches the polymerase, it "torpedoes" it, terminating transcription.

In the context of the human β-globin gene, the torpedo model implies that cleavage of the growing transcript at the CoTC site provides an entry site for a 5′→3′ exonuclease that will ultimately torpedo the polymerase. If so, then depleting cells of the relevant 5′→3′ exonuclease should interfere with proper termination. Proudfoot and colleagues tested this notion by using RNAi (Chapter 16) to "knock down" the level of the major human nuclear 5′→3′ exonuclease, **Xrn2.** Using this technique, they depleted the Xrn2 activity to about 25% of its normal value, then tested these cells for proper termination by nuclear run-on assay. They discovered that depletion of Xrn2 activity resulted in a two- to three-fold decrease in normal termination. That is, transcription was two- to three-fold more likely to continue beyond the normal termination site.

Proudfoot and colleagues considered the possibility that cleavage at the poly(A) site, and not the CoTC site, is the entry site for Xrn2. If this were the case, then RNA derived from the region between the poly(A) site and the CoTC site should be less depleted in Xrn2 knock-down cells than in untreated cells. But an RNase protection assay with a probe to measure the steady-state level of transcript from the region between the poly(A) site and the CoTC site showed no difference between Xrn2 knock-down and untreated cells.

Will any 5′-end in the CoTC region provide an entry site for Xrn2? Proudfoot and colleagues addressed this question by substituting a hammerhead ribozyme sequence for the normal CoTC sequence. Hammerhead ribozymes are self-cleaving RNAs, but they produce 5′-hydroxyl groups instead of the 5′-phosphates produced by CoTC. And nuclear run-on analysis showed that although the hammerhead ribozyme did cleave the growing β-globin transcript cotranscriptionally, the downstream RNA was not degraded, as it is in cells with the normal CoTC sequence. Thus, Xrn2 at least appears to require a 5′-phosphate group,

such as provided by CoTC, in order to begin degrading the downstream RNA.

How widespread is the torpedo mechanism for transcription termination? Jack Greenblatt, Steven Buratowski and their colleagues have found a 5′→3′ exonuclease called Rat1 that promotes transcription termination in yeast. There is no evidence for a CoTC element in yeast, so it is assumed that Rat1 gains access to the downstream RNA following cleavage at the poly(A) site, then chases the polymerase until it catches and torpedoes it.

SUMMARY Termination of transcription by RNA polymerase II occurs in two steps. First, the transcript experiences a cotranscriptional cleavage (CoTC) within the termination region downstream of the polyadenylation site. This step occurs before cleavage and polyadenylation at the poly(A) site and is independent of that process. Second, cleavage and polyadenylation occur at the poly(A) site, signaling the polymerase, which is still elongating RNA, to dissociate from the template. In certain genes, at least, this signal could be delivered by a "torpedo," as follows: The CoTC element downstream of the polyadenylation site in the human β-globin mRNA is a ribozyme that cleaves itself, generating a free RNA 5′-end. This cleavage is required for normal transcription termination, apparently because it provides an entry site for Xrn2, a 5′→3′ exonuclease that loads onto the RNA and "chases" the RNA polymerase by degrading the RNA. When it catches up to the polymerase, Xrn2 presumably "torpedoes" it, terminating transcription. A similar torpedo mechanism appears to operate in yeast.

Role of Polyadenylation in mRNA Transport

We have known since 1991 that polyadenylation plays a role in transport of mature mRNA out of the nucleus. That is when Max Birnstiel and colleagues demonstrated that transcripts of a bacterial neomycin gene transplanted into monkey COS1 cells remained in the nucleus. They reasoned that the lack of a polyadenylation signal in the bacterial gene would have left the transcripts without a mature 3′-end, and that might be the reason for defective transport to the cytoplasm.

To test this hypothesis, they provided the neomycin gene with the strong polyadenylation signal from a mammalian β-globin gene. This allowed for polyadenylation of the neomycin transcripts, which were then efficiently transported out of the nucleus into the cytoplasm.

In 2001, Patricia Hilleren and colleagues studied a strain of yeast carrying a temperature-sensitive mutation in the poly(A) polymerase gene. These cells could be shifted to the nonpermissive temperature to shut off polyadenylation of newly made transcripts. These workers focused their attention on transcripts of the *SSA4* gene, a heat-shock gene whose transcripts begin to accumulate at the time of the shift to the nonpermissive temperature. Then they showed by fluorescence in situ hybridization (FISH, Chapter 5) that the *SSA4* transcripts remained in small foci within the nucleus, presumably at or close to the site of their transcription. In wild-type cells, or in mutant cells at the permissive temperature, these transcripts could not be detected in the nucleus and had presumably been polyadenylated and transported to the cytoplasm. Again, it appeared that polyadenylation is required for active transport of mRNAs out of the nucleus. Without polyadenylation, transcripts didn't even seem to move far from their transcription site.

SUMMARY Polyadenylation is required for efficient transport of mRNAs from their point of origin in the nucleus to the cytoplasm.

SUMMARY

Caps are made in steps: First, an RNA triphosphatase removes the terminal phosphate from a pre-mRNA. Next, a guanylyl transferase adds the capping GMP (from GTP). Next, two methyl transferases methylate the N^7 of the capping guanosine and the 2′-O-methyl group of the penultimate nucleotide. These events occur early in the transcription process, before the chain length reaches 30. The cap ensures proper splicing of at least some pre-mRNAs, facilitates transport of at least some mature mRNAs out of the nucleus, protects the mRNA from degradation, and enhances the mRNA's translatability.

Most eukaryotic mRNAs and their precursors have a poly(A) about 250 nt long at their 3′-ends. This poly(A) is added posttranscriptionally by poly(A) polymerase. Poly(A) enhances both the lifetime and translatability of mRNA. The relative importance of these two effects seems to vary from one system to another.

Transcription of eukaryotic genes extends beyond the polyadenylation site. Then the transcript is cleaved and polyadenylated at the 3′-end created by the cleavage. An efficient mammalian polyadenylation signal consists of an AAUAAA motif about 20 nt upstream of a polyadenylation site in a pre-mRNA, followed 23 or 24 bp later by a GU-rich motif, followed immediately by a

U-rich motif. Many variations on this theme occur in nature, which results in variations in efficiency of polyadenylation. Plant polyadenylation signals also usually contain an AAUAAA motif, but more variation is allowed in this region than in an animal AAUAAA. Yeast polyadenylation signals are more different yet and rarely contain an AAUAAA motif.

Polyadenylation requires both cleavage of the pre-mRNA and polyadenylation at the cleavage site. Cleavage requires several proteins: CPSF, CstF, CF I, CF II, poly(A) polymerase, and the CTD of the RNA polymerase II largest subunit. One of the subunits of CPSF (CPSF-73) cleaves the pre-mRNA prior to polyadenylation. Short RNAs that mimic a newly created mRNA 3′-end can be polyadenylated. The optimal signal for initiation of such polyadenylation of a cleaved substrate is AAUAAA, followed by at least 8 nt. Once the poly(A) reaches about 10 nt in length, further polyadenylation becomes independent of the AAUAAA signal, and depends on the poly(A) itself. Two proteins participate in the initiation process: poly(A) polymerase and CPSF, which binds to the AAUAAA motif.

Elongation requires a specificity factor called poly(A)-binding protein II (PAB II). This protein binds to a preinitiated oligo(A) and aids poly(A) polymerase in elongating the poly(A) up to 250 nt or more. PAB II acts independently of the AAUAAA motif. It depends only on poly(A), but its activity is enhanced by CPSF.

Calf thymus poly(A) polymerase is probably a mixture of at least three proteins derived from alternative RNA processing. The structures of the enzymes predicted from these sequences include an RNA-binding domain, a polymerase module, two nuclear localization signals, and a serine/threonine-rich region. The latter region, but none of the rest, is dispensable for activity in vitro.

Poly(A) turns over in the cytoplasm. RNases tear it down, and poly(A) polymerase builds it back up. When the poly(A) is gone, the mRNA is slated for destruction. Maturation-specific polyadenylation of maternal mRNAs in the cytoplasm depends on two sequence motifs: the AAUAAA motif near the end of the mRNA, and an upstream motif called the cytoplasmic polyadenylation element (CPE), which is UUUUUAU or a closely related sequence.

Caps and poly(A) play a role in splicing, at least in removal of the introns closest to the 5′ and 3′ ends, respectively, of the pre-mRNA. Capping, polyadenylation, and splicing proteins all associate with the CTD during transcription.

The phosphorylation state of the CTD of Rpb1 in transcription complexes in yeast changes as transcription progresses. Transcription complexes close to the promoter contain phosphorylated serine 5, while complexes farther from the promoter contain phosphorylated serine 2. The spectrum of proteins associated with the CTD also changes. For example, the capping guanylyl transferase is present early in the transcription process, when the complex is close to the promoter, but not later. By contrast, the polyadenylation factor Hrp1 is present in transcription complexes both near and remote from the promoter. In addition to serines 2 and 5, serine 7 of the heptad repeat in the Rpb1 CTD is phosphorylated during transcription. This raises the number of combinations of phosphorylated and unphosphorylated serines in each repeat to eight, and raises the possibility of a CTD code that governs which genes are expressed. One piece of evidence for such a code is the fact that loss of serine 7 from the repeats prevents 3′-end processing of U2 snRNA transcripts, and therefore prevents expression of the U2 snRNA gene.

An intact polyadenylation site and active factors that cleave at the polyadenylation site are required for transcription termination, at least in yeast. Active factors that polyadenylate a cleaved pre-mRNA are not required for termination. Termination of transcription by RNA polymerase II occurs in two steps. First, the transcript experiences a cotranscriptional cleavage (CoTC) within the termination region downstream of the polyadenylation site. This step occurs before cleavage and polyadenylation at the poly(A) site and is independent of that process. Second, cleavage and polyadenylation occur at the poly(A) site, signaling the polymerase, which is still elongating RNA, to dissociate from the template. The CoTC element downstream of the polyadenylation site in the human β-globin mRNA is a ribozyme that cleaves itself, generating a free RNA 5′-end. This cleavage is required for normal transcription termination, apparently because it provides an entry site for Xrn2, a 5′→3′ exonuclease that loads onto the RNA and "chases" the RNA polymerase by degrading the RNA. When it catches up to the polymerase, Xrn2 presumably "torpedoes" it, terminating transcription. A similar torpedo mechanism appears to operate in yeast.

REVIEW QUESTIONS

1. You label a capped eukaryotic mRNA with ^{3}H-AdoMet and ^{32}P, then digest it with base and subject the products to DEAE-cellulose chromatography. Show the elution of cap 1 with respect to oligonucleotide markers of known charge. Draw the structure of cap 1 and account for its apparent charge.
2. How do we know that the cap contains 7-methylguanosine?
3. Outline the steps in capping.

4. Describe and show the results of an experiment that demonstrates the effect of capping on RNA stability.
5. Describe and give the results of an experiment that shows the synergistic effects of capping and polyadenylation on translation.
6. Describe and give the results of an experiment that shows the effect of capping on mRNA transport into the cytoplasm.
7. Describe and give the results of an experiment that shows the size of poly(A).
8. How do we know that poly(A) is at the 3′-end of mRNAs?
9. How do we know that poly(A) is added posttranscriptionally?
10. Describe and give the results of experiments that show the effects of poly(A) on mRNA translatability, mRNA stability, and recruitment of mRNA into polysomes.
11. With a simple sketch, summarize the polyadenylation process, beginning with an RNA that is being elongated past the polyadenylation site.
12. Describe and give the results of an experiment that shows that transcription does not stop at the polyadenylation site.
13. Describe and give the results of an experiment that shows the importance of the AAUAAA polyadenylation motif. What other motif is frequently found in place of AAUAAA? Where are these motifs found with respect to the polyadenylation site?
14. Describe and give the results of an experiment that shows the importance of the GU-rich and U-rich polyadenylation motifs. Where are these motifs with respect to the polyadenylation site?
15. Describe and give the results of an experiment that shows the effect of the Rpb1 CTD on pre-mRNA cleavage prior to polyadenylation.
16. Describe and give the results of an experiment that shows the importance to polyadenylation of poly(A) polymerase and the specificity factor CPSF.
17. Describe and give the results of an experiment that shows the effect on polyadenylation of adding 40 A's to the end of a polyadenylation substrate that has an altered AAUAAA motif.
18. Describe and give the results of an experiment that shows that CPSF binds to AAUAAA, but not AAGAAA.
19. Describe and give the results of an experiment that shows the effects of CPSF and PAB II on polyadenylation of substrates with AAUAAA or AAGAAA motifs, with and without oligo(A) added. How do you interpret these results?
20. Present a diagram of polyadenylation that illustrates the roles of CPSF, CStF, poly(A) polymerase (PAP), RNA polymerase II, and PAB II.
21. What part of the poly(A) polymerase PAP I is required for polyadenylation activity? Cite evidence.
22. Describe and give the results of an experiment that identifies the cytoplasmic polyadenylation element (CPE) that is necessary for cytoplasmic polyadenylation.
23. Describe and give the results of an experiment that shows that a capping enzyme binds to the RNA polymerase II CTD.
24. Describe and give the results of a Far Western blotting experiment that shows that a component of the U1 snRNP binds to the RNA polymerase II CTD.
25. Describe and give the results of ChIP analysis that shows: (a) that a capping enzyme associates with the RNA polymerase II CTD when it is close to the promoter but not when it is far from the promoter; and (b) that the phosphorylation state of the CTD changes as the RNA polymerase moves away from the promoter.
26. Describe and give the results of an experiment that shows that failure of polyadenylation results in failure of proper transcription termination. Is this behavior due to failure of polyadenylation per se, or is it due to failure of cleavage of the transcript at the polyadenylation site?
27. Describe and give the results of an experiment that indicates that transcription termination requires autocatalytic cleavage of the transcript, even as it is being elongated (cotranscriptional cleavage).
28. Present a torpedo model for transcription termination in eukaryotes.

ANALYTICAL QUESTIONS

1. You are studying a virus that produces mRNAs with extraordinary caps having a net charge of −4 instead of −5. You find these caps have the usual methylations of cap 1: the m^7G and the 2′-O-methyl on the penultimate nucleotide, but no additional methylations. Propose a hypothesis to explain the reduced negative charge and describe experiments to test your hypothesis. Describe sample positive results.
2. Design an experiment to demonstrate that CstF binds to the GU/U element of the cleavage and polyadenylation signal. How would you determine whether one or the other (GU-rich or U-rich) or both parts of this element are required for CstF binding?
3. You are working in a research laboratory that studies the biochemisty of mRNA processing. You have developed an in vitro assay for both splicing and polyadenylation. You produce in vitro the following radioactive mRNA substrates (see table, next page) that either include a 5′-cap or lack the 5′-cap. You incubate these radioactive mRNA substrates with HeLa nuclear extract for 20 min at 30°C and electrophorese the products on a high resolution gel. You then distinguish the splicing products based on their relative sizes in the gel. You count the amount of radioactivity found in the unprocessed mRNA (pre-mRNA), the amount with intron 1 removed (splice 1), the amount with intron 2 removed (splice 2), both introns

removed, and the amount of polyadenylated (poly A). You get the following results, where the number of pluses is related to the relative amount of radioactivity found in that band on the gel:

	Pre-mRNA	*Splice 1 only*	*Splice 2 only*	*Splice 1 and 2*	*Poly (A)*
RNA A uncapped	++	+	+++	+	+++
RNA A capped	+	+	+	+++	+++
RNA B uncapped	++++	+	+	+	+
RNA B capped	++	+++	+	+	+

Propose a hypothesis that explains all these results.

4. In yeast transcription complexes, the phosphorylation state of the CTD of Rpb1, as well as the spectrum of proteins associated with it, changes as transcription progresses. Currently the thought is that the shift in CTD phosphorylation from serine 5 to serine 2 may cause some RNA-processing proteins to leave the complex and possibly attract new proteins to the CTD (as depicted in Figure 15.32). Design and outline the experiments you would perform to demonstrate that the shift in CTD phosphorylation does indeed result in the release (or removal) of RNA-processing proteins as well as the addition of new RNA-processing proteins. Be sure to thoroughly explain your hypotheses to back up your experimental plans.

SUGGESTED READINGS

General References and Reviews

Barabino, S.M.L. and W. Keller. 1999. Last but not least: Regulated poly(A) tail formation. *Cell* 99:9–11.

Bentley, D. 1998. A tale of two tails. *Nature* 395:21–22.

Colgan, D.F. and Manley, J.L. 1997. Mechanism and regulation of mRNA polyadenylation. *Genes and Development* 11:2755–66.

Corden, J.L. 2007. Seven ups the code. *Science* 318:1735–36.

Manley, J.L. and Y. Takagaki. 1996. The end of the message—Another link between yeast and mammals. *Science* 274:1481–82.

Orphanides, G. and D. Reinberg. 2002. A unified theory of gene expression. *Cell* 108:439–51.

Proudfoot, N.J. 1996. Ending the message is not so simple. *Cell* 87:779–81.

Proudfoot, N.J., A. Furger, and M.J. Dye. 2002. Integrating mRNA processing with transcription. *Cell* 108:501–12.

Tollervey, D. 2004. Molecular biology: Termination by torpedo. *Nature* 432:456–57.

Wahle, E. and W. Keller. 1996. The biochemistry of polyadenylation. *Trends in Biochemical Sciences* 21:247–51.

Wickens, M. 1990. How the messenger got its tail: Addition of poly(A) in the nucleus. *Trends in Biochemical Sciences* 15:277–81.

Wickens, M. and T.N. Gonzalez. 2004. Knives, accomplices, and RNA. *Science* 306:1299–1300.

Research Articles

Bardwell, V.J., D. Zarkower, M. Edmonds, and M. Wickens. 1990. The enzyme that adds poly(A) to mRNA is a classical poly(A) polymerase. *Molecular and Cellular Biology* 10:846–49.

Barillà, D., B.A. Lee, and N.J. Proudfoot. 2001. Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences USA* 98:445–50.

Birse, C.E., L. Minvielle-Sebastia, B.A. Lee, W. Keller, and N.J. Proudfoot. 1998. Coupling termination of transcription to messenger RNA maturation in yeast. *Science* 280:298–301.

Dye, M.J. and N.J. Proudfoot. 2001. Multiple transcript cleavage precedes polymerase release in termination by RNA polymerase II. *Cell* 105:669–81.

Egloff, S., D. O'Reilly, R.D. Chapman, A. Taylor, K. Tanzhaus, L. Pitts, D. Eick, and S. Murphy. 2007. Serine 7 of the RNA polymerase II CTD is specifically required for snRNA Gene Expression. *Science* 318:1777–79.

Fitzgerald, M. and T. Shenk. 1981. The sequence 5′-AAUAAA-3′ forms part of the recognition site for polyadenylation of late SV40 mRNAs. *Cell* 24:251–60.

Fox, C.A., M.D. Sheets, and M.P. Wickens. 1989. Poly(A) addition during maturation of frog oocytes: Distinct nuclear and cytoplasmic activities and regulation by the sequence UUUUUAU. *Genes and Development* 3:2151–56.

Furuichi, Y., A. LaFiandra, and A.J. Shatkin. 1977. 5′-terminal structure and mRNA stability. *Nature* 266:235–39.

Furuichi, Y. and K.-I. Miura. 1975. A blocked structure at the 5′-terminus of mRNA from cytoplasmic polyhedrosis virus. *Nature* 253:374–75.

Furuichi, Y., S. Muthukrishnan, J. Tomasz, and A.J. Shatkin. 1976. Mechanism of formation of reovirus mRNA 5′-terminal blocked and methylated sequence m^7GpppG^mpC. *Journal of Biological Chemistry* 251:5043–53.

Gallie, D.R. 1991. The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. *Genes and Development* 5:2108–16.

Gil, A. and N.J. Proudfoot. 1987. Position-dependent sequence elements downstream of AAUAAA are required for efficient rabbit β-globin mRNA 3′-end formation. *Cell* 49:399–406.

Hamm, J. and I.W. Mattaj. 1990. Monomethylated cap structures facilitate RNA export from the nucleus. *Cell* 63:109–18.

Hirose, Y. and J.L. Manley. 1998. RNA polymerase II is an essential mRNA polyadenylation factor. *Nature* 395:93–96.

Hofer, E. and J.E. Darnell. 1981. The primary transcription unit of the mouse β-major globin gene. *Cell* 23:585–93.

Huez, G., G. Marbaix, E. Hubert, M. Leclereq, U. Nudel, H. Soreq, R. Salomon, B. Lebleu, M. Revel, and U.Z.

Littauer. 1974. Role of the polyadenylate segment in the translation of globin messenger RNA in *Xenopus* oocytes. *Proceedings of the National Academy of Sciences USA* 71:3143–46.

Izaurralde, E., J. Lewis, C. McGuigan, M. Jankowska, E. Darzynkiewicz, and I.W. Mattaj. 1994. A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell* 78:657–68.

Keller, W., S. Bienroth, K.M. Lang, and G. Christofori. 1991. Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3′ processing signal AAUAAA. *EMBO Journal* 10:4241–49.

Kim, M., N.J. Krogan, L. Vasiljeva, O.J. Rando, E. Nedea, J.F. Greenblatt, and S. Buratowski. 2004. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* 432:517–22.

Komarnitsky, P., E.-J. Cho, and S. Buratowski. 2000. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes and Development* 14:2452–60.

Mandel, C.R., S. Kaneko, H. Zhang, D. Gebauer, V. Vethantham, J.L. Manley, and L. Tong. 2006. Polyadenylation factor CPSF-73 is the pre-mRNA 3′-end-processing endonuclease. *Nature* 444:953–56.

McCracken, S., N. Fong, E. Rosonina, K. Yankulov, G. Brothers, D. Siderovski, A. Hessel, S. Foster, Amgen EST Program, S. Shuman, and D.L. Bentley. 1997. 5′-capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes and Development* 11:3306–18.

McDonald, C.C., J. Wilusz, and T. Shenk. 1994. The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location. *Molecular and Cellular Biology* 14:6647–54.

Morris, D.P. and A.L. Greenleaf. 2000. The splicing factor, Prp40, binds the phosphorylated carboxy-terminal domain of RNA polymerase II. *Journal of Biological Chemistry* 275:39935–43.

Munroe, D. and A. Jacobson. 1990. mRNA poly(A) tail, a 3′ enhancer of a translational initiation. *Molecular and Cellular Biology* 10:3441–55.

Sheets, M.D. and M. Wickens. 1989. Two phases in the addition of a poly(A) tail. *Genes and Development* 3:1401–12.

Sheiness, D. and J.E. Darnell. 1973. Polyadenylic acid segment in mRNA becomes shorter with age. *Nature New Biology* 241:265–68.

Teixeira, A., A. Tahirir-Alaoui, S. West, B. Thomas, A. Ramadass, I. Martianov, M. Dye, W. James, N.J. Proudfoot, and A. Akoulitchev. 2004. Autocatalytic RNA cleavage in the human β-globin pre-mRNA promotes transcription termination. *Nature* 432:526–30.

Wahle, E. 1991. A novel poly(A)-binding protein acts as a specificity factor in the second phase of messenger RNA polyadenylation. *Cell* 66:759–68.

Wei, C.M. and Moss, B. 1975. Methylated nucleotides block 5′-terminus of vaccinia virus mRNA. *Proceedings of the National Academy of Sciences USA* 72:318–22.

West, S., N. Gromak, and N.J. Proudfoot. 2004. Human 5′→3′ exonuclease Xrn2 promotes transcription termination at cotranscriptional cleavage sites. *Nature* 432:522–25.