1638–1645 *Nucleic Acids Research, 1996, Vol. 24, No. 9*

The 5′ and 3′ splice sites come together via a three dimensional diffusion mechanism

Zvi Pasman[1,2] and Mariano A. Garcia-Blanco[1,2,3,*]

Departments of [1]Molecular Cancer Biology, [2]Microbiology and [3]Medicine, Levine Science Research Center, Duke University Medical Center, Durham, NC 27710, USA

Received February 2, 1996; Revised and Accepted March 22, 1996

ABSTRACT

We present evidence that the splice sites in mammalian pre-mRNAs are brought together via a three dimensional diffusion mechanism. We tested two mechanisms for splice site pairing: a lateral diffusion ('scanning') model and the currently favored three dimensional diffusion ('jumping') model. Two lines of evidence that distinguish between these two models are presented. The first utilized bipartite splicing substrates tethered by double-stranded RNA stems predicted to provide either a moderate or severe block to splice site pairing via a scanning mechanism. Splice site pairing via a jumping mechanism was expected to be unaffected or affected minimally. The second approach utilized a flexible poly(ethylene glycol) moiety within the intron. This insertion was predicted to reduce scanning efficiency but not the efficiency of a three dimensional diffusion mechanism. The best explanation for the data with the bipartite RNAs is that splice site pairing occurs through three dimensional diffusion. Kinetic analysis of the poly(ethylene glycol) containing substrate showed that neither the lag phase nor the initial rates of mRNA production and spliceosome assembly were affected by this insertion. Therefore, both experimental approaches supported the three dimensional diffusion model of splice site pairing.

INTRODUCTION

The 5′ splice site and the polypyrimidine tract/branchpoint sequences are both defined very early in the splicing reaction. How are these two *cis* elements brought into apposition with each other? The large size of many mammalian introns makes this a difficult process to envision. Nuclear pre-mRNAs do not exist as free random coils, but rather, they are thought to be assembled into more compact ribonucleoprotein structures via interactions with hnRNP proteins (1). Nevertheless, the distance between splice sites must be enormous relative to the few angstroms required during the first transesterification reaction. The 5′ splice site and its associated factors can be brought together with the polypyrimidine tract/branchpoint and its interacting factors by one of two distinct mechanisms. The 'scanning' model, originally proposed by Sharp (2), postulates a mechanism in which a splice site-containing RNP complex diffuses laterally along the RNA, towards a second splice site specific complex. In this scenario, the scanning machinery could interact directly with intronic RNA or with the intron assembled into an RNP. The 'jumping' model posits a mechanism that involves the interactions of two splice-site containing RNP complexes, but in this case the association of the splice sites occurs independently of the RNA between the splice sites, and the rate at which the complexes associate is determined by three dimensional diffusion. This mode is operative during *trans* splicing in trypanosomes and nematodes (3–5). This mechanism is also operative during *in vitro trans* splicing in mammalian systems (6–8).

Several attempts were made to address the problem of splice site pairing. The use of tandem duplications of 5′ splice sites in a rabbit β-globin (9) or a human Gγ-globin (10) genes resulted in an almost exclusive use of the upstream 5′ splice site (or distal relative to the intron and 3′ splice site). Tandem duplications of 3′ splice sites resulted in splicing exclusively to the distal or downstream 3′ splice site in rabbit β-globin and exclusively to the proximal or upstream 3′ splice site in human Gγ-globin. Thus the two groups reached opposite conclusions: one favored a three dimensional diffusion mechanism (9), while the second advocated a 5′→3′ scanning model (10). Similar arguments were made in favor of scanning in the differential processing of the E3 transcription unit of adenovirus-2 (11). The same rationale was used to favor the three dimensional diffusion model in the exclusive utilization of exon α or exon β during alternative splicing of the Troponin T pre-mRNA (12), and in the exon skipping mutations in the dihydrofolate reductase gene in Chinese hamster ovary cells (13). In these experiments, the basic assumption that duplicated 5′ and 3′ splice sites of identical sequences were equivalent has since been shown not to be the case. The exclusive use of the proximal 3′ splice site in the Gγ-globin experiments could be explained by the presence of an almost complete exon downstream of that site (14). Similarly, the exclusive use of the distal 3′ splice site in the β-globin experiments could be explained by the absence of a full exon downstream of the proximal 3′ splice site (ibid.). The exon definition model of Berget and colleagues (15) and its subsequent experimental confirmation (16,17) also could explain the use of the distal splice sites in these experiments. Moreover, the more recent identification of exonic splicing enhancers (18–20) argues strongly that any assumption of splice site equivalence is naive. It is also known now that levels of ASF/SF2 and hnRNP A1 can influence the choice between tandem duplicated 5′ splice sites *in vitro*

*To whom correspondence should be addressed at: Department of Molecular Cancer Biology, Box 3686 DUMC, Durham, NC 27710, USA

(21–23) and *in vivo* (24,25). Therefore, this experimental approach did not distinguish one model from another.

An approach utilizing anti-sense 2′-*O*-methyl oligoribonucleotides was employed by Mayeda *et al.* (26). Oligonucleotides directed to the 5′ splice site or the branchpoint sequence of a human β-globin-derived pre-mRNA inhibited splicing *in vitro*, whereas oligonucleotides directed to other exonic or intronic sequences did not inhibit splicing. The lack of effect by the latter anti-sense oligonucleotides was taken to indicate a three dimensional diffusion mechanism. Two problems undermine this conclusion: first, since reaction times were very long (4 h) it is impossible to ascertain the relative rates of splicing in the presence and absence of oligos. Second, a scanning machinery may be capable of melting short duplexes such as those created by the dodecamer oligos added. That this was the case is suggested by the weak effect of an oligo placed downstream of the branchpoint, a sequence that is thought to be scanned late in the splicing reaction (27,28).

In vitro trans splicing was observed by Konarska *et al.* (29) and Solnick (30). In both studies significant RNA–RNA secondary structures were required to tether the precursors for a reasonably efficient reaction (15–30% of wild-type). RNAs not tethered by secondary structure were spliced at 0.2% of wild-type (29). Konarska *et al.* (ibid.) and Solnick (30) cautiously suggested that their results precluded an obligate scanning mechanism that requires continuity of all phosphodiester bonds between the splice sites. Recently, *in vitro trans* splicing not requiring obvious base pairing of the substrates was demonstrated by Chiara and Reed (6), Bruzik and Maniatis (7) and Eul *et al.* (8). *Trans* splicing in mammalian systems is consistent with splice site pairing through a three dimensional diffusion mechanism because this model requires functional interactions between complexes assembled independently at the splice sites irrespective of the intron between them. Consequently, the three dimensional diffusion model is currently favored, although the relatively low efficiency of this reaction and the requirement for a downstream 5′ splice site or exonic enhancer remain difficult to explain.

To test the three dimensional diffusion model directly, we designed splicing substrates that would block scanning, but not three dimensional diffusion. The first approach utilized bipartite splicing substrates predicted to have minimal effects on splice site pairing through jumping, and either moderate or severe effects on scanning. The best interpretation of the results with the bipartite RNAs is that splice site pairing occurs through three dimensional diffusion. The second approach utilized a poly(ethylene glycol) obstacle to a scanning machinery that due to its flexibility was not predicted to interfere with a reaction proceeding through three dimensional diffusion. The kinetics of spliceosome assembly and the appearance of splicing intermediates and products with this substrate also argued in favor of splice site pairing through three dimensional diffusion.

MATERIALS AND METHODS

Preparation of templates and *in vitro* transcription

pPIP.RNA A was prepared by cloning the following sequence into the *Xho*I site in pPIP7.A (31): 5′-TCGAGAAGCTTCACCTGGCCCGCGGTGATGCCTGAATTCGCGGCCGCAAGCTTC-3′.
pPIP.RNA B was prepared by cloning the complementary sequence into a *Sac*I–*Xho*I deletion of pPIP7.A. pPIP.RNA C was prepared by cloning the complementary sequence to ZP04 between the *Sph*I and *Hin*dIII sites of pPIP7.A. pPIP.RNA F was prepared by cloning the *Eco*RI fragment from pPIP.RNA C into pPIP.RNA B. pPIP.RNA F1i and pPIP.RNA F2i were prepared by PCR amplification of fragment 1074–1183 from the Adenovirus 2 major late promoter transcription unit with primers carrying *Xho*I and *Not*I restriction sites respectively, and cloning this fragment in the antisense orientation into pPIP.RNA F. pPIP.RNA F2i5′ss was prepared by inserting the same fragment in the sense orientation into the *Not*I site in pPIP.RNA F. DNA templates were prepared for transcription as follows: pPIP.RNA A was restricted with *Not*I. pPIP.RNA F, F1i, F2i and F2i5′ss were prepared by PCR amplification using the sequencing forward primer [#1211 New England Biolabs, (NEB)] and a reverse primer ZP16: 5′-GGGGGGTCCATTAACCCTCACTAAAGGG-3′, and restriction of the PCR product with *Ava*II. Restriction enzymes were purchased from NEB. All oligodeoxynucleotides used for cloning, RT and PCR reactions were synthesized by an ABI 382 DNA/RNA Synthesizer. All RNAs were transcribed using T7 RNA Polymerase (Stratagene) (32).

Splicing and RT–PCR of bipartite RNAs

Before splicing, RNAs were annealed at 98°C for 2 min and cooled slowly to 34°C, then placed on ice. Splicing reactions were incubated for 2 h as previously described (33). Nuclear extracts were prepared as described by Dignam *et al.* (34). Splicing reactions were stopped by adding an equal volume of high salt buffer (7 M urea, 0.5% SDS, 100 mM LiCl, 10 mM Tris–HCl pH 7.5, 10 mM EDTA) on ice, and the RNAs were extracted and precipitated. Each RT annealing reaction contained 100 mM KCl, 83 nM RNA and 166 nM RT primer. RT annealing reactions were placed at 98°C for 1 min, 50°C for 20 min, then on ice. RT extension reactions contained 250 mM Tris–HCl pH 8.3, 375 mM KCl, 15 mM $MgCl_2$, 10 mM DTT, 500 µM each dNTP and 5 U/µl Moloney murine leukemia virus RT (Gibco-BRL), and were incubated at 37°C for 1 h. For direct visualization of the RT products, [α-^{32}P]dCTP was added to a final concentration of 50 nCi/µl. RT reactions were treated with RNase A at a final concentration of 0.04 mg/ml for 20 min at 50°C, then phenol-extracted and precipitated. Labeled RT products were resolved on denaturing, Tris–borate–EDTA (TBE), 8 M urea, 10% acrylamide:bisacrylamide (27:5:1) gels. PCR reactions contained 10 mM Tris–HCl pH 8.8, 50 mM KCl, 1.5 mM $MgCl_2$, 0.001% (W/V) gelatin, 10% of the resuspended unlabeled RT reaction, 1µM reverse primer ZP18, 1 µM forward primer ZP20, 200 µM each dNTP, 2 nCi/µl [α-^{32}P]dCTP [New England Nuclear (NEN)], and 0.01 U/µl *Taq* DNA polymerase (Stratagene). Reactions were cycled 21 times for 1 min at 94°C, 1 min at 65°C, 3 min at 72°C and followed by a final 10 min extension at 72°C. DNAs were extracted, precipitated, and resolved on native 7% acrylamide:bisacrylamide (30:1), TBE gels. Electrophoresis was at 7 V/cm for 4 h. A Molecular Dynamics PhosphorImager was used for all quantifications.

Directed ligations and splicing of PIP.PE, PIP.DNA and PIP7.A

RNA E was transcribed as described (32), except that reactions contained 4 Ci/mmol [α-^{32}P]UTP. RNA D was transcribed in reactions containing 0.05 Ci/mmol [α-^{32}P]UTP and with 40× molar excess of dGMP or GMP over GTP. PE or DNA were phosphorylated using T4 polynucleotide kinase (NEB). Reactions contained 1.4 µM each RNA, 1.1 µM each bridging oligo and 0.8 µM

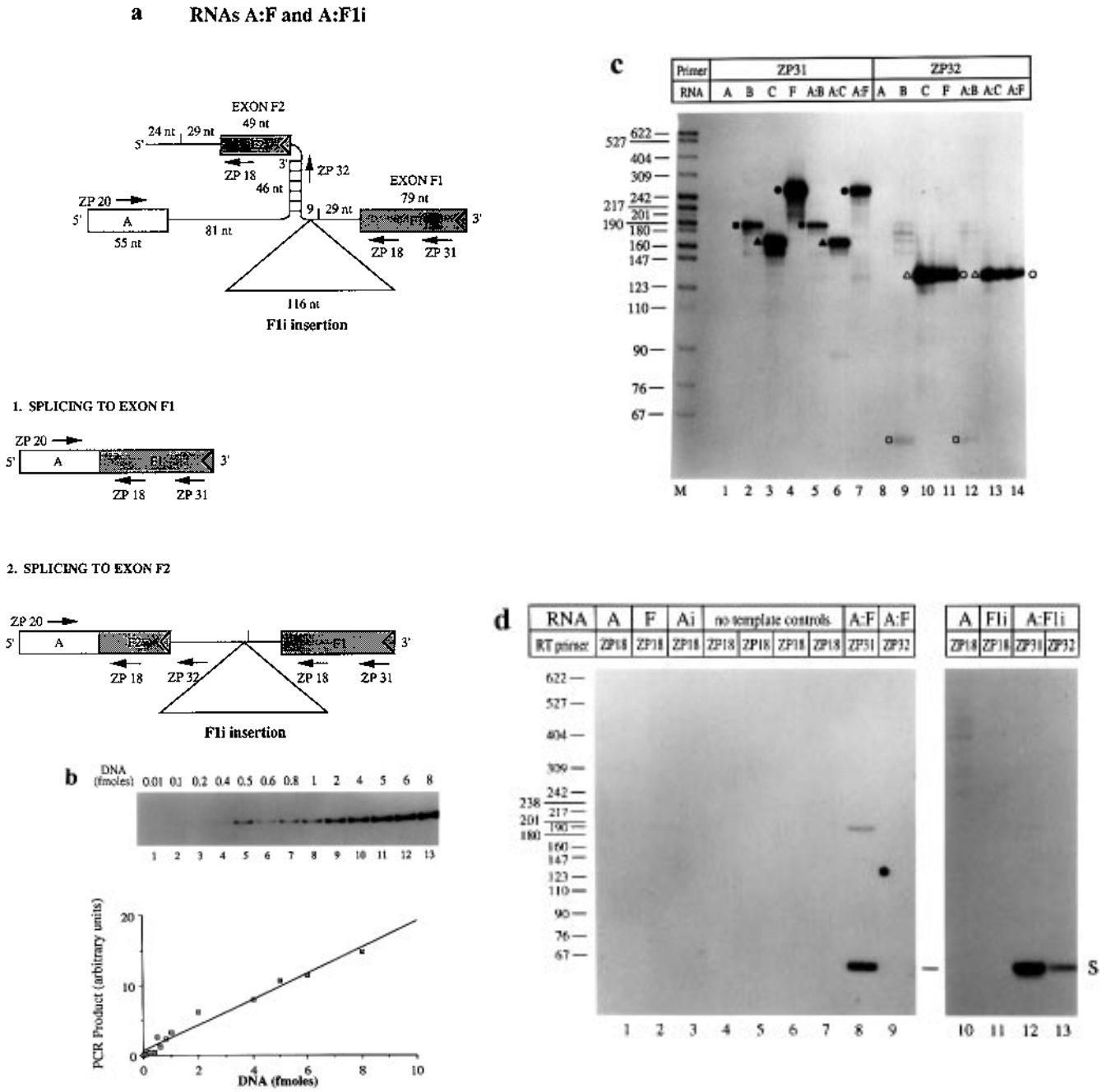

Figure 1. Splicing of bipartite RNAs. (**a**) Diagram of RNAs assayed for splicing and their possible splicing products. RNAs are drawn to scale and their lengths (in nucleotides) are indicated. Boxes indicate exons, lines indicate introns, and short vertical lines indicate branch points. Positions of reverse transcription (RT) primers ZP31 and ZP32, and polymerase chain reaction (PCR) primers ZP18 and ZP20 are indicated. RNA F1i contained a 116 nt insertion downstream of the stem forming sequence and upstream of the branchpoint. (**b**) PCR assay. PCR products were linear with respect to input DNA template over at least an 800-fold range. Primers were ZP18 and ZP20, and template was pPIP7.A (a plasmid carrying a contiguous sequence of exon A, the intron, and exon F1). (**c**) RT results using primer ZP31 or ZP32. Primers and substrates are indicated above each lane. Quantification of primer extension products and correction for the number of radioactive residues showed that ZP32 efficiency of extension from RNA F was 83% of ZP31 efficiency (compare lanes 4 and 7 with lanes 11 and 14). RNA B contained the same sequence as RNA F, except that it lacked the F2 exon and upstream sequence. RNA C contained the same sequence as in RNA F, except that it lacked sequences downstream of the stem-forming sequence. Major products expected from primer ZP31: RNA A: none, RNAs B, A:B: 189 nt (closed squares), RNA C, A:C: 166 nt (closed triangles), RNAs F, A:F: 260 nt (closed circles). Major products expected from primer ZP32: RNA A: none, RNAs B, A:B: 61 nt (open squares), RNAs C, A:C: 133 nt (open triangles), RNAs F, A:F: 133 nt (open circles). (**d**) RT–PCR results. The assay contained RNAs alone (lanes 1–3, 10 and 11; RNA Ai contained an unrelated 194 nt insertion downstream of the 5′ splice site and upstream of the stem forming sequence) or annealed RNAs A:F (lanes 8 and 9) and A:F1i (lanes 12 and 13). The band marked S represents spliced product (65 bp). Spliced and unspliced band identities were confirmed by direct sequencing of the PCR products (not shown). The faint band migrating near the 190 bp marker (lane 8) was the result of 'skipping over' the base of the stem by the RT enzyme. With RNA A:F1i, splicing to the F1 exon was 71% and splicing to the F2 exon was 29% of spliced RNA. Lanes 4–7 contained mock reactions containing no templates for the transcription (lane 4), splicing (lane 5), RT (lane 6) and PCR (7) reactions; each mock reaction was carried through all subsequent steps. When splicing to F2 was detected with ZP32, we were unable to detect similar levels of splicing to F2 with ZP31 as an RT and PCR primer. It is likely that RNA A remained annealed to the RNA produced by splicing to F2 and interfered with RT primed with ZP31. Therefore, the signal obtained with ZP31 as the RT primer represented splicing to F1 only.

kinased PE or DNA. Reactants were annealed at 98°C and slowly cooled to <10°C in water, then incubated under ligation conditions for 12 h at 16°C with 15–22 U/µl T4 DNA ligase (NEB) (35). PIP7.A was uniformly labeled in reactions containing 4 µCi/mmol [α-^{32}P]UTP. Splicing reactions were incubated at 30°C for the times indicated. To analyze splicing precursors, intermediates, and products, RNAs were purified as described above and resolved on denaturing 12–15% acrylamide:bisacrylamide (27:5:1), TBE, 8 M urea gels (33). To analyze splicing complexes, following incubation with nuclear extracts for the times indicated heparin was added to 0.5 mg/ml and reactions were incubated at 30°C for 5 min. Splicing complexes were resolved on native 4% acrylamide:bisacrylamide (80:1) Tris–glycine gels (36).

RESULTS

We constructed bipartite RNA splicing substrates predicted to interfere with a scanning mechanism, but not with a three dimensional diffusion mechanism (Fig. 1a). The double-stranded RNA stem in RNA A:F was expected to create a scanning block in splicing to exon F1 due to the discontinuity of the RNA at the base of the stem. We predicted, however, that this would be a weak block, as this substrate would potentially allow a scanning machinery to 'skip' over the base of the stem at some frequency, as it does for reverse transcriptase (37; see also Fig. 1d below). In contrast, splicing to exon F2 was expected to present a formidable block to single-stranded scanning because such a process would have to unwind the 46 base pair stem, switch scanned templates, and scan in both 5′→3′ and 3′→5′ directions.

A reverse transcription reaction followed by a polymerase chain reaction (RT–PCR) assay was used for detection and quantification of splicing efficiencies. The RT–PCR assay used was linear over at least an 800-fold range (Fig. 1b and data not shown). To differentiate splicing to exon F1 from splicing to exon F2, we used two different reverse transcription primers (Fig. 1a), which were equally efficient and gave the expected extension products (Fig. 1c). The reverse transcription primer ZP31 was used to detect splicing to F1, and primer ZP32 was used to detect splicing to F2 (see legend to Fig. 1). Splicing occurred almost exclusively (>99%) to exon F1 (Fig. 1d, lanes 8 and 9). This result suggested that a scanning mechanism was preferred over a three dimensional diffusion mechanism, and that the scanning machinery was inhibited in splicing to exon F2. We could not rule out, however, the possibility that exon F1 was a better splicing substrate in the particular configuration of RNA A:F (see Results below).

RNA F1i was synthesized to test the interpretation that the exclusive use of F1 above reflected splice site pairing via a scanning mechanism. The 116 nt insertion between the stem and the F1 branch point (Fig. 1a) was predicted to have minimal effects on scanning to the F1 splice site. Thus if the exclusive use of the F1 splice site in RNA A:F were due to scanning, we expected the splicing of RNA A:F1i to be identical to that of A:F. To our surprise, splicing assays with substrate A:F1i showed that 29% of spliced RNA was spliced to F2 (Fig. 1d, lanes 12 and 13). A different insertion in the same position of RNA F gave the same results, arguing that the effect of these insertions is not dependent on a particular sequence (not shown). A scanning model does not account well for these results. In contrast, a three dimensional diffusion model can fully explain them.

To address the possibility that exon F2 was an inferior splicing substrate due to its longer distance from the 3′ end of the RNA

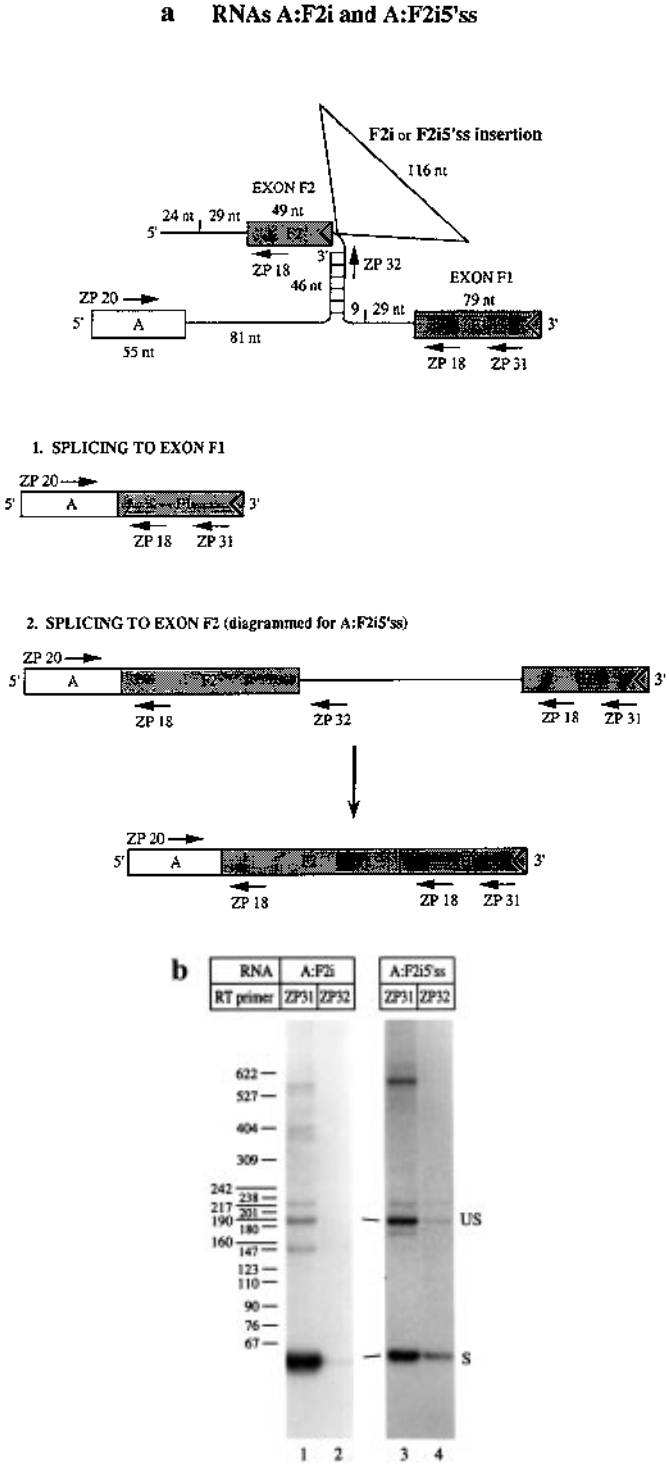

Figure 2. A 5′ splice site inserted downstream of the F2 3′ splice site stimulated splicing to F2. (**a**) Diagram of RNAs A:F2i, A:F2i5′ss and possible splicing products from A:F2i5′ss. RNAs are drawn to scale and their lengths (in nucleotides) are indicated. Note that splicing between AF2 and F1 (producing AF2F1) was not detected in significant levels with PCR reactions using primers ZP18 or ZP31. Also note that splicing between F2 and F1 was not measured in this assay. (**b**) RT–PCR of splicing reactions with A:F2i and A:F2i5′ss. The assay contained RNAs alone, which gave no signal (not shown) or annealed RNAs A:F2i (lanes 1 and 2) and A:F2i5′ss (lanes 3 and 4). The band marked S represents spliced product, US represents unspliced RNA. With RNA A:F2i, splicing occurred almost exclusively (>99%) to F1. With RNA A:F2i5′ss, splicing to the F1 exon was 68% and splicing to the F2 exon was 32% of spliced RNA.

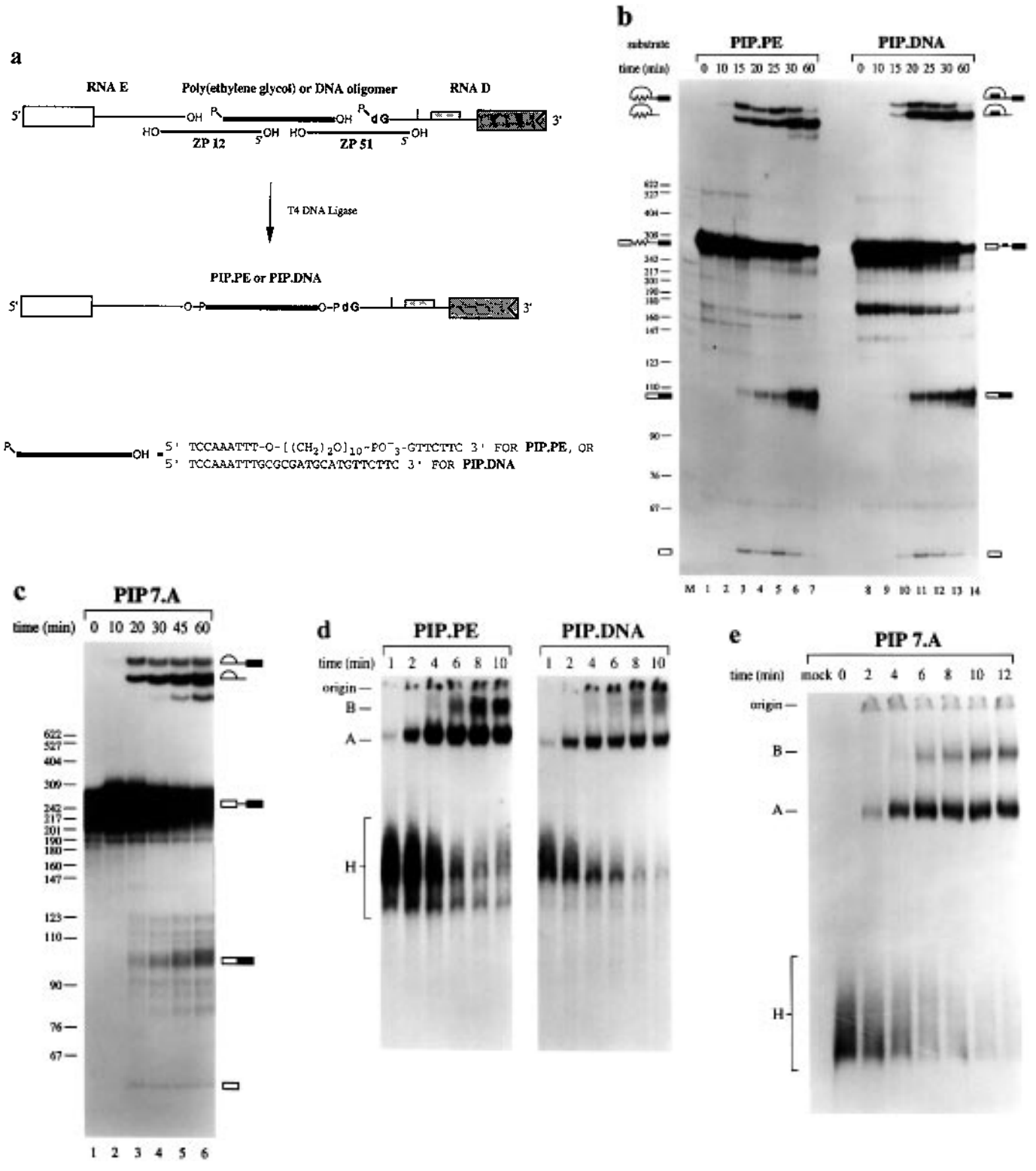

Figure 3. Splicing of PIP.PE, PIP.DNA and PIP7.A. (**a**) Construction of the splicing substrates PIP.PE and PIP.DNA by oligodeoxynucleotide-directed ligations. The oligodeoxynucleotide ZP12 was complementary to positions 122–141 of RNA E and 1–7 of PE or DNA (=ZP11). Oligodeoxynuceotide ZP51 was complementary to the seven 3′ end positions in PE or DNA, and to positions 1–20 of RNA D. T4 DNA ligase was used to generate PIP.PE or PIP.DNA (35). DNA–RNA ligation was enhanced 11-fold when priming transcription of RNA D with dGMP instead of GMP (not shown). (**b** and **c**) Results of splicing experiments with PIP.PE, PIP.DNA and PIP7.A. Substrates and incubation times are indicated above the lanes, and drawings representing reaction products and intermediates are shown beside the gels. Spliced product identities were confirmed by reverse transcription coupled to PCR amplification and sequencing of the PCR products (not shown). (**d** and **e**) Results of splicing complexes formation experiments with PIP.PE, PIP.DNA and PIP7.A. RNAs were incubated under splicing conditions and splicing complexes were resolved on native gels (29). Splicing substrates and incubation times are indicated above the lanes. H represents heterogeneous complexes, A represents the pre-spliceosome, and B represents the spliceosome.

(16), we synthesized RNAs F2i and F2i5′ss (Fig. 2a). RNA F2i did not contain a 5′ splice site, whereas the insertion in RNA F2i5′ss contained the L2 exon 5′ splice site from the adenovirus 2 major late transcription unit. It was demonstrated by many laboratories that a weak 3′ splice site could be made efficient by the insertion of a downstream 5′ splice site (16,17). The 5′ splice site sequence was not expected to affect scanning to either exon due to its location upstream of the 46 nt stem-forming sequence. Because of its ability to promote the assembly of splicing complexes on an upstream 3′ splice site (16,17), we predicted that a downstream 5′ splice site would enhance splicing to F2 via three dimensional diffusion.

Results of splicing assays with A:F2i and A:F2i5′ss are shown in Figure 2b. RNA A:F2i showed an almost exclusive use of the F1 exon (lanes 1 and 2), in the same manner previously obtained with RNA A:F. With RNA A:F2i5′ss, however, splicing to the F2 exon represented 32% of spliced RNA (Fig. 2b, lanes 3 and 4). A time course showed that splicing to F1 and F2 proceeded with similar kinetics with RNA A:F2i5′ss (not shown), consistent with the same mechanism operating in splicing to both exons. These results again argued against the scanning model. The best interpretation of the results obtained with all the bipartite splicing substrates is that the splice sites are brought together via a three dimensional diffusion mechanism.

We constructed another substrate in which splice site pairing via scanning was likely to become hindered, whereas splice site pairing via three dimensional diffusion would remain functional. We inserted a molecule consisting of 10 ethylene glycol repeats flanked by short DNA sequences (generous gift of A. Schepartz, Yale University) (38,39) into the intron, 24 nt upstream of the branch point (Fig. 3a). Since poly(ethylene glycol) (PE) does not resemble nucleic acids, we expected an RNA scanning machinery to have a diminished affinity for the ethylene glycol portion of PIP.PE, which was expected to reduce scanning efficiency. Due to the high flexibility of the PE moiety (38,39), PIP.PE was not expected to present a block to three dimensional diffusion. As a control, we used an oligodeoxynucleotide approximating the length of PE and containing identical sequences to the flanking DNA sequences in PE (PIP.DNA; Fig. 3a).

We incorporated PE or the control oligodeoxynucleotide into introns using the method described by Moore and Sharp (35), with modifications (see legend to Fig. 3a). Results of splicing assays with PIP.PE, PIP.DNA and PIP7.A are shown in Figure 3b and c. With all three substrates, the lariat intermediate began to accumulate concomitant with the appearance of the free 5′ exon after 10–15 min. mRNA and lariat product began to accumulate by 15–20 min in all three reactions. Quantification of mRNA levels showed that there were no differences in the initial rates of mRNA production with all three substrates (data not shown). Figure 3d and e shows the results of splicing complexes formation assays with PIP.PE, PIP.DNA and PIP7.A. Complex A (pre-spliceosome) appeared after 1–2 min of incubation with PIP.PE, PIP.DNA and the all-ribo control substrate PIP7.A. Complex B (spliceosome) began accumulating after 4–6 min of incubation with the three substrates. Figure 3 shows that the PE moiety did not interfere with the rate of splice site pairing in PIP.PE. Therefore, the kinetics of PIP.PE splicing, as assayed by the time of mRNA appearance, the initial rate of splicing catalysis, and the formation of splicing intermediates and products can be explained fully by splice site pairing through three dimensional diffusion.

DISCUSSION

Scanning models provided an attractive explanation for the efficient and precise removal of very large introns such as those found in the human c-*abl* (40) or in the *Drosophila ultrabithorax* (*ubx*) (41) genes. Given a three dimensional diffusion mechanism of splice site pairing and assuming that splicing of long introns (>10 kb) is efficient and precise, it is reasonable to assume that eukaryotes have evolved solutions to the problems these long introns present. One such solution was suggested for the alternatively spliced *ubx* 50 kb intron, where the splicing machinery was proposed to excise the intron in segments (42). Each splicing event in this intron may generate a novel 5′ splice site, which in turn may be used again. In some tissues, the entire intron may be removed by this mechanism (ibid.). A more common solution may be provided by an interaction between an elongating RNA polymerase II and factors that interact with 5′ splice sites (43). Greenleaf hypothesized that the C-terminal domain of the largest subunit of RNA polymerase II can interact with splicing factors such as SR proteins (ibid.). This would tether the polymerase to the nascent 5′ splice site occupied by U1 snRNP and SR proteins (44,45) such that upon synthesis of a 3′ splice site the two reactive groups for the first transesterification reaction would be in the vicinity of each other. Although according to this hypothesis the splicing machinery is ‘riding’ on the scanning polymerase, given our data we propose that the splicing machinery itself does not possess the scanning activity. Evidence for the association of SR proteins with the large subunit of RNA polymerase II is emerging from *in situ* immunolocalization studies (46).

A three dimensional diffusion mechanism of splice site pairing predicts a more error prone splicing machinery. mRNAs resulting from intramolecular (exon scrambling) and intermolecular (*trans* splicing) aberrant splicing may be produced more frequently than previously expected (47) (Fig. 4). Exon scrambling is predicted to occur particularly if splice sites are not immediately employed after being recognized, for example in pre-mRNAs that undergo alternative splicing and in exons preceding very large introns. The existence of scrambled exons (48–50) suggested that a three dimensional diffusion mechanism is operative in splice site pairing, although the mechanistic suggestion was undermined by the very low frequency of the observed scrambling events (0.01–1% in the *DCC* and c-*ets-1* genes and zero in the *p53* gene) (48–49). Why have these events not been recognized more frequently? First, exon scrambling is probably always associated with circularization as suggested by Nigro *et al.* (48) and Cocquerelle *et al.* (51), and first shown by Capel *et al.* (50). The circular *sry* transcripts and the presumed circular *DCC* and c-*ets-1* scrambled mRNAs are not polyadenylated and thus have been underrepresented in cDNA libraries. Second, cleavage and polyadenylation of circular RNAs can resolve the circles into potentially unstable linear RNA molecules due to their uncapped 5′ termini (52). Third, it is likely that proofreading functions have evolved to discard these aberrant transcripts, possibly functions similar to those that can sense the premature truncation of open reading frames (53–57).

In vivo trans splicing reactions were first observed in nematodes and trypanosomes, where a common leader segment is efficiently spliced to the body of many mRNAs (3–5). The expression of a *trans* spliced RNA was postulated to account for mRNAs encoding the μ variable region of a human transgene linked to endogenous

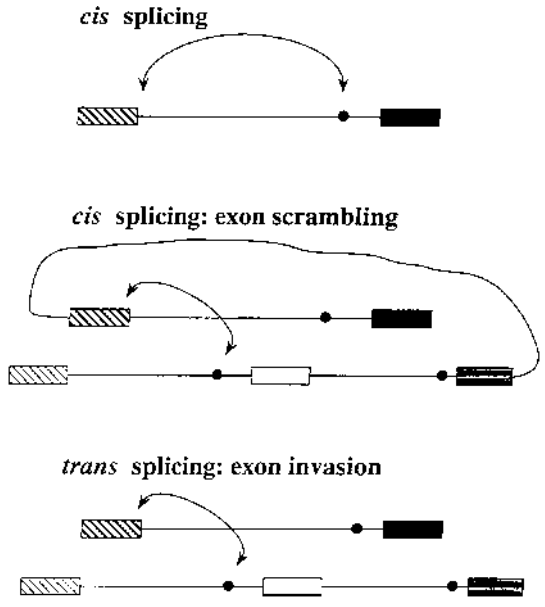

Figure 4. Possible modes of splice site pairing via a three dimensional diffusion mechanism. *Cis* splicing is the most common mode of interaction. The order of exons in the precursor is identical to the order of exons in the mRNA, and splicing of exons from one precursor generates one mRNA. In *Cis* splicing: exon scrambling, intramolecular interactions between the splice sites result in an mRNA whose exons are in a different order than the order of exons in the precursor. Presumably, exon scrambling proceeds through a circular intermediate. In *trans* splicing: exon invasion, splicing of exons from different precursors generates the mRNA. Genomic integration of *trans* spliced and scrambled mRNAs can result in exon shuffling.

murine γ1-constant region sequences in transgenic mice (58,59). Evidence for *trans* splicing into the c-*myb* mRNA in chicken and human thymus cells was obtained by Vellard *et al.* (60,61), and SV40 transcripts *trans* spliced to each other were observed in tissue cultured cells and in HeLa nuclear extracts (8). *Trans* splicing in mammalian cells is probably a rare event in pre-mRNAs involved in *cis* splicing (62). Nonetheless, a three dimensional diffusion mechanism of splice site pairing explains such events better than a scanning mechanism.

Nigro *et al.* (48) suggested that exon scrambling could be the source for exon shuffling. The same could be said for *trans* splicing (Fig. 4). Even a rare event can be captured and preserved by reverse transcription followed by genomic integration. This possibility may be termed 'exon invasion' (Fig. 4) and can illuminate a novel mechanism for evolutionary diversity. Whereas a strict scanning mechanism would forbid this type of exchange of genetic information, significant evolutionary events mediated by RNA–RNA interactions are more likely to occur in a splicing environment governed by interactions through three dimensional diffusion.

ACKNOWLEDGEMENTS

We thank A. Schepartz for the generous gift of the poly(ethylene glycol) tethered oligonucleotide, and L. Babis for the gift of adenovirus-2 genomic DNA. We are most grateful to J. A. Steitz and M. J. Moore for helpful suggestions and discussions. We thank all members of the Garcia-Blanco laboratory for stimulating discussions. We are also grateful to members of the RNA processing community, whose skepticism of the scanning model urged us to perform better experiments. This work was supported by a grant from the ACS to M.A.G.-B. Z.P. was supported in part by a training grant from the NIH to the Department of Microbiology, Duke University. M.A.G.-B. would also like to acknowledge the support of the Keck Foundation to the Levine Science Research Center.

REFERENCES

1 Matunis, E.L., Matunis, M.J. and Dreyfuss, G. (1993) *J. Cell Biol.*, **121**, 219–228.
2 Sharp, P.A. (1981) *Cell*, **23**, 643–646.
3 Sutton, R.E. and Boothroyd, J.C. (1986) *Cell*, **47**, 527–535.
4 Murphy, W.J., Watkins, K.P. and Agabian, N. (1986) *Cell*, **47**, 517–525.
5 Krause, M. and Hirsh, D. (1987) *Cell*, **49**, 753–761.
6 Chiara, M.D. and Reed, R. (1995) *Nature*, **375**, 510–513.
7 Bruzik, J.P. and Maniatis, T. (1995) *Proc. Natl Acad. Sci. USA*, **92**, 7056–7059.
8 Eul, J., Graessmann, M. and Graessmann, A. (1995) *EMBO J.*, **14**, 3226–3235.
9 Kuhne, T., Wieringa, B., Reiser, J. and Weissmann, C. (1983) *EMBO J.*, **2**, 727–733.
10 Lang, K.M. and Spritz, R.A. (1983) *Science*, **220**, 1351–1355.
11 Bhat, B.M., Brady, H.A., Pursley, M.H. and Wold, W.S.M. (1986) *J. Mol. Biol.*, **190**, 534–557.
12 Medford, R.M., Nguyen, H.T., Destree, A.T., Summers, E. and Nadal-Ginard, B. (1984) *Cell*, **38**, 409–421.
13 Mitchell, P.J., Urlaub, G. and Chasin, L. (1986) *Mol. Cell. Biol.*, **6**, 1926–1935.
14 Reed, R. and Maniatis, T. (1986) *Cell*, **36**, 681–690.
15 Robberson, B.L., Cote, G.J. and Berget, S.M. (1990) *Mol. Cell. Biol.*, **10**, 84–94.
16 Talerico, M. and Berget, S.M. (1990) *Mol. Cell. Biol.*, **10**, 6299–6305.
17 Kuo, H.-C., Nasim, F.-U.H. and Grabowski, P.J. (1991) *Science*, **251**, 1045–1050.
18 Watakabe, A., Tanaka, K. and Shimura, Y. (1993) *Genes Dev.*, **7**, 407–418.
19 Tian, M. and Maniatis, T. (1993) *Cell*, **74**, 105–114.
20 Lavigueur, A., La Branche, H., Kornblihtt, A.R. and Chabot, B. (1993) *Genes Dev.*, **7**, 2405–2417.
21 Ge, H. and Manley, J.L. (1990) *Cell*, **62**, 25–34.
22 Krainer, A.R., Conway, G.W. and Kozak, D. (1990) *Cell*, **62**, 35–42.
23 Mayeda, A. and Krainer, A.R. (1992) *Cell*, **68**, 365–375.
24 Caceres, J.F., Stamm, S., Helfman, D.M. and Krainer, A.R. (1994) *Science*, **265**, 1706–1709.
25 Yang, X., Nani, M.-R., Lu, S.-J., Rowan, S., Ben-David, Y. and Chabot, B. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 6924–6928.
26 Mayeda, A., Hayase, Y., Inoue, H., Ohtsuka, E. and Ohshima, Y. (1990) *J. Biochem.*, **108**, 399–405.
27 Smith, C.W.J., Porro, E.B., Patton, J.G. and Nadal-Ginard, B. (1989) *Nature*, **342**, 243–247.
28 Smith, C.W.J., Chu, T.T. and Nadal-Ginard, B. (1993) *Mol. Cell. Biol.*, **13**, 4939–4952.
29 Konarska, M.M., Padgett, R.A. and Sharp, P.A. (1985) *Cell*, **42**, 165–171.
30 Solnick, D. (1985) *Cell*, **42**, 157–164.
31 Gil, A., Sharp, P.A., Jamison, S.F. and Garcia-Blanco, M.A. (1991) *Genes Dev.*, **5**, 1224–1236.
32 Melton, D.A., Krief, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green, M.R. (1984) *Nucleic Acid Res.*, **12**, 7035–7056.
33 Grabowski, P.J., Padgett, R.A. and Sharp, P.A. (1984) *Cell*, **37**, 415–427.
34 Dignam, J.D., Lebovitz, R.M. and Roeder, R.G. (1983) *Nucleic Acid Res.*, **11**, 1475–1489.
35 Moore, M.J. and Sharp, P.A. (1992) *Science*, **256**, 992–997.
36 Konarska, M.M. and Sharp, P.A. (1986) *Cell*, **46**, 845–855.
37 Luo, G. and Taylor, J. (1990) *J. Virology*, **64**, 4321–4328.
38 Cload, S.T. and Schepartz, A. (1991) *J. Am. Chem. Soc.*, **113**, 6324–6326.
39 Richardson, P.L. and Schepartz, A. (1991) *J. Am. Chem. Soc.*, **113**, 5109–5111.
40 Bernards, A., Rubin, C.M., Westbrook, C.A., Paskind, M. and Baltimore, D. (1987) *Mol. Cell. Biol.*, **7**, 3231–3236.
41 Beachy, P.A., Helfand, S.L. and Hogness, D.S. (1985) *Nature*, **313**, 545–551.

42 Bomze, H.M. and Lopez, A.J. (1994) *Genetics*, **136**, 956–977.
43 Greenleaf, A.L. (1993) *TIBS*, **18**, 117–119.
44 Kohtz, J.D., Jamison, S.F., Will, C.L., Zuo, P., Luhrmann, R., Garcia-Blanco, M.A. and Manley, J.L. (1994) *Nature*, **368**, 119–124.
45 Jamison, S.F., Pasman, Z., Wang, J., Will, C., Luhrmann, R., Manley, J.L. and Garcia-Blanco, M.A. (1995) *Nucleic Acids Res.*, **23**, 3260–3267.
46 Bregman, D.B., Du, L., van der Zee, S. and Warren, S.L. (1995) *J. Cell Biol.*, **129**, 287–298.
47 Sharp, P.A. (1994) *Cell*, **77**, 805–815.
48 Nigro, J.M., Cho, K.R., Fearon, E.R., Kern, S.E., Ruppert, J.M., Oliner, J.D., Kinzler, K.W. and Vogelstein, B. (1991) *Cell*, **64**, 607–613.
49 Cocquerelle, C., Daubersies, P., Majerus, M.-A., Kerckaert, J.-P. and Baileul, B. (1992) *EMBO J.*, **11**, 1095–1098.
50 Capel, B., Swain, A., Nicolis, S., Hacker, A., Walter, M., Koopman, P., Goodfellow, P. and Lovell-Badge, R. (1993) *Cell*, **73**, 1019–1030.
51 Cocquerelle, C., Mascrez, B., Hetuin, D. and Baileul, B. (1993) *FASEB J.*, **7**, 155–160.
52 Moore, C.L. and Sharp, P.A. (1985) *Cell*, **41**, 845–855.
53 Maquat, L.E., Kinniburgh, A.J., Rachmilewitz, E.A. and Ross, J. (1981) *Cell*, **27**, 543–553.
54 Urlaub, G., Mitchell, P.J., Ciudad, C.J. and Chasin, L.A. (1989) *Mol. Cell. Biol.*, **11**, 3355–3364.
55 Cheng, J. and Maquat, L.E. (1993) *Mol. Cell. Biol.*, **13**, 1892–1902.
56 Dietz, H.C., Valle, D., Francomano, C.A., Kendzior, R.J. Jr, Pyeritz, R.E. and Cutting, G.R. (1993) *Science*, **259**, 680–683.
57 Pulak, R. and Anderson, P. (1993) *Genes Dev.*, **7**, 1885–1897.
58 Shimizu, A., Nussenzweig, M.C., Mizuta, T.-R., Leder, P. and Honjo, T. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 8020–8023.
59 Shimizu, A. and Honjo, T. (1993) *FASEB J.*, **7**, 149–154.
60 Vellard, M., Soret ,J., Viegas-Pequignot, E., Galibert, F., Van Cong, N., Dutrillaux, B. and Perbal, B. (1991) *Oncogene*, **6**, 505–514.
61 Vellard, M., Sureau, A., Soret, J,. Martinerie, C. and Perbal, B. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 2511–2515.
62 Bruzik, J.P. and Maniatis, T. (1992) *Nature*, **360**, 692–695.