MOLECULAR AND CELLULAR BIOLOGY, Nov. 1990, p. 6059–6063
0270-7306/90/116059-05$02.00/0

Vol. 10, No. 11

Splicing Precedes Polyadenylation during *Drosophila E74A* Transcription

MIMI F. LEMAIRE AND CARL S. THUMMEL*

Department of Human Genetics, Howard Hughes Medical Institute, University of Utah Medical Center, Salt Lake City, Utah 84132

Received 13 June 1990/Accepted 7 August 1990

The *E74* gene is one of a small set of early genes induced by the steroid hormone ecdysone at the onset of metamorphosis in the fruit fly, *Drosophila melanogaster*. This complex gene directs the synthesis of a 60-kilobase (kb) primary transcript that is spliced to form the 6-kb *E74A* mRNA. In a previous study, we found that ecdysone directly activates the *E74A* promoter and determined that RNA polymerase II transcribes this gene at a rate of approximately 1.1 kb/min. This elongation rate accounts for most of the 1-hour delay seen between the addition of ecdysone and the appearance of cytoplasmic *E74A* mRNA (C. S. Thummel, K. C. Burtis, and D. S. Hogness, Cell 61:101–111, 1990). We show here that nascent *E74A* transcripts are spliced, and we propose a model for the order of that splicing. This study provides, for the first time, direct biochemical evidence for splicing of a low-abundance cellular RNA before transcription termination and polyadenylation.

In *Drosophila melanogaster*, pulses of ecdysone trigger molting of the cuticle at the end of each larval instar as well as the dramatic metamorphosis that results in the formation of the adult fly. Careful observation of the giant salivary gland polytene chromosomes at the onset of metamorphosis has revealed a complex and highly reproducible cascade of more than 100 puffs that arise and regress in response to the hormone. Of these, six early puffs are induced directly by ecdysone and appear to encode key regulatory proteins that control this genetic hierarchy (3).

Our current studies concern the *Drosophila E74* gene, a complex gene isolated from the early puff at position 74EF in the polytene chromosomes (6). *E74* consists of three nested transcription units that derive from unique promoters but share a common polyadenylation site (Fig. 1). The *E74A* unit spans 60-kilobases (kb) of genomic DNA and directs the synthesis of a 6.0-kb mRNA. The *E74B* transcription units direct the synthesis of 4.8- and 5.1-kb mRNAs from two start sites located 40 kb downstream from the *E74A* start site (6). The protein encoded by the *E74A* mRNA is localized in the nucleus (L. Boyd and C. S. Thummel, unpublished data) and binds DNA in a site-specific manner (L. D. Urness and C. S. Thummel, Cell, in press), consistent with its performing a regulatory function.

Ecdysone directly induces *E74A* mRNA synthesis at the level of transcriptional initiation. This mechanism of induction, combined with the unusual length of the *E74A* transcription unit, allowed us to determine the rate of RNA polymerase movement along the DNA. This rate measurement, of approximately 1.1 kb/min, accounts for the 1-hour delay seen between the time of ecdysone addition and the first appearance of cytoplasmic *E74A* mRNA (21). These data led us to propose that the lengths of regulatory genes within a hierarchy may contribute toward controlling the timing of activation and repression of their corresponding target genes.

Our transcription rate measurement predicts that RNA polymerase should reach the end of the 60-kb *E74A* transcription unit approximately 55 min after the addition of ecdysone. The observation that cytoplasmic *E74A* mRNA can be detected approximately 5 min after transcription termination implies either that RNA processing occurs relatively rapidly after polyadenylation of the primary transcript (within minutes) or that the primary transcript is spliced as it is elongated. The *E74* gene is unique in that its length and mechanism of induction provide a means to capture sets of elongating RNAs and test them directly for exon excision. Taking advantage of this, we designed the following experiment to determine when *E74A* transcripts are spliced.

Organs were mass isolated from late third instar larvae and maintained in culture in the absence of ecdysone. After several hours, to allow recovery from possible exposure to the hormone in vivo, ecdysone was added at 15- or 30-min intervals to initiate *E74A* transcription. The tissues were then homogenized, nuclei were isolated, and nuclear RNA was purified away from contaminating DNA. Since RNA polymerase will not reach the 3′ end until approximately 55 min after the addition of ecdysone, any splicing detected prior to that time would have occurred on nascent primary transcripts.

Splicing proceeds through a two-step reaction, involving initial cleavage at the 5′ splice site and subsequent ligation to the corresponding 3′ splice site. Since these steps have only been resolved in vitro, we have limited our in vivo analysis of *E74A* splicing to detecting exon cleavage, assuming that ligation occurs essentially simultaneously. Our initial efforts were confounded by the low levels of *E74A* transcription (21). After attempting S1 nuclease protection assays using end-labeled single-stranded DNA probes (22) and uniformly labeled single-stranded DNA probes (21), we achieved the sensitivity required to detect the low levels of nuclear *E74A* RNA by using high-specific-activity RNA probes in RNase protection assays (12, 18). To conserve the small amounts of nuclear RNA that we recovered from larval organs, we used RNA probes that would allow us to detect several exons simultaneously. The probes were initially tested in RNase protection analysis of total yeast RNA and *Drosophila* pupal RNA to identify probes with a relatively low background. In this manner, three probes were selected, derived from different subcloned portions of genomic DNA. Taken together, these probes cover the five 5′-proximal *E74A* exons,

* Corresponding author.

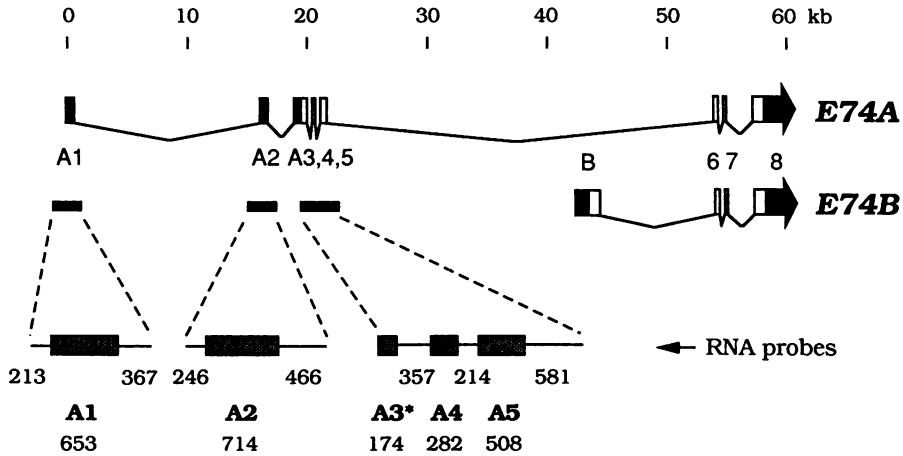

FIG. 1. *E74* gene structure and the positions and lengths of genomic fragments used to make RNA probes for RNase protection assays. The *E74A* and *E74B* transcription units are depicted along with a scale of genomic DNA. The nomenclature used for naming the transcripts and exons has been presented previously (6). Three genomic DNA fragments (solid black rectangles) were inserted into pBluescript (Stratagene) in the appropriate orientation to allow in vitro RNA synthesis with T3 RNA polymerase. A 1,500-base-pair (bp) *Hin*dIII fragment was inserted into the *Hin*dIII site of pBluescriptSK(+) to allow synthesis of the exon A1 RNA probe. This genomic fragment contains the 653-bp exon A1 flanked by 487 bp of upstream DNA and 367 bp of intron-encoding sequences. Digestion of the resultant plasmid with *Fok*I yielded a linear DNA template that was transcribed to generate a 1,233-nucleotide RNA probe, as shown at the bottom of the figure. The stippled box represents the exon, and the adjacent lines represent flanking sequences. A 1,596-bp *Eco*RI fragment was inserted into the *Eco*RI site of pBluescriptSK(−) to allow synthesis of the exon A2 RNA probe. This genomic fragment contains the 714-bp exon A2 flanked by 421 bp of upstream intron-encoding DNA and 461 bp of downstream intron-encoding sequences. Digestion of the resultant plasmid with *Xho*I yielded a linear DNA template that was transcribed to generate a 1,426-nucleotide RNA probe for exon A2, as shown in the figure. A 5,305-bp *Eco*RI fragment was inserted into the *Eco*RI site of pBluescriptSK(−) to allow synthesis of an RNA probe for exons A3, A4, and A5. This genomic fragment contains the 611-bp exon A3, the 282-bp exon A4, and the 508-bp exon A5 flanked by intron-encoding sequences. Digestion of the resultant plasmid with *Nru*I yielded a linear DNA template that was transcribed to generate a 2,116-nucleotide RNA probe for exons A3 to A5, as shown in the figure. *Nru*I cuts within the sequences encoding exon A3, truncating the RNA probe so that only 174 nucleotides of this exon would be protected against RNase digestion. This truncation is denoted by the asterisk in the figure.

A1 through A5, along with adjacent 5′- and 3′-flanking sequences (Fig. 1). We confined our study to the 5′-proximal *E74A* exons since the *E74B* units are actively transcribed at the time of ecdysone addition and thus obscure *E74A* transcripts that have elongated past the *E74B* promoters (21).

The six samples of nuclear RNA from the ecdysone time course were allowed to anneal with each of the exon-specific RNA probes. Samples of total yeast RNA and total *Drosophila* pupal RNA were used as controls. Single-stranded regions were digested with RNase T_1, and the protected probe sequences were fractionated by urea-acrylamide gel electrophoresis and visualized by autoradiography (12). Bands of the expected sizes were seen for each exon (Fig. 2). These experiments were repeated at least twice with different preparations of nuclear RNA. Although the elongating *E74A* primary transcript can be clearly detected by slot-blot hybridization (21), RNase-protected bands corresponding to the unspliced primary transcript were not detected in these experiments. Most likely, these bands are obscured by the relatively high background present at the top of the gel (Fig. 2). The detection of the primary transcript may be further compromised by the low level of unprocessed RNA relative to that of the spliced form.

Exons A1 and A2 were not detected in the 0-, 15-, or 30-min RNA samples. These exons could, however, be detected by 45 min after ecdysone addition, and they continued to accumulate through the end of the time course (Fig. 2A and B). longer exposures revealed no evidence of the A1 and A2 exons in the 30-min sample, even though the nascent *E74A* RNA contains both of these exons, as well as the downstream exons A3 through A5. Apparently, some change takes place between 30 and 45 min that allows exons A1 and A2 to be spliced. One possibility is that, due to either the synthesis of new RNA sequences or the splicing of other exons, a unique RNA conformation is formed during this 15-min interval that facilitates the splicing of A1 and A2.

Two abundant protected probe fragments, 280 and 445 nucleotides in length, that increase in intensity in proportion to exon A2 were also detected in Fig. 2B (marked by asterisks). These fragments are not due to differential splicing of exon A2 since they were not detected by our previous S1 analyses, with both end-labeled and uniformly labeled single-stranded DNA probes (6; C. S. Thummel, unpublished data). In addition, RNase protection analysis of total pupal RNA did not yield these fragments (Fig. 2B). It is possible that they represent unusually stable intron sequences or are due to abortively spliced forms of the *E74A* primary transcript that can only be detected in nuclear RNA samples. The length of one of these fragments corresponds to the size expected for the 3′ intron (443 nucleotides, after nuclease digestion). The 280-nucleotide fragment, however, is longer than expected for the 5′ intron (241 nucleotides, after nuclease digestion). This fragment does not derive from the use of an alternative splice acceptor since no consensus sequence can be found at the appropriate location. The deviation between the 280-nucleotide fragment and that predicted for the 5′ intron may be due to differences between the mobility of the double-stranded DNA markers and the protected RNA fragment.

Exons A3 and A4 can be clearly detected by 30 min after ecdysone addition, indicating that they are spliced before exons A1 and A2 (Fig. 2C). It is possible that splicing of exons A3 and A4 is a necessary prerequisite for splicing of A1 and A2. Splicing of exon A5 is difficult to see since it migrates just above a strong background band from the RNA probe (Fig. 2C). Exon A5 must, however, be spliced rela-

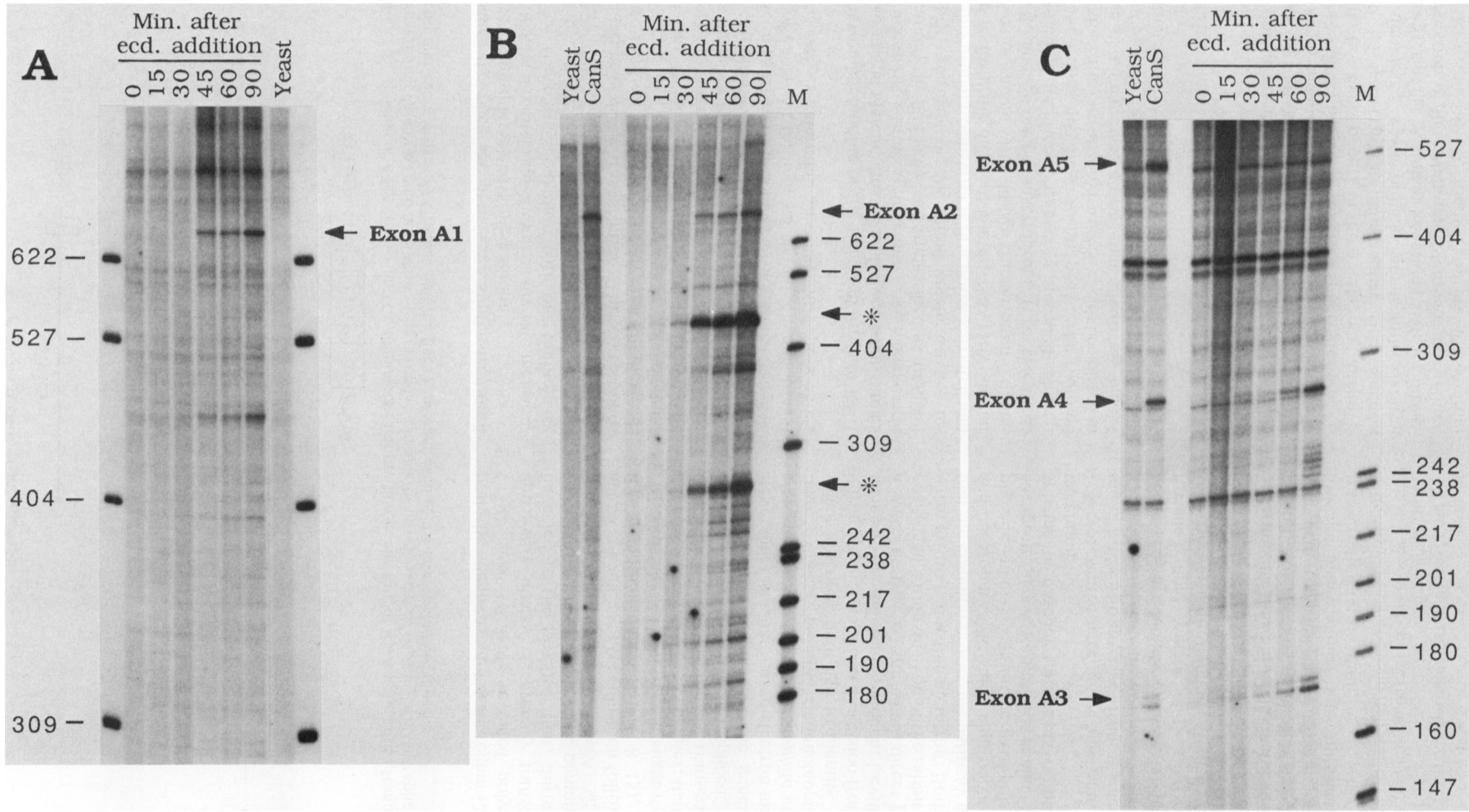

FIG. 2. RNase protection assays to detect splicing of exons A1 through A5. Mass-isolated late third instar larval organs were maintained in culture and treated with ecdysone (ecd.) for 0, 15, 30, 45, 60, or 90 min. Nuclei were isolated from the organs and RNA was extracted as described previously (21). (A) Fragments derived from the A1-specific RNA probe. (B) Fragments derived from the A2-specific RNA probe. (C) Fragments derived from the A3-, A4-, and A5-specific RNA probe. The asterisks with arrows mark fragments potentially derived from intron sequences. See Fig. 1 for a description of the different RNA probes. Size markers (in base pairs) were generated by digesting pBR322 DNA with *Msp*I and 5′ end labeling.

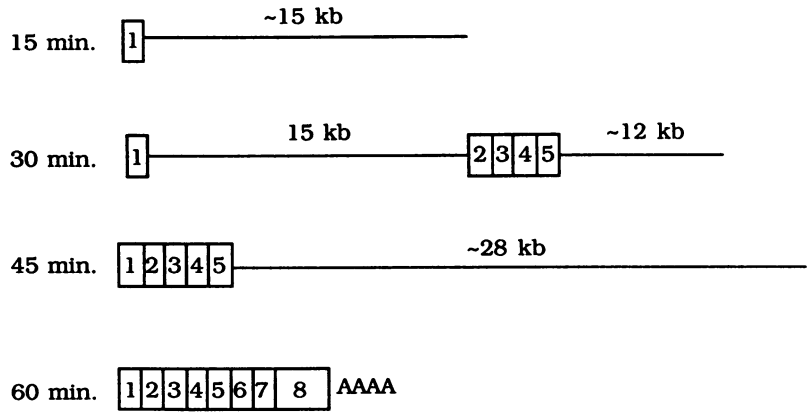

FIG. 3. Model for splicing of *E74A* transcripts. Combining our information on the rate of RNA polymerase II movement along *E74A* (21) with the splicing timing data presented here allows us to propose a model for how nascent *E74A* transcripts are spliced as a function of time. The boxes represent each of the *E74A* exons, and the lines represent each of the intervening sequences.

tively late since its splice acceptor is not synthesized until approximately 50 min after ecdysone addition.

Figure 3 depicts a model that combines the rate measurement of *E74A* transcript elongation with our preliminary splicing timing data presented above. We predict that exons A3 and A4 are the first to be spliced, between 15 and 30 min after the initiation of *E74A* transcription. This corresponds closely to the time when these exons are first synthesized. Splicing of exons A1 and A2, however, is delayed until between 30 and 45 min after transcription initiation. The splicing of exons A5 through A8 must occur between 45 and 60 min after transcription initiation and be complete by 60 min, when cytoplasmic *E74A* mRNA can first be detected (21). The use of short RNA probes that overlap the 5′ and 3′ boundaries of exons A1 through A5 should allow us to test this model as well as determine how efficiently these exons are spliced.

Most of our understanding of the temporal order of RNA polyadenylation and splicing has been derived from studies of the late adenovirus transcription unit. After adenoviral DNA replication, the late genes are transcribed as an intact 30-kb primary transcript (4, 9). Initial kinetic studies revealed that this primary transcript is first polyadenylated and then spliced to form the different late mRNAs (19). Similar kinetic studies of simian virus 40 late transcription (10) and adenovirus E2 and E4 early transcription (23) revealed that polyadenylation of the primary transcript preceded splicing. Polyadenylation does not, however, appear to be required for splicing since the late adenovirus RNAs can be spliced in the presence of cordycepin, a drug that effectively blocks polyadenylation (24). In addition, nonpolyadenylated globin mRNA synthesized in vitro can be efficiently spliced in vivo (13).

Further insights into the relationship between splicing and polyadenylation have emerged from analyses of complex transcription units in which overlapping RNAs are distinguished by unique combinations of splice sites and poly(A) addition sites. In some cases, such as alternative processing of the IgM heavy-chain transcripts, the selection of poly(A) sites appears to direct the use of different splice sites (7, 11). The opposite, however, occurs in the overlapping adenovirus E3 and L4 transcription units; splice site selection appears to determine which poly(A) site is utilized for a particular transcript (1). A similar situation exists for tissue-specific expression of the calcitonin/calcitonin gene-related peptide (CGRP) gene (8, 15). Analyses of pre-mRNA from both the adenovirus E3 early region (20) and the calcitonin/CGRP gene (2) have revealed polyadenylation of long, apparently unspliced, primary transcripts. These results have led to the proposal that, during E3 and calcitonin/CGRP transcription, splice site commitment on the nascent transcripts directs which poly(A) site will be used. The actual splicing event would then follow transcription termination and polyadenylation (1, 8, 15).

Taken together, these data suggest that most transcripts are processed in the same manner; that eucaryotic genes are transcribed into long primary transcripts which are subsequently polyadenylated and spliced. Recent experiments, however, have challenged the generality of this model. Beyer and Osheim (5) used electron microscopy to visualize elongating transcripts in *Drosophila* embryos. This allowed them to observe potential ribonucleoprotein association with splice junctions and also to identify what appeared to be intron loop formation and excision from nascent transcripts. Using a very different approach, Levitt et al. (16) found that a synthetic poly(A) site would direct polyadenylation when positioned within an exon of the rabbit β-globin gene but not when positioned within an intron. Both these results suggest that the common processing pathway utilized by viral RNAs and calcitonin/CGRP, in which polyadenylation appears to precede splicing, cannot be extrapolated to encompass all eucaryotic genes. Rather, Beyer and Osheim (5) suggest that splicing of nascent transcripts is more the rule than the exception.

We presented here direct biochemical evidence for splicing of nascent transcripts from a single-copy cellular gene. The adenovirus late transcription unit has been shown to be processed in a similar manner. S1 protection experiments revealed rapid splicing of the first and second 5′ leader sequences within minutes after their synthesis (14, 17). Because of the abundance of the late viral RNAs, pulse-labeled nuclear RNA could be used directly for S1 protection experiments, thus allowing kinetic studies of RNA processing in the absence of synchronous transcriptional initiation. The data reported here provide biochemical proof that low-abundance cellular transcripts can also be spliced before transcription termination.

We thank Pat Hurban and Rolf Karlstrom for their help during the initial stages of this project and Brenda Bass for critical comments on the manuscript.

This research was supported by the Howard Hughes Medical Institute.

LITERATURE CITED

1. **Adami, G., and J. R. Nevins.** 1988. Splice site selection dominates over poly(A) site choice in RNA production from complex adenovirus transcription units. EMBO J. **7**:2107–2116.
2. **Amara, S. G., R. M. Evans, and M. G. Rosenfeld.** 1984. Calcitonin/calcitonin gene-related peptide transcription unit: tissue-specific expression involves selective use of alternative polyadenylation sites. Mol. Cell. Biol. **4**:2151–2160.
3. **Ashburner, M., C. Chihara, P. Meltzer, and G. Richards.** 1974. On the temporal control of puffing activity in polytene chromosomes. Cold Spring Harbor Symp. Quant. Biol. **38**:655–652.
4. **Berget, S. M., and P. A. Sharp.** 1979. Structure of late adenovirus 2 heterogeneous nuclear RNA. J. Mol. Biol. **129**:547–565.
5. **Beyer, A. L., and Y. N. Osheim.** 1988. Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. Genes Dev. **2**:754–765.
6. **Burtis, K. C., C. S. Thummel, C. W. Jones, F. D. Karim, and D. S. Hogness.** 1990. The *Drosophila* 74EF early puff contains *E74*, a complex ecdysone-inducible gene that encodes two *ets*-related proteins. Cell **61**:85–99.
7. **Danner, D., and P. Leder.** 1985. Role of an RNA cleavage/poly(A) addition site in the production of membrane-bound and secreted IgM mRNA. Proc. Natl. Acad. Sci. USA **82**:8658–8662.
8. **Emeson, R. B., F. Hedjran, J. M. Yeakley, J. W. Guise, and M. G. Rosenfeld.** 1989. Alternative production of calcitonin and CGRP mRNA is regulated at the calcitonin-specific splice acceptor. Nature (London) **341**:76–80.
9. **Flint, S. J.** 1980. Structure and genomic organization of adenoviruses, p. 383–441. *In* J. Tooze (ed.), DNA tumor viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
10. **Ford, J. P., and M.-T. Hsu.** 1978. Transcription pattern of in vivo-labeled late simian virus 40 RNA: equimolar transcription beyond the mRNA 3′ terminus. J. Virol. **28**:795–801.
11. **Galli, G., J. Guise, P. W. Tucker, and J. R. Nevins.** 1988. Poly(A) site choice rather than splice site choice governs the regulated production of IgM heavy-chain RNAs. Proc. Natl. Acad. Sci. USA **85**:2439–2443.
12. **Gilman, M.** 1987. Ribonuclease protection assay, p. 4.7.1-4.7.8. *In* F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, and J. A. Smith (ed.), Current protocols in molecular biology. Green Publishing Associates and John Wiley & Sons, Inc., New York.
13. **Green, M. R., T. Maniatis, and D. A. Melton.** 1983. Human β-globin pre-mRNA synthesized in vitro is accurately spliced in *Xenopus* oocyte nuclei. Cell **32**:681–694.
14. **Keohavong, P., R. Gattoni, J. M. LeMoullec, M. Jacob, and J. Stévenin.** 1982. The orderly splicing of the first three leaders of the adenovirus-2 major late transcript. Nucleic Acids Res. **10**:1215–1229.
15. **Leff, S. E., R. M. Evans, and M. G. Rosenfeld.** 1987. Splice commitment dictates neuron-specific alternative RNA processing in calcitonin/CGRP gene expression. Cell **48**:517–524.
16. **Levitt, N., D. Briggs, A. Gil, and N. J. Proudfoot.** 1989. Definition of an efficient synthetic poly(A) site. Genes Dev. **3**:1019–1025.
17. **Mariman, E. C. M., R. J. van Beek-Reinders, and W. J. van Venrooij.** 1983. Alternative splicing pathways exist in the formation of adenoviral late messenger RNAs. J. Mol. Biol. **163**:239–256.
18. **Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. T. Green.** 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. **12**:7035–7056.
19. **Nevins, J. R., and J. E. Darnell, Jr.** 1978. Steps in the processing of Ad2 mRNA: poly(A)$^+$ nuclear sequences are conserved and poly(A) addition precedes splicing. Cell **15**:1477–1493.
20. **Sittler, A., H. Gallinaro, and M. Jacob.** 1986. In vivo splicing of the premRNAs from early region 3 of adenovirus-2: the products of cleavage at the 5′ splice site of the common intron. Nucleic Acids Res. **14**:1187–1207.
21. **Thummel, C. S., K. C. Burtis, and D. S. Hogness.** 1990. Spatial and temporal patterns of *E74* transcription during *Drosophila* development. Cell **61**:101–111.
22. **Weaver, R. F., and C. Weissman.** 1979. Mapping of RNA by a modification of the Berk-Sharp procedure: the 5′ termini of 15S β-globin mRNA have identical map coordinates. Nucleic Acids Res. **7**:1175–1193.
23. **Weber, J., J.-M. Blanchard, H. Ginsberg, and J. E. Darnell, Jr.** 1980. Order of polyadenylic acid addition and splicing events in early adenovirus mRNA formation. J. Virol. **33**:286–291.
24. **Zeevi, M., J. R. Nevins, and J. E. Darnell, Jr.** 1981. Nuclear RNA is spliced in the absence of poly(A) addition. Cell **26**:39–46.