Volume 3 · Number 1 · 1993
GENE EXPRESSION

Selection between a natural and a cryptic 5′ splice site: a kinetic study of the effect of upstream exon sequences

Lionel Domenjoud, Liliane Kister, Hélène Gallinaro, and Monique Jacob

Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS, Unité 184 de Biologie Moléculaire et de Génie Génétique de l'INSERM, Institut de Chimie Biologique, Faculté de Médecine, Strasbourg, France

To study the mechanism of selection of 5′ splice sites, we first analyzed the in vitro time course of appearance of intermediates and products of splicing at a natural and at a cryptic 5′ splice site. Our model system was a transcript derived from the early transcription unit 3 of adenovirus-2 harboring a cryptic 5′ splice site Dcr1, 74 nucleotides downstream of the natural site D1. When studied in isolation, the two sites have different kinetics of splicing, Dcr1 being spliced markedly more slowly than D1. The upstream exon, shown elsewhere to have a positive effect on the selection of D1, has no influence on these kinetics; thus, it does not affect selection by modifying the kinetics of splicing. Nevertheless, this exon is of crucial importance for the exclusive selection of D1. We demonstrate that the cryptic site is recognized in all cases, but that exons harboring a potential stem-loop structure (HP1) prevent Dcr1 usage. The data suggest that the upstream exon sequences play the role of a cis-acting selector for the natural 5′ splice site. The intrinsically rapid and efficient kinetics of splicing at the natural site and the selector function of the exon sequence may result in the exclusive use of the D1 site in the natural context.

One of the major problems in understanding the mechanism of splicing and its regulation is to know how splice sites are chosen. There are at least two levels of selection of splice sites. First, selection is required to discriminate between "natural" splice sites, whose use will generate functional mRNA, and "cryptic" splice sites, which are potentially usable sites but which generate aberrant messages. Such cryptic sites are found frequently in RNAs (Jacob and Gallinaro, 1989). Second, selection is required when two natural sites give rise to two different functional mRNAs. In this case of alternative splicing, tissue-specific or stage-specific factors are necessary for selection, which is not the case for the silencing of cryptic sites.

Our knowledge of the molecular mechanism of selection is quite limited. In the case of alternative splicing, the importance of splice site sequence and of sequence context has been stressed (see McKeown, 1990, for review; Cote et al., 1990; Talerico and Berget, 1990; Clouet d'Orval et al., 1991; Guo et al., 1991; Hedley and Maniatis, 1991; Ryner and Baker, 1991; Nemeroff et al., 1992). Similarly, we have shown that the selection between a natural and a cryptic 5′ splice site (or donor site) depends upon both the relative strength of the two sites and the sequence context (Domenjoud et al., 1991). The natural upstream exon is an important element to be considered in our transcript. When its size is reduced to less than 120 nucleotides,

Received July 1, 1992; revision accepted September 17, 1992.

Correspondence: Monique Jacob, Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS, Unité 184 de Biologie Moléculaire et de Génie Génétique de l'INSERM, Institut de Chimie Biologique, Faculté de Médecine, 11 rue Humann, 67085 Strasbourg Cédex, France Tel 33-88-37-12-55 Fax 33-88-37-01-48

 1052-2166/93/0301/083-12$2.00

the efficiency of splicing at the natural site is reduced, while the cryptic site is activated.

The effect of various parameters on splicing has been studied most often by determining splicing efficiency, which is the final result of a series of reactions. To obtain a better understanding of the underlying molecular mechanisms, we have adopted an approach that was expected to be more informative—the study of the kinetics of splicing. The two steps of the splicing reaction which transform the precursor first into intermediates (exon 1 and IVS–exon 2) and then into products (exon 1–exon 2 and IVS) are well known (see Green, 1991, for review). These two reactions occur in the spliceosome, a dynamic structure that is assembled in a stepwise manner and modified as the reaction proceeds (see Green, 1991, and Guthrie, 1991, for review). Thus, multiple reactions occur involving snRNPs and proteins, besides the precursor RNA. They do not by themselves modify the RNA molecules but may have repercussions on the kinetics of splicing.

Here, we report that the natural and the cryptic sites, in the absence of each other, have markedly different kinetics of splicing. The spliceosome is assembled more slowly at the cryptic site than at the natural site, and later changes in the conformation of the spliceosome are also impaired. The upstream exon structure does not alter these kinetics. The situation is dramatically different in the natural conditions of cis-competition, wherein upstream exon sequences confer resistance to competition to the natural site. Thus, the natural site predominates partly for kinetic reasons and partly because of the selector effect of the upstream exon.

Materials and methods

Construction of mutants

All DNA fragments to be transcribed were inserted in the polylinker of plasmid pSP64 (Melton et al., 1984). The wild-type fragment was a Pst I–Sac I fragment of 846 bp from the proximal part of the adenovirus-2 (Ad2) early region 3 transcription unit (E3), as described previously (Domenjoud et al., 1991). It contains the E3 intron 1 (395 bp) flanked by 259 bp of exon 1 and 192 bp of exon 2.

5′ deletion mutants (labeled 5′dl) were prepared by the action of the exonuclease Bal 31. For NY mutants, a synthetic oligonucleotide harboring a Hind III recognition site at the 3′ end and a Pst I site at the 5′ end was made double-stranded using a primer hybridizing to the Hind III site and the Klenow enzyme. After digestion with Hind III and Pst I, the DNA fragment was inserted in the residual polylinker of a 5′dl50 mutant (a 5′ deletion mutant which contains 7 nucleotides of the natural exon 1), opened by the same enzymes.

The transcribed version of the newly created exon was self-complementary for NY54:

5′-gaauacaagcuu**CGCGCGUACUGACGCGCG**cugcaggucGGCUCCG-3′

and linear for NY22:

5′-gaauacaagcuuAAUAAUACUGACGCGCGcugcaggucGGCUCCG-3′

(lower case letters: polylinker; upper case letter: from 5′ to 3′: synthetic oligonucleotide and natural sequence; bold characters: nucleotides involved in base-pairing).

Mutants labeled pm modify the sequence of the donor site D1 or Dcr1 and were originally prepared from the wild-type transcript by site-directed mutagenesis, according to Inouye and Inouye (1987). To introduce these point mutations in 5′dl or NY mutants, we used an exchange method. Restriction fragments harboring either D1 (Sac II–Apa I) or Dcr1 (Apa I–EcoR I) were excised from 5′dl or NY clones and replaced by the homologous fragments including the mutated sites from other transcripts. The sequence of the mutants was determined by using the dideoxynucleotide sequencing procedure (Winship, 1989).

RNA synthesis and in vitro splicing

Methods were as previously described (Domenjoud et al., 1991). Briefly, the pSP64 plasmids containing the inserted E3 fragments were linearized by EcoR I and transcribed in the presence of [α-^{32}P]CTP. With 100,000 cpm Cerenkov of RNA precursor per assay, the specific activity of the transcripts was sufficient to detect splicing intermediates clearly.

For in vitro splicing, the conditions were as established previously, except that Mg^{++} and polyvinyl alcohol concentrations were raised to 2.5 mM (instead of 2 mM) and 3.1% (instead of 2.6%), respectively. Three preparations of HeLa cell nuclear extract were used in these experiments. Splicing efficiency of the wild-type transcript was 70–75% in all three cases. Splicing products were analyzed in 5% polyacrylamide

gels (acrylamide/bisacrylamide, 20/1, w/w) containing 8 M urea. The bands corresponding to the different molecules were cut out and their radioactivity counted. The radioactivity of blank gel fragments at the same level as the molecules was also counted and subtracted after normalization to the same amount of total radioactivity in the lane. In our hands, this method was more reliable than the densitometric analysis of the autoradiograms. The radioactivity of intermediates (i: IVS–exon 2 and exon 1), products (p: IVS and exon 1–exon 2), and remaining precursor (pr) could thus be determined. As exon 1 is very short in several transcripts and runs out of the gels, our determinations were derived from IVS–ex2 measurements, irrespective of the transcript. The value of i was then calculated, taking into account the number of cytosines in IVS–ex2 and exon 1. The proportion of intermediates, products, and remaining precursor at each time was $(i \times 100)/(i + p + pr)$, or $(p \times 100)/(i + p + pr)$, or $(pr \times 100)/(i + p + pr)$, respectively. All experimental values are the average of duplicates. For the kinetic studies, these proportions were determined after 0, 10, 15, 20, 25, 30, 40, 50, 60, 90, 120, 150, 180, and 240 minutes of incubation for the same transcript. The time course curves were drawn, and the time lags were estimated as the intersection of the maximum slope of the curves with the axis of abscissae. To optimize comparisons, transcripts to be compared were labeled using the same CTP batch; they were used fresh (within a week) and spliced in parallel with the same nuclear extract preparation.

Results

Characteristics of the transcripts

As in previous work, we have studied the splicing of the first intron of the early region 3 (E3) from adenovirus-2. The characteristics of the wild-type transcript used for our in vitro studies have already been described (Domenjoud et al., 1991). Briefly, with the exception of vector polylinker sequences, it contains only natural contiguous E3 sequences and all the cis-acting elements required for an efficient splicing at the natural donor site D1 and for the silencing of a cryptic donor site Dcr1 located 74 nucleotides downstream. The hybrids that can be formed between the 5′ extremity of U1 RNA and D1 or Dcr1 are 7 or 6 bp long, respectively; this rules out the possibility that the site's strength is the only parameter of selection of D1. Dcr1 can be induced either by suppression of D1 or by modifications of the upstream exon sequence. The same acceptor site is used with both donor sites.

To establish the kinetics of splicing independently at the two sites, D1 or Dcr1 were suppressed by mutation of their first intronic G (Fig. 1B). To test the influence of exon sequences, 5′ deletion or 5′ deletion-insertion mutants were constructed so that the upstream exon has different primary or secondary structures (Fig. 1A). The presence of a potential hairpin structure HP1 in the 120 nucleotide exonic segment,

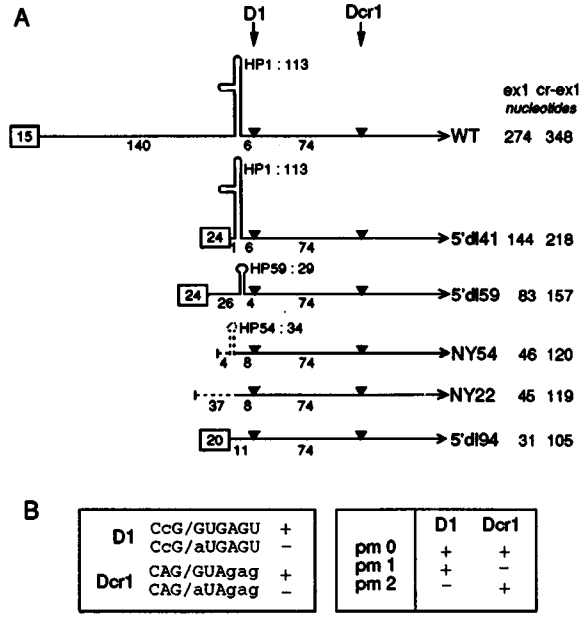

Figure 1. Schematic representation of the 5′ part of the transcript. **A.** The various exons. The presence of a potential hairpin loop structure HP1 in the WT and 5′dl41 exons was demonstrated previously (Domenjoud et al., 1991). The residual exon of 5′dl59 is able to fold in a smaller hairpin loop structure (HP59), as shown by an enzymatic study (our unpublished data). The sequence and potential secondary structure (HP54) of NY54 and NY22 (dotted lines: inserted artificial sequence) are described in Materials and Methods. The size in nucleotides of the plasmidic sequences (boxed) and of the various exon or intron segments is indicated. On the right, the total length of the exon sequence in the case of splicing at D1 (ex1) or Dcr1 (cr-ex1) is also shown. **B.** The donor site sequences. Left: the sequences of the natural (+) and suppressed (–) D1 and Dcr1 (lower case letters: nucleotides that do not hybridize to U1 RNA). Right: the combinations of D1 and Dcr1 in the pm series of transcripts. pm0 is the natural combination.

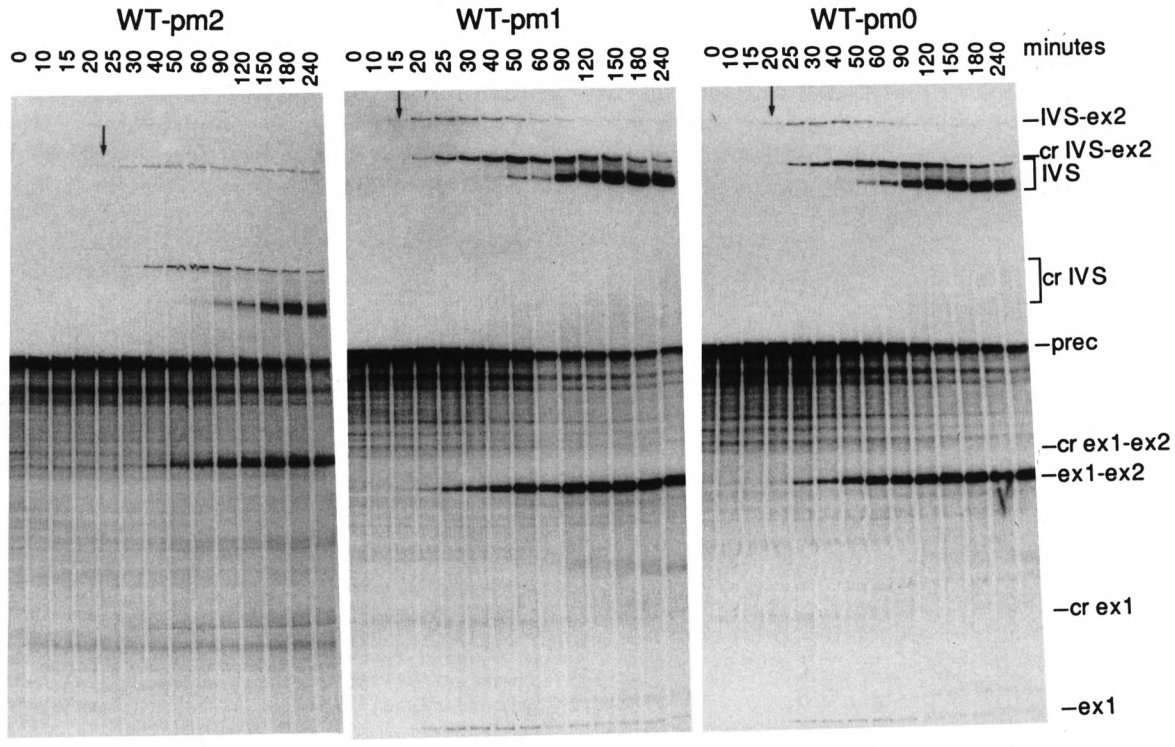

Figure 2. Time course of appearance of intermediates and products of splicing of transcripts WT-pm1, WT-pm0, and WT-pm2. The autoradiograms of the gels are shown. The three transcripts were analyzed in the same experiment (i.e., labeled with the same CTP batch and spliced with the same nuclear extract). The identity of the various RNA species (cr: cryptic), as indicated in the right margin, was documented previously (Domenjoud et al., 1991). The free lariat introns (IVS) always appear as a large heterogenous band due to the trimming of the tail of the lariat that occurs systematically in our in vitro studies. The figure shows that this trimming is time-dependent. Examination of the evolution of IVS–ex2 clearly indicates that the intermediate accumulates and is used rapidly in the WT-pm1 and WT-pm0 transcripts, whereas it forms a plateau in the WT-pm2 transcript, as illustrated in the graphs of Figure 3. The arrows indicate the lane (the time) at which IVS–ex2 is already clearly visualized; a faint band, not clearly visible on the reproduction, is sometimes detected in the preceding lane by the direct examination of the gels and also by counting the radioactivity at this level. The signal levels at maximum were 315 cpm (at 25 minutes) for WT-pm1, 228 cpm (at 30 minutes) for WT-pm0, and 153 cpm (at 40 minutes) for WT-pm2. After 4 hours of incubation, the signals were 34, 45, and 94 cpm, respectively.

shown to be necessary for the exclusive selection of the natural site, has been demonstrated previously (Domenjoud et al., 1991). HP1 is conserved in the 5′dl41 exon that lacks the exonic upstream sequences. A further truncation of exon 1 disrupts HP1, but a small artificial hairpin, HP59, can be constructed from the residual sequence, and an enzymatic study showed that HP59 exists in the 5′dl59 transcript in solution (H. Gallinaro, unpublished data). The NY54 and NY22 exons are artificial exons (see Materials and Methods); NY54 can be folded in a small hairpin structure, HP54. The transcripts are named first by the exon designation (Fig. 1A), followed by the D1–Dcr1 combination (Fig. 1B).

Splicing at the natural donor site in the absence of the cryptic site

The pm1 series of transcripts, where the cryptic site was suppressed by a point mutation, was used to study splicing at the natural site. A typical time course of appearance of intermediates and products of transcript WT-pm1 is shown in Figure 2. The other data are summarized graphically in Figure 3A and B. Note that the scale of ordinates is expanded 10 times for intermediates relative to products.

The curves for the WT-pm1, 5′dl41-pm1, 5′dl59-pm1, NY54-pm1, and NY22-pm1 transcripts are superimposable, that is, they do not differ from each other more than do those from

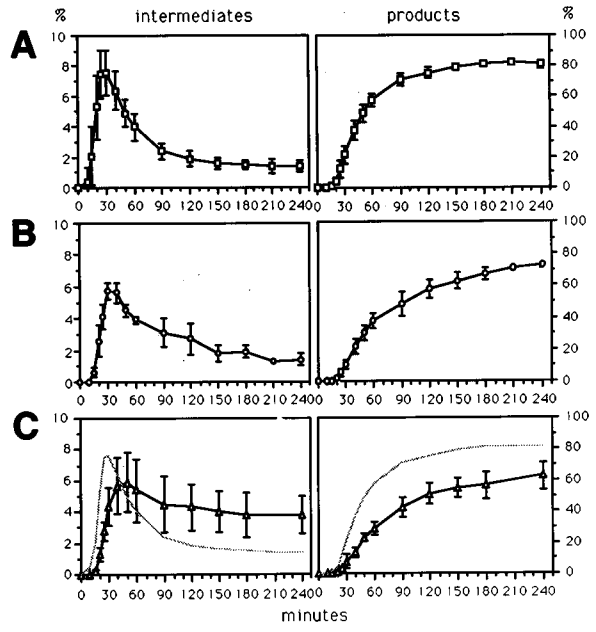

Figure 3. Time course of appearance of intermediates and products of splicing. Means and standard errors are shown. Note the change of the scale of ordinates between intermediates and products. **A.** Splicing at D1 of the WT-pm1, 5′dl41-pm1, 5′dl59-pm1, NY54-pm1, and NY22-pm1 transcripts (8 determinations). **B.** Splicing at D1 in the 5′dl94-pm1 transcript (3 determinations). **C.** Splicing at Dcr1 of the WT-pm2, 5′dl41-pm2, 5′dl59-pm2, and 5′dl94-pm2 transcripts (7 determinations). The light gray curves reproduce the curves of **A.**

the same transcript in different gels or experiments; therefore, only the composite curves are presented (Fig. 3A). The intermediates are detected after a time lag of about 12 minutes (see also arrow in Fig. 2). Thereafter, a linear accumulation rising to a maximum at about 30 minutes is observed, followed by an exponential decrease. After 3–4 hours of incubation, a plateau is reached with 1–2% of the transcripts remaining in the form of intermediates. The products of the reaction become detectable 5 minutes later than the intermediates. Thereafter, they accumulate to reach a plateau representing the amount of fully spliced transcripts, which at 3–4 hours is equivalent to 79–85% of the initial transcript.

The transcript 5′dl94-pm1 behaves differently (Fig. 3B). The lag periods are longer (+2 minutes), and the peak of intermediates is lower and appears flattened, while the plateau of intermediates remains unchanged at 1–2%. The amount of products formed at 3–4 hours is lower and reaches only ~70% of the initial transcript. It is possible that the reactions are not entirely terminated at 4 hours.

Thus, the shortening of exon 1 from 274 to 45 nucleotides does not affect the rate or efficiency of splicing; a fortiori, there is no influence of exon structure. In contrast, there is a dramatic transition between the transcripts harboring the 45 nucleotide NY22 exon and the 31 nucleotide 5′dl94 exon. Since these two exons differ only by the presence or absence of an artificial sequence that is not likely to play a specific role, we infer that a 31 nucleotide exon is not sufficient to insure an optimal rate of splicing, while a 45 nucleotide exon is sufficient. The minimum upstream exon length required for full splicing efficiency of other transcripts has been proposed to be 30–50 nucleotides (Parent et al., 1987; Adami et al., 1989; Mayeda and Oshima, 1990). It seems plausible that a minimum exon size is required for the optimal binding of splicing factors at the donor site, and that this size is between 31 and 45 nucleotides in the case of our transcript. Thus, provided that exon length is sufficient, the data demonstrate that the previously observed positive influence of exon 1 on the selection of D1 (Domenjoud et al., 1991) is not due to an enhancing effect on the kinetics or efficiency of splicing at this site.

Splicing at the cryptic donor site in the absence of the natural site

The effect of exon 1 on splicing at Dcr1 was studied with transcripts WT-pm2, 5′dl41-pm2, 5′dl59-pm2, and 5′dl94-pm2, in which D1 was suppressed by a point mutation at G1 (Fig. 1). The time course of appearance of intermediates and products was similar, and the composite curves are shown in Figure 3C. The curve of pm2 intermediates is clearly different from that of pm1. The time lag is 16 minutes instead of 12 (compare also the position of arrows in Fig. 2). No sharp peak is formed, partly because the amount of intermediates remaining after 3–4 hours is 2–3 times higher for Dcr1 than for D1. The appearance of products is also retarded, and their final plateau is lower (55–65% instead of 79–85%). The similarity of the kinetics of splicing irrespective of exon 1 structure indicates that this exon does not favor selection of D1 by inhibiting splicing at Dcr1. It

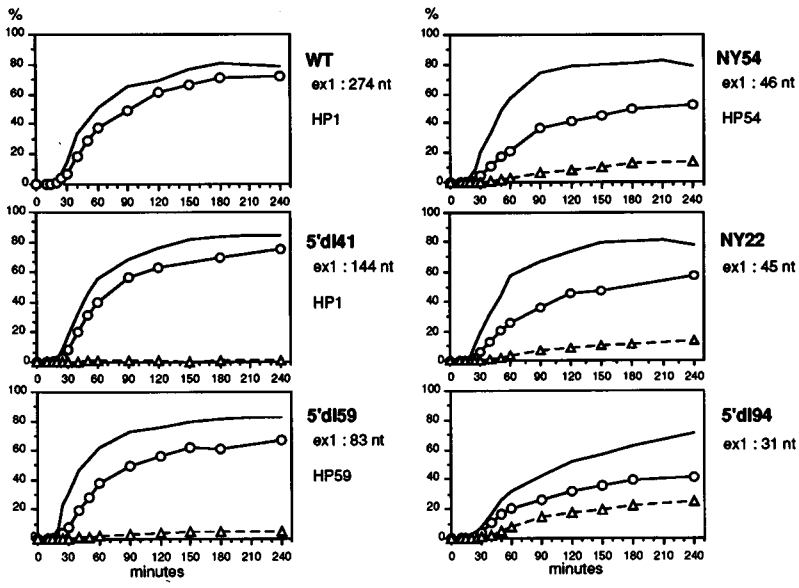

Figure 4. Influence of exon 1 on splicing at D1 and Dcr1. The curves of appearance of the products of splicing are shown for each transcript. Solid lines with circles: splicing at D1 in the pm0 transcripts; dashed lines with triangles: splicing at Dcr1 in the pm0 transcripts; solid lines without symbols: splicing at D1 in the pm1 transcripts (data summarized in Figs. 3A and B). The size of exon 1 and the presence of a hairpin structure (see Fig. 1) are indicated for each transcript.

is worth noting that, in contrast to the pm1 series, there is no significant reduction of splicing efficiency in the 5'dl94-pm2 transcript relative to the other transcripts of the pm2 series. This may be related to the size of the 5'dl94 exons: 105 nucleotides for the cryptic exon (pm2) against 31 nucleotides for the natural exon (pm1).

A comparison of the D1 and Dcr1 data shows that they are much more dispersed for Dcr1 than for D1 (compare standard errors in Fig. 3A and C). As conditions were the same and the experiments often performed in parallel, the dispersion of the results is likely to be a particular feature of Dcr1. We assume that Dcr1, which is not an authentic donor site and has a different sequence context than D1, might be more sensitive to small differences of the in vitro system. It is evident, however, that splicing at Dcr1 is less rapid and efficient than splicing at D1.

Cis-competition: splicing efficiency at the natural and cryptic 5′ splice sites

The time course of appearance of intermediates and products of splicing at D1 and Dcr1 was established for the various transcripts harboring the two sites (pm0). In Figure 4, we compare the appearance of fully spliced products from the various pm0 transcripts to that from the homologous pm1 transcripts in which Dcr1 was suppressed. With the WT and 5'dl41 exons, splicing efficiency at D1 is reduced when the cryptic site is present (from 83 to 75% of the initial transcript, after 3–4 hours of incubation). A difference is also observed in the appearance of intermediates (Fig. 5); in particular, the time lag is lengthened, while the height of the peak and rate of the reactions are reduced. As these differences are small, we performed an experiment in which transcripts WT-pm1 and

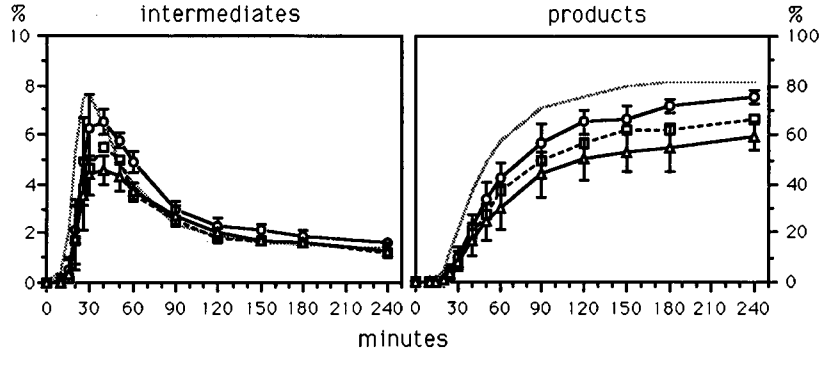

Figure 5. Effect of exon size and structure on splicing at D1 in pm0 transcripts. Curves showing the time course of appearance of intermediates and products are shown. Circles: mean and standard errors for WT-pm0 and 5'dl41-pm0; squares with dashed lines: 5'dl59-pm0; triangles: mean and standard errors for NY54-pm0 and NY22-pm0. Light gray curves: splicing in the pm1 transcripts is shown for comparison.

Table 1. Some characteristics of the time course curves.

The time lags and slopes of the curves were determined graphically. Lag i: the time required for reactions *a*, *b*, and *c*—i.e., for spliceosome assembly (see Fig. 6); lag p − i : the time required for reactions 1, *d*, and *e*—i.e., for reaction 1 and transition to reaction 2. Rate: the overall rate of splicing as estimated from the maximum slope of the curve of products. Distribution: the proportion of fully spliced transcripts (p), of intermediates (i), and of unused precursor were determined at 3–4 hours, i.e., at the end of the reaction. The sum of the amounts of fully spliced transcripts (p) and of intermediates (i) relative to the initial precursor was indicative of the proportion of precursors committed to splicing either at D1 (pr1) or at Dcr1 (pr2). The unspliced precursor was considered to be in inactive complexes (pr0). When individual values from different transcripts were close and the number of experiments sufficient (n > 2), the means and standard errors are presented. Column n: number of determinations. nd: non-determined because of a low amount of material.

		Splicing at D1				*Splicing at Dcr1*						
		Lag (minutes)		*Rate (% per*	*i1*	*Lag (minutes)*		*Rate (% per*	*i2*	*Distribution (%)*		
Transcripts	*n*	*i*	*p − i*	*10 min)*	*(%)*	*i*	*p − i*	*10 min)*	*(%)*	*pr1*	*pr2*	*pr0*
WT-pm1	2											
5′dl41-pm1	1	11.9	4.8	15.5	1.4					83.0		17.0
5′dl59-pm1	2	±1.5	±1.0	±2.4	±0.4					±1.8		±1.8
NY54-pm1	1											
NY22-pm1	2											
WT-pm0	2	15.0	5.5	11.8	1.7	–	–	–	–	75.8		24.2
5′dl41-pm0	1	14	8	11.3	1.8	nd	nd	nd	0.3	74.1	1.6	24.3
5′dl59-pm0	1	14	7	9.7	1.4	nd	nd	nd	0.7	65.6	5.5	28.9
NY54-pm0	2	15.5	7.0	7.1	1.6	19.2	14.5	1.0	0.9	58.4	12.6	29.0
NY22-pm0	2	16.0	5.0	7.8	1.4	18.5	12.5	1.1	1.0	59.1	12.9	28.0
WT-pm2	3											
5′dl41-pm2	1					15.9	6.9	8.9	3.7		64.2	35.8
5′dl59-pm2	1					±1.3	±3.1	±1.5	±1.3		±9.2	±9.2
5′dl94-pm2	2											
5′dl94-pm0	1	15	8	5.6	1.2	19	8	2.4	1.7	41.3	25.4	33.3
5′dl94-pm1	3	14.0	5.3	9.9	1.6					71.9		28.1
		±1.7	±1.2	±1.2	±0.4					±0.9		±0.9

WT-pm0 were studied in parallel and confirmed that the decrease was not due solely to experimental variations (Fig. 2). This difference between the pm1 and pm0 transcripts was further confirmed in several other experiments in which only splicing efficiencies were determined. We conclude that the difference is significant, that splicing at D1 is impeded by the presence of Dcr1, and therefore that Dcr1 is recognized in our in vitro system.

With the WT exon, there is no detectable splicing at Dcr1, although we estimate that we would have detected its products if 0.5% of the precursor had been spliced at Dcr1. However, the 5′dl41 exon, which lacks 139 E3 nucleotides upstream of HP1, gives rise to a low amount of splicing at Dcr1 (1–2%) relative to the WT exon. Thus, these 139 nucleotides are likely to contain an element partially responsible for the silencing of Dcr1.

The differences between the pm1 and pm0 transcripts are accentuated when exon 1 is further truncated. In 5′dl59, the rates of the reactions are reduced relative to WT-pm0 and 5′dl41-pm0 (Fig. 4), and the 3–4 hour splicing efficiency is 66% as opposed to 74–76%. Simultaneously, the transcripts become more permissive for splicing at Dcr1 (5%). A more marked effect is observed with transcripts NY: splicing efficiencies at D1 and Dcr1 are 59% and 13%, respectively (Table 1). The presence or absence of the small stem-loop structure HP54 does not influence splicing of the NY transcripts. A decrease in the formation of intermediates is also observed, more marked for NY54 and NY22 than for 5′dl59 (Fig. 5). As the time course of splicing of the pm1 and pm2 transcripts is not modified with the same exon 1 changes (Fig. 3), we conclude that the 120 nucleotide segment, containing HP1 and located at the 3′ extremity of exon 1, influences splicing by interfering during cis-competition.

The 5′dl94 truncated exon is too short to allow optimal splicing efficiency in the pm1 transcript (Fig. 3B). The presence of Dcr1 in the 5′dl94-pm0 transcript promotes a further decrease in splicing at D1 (from 72 to 41%) and a marked induction of splicing at Dcr1 (25%).

Thus, when splicing at D1 is impaired due to the suboptimal size of exon 1, the transcript becomes more permissive to splicing at Dcr1. The effects of suboptimal exon size and of cis-competition appear additive.

In summary, the data show that upstream exon sequences, which have no influence on splicing at D1 in the absence of Dcr1, clearly affect this splicing in the presence of the cryptic site. Thus, exon 1—in particular, the sequence harboring the potential stem-loop structure HP1—confers resistance to competition to the natural site D1 and therefore acts as a selector for this site.

Variations in the kinetic curves and their interpretation

Comparison of the gels (Fig. 2) and kinetic curves (Figs. 3, 4, and 5) of the various transcripts reveals marked differences in the appearance and utilization of splicing intermediates, as well as in the appearance and yield of fully spliced products. This repercussion of the structure of the transcripts on the kinetics of splicing is likely to derive from its influence on one or the other of the various steps in the assembly of splicing complexes. In Figure 6, we present a simplified scheme of mammalian splicing complex assembly that should be helpful for interpreting the kinetic study.

The precursor is first coated with proteins (reaction *a*). Then, these early complexes are committed to splicing (Legrain et al., 1988; Michaud and Reed, 1991; reaction *b*), giving rise to commitment complexes (pr1 and pr2, for D1 and Dcr1, respectively). Thereafter, active type 1 spliceosomes are made from commitment complexes (reaction *c*) in which splicing reaction 1 proceeds. This period, rich in crucial events as far as spliceosome assembly is concerned, appears as the time lag of the curve of intermediates in our kinetic studies. The length of this time lag differs according to the transcripts, as illustrated for instance in Figure 2. There is clearly a delay in the appearance of D1 intermediates in WT-pm0 or of Dcr1 intermediates in WT-pm2 relative to D1 intermediates in WT-pm1 (arrows). The time lags were estimated for the various transcripts (Table 1). The data show that the time lag of D1 intermediates is lengthened, relative to the pm1 series, by about 2–3 minutes when exon 1 is truncated (5′dl94 exon) or when Dcr1 competes with D1 (pm0 transcripts). The time lag of the Dcr1 intermediates is longer than that of the D1 intermediates by about 4 minutes (pm2 series) and is further lengthened in the presence of D1. Thus, spliceosome assembly is slowed down when exon 1 is too short, or in the conditions of cis-competition. Furthermore, spliceosomes may assemble more or less rapidly, according to the donor site. The individual reactions that are affected by the various structural changes of the precursor remain to be determined.

Once splicing reaction 1 is completed, some conformational changes are necessary to allow the onset of the second splicing reaction (reactions *d* and *e*, Fig. 6). This is suggested by the difference in the time required for the appearance of intermediates and products, which is about 5 minutes in the WT-pm1 transcripts (Table 1, lag p − i). This time is not significantly modified upon truncation of exon 1 (5′dl94-pm1) and slightly increased in the presence of the cryptic site (pm0 transcripts). For splicing at Dcr1, lag p − i is 2 minutes longer than at D1 and is further lengthened in the pm0 transcripts. Though small, these differences were reproducible and indicate that splicing can be influenced not only by defective early spliceosome assembly, but also by some defects in later reactions that can be detected by the analysis of the time course curves.

The curves of intermediates never tend to zero, and a plateau is reached in all cases. This shows that intermediates persist under a form different from spliceosomes of type 1 and 2 proper—a form we propose to designate "intermediate complexes." The plateau of the curve indicates that equilibrium is reached between formation and utilization of the intermediates. The level of the plateau is 2–3 times higher for Dcr1 than for D1 (Table 1, i2 and i1), which suggests that utilization of the Dcr1 intermediates is slower relative to formation in the pm2 transcripts than that of D1 in the pm1 transcripts, in agreement with the lengthening of the time lag p − i. The level of the plateau does not significantly change as long as the formation of intermediates is active enough (compare i1 of the pm1 and pm0 transcripts in Table 1) and is reduced when the rate of formation is low, as expected (compare i2 of the pm2 and pm0 transcripts). The persistance of intermediate complexes when no more products are made may be due to conformational changes of these

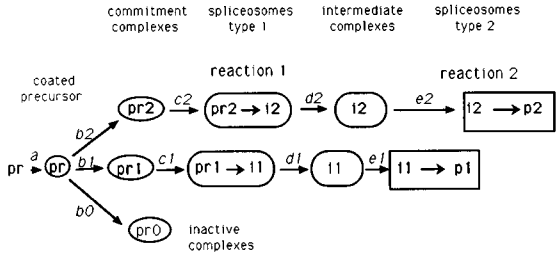

Figure 6. Schematic representation of splicing complexes in the course of splicing. pr: precursor; i: intermediates; p: products. Reactions 1 and 2 are the two splicing reactions. The different frames correspond to different complexes. Reactions *a* to *e* do not affect the RNA molecules but do affect the composition or conformation of the complexes; they include an undetermined number of individual events. 1 stands for D1 and 2 for Dcr1; 0 indicates inactive complexes. The differences in lengths of the arrows in the D1 and Dcr1 pathways indicate the reactions that are lengthened for Dcr1 relative to D1 (*b2*, *b0*, and *e2*; not to scale), as determined in the experiments of Figure 2.

complexes or to an inhibition of reaction *e*, whatever the cause.

As not all the precursor is engaged in splicing, we postulate that "inactive complexes" (prO) can be assembled in addition to the commitment complexes. We observe a correlation between the increase of the proportion of prO and that of the time lag of intermediates (Table 1). Therefore, we propose that the assembly of prO follows kinetic rules of its own (reaction *b0* in Fig. 6), and that competition for the precursor occurs during early assembly. It cannot be ruled out that the assembly of prO is partially reversible, which may explain the low increase of the plateau of the curves of products in some transcripts.

Splicing efficiency can be estimated as the proportion of precursor engaged in splicing, that is, the sum of intermediates and products formed at the end of the reaction. These proportions, after 3–4 hours of incubation, are indicated in Table 1 (pr1 for D1 and pr2 for Dcr1). They are representative of the proportion of precursor that was committed to splicing at D1 and Dcr1 during early spliceosome assembly (reaction *b*). The exact values may be somewhat different, as when the reactions are not entirely terminated at 3–4 hours; however, such values are indicative of earlier events. The maximal slope of the curve of products (which is expressed as the percentage of fully spliced transcripts made in 10 minutes) can be considered an index of the overall rate of the splicing reaction and varies, as do pr1 and pr2. As already stated, the rate and efficiency of splicing at D1 are dependent on the structure of exon 1 in the pm0 transcripts and independent of it in the pm1 transcripts.

Thus, the analysis of the time course curves allows the specification of the period at which a given structural feature may modify the use of a 5′ splice site and may thereby orient further investigation. It also allows a relatively easy comparison of transcripts harboring several splice sites and various modifications.

The function of the upstream exon

One of the major results of the kinetic study was to show that two 5′ splice sites, whose capacity to hybridize to U1 RNA is similar, may have very different kinetics of splicing when studied in the absence of each other. The difference is due both to a slower early spliceosome assembly and to a slower transition from spliceosomes carrying out splicing reaction 1 to those carrying out reaction 2 (Fig. 3 and Table 1). These characteristics of the time course curves lead us to predict that, when the two sites are present in a same transcript (pm0), D1 will be used much more rapidly than Dcr1. The 4 minute difference in the time lag of appearance of intermediates allows the assembly of a relatively large number of complexes committed to splicing at D1 (pr1) and a concomitant increase in the number of inactive complexes (prO) in which the precursor is sequestered. Thus, the proportion of precursor available for splicing at Dcr1 will be lowered and, in turn, will lead to the formation of a relatively small number of complexes committed to splicing at Dcr1 (pr2). As the transition between spliceosomes carrying out reactions 1 and 2 is also longer, the final splicing efficiency at Dcr1 is expected to be low under conditions of cis-competition between the sites. Comparison of splicing at D1 in the pm0 and pm1 transcripts and at Dcr1 in the pm0 and pm2 transcripts indeed indicates that Dcr1 is more affected by competition than D1 (Table 1).

To determine the effect of exon sequences on selection, the curves of appearance of intermediates and products of splicing in the pm0 transcripts with the various exons 1 can be compared (Figs. 4 and 5, and Table 1). We first observe that, compared to the pm1 or pm2 transcripts, as appropriate, the time lags are

lengthened, irrespective of the exon or the site. This lengthening is accompanied by an increase in the proportion of the inactive complexes pr0, relative to that found in the pm1 transcripts, but smaller than that found in the pm2 transcripts. These observations are in agreement with the proposal that the assembly of pr0 complexes competes with that of pr1 and pr2, and that splicing at Dcr1 in the pm0 transcripts starts after a large amount of precursor has already been committed to splicing at D1 (or to inactivity). The similarity between the time lag changes, irrespective of the exon, also implies that their lengthening is not the consequence of modification of exon structure, but that it is related to the presence of both sites in a same transcript.

In contrast to the time lags, the overall rate of the reactions (i.e., the maximal slope of the curve of products, used as an index) is affected by exon 1 in the pm0 transcripts, as is the splicing efficiency at each site. The rate of splicing at D1 is reduced in the pm0 relative to the pm1 transcripts and gradually decreases when exon 1 becomes shorter. Conversely, splicing at Dcr1, which is not detectable in the WT-pm0 transcript, is induced when exon 1 is shortened. However, the effect of exon size on splicing rate and distribution of the precursor between the two sites is not linear. For instance, the 139 nucleotide shortening of the WT to the 5′d141 exon has only a small effect, while further truncations (of 61, then 38 nucleotides) dramatically affect the distribution of the precursor. Thus, D1 appears to be "protected" against cis-competition in the transcripts WT-pm0 and 5′d141-pm0, and this protection correlates with the presence of the 120 nucleotide sequence that can be folded into the stem-loop structure HP1. Transcript 5′d159-pm0, which is permissive for Dcr1 (pr2: ~5%), harbors HP59, a hairpin structure shorter than HP1, which might also have some protective effect. In contrast, HP54 has no influence. The two NY transcripts, whose exons are made of artificial sequences, are more permissive for Dcr1 (pr2: ~13%) than 5′d159-pm0, and, conversely, D1 is less protected. It is conceivable that there is no protection of D1 by exon 1 in these transcripts, and that their splicing pattern is the result of competition alone. If so, this would indicate that about 4.5 times less precursor is used for splicing at Dcr1 than for splicing at D1 in the conditions of cis-competition.

The data underline the importance of the structure of exon 1 in the selection of the natural site D1. As it does not exert its influence by modifying the kinetics of splicing directly, we conclude that the observed kinetic changes are due to its selector effect during cis-competition between the sites, and that the secondary structure of the exon is involved in the selection.

Discussion

Our aim is to understand the mechanisms underlying 5′ splice site selection. Our model system allows the study of one of the simplest and most widespread cases of 5′ splice site selection, that is, the selection of a natural 5′ splice site in the presence of a nearby, potentially functional but normally silent cryptic site. We started our study by a search for cis-acting elements that could influence selection. Though the strength of the 5′ splice site is generally an important element in selection (Domenjoud et al., 1991; Kuo et al., 1991; Tacke and Goridis, 1991), for our particular transcript it is not. The two sites have similar capacities to hybridize to U1 RNA—that is, they have nearly the same strength (stability of the donor site–U1 RNA hybrid: −6.9 kcal/mol for D1 and −6.2 kcal/mol for Dcr1). On the other hand, we have shown that the upstream exon is required for the exclusive selection of the natural site and the silencing of the cryptic site (Domenjoud et al., 1991).

For a better understanding of the mechanism of selection, we determined the time course of appearance of intermediates and products of splicing. As shown here, it is an experimental approach more informative than simply determining splicing efficiency. We studied splicing at the natural and at the cryptic sites, first independently and then in the natural conditions of cis-competition. To determine whether a given cis-acting element, such as the upstream exon, influences splicing at one of the sites, we analyzed transcripts with different upstream exon structures.

Our results demonstrate first that *the rate and efficiency of splicing are higher for the natural donor site than for the cryptic site.* The analysis of the time course curves indicates that at least early spliceosome assembly and transition between the two forms of mature spliceosomes (type 1 and 2) are impeded in the Dcr1 relative to the D1 transcripts. The cause of this difference in behavior is unknown. It is too large to be attrib-

uted solely to the small difference in strength. Since the differences persist when a large part of exon 1 is deleted, it seems likely that structural features of the precursor in regions other than exon 1 are involved. In fact, *the structure of exon 1 has little influence on splicing at either site.* We may conclude that its positive effect on the selection of D1 (Domenjoud et al., 1991) is due neither to an enhancement of splicing at D1, nor to an inhibition of splicing at Dcr1. Indeed, a positive effect of the exon on the kinetics of splicing should have been displayed even in the absence of the other site.

Based on analysis of the time course curves, it can be predicted that the natural site will largely predominate over the cryptic site in transcripts harboring both sites. This is indeed the case; in addition, our data show that *the cryptic site is a cis-competitor for the natural site.* The presence of the cryptic site lengthens the time of spliceosome assembly at the natural site, and it does so even when it is not itself used significantly (as in the WT transcript). Thus, the cryptic site is indeed recognized by trans-acting factors. If U1 snRNP is the recognition factor, it would not be surprising that the two sites are recognized similarly, since they have similar capacities to hybridize to U1 RNA. Thereafter, a structural feature of the coated transcript may displace the U1 snRNP (or another factor) bound to the cryptic site. This event is time-dependent, irrespective of the structure of exon 1. The data also indicate that the cryptic site has a negative cis-acting effect on the use of the natural site, at least in our in vitro system. Whether this effect also exists in vivo and in other systems remains to be demonstrated.

The relative usage of D1 and of Dcr1 is dependent on the upstream exon sequences. In the transcripts harboring the short NY exons that consist mainly of artificial sequences, the cryptic site is used about 4.5 times less actively than the natural site. Although there is no direct evidence that the splicing pattern of the NY transcripts is solely due to cis-competition between two sites with different splicing kinetics, these transcripts may nevertheless serve as a reference for the study of the exon sequence effect. Relative to the NY exons, the WT, 5′dl41, and to a lesser extent 5′dl59 exons "protect" the natural site against competition by the cryptic site. Therefore, we conclude that the *upstream exon acts as a selector in splicing* without directly interfering with the kinetics of splicing. The data allow us to define partially the elements involved in the selection. *The potential stem-loop structure HP1 appears as a key element*, since it is only in its presence that the natural site is efficiently spliced and the cryptic site almost entirely silenced. The other potential hairpin structure, HP59, has a less marked effect. Finally, the 139 nucleotide sequence upstream of HP1 has a modest function in the silencing of the cryptic site, possibly a consequence of the tertiary structure of the precursor RNA (our unpublished data). These elements, as well as analogous elements in other gene transcripts, might be targets for the trans-acting factors that have been shown to have a function in the selection of 5′ splice sites (Harper and Manley, 1991; Mayeda and Krainer, 1992).

The exclusive selection of D1 and the silencing of Dcr1 in the natural context of the transcript may be explained by a combination of the kinetic characteristics of the two sites and by the selector effect of the exon sequences, both of which favor the natural site. The molecular basis of the selector effect remains to be established.

Acknowledgments

We thank Dr. G. Richards for critical reading of the manuscript. We are also grateful to the cell culture group for providing HeLa cells, to A. Staub for oligonucleotide synthesis, and to the secretarial staff for help in preparing the manuscript.

This work was supported by grants from the Centre National de la Recherche Scientifique, the Institut National de la Santé et de la Recherche Médicale, the Centre Hospitalier Universitaire Régional, the Fondation pour la Recherche Médicale, and the Association pour la Recherche contre le Cancer.

The costs of publishing this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC Section 1734 solely to indicate this fact.

Lionel Domenjoud is currently at the Laboratoire d'Enzymologie et de Génie Génétique, Boulevard des Aiguillettes, 54200 Vandoeuvre les Nancy, France.

References

G. R. Adami, C. W. Marlor, N. L. Barrett, and G. G. Carmichael (1989), J Virol 63, 85–93.

B. Clouet d'Orval, Y. d'Aubenton-Carafa, J. Marie, and E. Brody (1991), J Mol Biol 221, 837–856.

G. J. Cote, I. N. Nguyen, S. M. Berget, and R. F. Gagel (1990), Mol Endocrinol 4, 1744–1749.

L. Domenjoud, H. Gallinaro, L. Kister, S. Meyer, and M. Jacob (1991), Mol Cell Biol 11, 4581–4590.

M. R. Green (1991), Annu Rev Cell Biol 7, 559–599.
W. Guo, G. J. Mulligan, S. Wormsley, and D. M. Helfman (1991), Genes Dev 5, 2096–2107.
C. Guthrie (1991), Science 253, 157–163.
J. E. Harper and J. L. Manley (1991), Mol Cell Biol 11, 5945–5953.
M. L. Hedley and T. Maniatis (1991), Cell 65, 579–586.
S. Inouye and M. Inouye (1987), in Synthesis and Applications of DNA and RNA (S. A. Narang, ed.), Academic Press, Inc., Orlando, FL, pp. 181–206.
M. Jacob and H. Gallinaro (1989), Nucleic Acids Res 17, 2159–2180.
H. C. Kuo, F. U. H. Nasim, and P. J. Grabowski (1991), Science 251, 1045–1050.
P. Legrain, B. Seraphin, and M. Rosbash (1988), Mol Cell Biol 8, 3755–3760.
A. Mayeda and Y. Oshima (1990), Nucleic Acids Res 18, 4671–4676.
A. Mayeda and A. R. Krainer (1992), Cell 68, 365–375.
M. McKeown (1990), Genet Eng 12, 139–181.
D. A. Melton, P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green (1984), Nucleic Acids Res 12, 7035–7056.
S. Michaud and R. Reed (1991), Genes Dev 5, 2534–2546.
M. E. Nemeroff, U. Utans, A. Krämer, and R. M. Krug (1992), Mol Cell Biol 12, 962–970.
A. Parent, S. Zeitlin, and A. Efstratiadis (1987), J Biol Chem 262, 11284–11291.
L. C. Ryner and B. S. Baker (1991), Genes Dev 5, 2071–2085.
R. Tacke and C. Goridis (1991), Genes Dev 5, 1416–1429.
M. Talerico and S. M. Berget (1990), Mol Cell Biol 10, 6299–6305.
P. R. Winship (1989), Nucleic Acids Res 17, 1266.