THE JOURNAL OF BIOLOGICAL CHEMISTRY

Vol. 269, No. 38, Issue of September 23, pp. 23590-23596, 1994
Printed in U.S.A.

Identification of Exon Sequences Involved in Splice Site Selection*

(Received for publication, April 26, 1994, and in revised form, June 22, 1994)

Zbigniew Dominski‡ and Ryszard Kole§

From the Department of Pharmacology and Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina 27599

The involvement of exon sequences in splice site selection was studied *in vivo* in HeLa cells transfected with a series of model three exon-two intron pre-mRNAs which differed only in the sequence of their internal exons. When the majority of the human globin-derived 175-nucleotide internal exon (DUP175) was replaced with a sequence from the yeast URA3 gene (DUP184), the splicing pathway changed from complete inclusion of the internal exon in DUP175 to its predominant skipping in the DUP184 construct. Skipping of the exon was reversed by increasing the strength of its flanking splicing elements indicating that exon sequences exert their effect only in the presence of relatively weak splicing signals. A series of block mutations in the internal exon of DUP184 showed that a stretch of 6 cytidine nucleotides increased the inclusion of the DUP184 internal exon about 7-fold. Mutations generating purine-rich sequences (AAG and GAAG) at the 3′ end of the exon led to complete exon inclusion while stepwise insertion of sequences from the internal exon of DUP175 into the DUP184 background increased exon inclusion 5-fold. Combination of the stretch of cytidines with sequences derived from DUP175 exon resulted in complete exon inclusion indicating that diverse signals within exons may cooperate with each other in affecting splice site selection.

Splicing of pre-mRNA depends on the presence of relatively short sequence elements, *i.e.* the 3′ splice site, the 5′ splice site, the branch point sequence and the polypyrimidine (poly-Y) tract, encompassing the region between the branch point sequence and the 3′ splice site. In addition to authentic splicing elements, similar sequence motifs occur in many locations within exons and introns but under normal conditions they remain silent. These cryptic splicing elements frequently become activated in transcripts carrying point mutations or deletions within the authentic elements. Moreover, in alternative splicing a given splice site may be selected or ignored depending on the cell type or its physiological state. This apparent flexibility of the splicing machinery raises the question of molecular mechanisms involved in ignoring cryptic splice sites in constitutive splicing and in selection of certain splice sites over others in alternative splicing (for review, see Refs. 1–4).

A number of determinants were demonstrated to be involved in splice site selection. They include the concentrations of constitutive (5–10) and gene specific splicing factors (11–14), the characteristics of cis-acting sequence elements in pre-mRNA, such as the match of splice sites (15–19) or the branch point to the consensus sequence (20–23), the proximity of splice sites to each other or to the branch point (24–28) as well as the length and composition of the poly-Y tracts (29–35). There is evidence that selection of splice sites is also affected by the secondary structure of pre-mRNA (36–40) and by coupling of splicing to transcription *in vivo* (41–43).

Initially, exons were not considered to be involved in splice site selection. However, a growing body of evidence indicates that exons constitute an important cis-acting sequence element in splicing. It was found that exon length affects splicing efficiency in monointronic pre-mRNAs (44–47) and modifies splicing patterns in multi-intron transcripts (10, 29, 30, 48, 49). Early reports noted that exon sequences also play a role in splice site selection (20, 50). Subsequently it was shown that in alternative splicing even one or two point mutations within exon sequences can alter full recognition of the exon to its predominant skipping (51, 52). Specific exon sequences were found essential for sex-specific splicing in *Drosophila* (11, 53), for proper splicing in certain retroviruses (54, 55), and for a number of other pre-mRNAs (56–60). The cumulative evidence led to a proposal (61), supported by subsequent observations (16, 30, 49, 62), that exons and not introns constitute the element in pre-mRNA initially recognized by transacting splicing factors.

If exons are recognized by the splicing factors, they must contain some negative or positive sequence elements which contribute to splice site selection. However, except for purine-rich sequences that have been recently identified in some alternatively spliced exons (63–67), there appear to be no generally conserved sequences within exons which may serve as recognition signals. Thus it remains unclear which internal exon sequences affect splice site selection. To address this question we took advantage of model three-exonic pre-mRNAs in which alterations in splicing patterns are easily detectable, being reflected in either skipping or inclusion of the internal exon. This system has already allowed us to demonstrate that the length of exons (29) as well as their interactions with other sequence elements in pre-mRNA affect splice site selection both *in vitro* and *in vivo* (30). In this report we extend these studies to identify exon sequences which contribute to splice site selection.

MATERIALS AND METHODS

Construction of Plasmids and Site-directed Mutagenesis—All recombinant DNA manipulations were carried out according to standard protocols (68). In all plasmids the modified gene fragments were inserted between the cytomegalovirus immediate early promoter and a human β-globin poly(A) site (29, 30, 69). The construction of DUP175, DUP51, and DUP33 clones, which are based on a duplicated fragment of human β-globin gene and contain three exons and two introns, was described in detail previously (29, 30; DUP175 was termed DUP171 in these reports). DUP54 was obtained by Bal31 nuclease digestion of plasmid DUP175 linearized at the unique *Nco*I site in the internal exon. A 151-nucleotide long *Nco*I-*Bsm*I fragment of *Saccharomyces cerevisiae*

* This work was supported in part by United States Public Health Service Grant GM32994 from the National Institutes of Health (to R. K.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Recipient of the postdoctoral fellowship from Cooley's Anemia Foundation.

§ To whom reprint requests should be addressed.

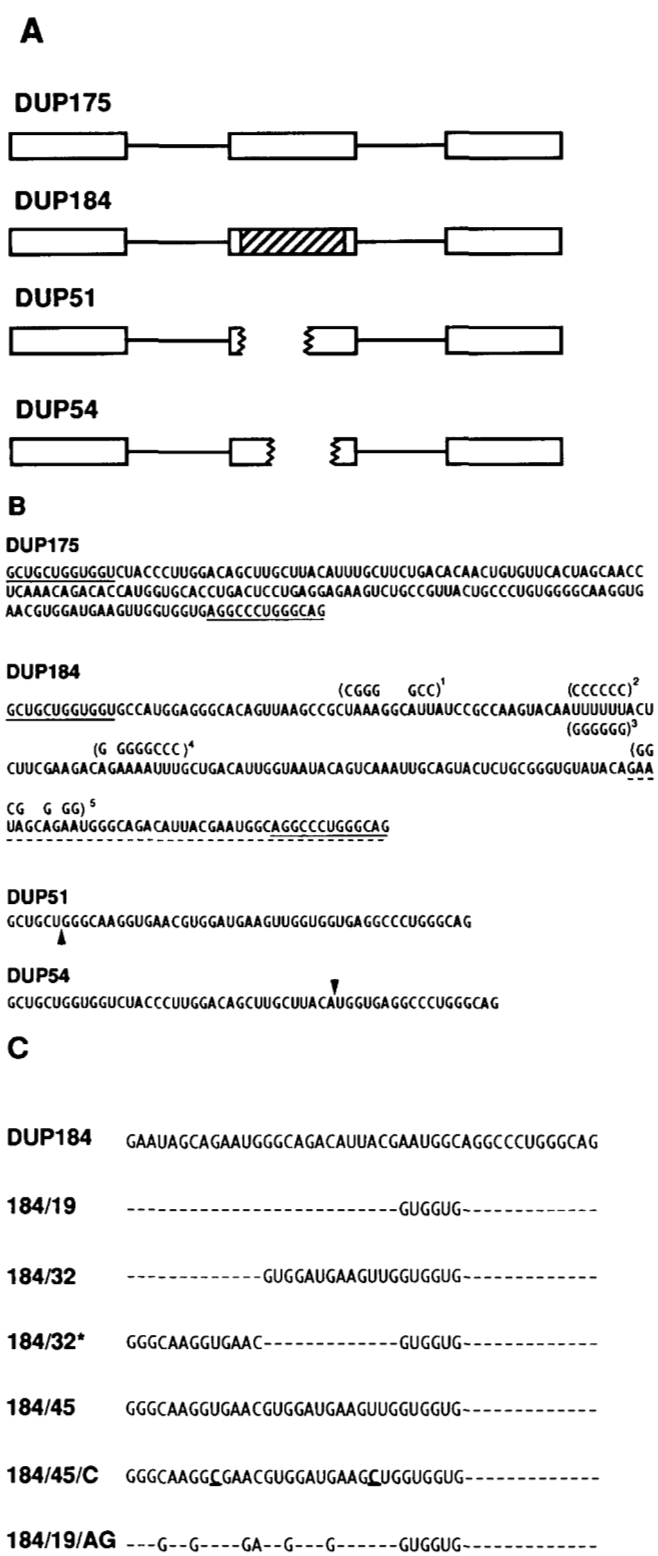

FIG. 1. *Panel A*, the structure of RNA transcripts. RNA was expressed transiently *in vivo* in transfected HeLa cells from plasmids carrying an immediate early cytomegalovirus promoter and appropriate fragments of the human β-globin gene. Construction of the plasmids has been described previously (29, 30) and under "Materials and Methods." *Boxes* represent exon sequences, *lines*, intron sequences. *Hatched area* in DUP184 represents a fragment of *S. cerevisiae* URA3 gene that replaced most of the β-globin sequence in the internal exon of the DUP175 construct. *Jagged lines* represent end points of the internal deletions in DUP175 that result in DUP51 and DUP54 constructs. *Panel B*, sequences of the internal exons. Sequences of DUP175 RNA that remain in the second exon of DUP184 pre-mRNA are *underlined* by a *solid line*. Sequences shown in *parentheses* were introduced into the second exon URA3 gene was inserted in both orientations into the *Nco*I site of DUP33 to construct DUP184 and DUP184rev, respectively. The remaining constructs tested in this study were prepared by site-specific mutagenesis of the DUP184 clone (29, 70). DUP184/Y2 and DUP184/Y3 have strengthened polypyrimidine tracts in the intron upstream of the internal exon. DUP184/BP and DUP184/5'ss have the consensus branch point sequence in the upstream intron and the consensus 5' splice site at the 3' end of the internal exon, respectively. The sequences of the mutations within the internal exon of DUP184 are shown in Fig. 1, *B* and *C*. The structure of the clones was verified by DNA sequencing.

Analysis of in Vivo Splicing—For analysis of *in vivo* splicing all plasmids were transfected into HeLa cells by electroporation and allowed 40–48 h for expression (71). RNA was isolated thereafter by the guanidinium/cesium chloride method (68) or using a Tri-reagent (MRC, Cincinnati, OH). Moloney murine leukemia virus reverse transcriptase (Life Technologies Inc.) and ^{32}P kinased 25-mer, complementary to nucleotides 76–100 of the third exon, were used in primer extension assay to analyze the endogenously synthesized and spliced RNA (30). After reverse transcription, cellular RNA was hydrolyzed in 0.2 N NaOH for 1 h at 65 °C. Products of primer extension were analyzed on 8% polyacrylamide, 7 M urea sequencing gels. To quantitate the results, the amount of material in primer extension bands was determined by densitometry. Exon inclusion was expressed as percentage of the product with included internal exon relative to the sum of products with the internal exon skipped and included. All autoradiograms were captured by DAGE MTI CCD72 video camera (Michigan City, IN) and the images were processed using NIH Image 1.43 and MacDraw Pro 1.0 software. The final figures were printed on a Sony dye sublimation printer.

RESULTS

Exon Sequence Affects Selection of Splice Sites—We have previously reported the use of a model pre-mRNA splicing substrate DUP175, based on a duplicated fragment of the human β-globin gene containing the first β-globin intron and flanking exon sequences (29). The duplication yields a construct containing two identical introns and two pairs of identical 3' and 5' splice sites (Fig. 1*A*). The resulting three-exonic pre-mRNA undergoes efficient splicing *in vitro* and *in vivo* leading to full and accurate removal of both introns and ligation of all three exons (Ref. 29, see also Fig. 2, *lane 1*). However, when the majority of the second exon in this construct was replaced with *S. cerevisiae* URA3 gene fragment (Fig. 1*A*, DUP184), the *in vivo* splicing pattern of the pre-mRNA (Fig. 2, *lane 4*) shifted toward exon skipping. Quantitative analysis showed that for DUP184 pre-mRNA only 8% of the spliced product contains the internal exon (Table I). This shift occurred in spite of the fact that the pre-mRNA retained the original splice sites, their immediately adjacent exon sequences, and a similar length of the internal exon (184 *versus* 175). These results suggested that substitution of the internal sequences within the exon either removed positive sequence elements present in DUP175 or introduced negative sequence elements originating from the URA3 gene fragment. The series of experiments described below was intended to discriminate between these possibilities.

Substitution of Splicing Elements Flanking the Suboptimal Exon—To determine whether the sequence of the 184-nucleotide exon represents a major block to the splicing machinery, we mutated the splicing elements flanking the exon to increase their strength. The mutations included purine-to-pyrimidine substitutions in two (DUP184/Y2) or three (DUP184/Y3) different positions in the polypyrimidine tract of the upstream intron, resulting in 15 contiguous pyrimidines, and a change in the branch point sequence, which modified the existing se-

of DUP184 by site-specific mutagenesis and the resulting constructs are termed DUP184/GC1, DUP184/GC2, etc., as numbered by the superscripts at the parentheses. Sequence of DUP184 RNA that was further mutated as shown in *panel C* is *underlined* by a *dashed line*. *Arrows* in DUP51 and DUP54 indicate the end points of the internal deletions in DUP175. *Panel C*, sequence of mutations within the 45 nucleotides at the 3' end of the internal exon in DUP184. *Dashes* indicate DUP184 sequence which remains unchanged in the mutants.

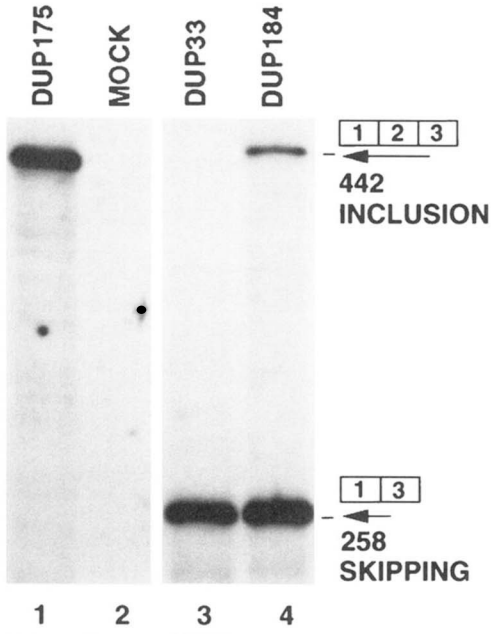

FIG. 2. **Internal exon of DUP184 pre-mRNA is mostly skipped during splicing *in vivo*.** Total RNA isolated from HeLa cells electroporated with DNA constructs indicated above each lane was analyzed by primer extension as described under "Materials and Methods." The size and structure of the corresponding primer extension products are shown. Mock (*lane 2*) is a negative control carried out with RNA isolated from HeLa cells electroporated without exogenous DNA. *In vivo* splicing of DUP175 (*lane 1*) and DUP33 (*lane 3*) was analyzed previously (29, 30) and served as a control for internal exon inclusion and skipping, respectively. *Lane 4*, *in vivo* splicing of DUP184 pre-mRNA. The same method of analysis was used in all subsequent figures.

TABLE 1
Quantitative analysis of the efficiency of exon inclusion

The amount of the material in the primer extension products with skipped or included internal exon or with the internal exon skipped was quantitated by densitometry of the autoradiograms. The results are expressed as percent of the product with included internal exon relative to the sum of both products.

Construct	Internal exon inclusion
	%
DUP175	100
DUP33	0
DUP184	8
DUP184/Y2	100
DUP184/Y3	100
DUP184/BP	70
DUP184/5′ss	96
DUP184/GC2	55
DUP184/GC3	7
DUP184/GC4	2
DUP184/GC5	2
DUP184/GC1/4	3[a]
DUP184/GC3/4/5	2
DUP51	100
DUP54	34
DUP184/19	1
DUP184/32	5
DUP184/32*	11
DUP184/45	43
DUP184/45/GC2	90
DUP184/45/GC3	40
DUP184/45/C	66
DUP184/19/AG	100

[a] The band visible in Fig. 4, *lane 9*, does not comigrate with the extension product of spliced RNA.

quence to the yeast consensus UACUAAC sequence (DUP184/BP). In addition, a consensus 5′ splice site, AG/GUAAGU, was created downstream of the internal exon by mutation of the existing AG/GUUGGU sequence.

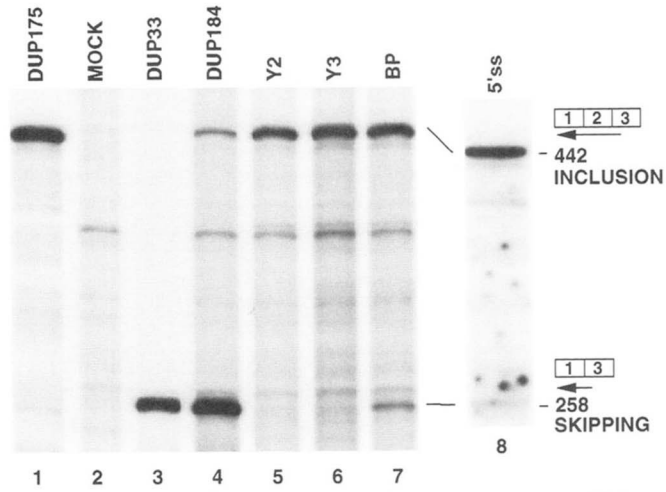

FIG. 3. **Strong splicing sequence elements promote efficient inclusion of the suboptimal exon.** Sequences in the introns upstream and downstream from the second exon in DUP184 were mutated to increase the strength of the splicing elements flanking the exon as described previously (29, 30) and under "Materials and Methods." The mutations included two (*lane 5*, construct DUP184/Y2) and three (*lane 6*, DUP184/Y3) purine to pyrimidine substitutions in the upstream polypyrimidine tract, resulting in 15 contiguous pyrimidines, a mutation in the upstream branch point resulting in the UACUAAC branch point sequence (*lane 7*, DUP184/BP), and a mutation in the downstream 5′ splice site resulting in a consensus 5′ splice site, CAG/GUAAGU (*lane 8*, DUP184/5′ss). The controls (*lanes 1–4*) are as described in the legend to Fig. 2.

Primer extension analysis of the *in vivo* spliced RNA shows a substantial increase in the amount of inclusion product generated by these changes (Fig. 3). Either double (Y2) or triple (Y3) mutations in the poly-Y tract upstream of the internal exon (Fig. 3, *lanes 5* and *6*, respectively) or the creation of a consensus 5′ splice site downstream (Fig. 3, *lane 8*) are sufficient to direct essentially complete inclusion of the exon. The consensus branch point sequence is not as effective and leads to approximately 70% of exon inclusion (Fig. 3, *lane 7*, and Table I). The above data indicate that poor recognition of the internal exon is markedly improved by increasing the strength of the splicing elements around the modified 184-nucleotide exon.

Mutational Scanning of the 184-Nucleotide Exon for Negative Sequence Elements—Several regions in the internal sequence of the DUP184 exon were subjected to sequence-specific mutagenesis to test for the existence of negative elements that may prevent its inclusion in the final spliced product. A number of short tracts of adenosine and uridine nucleotides located throughout the exon were replaced with guanosine and cytidine nucleotides resulting not only in a change of sequence at several positions in the exon but also in its overall GC content (Fig. 1*B*; the rationale for this approach is presented under "Discussion"). Primer extension analysis of RNA from HeLa cells transiently expressing the modified transcripts shows that mutations termed GC3, GC4, and GC5 do not promote inclusion of the internal exon (Fig. 4, *lanes 6–8*). A combination of these three mutations in a single transcript (Fig. 4, *lane 10*) or a combination of mutation GC4 with mutation GC1 (Fig. 4, *lane 9*) were also without effect as were three additional mutations located between GC4 and GC5 (not shown). The only detectable effect of these mutations is the generation of a spliced product resulting apparently from the activation of a cryptic splice site near the center of the exon in GC5 pre-mRNA (Fig. 4, *lane 8*). This effect is much less pronounced in the triple mutant (Fig. 4, *lane 10*). Thus, these mutations did not uncover any major negative elements in the DUP184 exon preventing its inclusion. The apparent lack of strong negative elements is also

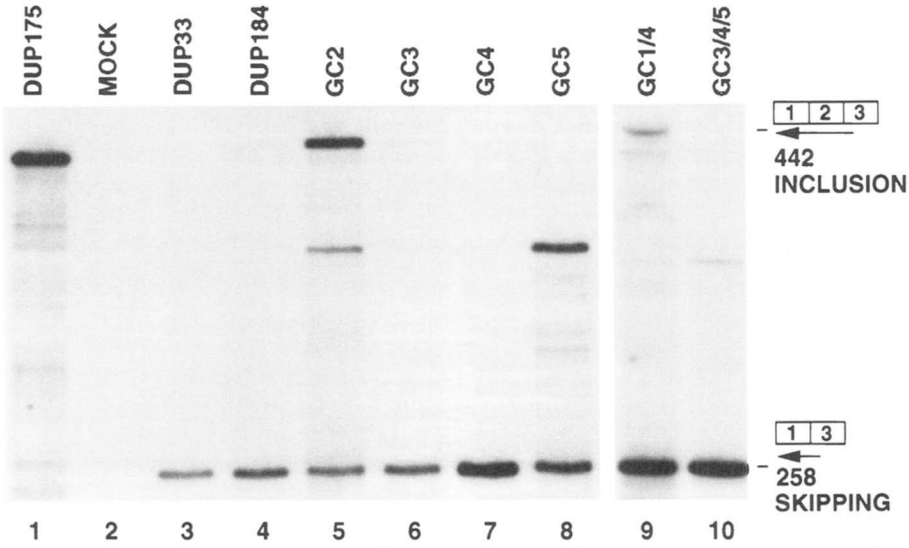

FIG. 4. **Mutational scanning of the 184-nucleotide exon.** Sequences shown in *parentheses* in Fig. 1*B* were introduced into the internal exon of DUP184 by site-specific mutagenesis. *Lanes 1–4,* controls; *lane 5*, DUP184/GC2; *lane 6*, DUP184/GC3; *lane 7*, DUP184/GC4; *lane 8*, DUP184/GC5. In *lanes 9* and *10* mutations GC1 and GC4, and GC3, GC4 and GC5, respectively, were combined in a single transcript. The additional band visible in *lanes 5, 8*, and *10* corresponds to the product of splicing to a cryptic 3′ splice site in the internal exon (see text).

shown by the fact that DUP184rev pre-mRNA, which contains the yeast DNA fragment in reverse orientation and therefore has a different sequence, undergoes full skipping of the internal exon.[1]

Surprisingly, mutation GC2 which alters the AATTTTTTA sequence, located 60 nucleotides downstream from the 3′ splice site, to AACCCCCCA led to significant inclusion of the internal exon in the spliced product and the activation of the same cryptic splice site as mutation GC5 (Fig. 4, *lane 5*). Quantitative analysis of the data showed that inclusion of the exon increased approximately 7-fold (Table I). Note that the GC2 and GC3 constructs differ in that the same short tract of thymidines is replaced by cytidines and guanosines, respectively (Fig. 1*B*). Thus, since the GC3 mutation was neutral, it appears that inclusion of the exon is enhanced by a positive signal provided by the cytidine tract in the GC2 mutation and not by the removal of a negative sequence element.

Positive Sequence Elements in β-Globin DUP175 and DUP51 Internal Exons—The lack of negative elements in the 184-nucleotide exon prompted us to search for positive sequence elements in the fully included DUP175 exon. As a result of deletion/insertion steps leading to the construction of DUP184, this clone and the parent DUP175 construct share a common sequence of 12 and 13 nucleotides at the 5′ and 3′ ends of the exon, respectively. Since the 184-nucleotide exon undergoes predominant skipping, these two short regions clearly do not contain sequences required for efficient exon recognition. In contrast, *in vivo* splicing of a deletion mutant of DUP175, termed DUP51, leads to full inclusion of the 51-nucleotide long internal exon (Fig. 5, *lane 1*) in spite of the removal of 124 nucleotides of the internal exon sequence. Sequence comparison of the internal exons in DUP175 and DUP51 constructs shows that they share 45 nucleotides of the sequence at the 3′ end of the internal exon (Fig. 1, *A* and *B*) and only 6 nucleotides at the 5′ end. Thus, it seems possible that the 45 nucleotides contain positive signals necessary for efficient exon recognition. This idea is reinforced by an additional deletion mutant of DUP175, *i.e.* DUP54. The second exon of this mutant retains only 18 nucleotides from the 3′ end of the DUP175 exon, the remainder originating from the 5′ end (Fig. 1*B*). Splicing of DUP54 pre-mRNA leads to predominant skipping of the internal exon (Fig. 5, *lane 2*; 34% inclusion of the exon, Table I), narrowing down the positively acting sequence to one between 18 and 45 nucleotides from the 3′ end of the DUP175 internal exon. Note that skipping of the DUP54 exon is unlikely to be due to negative elements since its entire sequence is present in the efficiently included second exon of DUP175 pre-mRNA (see Fig. 1*B*).

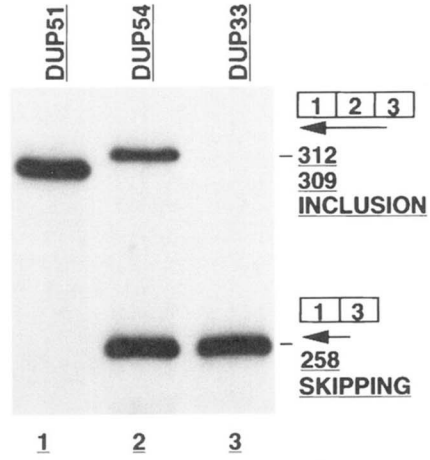

FIG. 5. **Different pathways of *in vivo* splicing for DUP51 and DUP54 pre-mRNAs.** *Lane 1*, DUP51; *lane 2*, DUP54; *lane 3*, DUP33 as control.

Reconstitution of the 175-Nucleotide Exon Sequences in the DUP184 Background—To test whether the putative positive sequence from the DUP175 exon can function in a different sequence context and promote the inclusion of the 184-nucleotide exon and to more closely identify the sequences within the 45-nucleotide fragment responsible for exon recognition, a stepwise reconstitution of the 3′ end of the DUP175 exon in the DUP184 background was undertaken. Addition of 6 nucleotides (GUGGUG) of the 175-nucleotide exon immediately upstream of the common sequence at the 3′ end, to create a 19-nucleotide sequence shared by both exons (Fig. 1*C*, DUP184/19), did not enhance exon recognition *in vivo* (DUP184/19; Fig. 6*A*, *lane 3*), as shown by primer extension of RNA from transfected HeLa cells. This was to be expected since 18 nucleotides of this sequence are present in the poorly recognized DUP54 exon (Fig. 1*B* and Fig. 5, *lane 2*). Further extension of the common se-

[1] J. Wozniak and R. Kole, unpublished results.

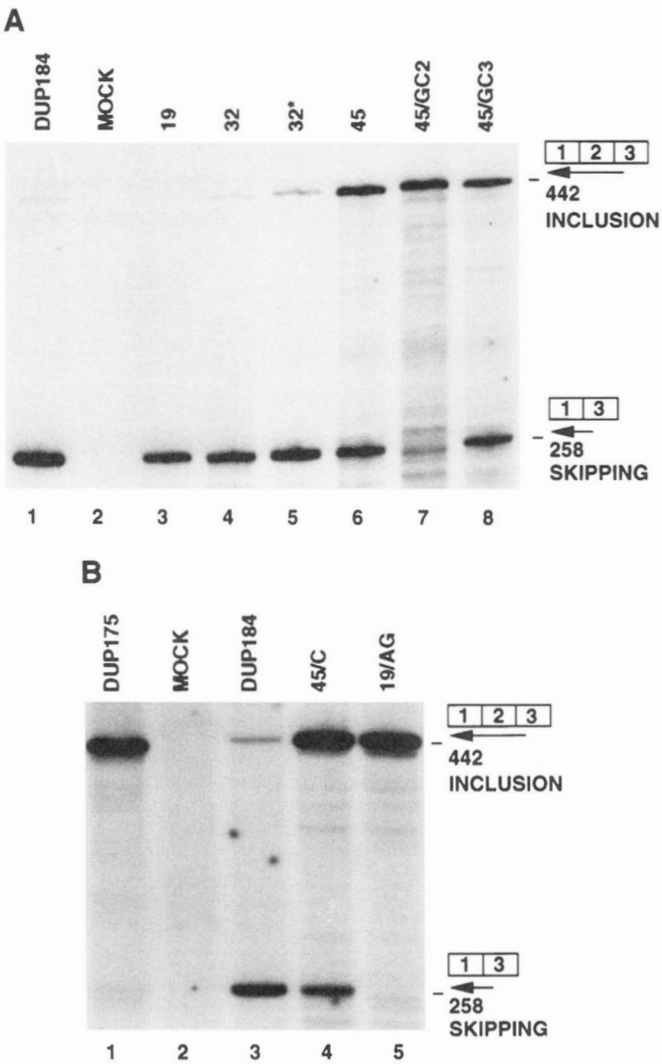

FIG. 6. **Mutations in the 45-nucleotide sequence at the 3′ end of the DUP184 internal exon affect splicing pathways.** *Panel A, lanes 3–6*, sequences at the 3′ end of the internal exon in DUP184 pre-mRNA are replaced with their equivalents from the DUP175 exon; the actual sequences are shown in Fig. 1*C*. *Lane 3*, DUP184/19; *lane 4*, DUP184/32; *lane 5*, DUP184/32*; *lane 6*, DUP184/45. In *lanes 7* and *8*, the DUP184/45 construct is combined with mutations GC2 and GC3, respectively. *Lanes 1* and *2*, controls. *Panel B, lane 4*, cryptic splice sites located in the 26-nucleotide insert are mutated (DUP184/45/C); *lane 5*, purine nucleotides are inserted into the URA3 background of DUP184/19 resulting in DUP184/19/AG. See Fig. 1*C* for actual sequence. *Lanes 1–3*, controls.

quence by replacement of an additional 13 nucleotides to the total of 32 nucleotides in DUP184/32 shows some stimulatory effect on exon inclusion (Fig. 6*A, lane 4*), especially when compared to DUP184/19, its parent construct (Fig. 6*A, lane 3*). A substantial increase in exon inclusion is observed when another 13-nucleotide region located further upstream is replaced by a corresponding sequence from the 175-nucleotide exon (DUP184/45; Fig. 6*A, lane 6*). The addition of 26 nucleotides in two 13-nucleotide steps extended the common sequence of DUP184 and DUP175 exons to the 45 nucleotides apparently required for exon recognition in DUP175 and DUP51 constructs. Note that the addition of the last 13 nucleotides in the DUP184/45 construct is not solely responsible for the enhanced inclusion of the internal exon. When the same 13 nucleotides were inserted in the DUP184/19 mutant, leaving a 13-nucleotide spacer of the yeast sequence (Fig. 1*C*, *DUP184/32**), the effect on exon inclusion was less pronounced (Fig. 6*A, lane 5*). The synergistic effect of both additions observed in DUP184/45 suggests a certain cooperation between sequence elements present in this region. Quantitation of the data shows that inclusion of the exon in DUP184/45 is increased to 43% (Table I), approximately 5-fold compared to the DUP184 construct. This result is likely caused by the positive effects of sequences derived from the DUP175 exon rather than by the removal of a negative element from DUP184 since mutation GC5, which overlaps the 13-nucleotide fragment introduced in the DUP184/32* construct does not enhance exon recognition (see Fig. 4, *lane 8*). Interestingly, when both positively acting elements of DUP184/45 and GC2 are combined in DUP184/45/GC2, they promote highly efficient, almost complete inclusion of the internal exon (Fig. 6*A, lane 7*, and Table I). The sequence specificity of this effect is supported by the fact that the GC3 mutation in the background of the DUP184/45 nucleotide substitution does not enhance exon inclusion (Fig. 6*A, lane 8, cf. lane 6*).

Analysis of Exon Sequence in DUP184/45 Pre-mRNA—In an attempt to further characterize the nature of the exon recognition element residing in the β-globin-derived 26-nucleotide fragment inserted in the DUP184/45 pre-mRNA, we have introduced additional mutations in this area and analyzed their effect on splicing of the resulting constructs *in vivo*. The choice of mutations was prompted by two features of the insert (see DUP184/45 in Fig. 1*C*).

First, the 26-nucleotide insert contains two GU dinucleotides in a sequence context compatible with the 5′ splice site sequence and includes a cryptic 5′ splice site, activated in certain thalassemias, located 38 nucleotides upstream from the 3′ end of the exon (72). It could be hypothesized that these sequences attract U1 small nuclear ribonucleoprotein and/or other splicing factors and facilitate subsequent recognition of a downstream, authentic 5′ splice site (64). However, our results do not support this hypothesis. Splicing of pre-mRNA constructs in which the GU dinucleotides were mutated to GC (Fig. 1*C*, *DUP184/45/C*) led to a slight improvement in exon recognition (Fig. 6*B, lane 4*) relative to DUP184/45 (Fig. 6*A, lane 6*, and Table I), suggesting that cryptic 5′ splice sites adjacent to the authentic one may interfere with the proper assembly of the splicing factors (25, 73).

Second, the insert also contains one AAG and one GAAG sequence, purine-rich elements suggested by Watakabe *et al.* (63, 64) to be exon recognition sequences. To test the possibility that purine-rich elements are responsible for the positive effect of this region on exon inclusion, we have replaced six pyrimidine with six purine nucleotides within the *S. cerevisiae* sequence of the DUP184/19 exon (DUP184/19/AG, Fig. 1*C*). Without additional purines this exon is not included in the spliced product, as seen in Fig. 6*A, lane 3*. Insertion of purines resulted in an RNA with a purine-rich stretch containing two GAAG sequence elements. This modification led to 100% inclusion of the internal exon (Fig. 6*B, lane 5*, and Table I), even higher than that for the DUP184/45 construct (Fig. 6*A, lane 6*).

DISCUSSION

To study the involvement of exon sequences in splice site selection we utilized derivatives of DUP175 which differ only in the internal sequence within the second exon of the pre-mRNA. Splicing of two of the constructs, DUP175 and DUP51, led to complete inclusion of the internal exon, whereas for DUP54 and DUP184 the splicing pattern was characterized by predominant skipping of the exon. These results indicate that exon sequences play a role in splice site selection presumably via exon recognition by the splicing machinery. As shown by the results of mutations in the introns of DUP184 pre-mRNA, the contribution of exon sequence is relatively weak since its effect is easily overcome by a strong polypyrimidine tract, a consen-

sus branch point, or a 5′ splice site. In other words, the effect of exon sequences manifests itself only in the presence of relatively weak splicing elements. Thus, it is probably no coincidence that alternatively spliced exons, such as in *Drosophila* doublesex (11, 14, 53), bovine growth hormone (57), or Rous sarcoma virus (55) genes, are flanked by suboptimal splice sites and appear to have additional sequences regulating splicing via interactions with specific protein factors (12–14). However, our experiments in which splicing pathways of artificial pre-mRNAs were modified by exon sequences indicate that the effect of exon sequences is not limited to exons which undergo alternative splicing in nature.

Which sequence elements or their absence are responsible for poor inclusion of the internal exons in DUP184 and DUP54 pre-mRNAs? At the outset of this series of experiments we have considered a possibility that the difference in the overall GC content of the internal exons (55% in DUP175 *versus* 47% in DUP184) may be responsible for the difference in the splicing pathways. This concept was based on a GenBank data base survey which showed a higher GC content in exons of human genes relative to introns (74). A correlation between a higher GC content and exon recognition in splicing of plant pre-mRNAs was also confirmed experimentally (75, 76). Nonetheless, our results did not support this hypothesis. Although the mutations analyzed in Fig. 4 significantly increased the GC content of the DUP184 exon (58% in DUP184/GC3/4/5), they did not reverse exon skipping. Furthermore, the difference in the GC content between DUP51 *versus* DUP54 (63 and 61%, respectively) is insignificant but they vary in splicing pathways. However, the fact that the above sequence modifications in several different regions within the 184-nucleotide exon did not enhance its inclusion argues against the existence of strongly negative sequence elements within this exon. The same consideration seems to rule out internal secondary structure of the exon as the cause for poor exon recognition by the splicing machinery.

Mutations in the DUP184 exon identified positive sequence elements, *i.e.* the short oligo(C) sequence in DUP184/GC2 and the sequences originating from the DUP175 exon in DUP184/32 and −32*. The effect of these sequences is additive, as clearly seen in the splicing of DUP184/45 and DUP184/45/GC2 pre-mRNAs. In the latter construct exon inclusion is almost complete. The positive effect of the 26-nucleotide sequence responsible for efficient exon inclusion in DUP184/45 is not restricted to the 175- and 184-nucleotide exons. This element facilitates with similar efficiency the inclusion of the 184-nucleotide exon containing a fragment of the URA3 gene inserted in the reversed orientation and of the internal exon in DUP51 pre-mRNA, suggesting that it is able to induce inclusion in different exon "contexts."

What is the nature of exon sequences that enhance exon inclusion? The stimulatory effect of the oligo(C) tract created by a block mutation in DUP184/GC2 suggests a novel exon recognition element. This element may be recognized by poly(C)-binding proteins, like heterogeneous nuclear ribonucleoproteins J and K (77, 78), or other heterogeneous nuclear ribonucleoproteins found in the spliceosome that interact with the pre-mRNAs in a sequence specific manner (79). The fact that the DUP184/GC2 mutation also activates a cryptic splice site is probably insignificant for exon recognition since the same cryptic splice site is activated in the DUP184/GC5 mutant without any positive effect on exon inclusion. The positive role of 5′ cryptic splice sites in exon recognition, which seemed a possibility in splicing of the DUP184/32* and DUP184/45 constructs, was not supported by analysis of DUP184/45/C and DUP184/19/AG pre-mRNAs. Although these mutants lack cryptic splice sites, the internal exon is efficiently included in the final spliced product.

Splicing of DUP175, DUP51, DUP184/45, −184/32, −184/32*, −184/AG, and −184/45/C pre-mRNAs leads to enhanced inclusion of the internal exon. The feature that appears to characterize these constructs is the presence of AAG and/or GAAG sequences close to the 3′ end of the internal exon. Similar sequences have been recently shown to act as exon recognition elements in exon M2 of the IgM gene (63, 64), exon 5 of cardiac troponin T gene (65), exon ED1 of human fibronectin gene (66), and in the last exon of bovine growth hormone (67). We note that in our constructs relatively minor differences in the purine-rich sequence have a major effect on exon inclusion. In DUP184 pre-mRNA the 45-nucleotide region at the 3′ end of the internal exon contains 30 purines and, as can be seen in Fig. 1*C*, two GAA, one AGAA, and one AGA element. The corresponding region in DUP184/45 contains 31 purines and one AAG, one GAA, and one GAAG element but the splicing pathways of these two pre-mRNAs are significantly different. Various synthetic polypurine sequences were in fact recently shown to stimulate exon recognition to a different extent (80). Interestingly, among all synthetic repeats studied (80) the AAG trinucleotide, represented twice in the 45-nucleotide region of DUP175 and absent in the corresponding region of DUP184, appears to have the strongest stimulatory effect. Clearly, this remarkable sequence specificity must be reflected in complex interactions with appropriate splicing factors. *In vitro* splicing experiments are in progress to identify factors interacting with these and other signal elements described in this report.

Short internal exons are rare in nature (81) and tend to be skipped during splicing *in vitro* and *in vivo* (29, 30, 48, 49). The fact that certain exon sequences play a role in splicing raises the possibility that it is the lack of these exon recognition sequences and not the short length of the exons *per se* that is responsible for their skipping. However, Xu *et al.* (65) showed that insertion of purine-rich sequences into short internal exons of pre-mRNAs similar to DUP33 or DUP23 constructs (29, 30) is not sufficient to promote exon inclusion. Furthermore, the difference in exon inclusion between DUP33 and DUP54 clones (Fig. 5), both lacking the positive sequence elements discussed above, also suggests that exon length is an important factor in splice site selection. Comparison of our previous results (29, 30) and those of Xu *et al.* (65) also suggests that purine-rich sequences constitute weaker splicing elements than the strong splice sites or branch points since only the latter effectively restore inclusion of short exons into spliced RNA.

Although exon inclusion was enhanced in the DUP184/45 transcript it remained still below that seen in DUP175 and DUP51. This would suggest that either additional positive sequence signals exist in the latter constructs or that some negative elements, unidentified by the mutational scan, remain in the 184-nucleotide exon. Clearly, the oligo(C) sequence introduced by mutagenesis into the internal exon of DUP184/GC2 represents a positive element since it promoted exon inclusion and further enhanced the effects of purine-rich elements as seen in the DUP184/45/GC2 construct. These results suggest a diversity of exon recognition signals that similarly to other splicing elements (29, 30) cooperate with one another in splice site selection.

Acknowledgments—We thank Elizabeth Smith for excellent technical assistance and summer students Izabella Naruszewicz, Erica Barber, and Lisa Sawer for their help. We also thank our colleagues for comments on the manuscript.

REFERENCES

1. Green, M. R. (1991) *Annu. Rev. Cell Biol.* **91,** 559–595
2. Maniatis, T. (1991) *Science* **251,** 33–34
3. McKeown, M. (1990) in *Regulation of Alternative Splicing: Genetic Engineering* (Setlow, J. K., ed) pp. 139–181, Plenum Press, New York
4. Moore, M. J., Query, C. C., and Sharp, P. A. (1993) in *Splicing of Precursors to mRNAs by the Spliceosome* (Gesteland, R. F., and Atkins, J. F., eds) pp. 303–358, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
5. Fu, X-D., Mayeda, A., Maniatis, T., and Krainer, A. R. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89,** 11224–11228
6. Ge, H., and Manley, J. L. (1990) *Cell* **62,** 25–34
7. Harper, J. E., and Manley, J. L. (1991) *Mol. Cell. Biol.* **11,** 5945–5953
8. Harper, J. E., and Manley, J. L. (1992) *Gene Exp.* **2,** 19–29
9. Mayeda, A., and Krainer, A. R. (1992) *Cell* **68,** 365–375
10. Mayeda, A., Helfman, D. M., and Krainer, A. R. (1993) *Mol. Cell. Biol.* **13,** 2993–3001
11. Hedley, M. L., and Maniatis, T. (1991) *Cell* **65,** 579–586
12. Hoshijima, K., Inoue, K., Higuchi, I., Sakamoto, H., and Shimura, Y. (1991) *Science* **252,** 833–836
13. Inoue, K., Hoshijima, K., Higuchi, I., Sakamoto, H., and Shimura, Y. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89,** 8092–8096
14. Tian, M., and Maniatis, T. (1992) *Science* **256,** 237–240
15. Grabowski, P. J., Nasim, F. H., Kuo, H-C., and Burch, R. (1991) *Mol. Cell. Biol.* **11,** 5919–5928
16. Kuo, H-C., Nassim, F. H., and Grabowski, P. J. (1991) *Science* **251,** 1045–1050
17. Lear, A. L., Eperon, L. P., Wheatley, I. M., and Eperon, I. C. (1990) *J. Mol. Biol.* **211,** 103–115
18. Zhuang, Y., and Weiner, A. M. (1986) *Cell* **46,** 827–835
19. Zhuang, Y., Lueng, H., and Weiner, A. M. (1987) *Mol. Cell. Biol.* **7,** 3018–3020
20. Reed, R., and Maniatis, T. (1988) *Genes & Dev.* **2,** 1268–1276
21. Wu, J., and Manley, J. L. (1989) *Genes & Dev.* **3,** 1553–1561
22. Zhuang, Y., and Weiner, A. M. (1989) *Genes & Dev.* **3,** 1545–1552
23. Zhuang, Y., Goldstein, A. M., and Weiner, A. M. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86,** 2752–2756
24. Cuningham, S. A., Else, J. A., Potter, B. V. L., and Eperon, I. C. (1991) *J. Mol. Biol.* **217,** 265–281
25. Nelson, K. K., and Green, M. R. (1988) *Genes & Dev.* **2,** 319–329
26. Nelson, K. K., and Green, M. R. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 6253–6257
27. Reed, R., and Maniatis, T. (1986) *Cell* **46,** 681–690
28. Smith, C. W. J., and Nadal-Ginard, B. (1989) *Cell* **56,** 749–758
29. Dominski, Z., and Kole, R. (1991) *Mol. Cell. Biol.* **11,** 6075–6083
30. Dominski, Z., and Kole, R. (1992) *Mol. Cell. Biol.* **12,** 2108–2114
31. Freyer, G. A., O'Brien, J. P., and Hurwitz, J. (1989) *J. Biol. Chem.* **264,** 14631–14637
32. Fu, X-Y., Ge, H., and Manley, J. (1988) *EMBO J.* **7,** 809–817
33. Mullen, M. P., Smith, C. W. J., Patton, J. G., and Nadal-Ginard, B. (1991) *Genes & Dev.***5,** 642–655
34. Patterson, B., and Guthrie, C. (1991) *Cell* **64,** 181–187
35. Reed, R. (1989) *Genes & Dev.* **3,** 2113–2123
36. Clouet-d'Orval, B., d'Aubenton-Carafa, Y., Sirand-Pugnet, P., Gallego, M., Brody, E., and Marie, J. (1991) *Science* **252,** 1823–1828
37. Clouet-d'Orval, B., d'Aubenton-Carafa, Y., Brody, J. M., and Brody, E. (1991) *J. Mol. Biol.* **221,** 837–856
38. Domenjoud, L., Gallinaro, H., Kister, L., Meyer, S., and Jacob, M. (1991) *Mol. Cell. Biol.* **11,** 4581–4590
39. Libri, D., Goux-Pelletan, M., Brody, E., and Fiszman, M. Y. (1990) *Mol. Cell. Biol.* **10,** 5036–5046
40. Solnick, D., and Lee, S. I. (1987) *Mol. Cell. Biol.* **7,** 3194–3198
41. Aebi, M., Horning, H., Padgett, R. A., Reiser, J., and Weissmann, C. (1986) *Cell* **47,** 555–565
42. Beyer, A. L., and Osheim, Y. N. (1988) *Genes & Dev.***2,** 754–765
43. Eperon, L. P., Graham, I. R., Griffiths, A. D., and Eperon, I. C. (1988) *Cell* **54,** 393–401
44. Furdon, P. J., and Kole, R. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83,** 927–931
45. Furdon, P. J., and Kole, R. (1988) *Mol. Cell. Biol.* **8,** 860–866
46. Parent, A., Zeitlin, S., and Efstratiadis, A. (1987) *J. Biol. Chem.* **262,** 11284–11291
47. Turnbull-Ross, A. D., Else, A. J., and Eperon, I. C. (1988) *Nucleic Acids Res.* **16,** 395–411
48. Black, D. L. (1991) *Genes & Dev.* **5,** 389–402
49. Sterner, D., and Berget, S. M. (1993) *Mol. Cell. Biol.* **13,** 2677–2687
50. Somasekhar, M. B., and Mertz, J. E. (1985) *Nucleic Acids Res.* **13,** 5591–5609
51. Cooper, T. A. (1992) *J. Biol. Chem.* **267,** 5330–5338
52. Cooper, T. A., and Ordahl, C. P. (1989) *Nucleic Acids Res.* **17,** 7905–7921
53. Nagoshi, R. N., and Baker, B. S. (1990) *Genes & Dev.* **4,** 89–97
54. Fu, X-D., Katz, R. A., Skalka, A. M., and Maniatis, T. (1991) *Genes & Dev.* **5,** 211–220
55. Katz, R. A., and Skalka, A. M. (1990) *Mol. Cell. Biol.* **10,** 696–704
56. Cote, G. J., Stolow, D. T., Peleg, S., Berget, S. M., and Gagel, R. F. (1992) *Nucleic Acids Res.* **20,** 2361–2366
57. Hampson, R. K., La Follette, L., and Rottman, F. (1989) *Mol. Cell. Biol.* **9,** 1604–1610
58. Streuli, M., and Saito, H. (1989) *EMBO J.* **8,** 787–796
59. Tsai, A. Y. M., Streuli, M., and Saito, H. (1989) *Mol. Cell. Biol.* **9,** 4550–4555
60. Wakamatsu, N., Kobayashi, H., Miyatake, T., and Tsuji, S. (1992) *J. Biol. Chem.* **267,** 2406–2413
61. Robberson, B. L., Cote, G. J., and Berget, S. M. (1990) *Mol. Cell. Biol.* **10,** 84–94
62. Talerico, M., and Berget, S. M. (1990) *Mol. Cell. Biol.* **10,** 6299–6305
63. Watakabe, A., Sakamoto, H., and Shimura, Y. (1991) *Gene Exp.* **1,** 175–184
64. Watakabe, A., Tanaka, K., and Shimura, Y. (1993) *Genes & Dev.* **7,** 407–418
65. Xu, R., Teng, J., and Cooper, T. (1993) *Mol. Cell. Biol.* **13,** 3660–3674
66. Lavigueur, A., La Branche, H., Kornblihtt, A. R., and Chabot, B. (1993) *Genes & Dev.* **7,** 2405–2417
67. Dirksen, W. P., Hampson, R. K., Sun, Q., and Rottman, F. M. (1994) *J. Biol. Chem.* **269,** 6431–6436
68. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, pp. 188–209, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
69. Davis, M. G., and Huang, E-S. (1988) *Biotech. Appl. Biochem.* **10,** 6–12
70. Kunkel, T. A., Roberts, J. D., and Zakour, R. A. (1987) *Methods Enzymol.* **154,** 367–382
71. Potter, H. (1994) in *Current Protocols in Molecular Biology* (Ausubel, P., ed) pp. 9.3.1–9.3.6, J. Wiley and Sons, New York
72. Schwartz, E., and Benz, E. J. (1991) in *Hematology, Basic Principles and Practice* (Hoffman, R., Benz, E. J., Shaltiel, S. J., Furie, B., and Cohen, H. J., eds) pp. 368–392, Churchill Livingstone, New York
73. Siebel, C. W., Fresco. L. D., and Rio, D. C. (1992) *Genes & Dev.* **6,** 1386–1401
74. Brunak, S., Engelbrecht, J., and Knudsen, S. (1991) *J. Mol. Biol.* **220,** 49–65
75. Goodall, G., and Filipowicz, W. (1989) *Cell* **58,** 473–483
76. Goodall, G., and Filipowicz, W. (1991) *EMBO J.* **10,** 2635–2644
77. Matunis, M. J., Michael, W. M., and Dreyfuss, G. (1992) *Mol. Cell. Biol.* **12,** 164–171
78. Swanson, M. S., and Dreyfuss, G. (1988) *Mol. Cell. Biol.* **8,** 2237–2241
79. Bennett, M., Pinol-Roma, S., Staknis, D., Dreyfuss, G., and Reed, R. (1992) *Mol. Cell. Biol.* **12,** 3165–3175
80. Tanaka, K., Watakabe, A., and Shimura, Y. (1994) *Mol. Cell. Biol.* **14,** 1347-1354
81. Hawkins, J. D. (1988) *Nucleic Acids Res.* **16,** 9893–9908