Splicing of a *C.elegans* myosin pre-mRNA in a human nuclear extract

Stephen C.Ogg^{1*}, Philip Anderson² and Marvin P.Wickens¹

¹Department of Biochemistry and ²Department of Genetics, College of Agricultural and Life Sciences, University of Wisconsin, Madison, WI 53706, USA

Received September 15, 1989; Revised and Accepted November 20, 1989

ABSTRACT

Splicing of mammalian introns requires that the intron possess at least 80 nucleotides. This length requirement presumably reflects the constraints of accommodating multiple snRNPs simultaneously in the same intron. In the free-living nematode, *C. elegans*, introns typically are 45 to 55 nucleotides in length. In this report, we determine whether *C. elegans* introns can obviate the mammalian length requirement by virtue of their structure or sequence. We demonstrate that a 53 nucleotide intron from the *unc-54* gene of *C. elegans* does not undergo splicing in a mammalian (HeLa) nuclear extract. However, insertion of 31 nucleotides of foreign, prokaryotic sequence into the same intron results in efficient splicing. The observed splicing proceeds by the same two-step mechanism observed with mammalian introns, and exploits the same 3' and 5' splice sites as are used in *C. elegans*. The branch point used lies in the inserted sequence. We conclude that *C. elegans* splicing components are either fewer in number or smaller than their mammalian counterparts.

INTRODUCTION

Splicing of nuclear pre-mRNAs occurs in two-steps (1–4). In the first, the precursor is cleaved at the 5' splice site. This is accompanied by formation of a branched nucleotide in the intron, in which the 5' terminal nucleotide of the intron is joined to an internal nucleotide via a 2'-5' phosphodiester bond (the 'lariat' structure). In the second step, the precursor is cleaved at the 3' splice site, and the two exons are ligated (rev. in 1,2).

Efficient and accurate splicing of metazoan nuclear pre-mRNAs requires several distinct features in the precursor, each of which is highly conserved (rev. in 1,2). These include the conserved GU and AG dinucleotides present at the 5' and 3' splice sites, respectively; a polypyrimidine tract near the 3' splice site; and a minimal length of approximately 80 residues. The minimal length requirement is presumed to result from the fact that several large factors, both snRNPs and non-snRNPs, must simultaneously interact with each other and with the intron, and that 80 nucleotides are needed to accommodate the resulting large structure.

In *C. elegans*, introns generally are short: more than 90% of all known *C. elegans* introns are between 45 and 55 nucleotides long (5). The 5' and 3' splice sites of nematode introns are highly conserved, and their sequences are similar to those of mammalian splice sites (5). How does *C. elegans* accommodate its short introns? Although the internal features of *C. elegans* introns appear not to be highly conserved, the possibility exists that their structure might obviate the requirement for one or more splicing factors. Alternatively, perhaps *C. elegans* snRNPs and non-snRNP splicing factors are either smaller or fewer in number than their mammalian counterparts.

In this report, we test whether the difference between nematodes and mammals with respect to the splicing of short introns is a consequence of *C. elegans* introns or of the *C. elegans* splicing apparatus. To do so, we examine the splicing of a typically short nematode intron in a mammalian nuclear extract *in vitro*.

MATERIALS AND METHODS

Enzymes and chemicals

SP6 RNA polymerase was obtained from Promega Biotech (Madison, WI). All other enzymes were obtained from New England Biolabs (Beverly, MA), unless noted otherwise.

Plasmids for transcription *in vitro*

Plasmid pTR #106, which was used as a template for transcription of NEM 53 RNA, was prepared by cloning a *TaqI* fragment of *unc-54* (coordinates 3636 to 3736 of *unc-54*; ref. 19) into the *AccI* site of pSP64. pTR #107, which was used as a template for transcription of NEM 53+31 RNA, was prepared by cloning a 31 nt *Sau3A* fragment of pBR322 (coordinates 1097 to 1127 of pBR322) into the *BclI* site of TR #106. The template used for the synthesis of human β -globin RNA was pSP64-H- β -Delta6 (20).

Preparation of RNA substrates

RNA was prepared by transcription of linear DNA *in vitro* using SP6 polymerase (6), except that the reactions contained 0.1 mM GTP and 1.5 mM G(5')ppp(5')G (Pharmacia). RNA was radiolabeled by including 30 to 300 microcuries of a ³²P nucleoside triphosphate (Amersham Corp., Arlington Heights,

* Present address: Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143, USA

III). Full-length RNAs were purified by electrophoresis through polyacrylamide gels containing 7M urea.

Structure of RNA substrates

RNAs were transcribed from DNA which had been digested with either *Bam*HI or with *Rsa*I. Both types of transcript contain 39 nt of exon 1 and either 53 (NEM 53 RNA) or 84 (NEM 53+31 RNA) nt of intron sequence. Transcripts prepared from *Bam*HI-cleaved DNA contain 122 nucleotides of the second exon, and end after oligonucleotide 48 (Figure 4). Transcripts prepared from *Rsa*I-cleaved DNA contain 40 nucleotides of the second exon, and end at oligonucleotide 34 (Figure 4). 'BamHI RNAs' were used for all experiments unless specified otherwise. Human β -globin RNAs were prepared by transcription of *Bam*HI-cleaved pSP64-H- β -Delta6 (20). These transcripts contain 142 nts of exon 1, 130 nts of the first intron, and 210 nts of exon 2.

Splicing in vitro

Splicing *in vitro* was carried out using crude HeLa cell nuclear extract prepared by the method of Dignam et al. (21), except that $MgCl_2$ was omitted from Buffers A and D. Splicing reaction conditions were similar to those described elsewhere (20). RNA (20 fmol or less) was incubated for various times at 30°C in a 12.5 microliter reaction mixture which contained 7.5 microliter of extract (equivalent to approx. 7×10^5 cells). Each reaction contained 2.6% polyvinyl alcohol, 3.2 mM $MgCl_2$, 0.5 mM ATP and 20 mM phosphocreatine. After incubation, RNA was isolated from the reaction as described by Krainer et al (20).

Analysis of RNA sequence

For RNase T1 fingerprinting (9,10), RNAs were mixed with 20 μ g yeast RNA and digested with 10 units of RNase T1 (Calbiochem, La Jolla, CA) for 30 mins at 37° in 10 mM Tris-HCl, pH 8.0, in a total volume of 3 microliters. The resulting oligonucleotides were separated by high voltage electrophoresis on cellulose acetate at pH 3.5, followed by homochromatography on Cel 300 polyethyleneimine plates (Brinkmann Instruments, Westbury, N.Y.).

Chemical sequencing and modification interference experiments were performed as described (13).

Electrophoresis

RNAs were analyzed by electrophoresis through either 6% or 15% polyacrylamide gels (22), or by electrophoresis through two-dimensional gels, in which the first dimension was 6% and the second dimension was 15% polyacrylamide (3).

RESULTS

Structure of substrates

The structures of the two nematode RNA substrates (designated NEM 53 and NEM 53+31) used in this report are diagrammed in Figure 1. Both RNAs contain a 53 nt intron of the *unc-54* myosin heavy chain gene (intron number 5) flanked by portions of the adjacent exons (see Methods). The two RNAs are identical, except that, in NEM 53+31, a 31 nucleotide sequence from pBR322 has been inserted into the middle of the natural intron, lengthening it to 84 nucleotides. (Sequences are in Figure 4.) Both RNAs were prepared by transcription *in vitro* (6) and possess a 5' terminal cap (7).

FIGURE 1. Splicing kinetics of the natural *C. elegans* myosin intron and of an intron lengthened by 31 nucleotides. NEM 53, the natural precursor, containing an intron of 53 nt; NEM 53+31, identical to NEM 53 except for an insertion of 31 nucleotides of prokaryotic sequence 18 nt upstream of the 3' splice site (see Figure 4 for sequence). 32 P-labeled RNAs were incubated in the mammalian extract for the times indicated (1 to 6 hours). RNA was prepared and analyzed by electrophoresis through a 6% polyacrylamide gel. The positions of the precursors are indicated, as are each of the five bands generated from NEM 53+31 RNA (RNAs A to E). Markers, indicated on the right of the autoradiogram, are an end-labeled *Msp*I digest of pBR322.

Splicing of NEM 53 and NEM 53+31 RNAs

32 P-labeled NEM 53 and NEM 53+31 RNAs were incubated in a crude HeLa cell nuclear extract, under conditions in which splicing of mammalian introns occurs. At 1 hr intervals, samples of the splicing reaction were removed. RNA was prepared from each sample and analyzed by electrophoresis on a 6% polyacrylamide gel. The results are presented in Figure 1.

NEM 53 RNA changes little during the 6 hour incubation: it appears only to be slowly degraded. In contrast, NEM 53+31 RNA generates a collection of discrete bands, with apparent lengths ranging from 43 to 217 nt. These are designated RNAs A to E in Figure 1. Based on their electrophoretic mobility and additional evidence (see below), RNA E, the 43 nt species, likely corresponds to exon 1, and RNA B, the 163 nt species, likely corresponds to ligated exons. The remaining bands cannot readily be assigned as splicing intermediates or products based on their apparent lengths because lariat-containing RNAs migrate anomalously on polyacrylamide gels (3,8).

FIGURE 2. NEM 53+31 RNA generates molecules which exhibit anomalous mobility on high percentage polyacrylamide gels. The protocol is as in Figure 1, except that earlier time points are used, and the RNA is analyzed by electrophoresis through both low (6%) and high (15%) percentage polyacrylamide gels containing 7M urea. Relative to the markers (indicated to the right of each autoradiogram), bands B and E do not change in mobility; bands A, C and D do. The identity of each band on the high percentage gel was determined by T1 RNase fingerprinting (not shown).

To determine which of the RNAs generated from NEM 53+31 RNA contained lariats, we repeated the analysis shown in Fig. 1, but analyzed the RNAs by electrophoresis through both 6% and 15% polyacrylamide gels. RNAs containing lariats migrate more slowly on higher percentage gels than their length would predict. The data, presented in Fig. 2, demonstrate that, relative to markers of known length, all RNAs other than species B and E (the 43 and 163 nt species) migrate more slowly through the higher percentage gel. RNAs B and E migrate identically on the two gels, as expected from linear molecules.

To confirm these data and to identify each band on the higher percentage gel, RNAs A to E were purified, then analyzed by electrophoresis through both high (15%) and low (8%) percentage gels. The data confirm that RNAs B and E do not shift in mobility, but that RNAs A, C and E do, and so are likely to contain lariats (data not shown). The position of each band on the 15% gel, derived from this analysis, is indicated in Figure 2.

NEM 53 RNA does not generate RNA B, which corresponds to ligated exons. Although NEM 53 RNA generates a band that comigrates with Band B from NEM 53+31, RNA fingerprinting

FIGURE 3. NEM 53+31 RNA generates lariat containing molecules, while NEM 53 RNA does not. 32 P-labeled RNA was incubated in the mammalian extract. After 2 hours, RNA was prepared and analyzed by electrophoresis through a two-dimensional gel containing 6% polyacrylamide in the first dimension and 15% in the second. Arrowheads indicate the positions of lariat-containing molecules.

demonstrates that the NEM 53+31 product is ligated exons (see below), but that the NEM 53 product is not (not shown). Thus a nematode intron containing 53 nucleotides does not undergo complete splicing in the mammalian extract. To examine whether it undergoes the first step in splicing, we tried to detect any possible lariat-containing molecules produced from NEM 53 RNA. We used a sensitive two-dimensional gel assay, in which the first dimension is a gel containing a low percentage of polyacrylamide (6%) and the second a gel containing a high percentage (15%; ref. 3). Lariat-containing molecules shift off the diagonal in the second dimension. Analysis of NEM 53+31 RNA confirms the existence of lariat-containing molecules (indicated by arrowheads in Fig 3). In contrast, virtually all the NEM 53 radioactivity is on the diagonal, even following an over-exposure of the autoradiogram. (The small amount of radioactivity off the diagonal near the origin is not reproducible.) These data strongly suggest that NEM 53 RNA, containing the natural 53 nucleotide intron, is inert and cannot support lariat formation in the mammalian extract.

We conclude that while NEM 53+31 RNA appears to be competent for splicing, NEM 53 RNA is not. The two RNAs differ dramatically in their behavior in the HeLa nuclear extract: NEM 53+31 generates discrete and reproducible bands, but NEM 53 RNA does not. Furthermore, NEM 53 RNA fails to undergo even the first step of splicing, as judged by its failure to produce lariat containing molecules. In contrast, NEM 53+31 RNA appears to produce both intermediates (lariat-containing molecules) and final products (ligated exons). These data suggest that NEM 53 RNA undergoes splicing via the same two step mechanism observed with mammalian pre-mRNAs.

Identity of NEM 53+31 products

The data presented in Figs 1 to 3 are consistent with splicing of NEM 53+31 RNA at the same splice sites as are used in *C. elegans*, via a lariat intermediate. To establish whether this inference is correct, we characterized the sequence and structure

FIGURE 4. Sequence and T1 RNase digestion products of NEM 53+31 RNA. Uppercase letters, exon sequence; lower-case letters, intron sequence; shaded box, inserted pBR322 sequence.

of each of the RNAs generated from a NEM 53+31 precursor. These RNAs, designated RNA A through E in Fig. 2, were characterized by RNase T1 fingerprinting and chemical sequencing. We first discuss linear RNAs, then RNAs which contain lariats. For reference, the sequence of NEM 53+31 RNA, and of each of its T1 oligonucleotides, is given in Fig. 4.

Ligated exons (RNA B). To determine the sequence of the 163 nt RNA, NEM 53+31 RNA was 3' end-labeled using ^{32}P Cp and T4 RNA ligase. It was then modified on average once per molecule using diethylpyrocarbonate (DEPC); this reagent modifies purines. The collection of modified RNAs was then incubated in the extract, and the RNA B product purified by electrophoresis. RNA B was then treated with aniline, which breaks the RNA chain at phosphates corresponding to sites of modification. The aniline treated RNA was then analyzed by electrophoresis. For comparison, NEM 53+31 RNA which had not been incubated in the extract, but which had been treated with DEPC and aniline, was analyzed in parallel.

The results are shown in Figure 5. RNA B yields a sequence which corresponds to ligated exons. It yields the same sequencing 'ladder' in exon 2 as does the precursor, but diverges from the precursor 'ladder' at the 3' splice site. Thereafter, the sequence of RNA B is that of exon 1.

In mammalian splicing, the phosphate between exon 1 and exon 2 is derived from the 3' splice site (4). We tested whether this was also true in the splicing of NEM 53+31 RNA, and, in the process, confirmed its identity as ligated exons. NEM 53+31 RNA was prepared using ^{32}P -CTP. RNA B was purified, digested with RNase T1, and the oligonucleotides separated in two dimensions by electrophoresis and homochromatography (9,10). A portion of the resulting fingerprint is shown (Figure

FIGURE 5. RNA B corresponds to exons ligated at the same positions as in *C. elegans*. (A) Chemical sequencing of RNA B demonstrates that it is ligated exons. See text for details. White letters are exon sequences, while black letters are intron sequences. Purines, which are detected as bands in the sequencing ladder in this experiment, are indicated in uppercase letters; pyrimidines are indicated in lower-case letters. Asterisk, 3'-terminal ^{32}P -pCp. (B) RNase T1 fingerprinting of Band B reveals that the phosphate at the ligation site is derived from the 3' splice site. The sequence of a ^{32}P -C labeled precursor is shown. In this experiment, the precursor extended to the end of oligonucleotide 34 (an 'RsaI run-off'). Asterisks indicate the position of a ^{32}P molecule. Exon sequences are contained in boxes, while intron sequences are not. Both the precursor and ligated exons will generate ApApGp after digestion with RNase T1. If the phosphate at the 3' splice site is contained in ligated exons, then this oligonucleotide will be labeled (ApApG*); if not, it will be undetectable. Shown at the bottom are two-dimensional separations of RNase T1 oligonucleotides derived from the Pre-mRNA and from Band B. For reference, we indicate oligonucleotides 5, 6, 9 and 22 (see Figure 4 for sequence). In the analysis of Band B, AAG* contains less radioactivity than the G* (spot 5) because Band B is contaminated by precursor (not shown).

5). If the phosphate at the ligation site were derived from the C at the 3' splice site, then RNA B should generate a new labeled oligonucleotide, ApApG* (see Fig. 5 diagram). This was found to be the case. As expected, RNA B is missing all T1 oligonucleotides of the intron, but contains all oligonucleotides of exons 1 and 2 (not shown).

Exon 1 (RNA E). The 43 nt RNA (RNA E) corresponds in mobility to that expected for exon 1. (The cap structure retards mobility by an amount equivalent to 3 or 4 nucleotides (ref. 11; S.O., unpublished). To confirm its identity, ^{32}P -C-labeled RNA E was analyzed by RNase T1 fingerprinting. As expected, it contains only oligonucleotides derived from exon 1 (data not shown).

Lariat-intron/exon 2 (band A). A partial sequence of RNA A was deduced by chemical sequencing (12,13). 3' end-labeled NEM 53+31 RNA was modified with DEPC, incubated in extract, and RNA A purified by electrophoresis. After cleavage with

FIGURE 6. Band A corresponds to a branched intron/exon 2 intermediate. At the left is shown the experimental strategy. *, 3'-terminal ^{32}P -pCp; X, site of DEPC modification. As indicated, RNAs modified downstream of the branch point will generate a 'normal' sequencing ladder after aniline cleavage (top three molecules in diagram); RNAs modified upstream of the branch point will all generate Y-structures after aniline cleavage (bottom three molecules in the diagram). These Y-structure RNAs will all comigrate in the 6% polyacrylamide gel. Thus the sequencing ladder will contain a large gap between the branch point and 'full length' RNAs. On the right are shown the experimental results. The deduced sequence of the band A is shown (white letters, exon sequences; black letters, intron; uppercase letters, purines [detected in this experiment]; lowercase letters, pyrimidines [not detected]).

aniline, the ladder generated by RNA A was compared to that generated by the pre-mRNA. As illustrated in Fig. 6, the sequence should extend from the 3' terminal label to the site of the lariat, and then should end: RNAs modified downstream of the branch point will generate a sequencing ladder, but RNAs modified upstream of the branch point will all generate Y-structures containing the same number of nucleotides, and so will comigrate in the gel. As shown in Figure 6, the 'ladder' derived from band A includes all of exon 2 and extends to a position 20 nt upstream of the putative 3' splice site. At this position, the ladder stops. We infer that the branch point lies at or between the purines 20 to 26 bases upstream of the 3' splice site. Thus the branch is formed in pBR322 sequences that were inserted into NEM 53 to generate NEM 53+31.

In Figure 6, any DEPC modification that blocked the formation of RNA A would be seen as a gap in the ladder derived from that RNA (13). No such gap is observed at or near the 3' splice site, suggesting that no single purine nucleotide in this region is required for the first step of splicing.

Excised, intact intron (RNA C) and excised, trimmed intron (RNA D). ³²P-C-labeled RNA was used to generate RNAs C and D, which were then analyzed by T1 fingerprinting (Figure 7). ³²P-C-labeled pre-mRNA was analyzed in parallel, to provide markers for each oligonucleotide. RNAs C and D contain oligonucleotides 19, 20, 22 and 23, all of which are derived from the intron, and lack oligonucleotides from exon 1 and 2.

RNA C contains two additional oligonucleotides. One, designated 25_{OH}, likely corresponds to oligonucleotide 25 bearing a 3' hydroxyl rather than a phosphate: 25_{OH} migrates more slowly than 25 both in the first dimension and on 20% polyacrylamide gels, by an amount expected from this single change (data not shown). The other, designated oligonucleotide 24Y, is likely derived from oligonucleotide 24, but contains in addition a branched nucleotide: 24Y migrates much faster than 24 in the first dimension, presumably due to the branch structure. Based on our previous assignment of the branch point (Figure

FIGURE 7. Bands C and D are derived from excised introns. Two-dimensional separations of the RNase T1 oligonucleotides derived from the precursor (which extends to oligonucleotide 34 [an 'RsaI-run-off']), and from Bands C and D are shown. Spots labeled 24Y and 24Y' are presumed to contain the branch point; spot 25_{OH} is spot 25 with a 3' terminal hydroxyl. (See Figure 4 for the sequence of all other oligonucleotides). The identities of spots 24Y, 24Y' and 25_{OH} were confirmed by electrophoresis of the purified oligonucleotides through 20% polyacrylamide gels with and without phosphatase treatment (not shown).

FIGURE 8. NEM 53+31 RNA is spliced with a rate similar to that of human β -globin pre-mRNA. Experimental details as in Figure 1.

6), oligonucleotide 24 is indeed expected to contain the lariat.

RNA D contains one unique oligonucleotide, designated 24Y'. We believe that 24Y' is derived from 24, contains a branched nucleotide, but lacks sequences downstream of the branch point, including oligonucleotide 25. We presume that RNA D is generated by degradation of the excised intron (RNA C) in the extract, consistent with the kinetics with which they appear during the splicing reaction (Figure 2).

Splicing kinetics of nematode vs. mammalian pre-mRNAs

The results presented above demonstrate that NEM 53+31 RNA is spliced by the same mechanism as most mammalian RNAs, using the same splice sites as are used in *C. elegans*. Furthermore, as shown in Figure 8, the rate with which NEM 53+31 RNA is spliced is not less than 50% that of human β -globin pre-mRNA. Splicing intermediates first appear at 15 mins for globin and at 30 mins for NEM 53+31. The final products of splicing first appear at 30 mins in both cases.

DISCUSSION

The experiments reported here demonstrate that a typically short (53 nt) *C. elegans* intron is not spliced in a mammalian extract. Lengthening the intron by 31 nucleotides enables the intron to be spliced efficiently by mammalian splicing components, at the same sites as those that are used in *C. elegans*. The splicing of the lengthened intron involves the formation of lariat-containing intermediates. Thus the *C. elegans* intron behaves the same as

do mammalian introns, in that splicing requires a minimum intron length (14,15; rev. in 1,2)

The branch point that is used in NEM 53+31 lies within the inserted 31 nucleotide pBR322 sequences. In principle, the reason that the lengthened intron can support splicing could be that the inserted prokaryotic sequence contains a specific site that is competent for formation of a branch, rather than it simply lengthening the intron. This alternative explanation is unlikely. Although a mammalian branch point consensus sequence has been derived—YNYURAY—that sequence is only modestly conserved (16), and a variety of sequences support branch formation *in vivo* and *in vitro* (rev. in 1,2). Furthermore the inserted prokaryotic sequence at which the branch forms (GCCUAAC) is no better matched to the mammalian consensus than is the natural intron 19 nt from the 3' splice site (UUUUGAU).

C. elegans introns contain no discernible consensus sequence at the branch point (5). This sequence degeneracy suggests that, in *C. elegans*, recognition of a specific sequence at the branch point is unlikely. Consequently, the failure of NEM 53 RNA to undergo splicing *in vitro* is unlikely to be due simply to lack of a specific *C. elegans* factor which interacts with the branch point.

We conclude that short *C. elegans* introns possess no special properties which obviate the length requirement exhibited by most mammalian introns. In this sense, the *C. elegans* intron we have examined differs from the exceptionally short, small t intron of SV40, which, although only 66 nts in length, supports splicing *in vitro* (17,18). We infer that *C. elegans* splicing components (snRNPs and non-snRNP factors) are able to accommodate short introns either because the components are smaller in size or fewer in number, or both. *C. elegans* possesses all of the known snRNAs involved in splicing (U1, U2, U4, U5 and U6), and all are of approximately the same length as in mammalian cells (23). An *in vitro* splicing system using extracts of *C. elegans* will be needed to determine how these components support the splicing of short introns in *C. elegans*.

ACKNOWLEDGMENTS

We are grateful to Brian Carr for the construction of plasmid DNAs used in this study. We also thank the members of the Wickens lab for their helpful comments and advice. This work was supported by an NIH grant (GM31892), a Research Career Development Award (GM00521), and a Steenbock Career Development Award to M.P.W.

REFERENCES

1. Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S., and Sharp, P.A. (1986) *Ann. Rev. Biochem.* 55, 1119–1150.
2. Green, M.R. (1986) *Ann. Rev. Genetics* 20, 671–708.
3. Ruskin, B., Krainer, A.R., Maniatis, T., and Green, M.R. (1984) *Cell* 38, 317–331.
4. Padgett, R.A., Konarska, M., Grabowski, P.J., Hardy, S.F., and Sharp, P.A. (1984) *Science* 225, 898–903.
5. Blumenthal, T.J., and Thomas, J. (1988) *Trends in Genetics* 4, 305–307.
6. Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K., and Green, M.R. (1984) *Nuc. Acids Res.* 12, 7035–7056.
7. Konarska, M.M., Padgett, R.A., and Sharp, P.A. (1984) *Cell* 38, 731–736.
8. Grabowski, P.J., Padgett, R.A., and Sharp, P.A. (1984) *Cell* 37, 415–427.
9. Barrell, B.G. (1971) In *Cantonit, G.L. and Davies, D.R. (eds.), Proc. Nuc. Acids Res.*, Harper and Row, New York, 2, 751–759.
10. Brownlee, G.G. (1982) *Determination of sequences in RNA*. Elsevier/North-Holland Pub. Co., New York.
11. Nielsen, D.A. and Shapiro, D.J. (1986) *Nuc. Acids Res.* 14, 5936.

12. Peattie, D.A. (1979) *Proc. Natl. Acad. Sci.* 76, 1760–1764.
13. Conway, L. and Wickens, M. (1987) *EMBO J.* 6, 4177–4184.
14. Wieringa, B., Hofer, E., and Weissmann, C. (1984) *Cell* 37, 915–925.
15. Ruskin, B., Greene, J.M., and Green, M.R. (1985) *Cell* 41, 833–844.
16. Reed, R., and Maniatis, T. (1985) *Cell* 41, 95–105.
17. Fu, X., Colgan, J.D., and Manley, J.L. (1988) *Mol. Cell Biol.* 8, 3582–3590.
18. Noble, J.C.S., Prives, C., and Manley, J.L. (1988) *Genes and Dev.* 2, 1460–1475.
19. Karn, J., Brenner, S., and Barnett, L. (1983) *Proc. Natl. Acad. Sci.* 80, 4253–4257.
20. Krainer, A.R., Maniatis, T., Ruskin, B., and Green, M.R. (1984) *Cell* 36, 993–1005.
21. Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983) *Nuc. Acids Res.* 11, 1475–1489.
22. Sanger, F., and Coulson, A.R. (1978) *FEBS Letters* 87, 107–112.
23. Thomas, J.D., Conrad, R.C. and Blumenthal, T. (1988) *Cell* 54, 533–539.