Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model

Charles C. Query, Melissa J. Moore, and Phillip A. Sharp[1]

Center for Cancer Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307 USA

Selection of the nucleophile for the first step of nuclear pre-mRNA splicing was probed by site-specific incorporation into splicing substrates of nucleotides modified at the 2′ position. The differing abilities of ribose, 2′-deoxyribose, and arabinose nucleotides to base-pair within an RNA · RNA duplex and to contribute a nucleophilic 2′-OH group were exploited to analyze the paired/unpaired disposition of the branch site nucleotide. The results provide direct evidence for a bulged duplex model in which either of two adjacent purines within the consensus branch site sequence may shift into a bulged position and contribute the 2′-OH group for the first step of splicing. Furthermore, the presence of a consensus branch site that cannot present a reactive nucleophile suppresses splicing, including the use of cryptic branch sites elsewhere. We conclude that the branch site region base-pairing with U2 snRNA determines the first step nucleophile and persists at the time of the first transesterification reaction.

[*Key words*: RNA splicing; transesterification; nucleophile; site-specific modifications; arabinose; duplex stability]

Received November 15, 1993; revised version accepted January 11, 1994.

The splicing of intervening sequences from nascent RNA transcripts occurs via a two-step mechanism. The first step, nucleophilic attack at the 5′ splice site phosphate by the 2′-OH of a residue within the intron, the branch site, generates a free 5′ exon (E1) and a lariat form of the intron–3′ exon (IVS–E2). The second step results in ligated exons (E1–E2) and released lariat intron (IVS). Both steps are single in-line transesterification reactions (Maschhoff and Padgett 1993; Moore and Sharp 1993), catalyzed by the spliceosome, a 50–60S complex composed of the small nuclear RNAs (snRNAs)U1, U2, U4, U5, and U6, as well as protein components (for review, see Guthrie 1991; Moore et al. 1993).

A critical question has been how the sites on the pre-mRNA substrate for the chemical events of splicing are determined. For example, the processes by which the exact branch site nucleotide is chosen and how the 2′-OH of that nucleotide is activated as a nucleophile are not well understood. In the yeast *Saccharomyces cerevisiae*, the site of lariat formation is within the highly conserved sequence, 5′-UACUAAC-3′. In mammals, the consensus sequence is the more degenerate YURAC, but the UACUAAC prototype sequence is still optimal (Reed and Maniatis 1988; Zhuang et al. 1989). In both yeast (Domdey et al. 1984; Rodriguez et al. 1984; Lin et al. 1985) and mammals (Ruskin et al. 1984; Konarska et al. 1985; Hornig et al. 1986) it is the most 3′ adenosine (underlined above and throughout) at which lariat formation primarily occurs, although in mammalian cell extracts, the immediately 5′ adjacent nucleotide has been documented as an alternative branch site (Hornig et al. 1986; Noble et al. 1987; 1988). The branch region is recognized, in part, by base-pairing to a sequence within U2 snRNA (5′-GUAGUA-3′); this interaction has been proven in vivo by compensatory mutations in both yeast (Parker et al. 1987) and mammals (Wu and Manley 1989; Zhuang and Weiner 1989) at the positions in bold face type in UACUAAC · **GUAGUA**, collectively. However, although these compensations elegantly demonstrated Watson–Crick base-pairing at three positions, they did not provide any information as to the paired/unpaired disposition of the branch adenosine or of nucleotides 3′ to it. On the basis of the sequence complementarity (Parker et al. 1987) and in analogy to the potentially homologous group II self-splicing introns (Sharp 1985; Cech 1986; for review, see Jacquier 1990; Moore et al. 1993), it has been proposed that the nucleotide that is to become the branch site is bulged (i.e., having no base opposite it) from the UACUAAC · GUAGUA helix. Although an obvious possibility, this model has not been tested previously. It has not been shown whether the branch site · U2 snRNA pairing is necessary for spliceosomal assembly, for the chemical step(s) subsequent to assembly, or for both. An analogous base-pairing inter-

[1]**Corresponding author.**

action between the 5′ splice site and U1 snRNA is required only transiently for spliceosomal assembly (Wyatt et al. 1992; Konforti et al. 1993 and references therein), whereas the branch site · U2 snRNA interaction might play a direct role in identification and activation of the nucleophile for the first chemical step. In this study the process of branch nucleophile selection and the bulged duplex model were directly tested by incorporation of site-specific modifications within the branch sequence of a substrate pre-mRNA.

Results

Incorporation of modified nucleotides at the branch site

The splicing substrate used in the present study, PIP85.B, contains a consensus branch site sequence, UGCUGAC, at positions 149–155 of the intron. Primer extension analysis of all-ribose lariat intron species confirmed that the major branch site PIP85.B RNA is adenosine(154) (data not shown). To incorporate modified nucleotides and/or ^{32}P-label (‡) site-specifically within this sequence, a short RNA, RNA(146–156) (Fig. 1A), was transcribed using T7 RNA polymerase and the appropriate NTPs. Use of 2′-deoxy-ATP or ara-ATP resulted in a single modified adenine nucleotide at position 154. There was no detectable contamination of the modified RNAs with ribo-adenosine, as determined by digestion of RNAs(146–156) with ribonucleases (Fig. 1B). To generate the full-length E1–IVS–E2 RNAs containing one modified position, RNAs(146–156) and separately transcribed flanking sequences, RNA(1–145) and RNA(157–234), were joined using T4 DNA ligase (Fig. 1A; Moore and Sharp 1992) and analyzed for splicing as in Figure 2 and Table 1.

In unmodified RNA the preferred branch site is adenosine(154)

To quantitate precisely the usage of position 154 as the branch relative to adjacent positions, substrate and lariat RNAs containing a single ^{32}P label at position 154 were digested with nuclease P1. Nuclease P1 cleaves 5′ to any nucleotide, producing 5′-monophosphate nucleotides, but it does not cleave within branched trinucleotides (Konarska et al. 1985; Reilly et al. 1989). Thus, whereas digestion of singly labeled all-ribose E1–IVS–E2 RNA produced labeled adenosine-5′-monophosphate, $^{\ddagger}pA^{OH}$ (where the rightward-pointing -OH on adenosine indicates ribose; Fig. 3, lane 1), digestion of the IVS–E2 RNA produced two more slowly migrating bands containing almost all of the label (lane 3). These branched oligonucleotides could have been either $^{\ddagger}pA^{pG}pC$ or $pG^{pG}{}^{\ddagger}pA^{OH}$ (where superscripts indicate the 2′ constituents). Because the vast majority of the [^{32}P]phosphate was sensitive to treatment with phosphatase (lane 4), the predominant species was identified as $^{\ddagger}pA^{pG}pC$. The small amount of ^{32}P-label in the branched trinucleotides resistant to phosphatase treatment identified the minor branched trinucleotide as $G^{pG}{}^{\ddagger}pA^{OH}$ (lane 4). The ratio of $^{\ddagger}pA^{pG}pC$ to $G^{pG}{}^{\ddagger}pA^{OH}$ reflects the relative usage in branch formation of adenosine(154) to guanosine(153). For the sequence UGCUGAC, this ratio is 60:1 (Table 2).

Splicing of E1–IVS–E2 RNAs containing 2′-deoxyadenosine at position 154

The 2′-OH of the branch nucleotide is the nucleophile for the first step of splicing (Moore and Sharp 1993). The consequences of removing this preferred nucleophile were determined by examining the splicing of E1–IVS–E2 RNA containing a single 2′-H at position 154. Precedents for these experiments are manyfold: The nucleophile position of the hammerhead ribozyme (Uhlenbeck 1987) previously has been substituted with 2′-H (Perreault et al. 1990) or with 2′-O-methyl (Koizumi et al. 1989; Paolella et al. 1992) with the effect of abolishing detectable cleavage activity. Substitution of the branch site nucleophile could potentially have blocked the first step of splicing, generating spliceosomal complexes

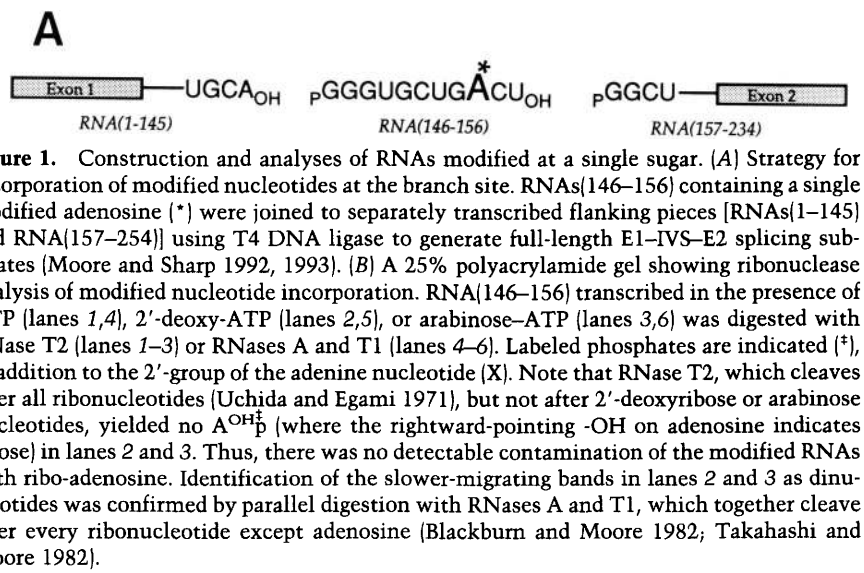

Figure 1. Construction and analyses of RNAs modified at a single sugar. (*A*) Strategy for incorporation of modified nucleotides at the branch site. RNAs(146–156) containing a single modified adenosine (*) were joined to separately transcribed flanking pieces [RNAs(1–145) and RNA(157–254)] using T4 DNA ligase to generate full-length E1–IVS–E2 splicing substrates (Moore and Sharp 1992, 1993). (*B*) A 25% polyacrylamide gel showing ribonuclease analysis of modified nucleotide incorporation. RNA(146–156) transcribed in the presence of ATP (lanes *1*,*4*), 2′-deoxy-ATP (lanes *2*,*5*), or arabinose–ATP (lanes *3*,*6*) was digested with RNase T2 (lanes *1–3*) or RNases A and T1 (lanes *4–6*). Labeled phosphates are indicated (‡), in addition to the 2′-group of the adenine nucleotide (X). Note that RNase T2, which cleaves after all ribonucleotides (Uchida and Egami 1971), but not after 2′-deoxyribose or arabinose nucleotides, yielded no $A^{OH}{}_p^{\ddagger}$ (where the rightward-pointing -OH on adenosine indicates ribose) in lanes *2* and *3*. Thus, there was no detectable contamination of the modified RNAs with ribo-adenosine. Identification of the slower-migrating bands in lanes *2* and *3* as dinucleotides was confirmed by parallel digestion with RNases A and T1, which together cleave after every ribonucleotide except adenosine (Blackburn and Moore 1982; Takahashi and Moore 1982).

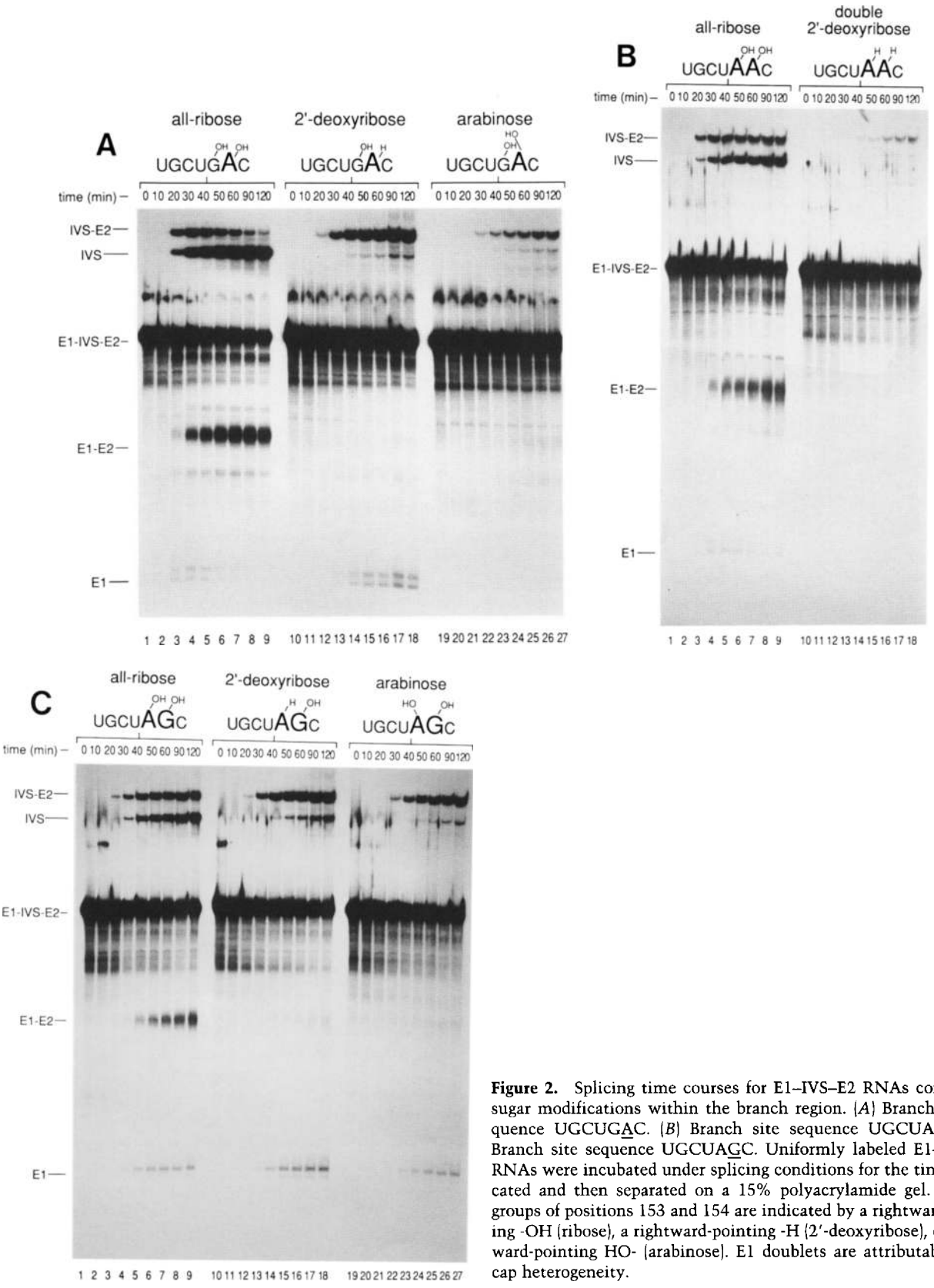

Figure 2. Splicing time courses for E1–IVS–E2 RNAs containing sugar modifications within the branch region. (*A*) Branch site sequence UGCUGAC. (*B*) Branch site sequence UGCUAAC. (*C*) Branch site sequence UGCUAGC. Uniformly labeled E1–IVS–E2 RNAs were incubated under splicing conditions for the times indicated and then separated on a 15% polyacrylamide gel. The 2′-groups of positions 153 and 154 are indicated by a rightward-pointing -OH (ribose), a rightward-pointing -H (2′-deoxyribose), or a leftward-pointing HO- (arabinose). E1 doublets are attributable to 5′ cap heterogeneity.

Table 1. *Relative rates and yields for splicing of modified substrates*

Sequence (2′-group X)	Relative rates			
	first step		second step	
	E1	IVS–E2	IVS	E1–E2
UGCUGAC (A: X)				
–OH	1.0	1.0	1.0	1.0
–H	0.52	0.34	0.00	0.04
HO–ara	0.15	0.10	0.02	0.05
GUCGUAC	0.58	0.25	0.05	0.07
UGCUGGC	0.81	0.36	0.07	0.05
UGCUAAC (A: X X)				
–OH, –OH	1.0	1.0	1.0	1.0
–H, –H	0.06	0.04	0.02	0.00

Relative rates for the first and second steps of splicing using RNAs site-specifically modified at the branch site. Polyacrylamide gels were quantitated using a Molecular Dynamics PhosphorImager and ImageQuant Software version 3.22. Rates were calculated independent of the specific site of branch formation. The relative amount of RNA in each band was expressed as a percentage of the total obtained by summing the values of E1–IVS–E2, E1, IVS–E2, IVS, and E1–E2 at that time point. For each band in every lane an individual background value was determined from the area in the same lane immediately above or below that band. Initial rates were chosen from the linear portion of each curve (four time points) and normalized to the respective all-ribose or wild-type sequence rate. Because of differing lag periods for the modifications, the time points comprising the linear portion of each curve varied.

stalled just prior to the first chemical event. However, the kinetics of splicing of this modified substrate showed only a 10-min lag in the appearance of first step products (E1 and IVS–E2; Fig. 2A, lanes 10–18), after which both E1 and IVS–E2 were generated at rates within two fold of those of the all-ribose RNA (Table 1). The E1 RNAs generated from both 2′-H substituted and all-ribose E1–IVS–E2 had identical mobilities (Fig. 2A, cf. lanes 10–18 and 1–9), indicating no shift in the site of 5′ exon cleavage. The almost complete lack of mature lariat IVS and spliced product E1–E2 in lanes 10–18, however, indicated that the rate of the second step of splicing was significantly reduced for the 2′-H modified RNA (Table 1).

The above effects were partially explained upon determination that the position of branch formation in the 2′-H-modified RNA was guanosine(153). Nuclease P1 digestion of lariats containing a single [^{32}P]2′-deoxyadenosine residue (Fig. 3, lane 9) yielded a species of lower mobility than the all-ribose $\ddagger pA^{pG}pC$ branch (lane 3). Assuming a usual 2′–5′ branch linkage, the labeled species in lane 9 could only be $pG^{pG}\ddagger pA^{H}$, where guanosine(153) formed the core of the branch. This identification was confirmed by phosphatase treatment, which yielded a ^{32}P-labeled species of even lower mobility, $G^{pG}\ddagger pA^{H}$ (lane 10), instead of $\ddagger p$. Therefore, substitution of a 2′-H for the 2′-OH at the preferred branch position resulted in efficient branch formation at the adjacent 2′-OH of guanosine(153). The subsequent block at the second step is consistent with previous reports that utilization of a guanosine residue as the branch site inhibits progression through the second step (Hornig et al. 1986; Freyer et al. 1987).

Two 2′-deoxynucleotide positions: 2′-deoxyadenosines or 2′-deoxycytidines

The above data demonstrate that within the consensus branch site sequence either of two adjacent positions, 153 or 154, may be used relatively efficiently as the nucleophile for the first step of splicing. To determine whether any other 2′-OH groups in the branch sequence could be activated for branch formation, a double 2′-deoxyadenosine substitution was prepared within the branch sequence UGCUAAC. Splicing of this doubly modified RNA showed a >30-min lag before the generation of any splicing intermediates (Fig. 2B, lanes 10–18) relative to the all-ribose control (lanes 1–9), after which E1 and IVS–E2 were produced at significantly reduced rates (down ~20-fold; Table 1). The total first step yield was only 5% compared with all-ribose RNA, and the mobility of the small amount of lariats formed was consistent with branching to cryptic branch sites (at positions 157 and 158; see below).

The assembly of splicing complexes was also examined (Fig. 4). For unmodified RNA, the first stable splicing-specific complex to form, A, is generated by binding of U2 snRNP to the branch region of the pre-mRNA. Next, a larger complex, B, is formed by association of complex A with U4/5/6 tri-snRNP (small nuclear ribonucleoprotein) particle. Complex C, which follows formation of complex B, contains splicing intermediates, and its appearance correlates with both the destabilization of U4 snRNP and the timing of the first chemical event (for review, see Moore et al. 1993). The E1–IVS–E2 RNA containing two adjacent 2′-deoxyadenosines assembled into complexes A and B with slightly reduced kinetics compared with the all-ribose RNA (Fig. 4, cf. lanes 10–18 and 1–9). At later time points, there was an accumulation of RNA in all complexes, consistent with the relative inability of the RNA to undergo either step of splicing. These observations suggest that U2 snRNP recognition of this doubly substituted site was relatively efficient and that the assembled complexes were unable to utilize an alternative nucleophile.

Thus, the double 2′-deoxyadenosine substitution resulted in a significant block to the first step, without utilization of immediately adjacent nucleotides; furthermore, cryptic site usage relative to all-ribose RNA was only minimally activated, in contrast to the strong activation observed after mutating the branch consensus (see below). Together with the data from the single-deoxyribose experiments, these results demonstrate that only two adjacent purine positions (153 and 154) may be used as alternative nucleophiles for the first step.

To assess whether 2′-OH modifications within the

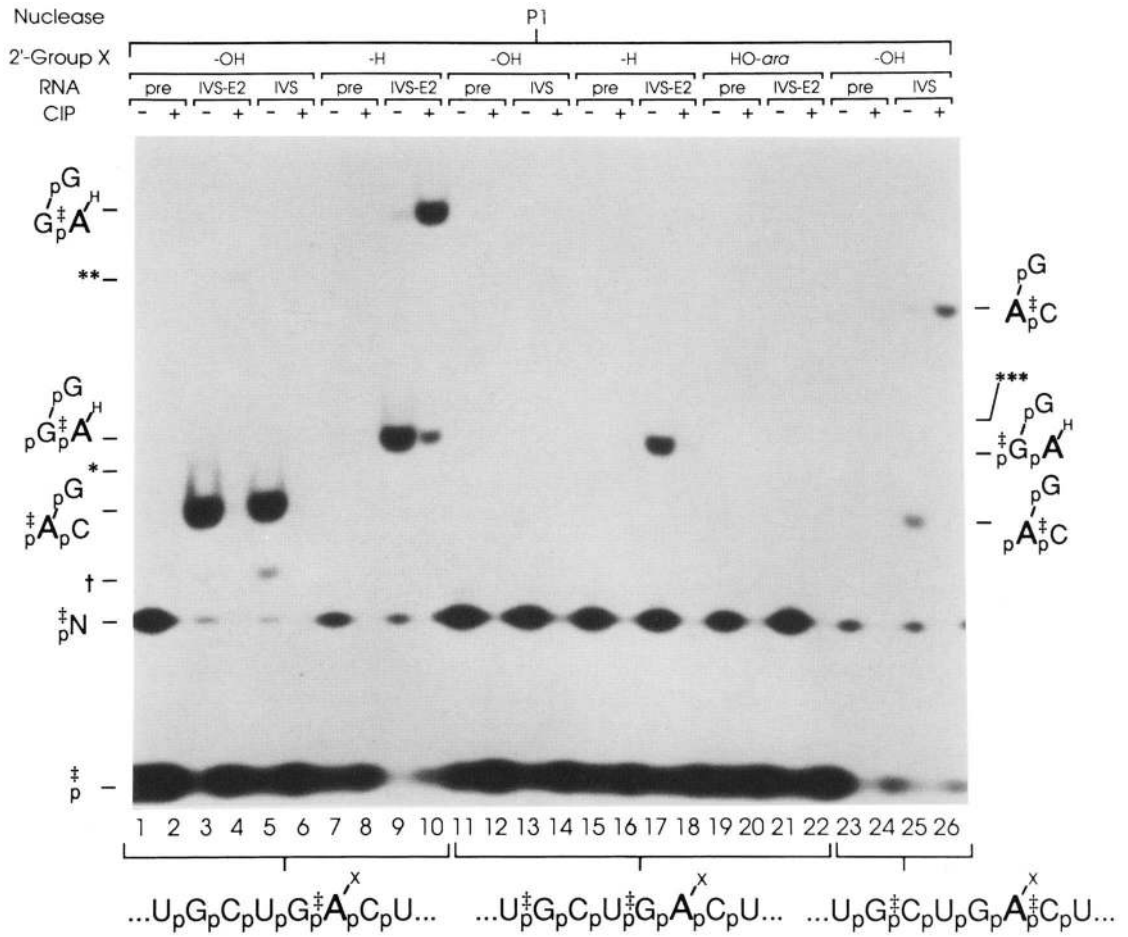

Figure 3. Nuclease analysis of branch site usage. A 25% polyacrylamide gel of site-specifically labeled splicing intermediates and products digested with nuclease P1. E1–IVS–E2 RNAs were constructed to contain ^{32}P-labeled phosphates (‡) as indicated below the lanes. The 2′-group of the adenine nucleotide (X) is indicated above. After incubation of E1–IVS–E2 RNAs under splicing conditions for 120 min, all species were purified on a 15% polyacrylamide gel. Subsequently, the E1–IVS–E2 (pre), IVS–E2, and IVS RNAs were digested with nuclease P1, and a portion of each sample was additionally treated with calf intestinal phosphatase (CIP) to remove 5′ phosphates (+). Structures of the predominant oligonucleotides are indicated. Three minor products, discussed in the text, are indicated by asterisks: * represents $pG^{pG}\ddagger pA^{OH}$ in lane *3*; ** represents $G^{pG}\ddagger pA^{OH}$ in lane *4*; and *** represents $\ddagger pG^{pG}pA^{HO}$ in lane *21*. † Indicates an unidentified minor band in lanes *3* and *5*. In addition to the analysis described in the text, the site of branch formation in RNAs containing 2′-deoxyadenosine(154) was determined independently using an [α-^{32}P]GTP labeling scheme (lanes 15–18). In this case, the labeled branch trinucleotide was $\ddagger pG^{pG}pA^{H}$ (lane *16*). Lanes *23–26* demonstrate lack of phosphodiesterase activity during the phosphatase treatment; as expected, the internal [^{32}P]phosphate in the $pA^{pG}\ddagger pC$ trinucleotide (lane *25*) was resistant to phosphatase treatment, resulting in labeled $A^{pG}\ddagger pC$ (lane *26*). ‡pN indicates the labeled 5′ monophosphate nucleotides $\ddagger pA^{OH}$ (lnes *1, 3, 5*), $\ddagger pA^{H}$ (lanes *7, 9*), ‡pG (lanes *11, 13, 15, 17, 19, 21*), and ‡pC (lanes *23,25*), which comigrated on this gel.

consensus branch region not involving the branch nucleophile also affected the first step, a double 2′-deoxycytidine substitution was prepared within the UGCUGAC branch sequence. A splicing time course of the resultant E1–IVS–E2 RNA showed no significant difference in the rate of either step of splicing compared with the all-ribose control (data not shown). This is consistent with the effects of substitutions of the 2′-OH at the preferred branch position being specific to the first chemical step or to specific recognition of the 2′-OH nucleophile, rather than a general effect of 2′-OH removal in the branch region.

Splicing of E1–IVS–E2 RNAs containing arabinosyladenine at position 154

The alternative use of two adjacent purine positions for branch formation suggests a dynamic model where one purine base-pairs with the complementary uridine in U2 snRNA while the other nucleotide occupies the bulged position, thereby being specified for reactivity (see Fig. 6, below). Incorporation of arabinosyladenine (the C2′ epimer of ribo-adenosine) offered a test of this model. Previous molecular modeling studies had suggested that an arabinosyl nucleotide should not base-pair into an

Table 2. *Summary of alternative utilization of branch positions*

Modification	Sequence UGCU--C	Total yield of lariats	Percent first position (153)	Percent second position (154)	Relative use of cryptic sites[a]
—	AA	1.0	4	96	–
—	GA	1.0	2	98	–
—	AG	1.0	50	50	–
—	GG	0.60	25	75	+/–
2′-deoxyribose	G(H)A	0.49	95	0	–
arabinose	G(HO)A	0.12	5	0	+
2′-deoxyribose	A(H)G	0.88	0	95	–
arabinose	A(HO)G	0.53	N.D.	4	++
2′-deoxyribose	A(H,H)A	0.05	0	0	+/–
—	GUCGUAC	0.42	N.D.	N.D.	++++

Alternative utilization of branch positions was determined using substrate RNAs containing site-specific [^{32}P]phosphates and the sequences and modifications shown. These RNAs were spliced in vitro, and the products of the reactions separated and purified. IVS–E2 and IVS RNAs were digested using ribonucleases or nuclease and analyzed in polyacrylamide gels as in Fig. 3. The amount of RNA in each band was quantitated using a Molecular Dynamics PhosphorImager. (+ and –) Total level of cryptic site usage relative to the GUCGUAC sequence (++++); (++) 25–50% relative usage; (+) 5–25%; (+/–) 1–5%; (–) <1%. (N.D.) Not determined.
[a]Although the cryptic branch sites constituted a majority of the branches detected in the two arabinose substitution cases, they do not represent strong activation of the cryptic sites, because the overall efficiency of splicing was low (cf. Total yield of lariats).

A-form RNA duplex (Gao et al. 1991). Therefore, it might preferentially occupy the bulged position when present in the branch site · U2 snRNA duplex. In this case, the adjacent ribose nucleotide would be preferentially paired within the branch site · U2 snRNA duplex and, according to this model, not available as the first step nucleophile.

To determine directly the effect of arabinosyladenine within the sequence being examined in this study, the melting temperatures of two RNA duplexes, one containing a single ribo-adenosine and the other containing arabinosyladenine, were compared (Fig. 5). For completely complementary duplexes the arabinose-containing RNA had a lower T_m by 10°C, suggesting destabilization of the helix. RNA duplexes containing ribo-adenosine or arabinosyladenine at a potentially bulged position, however, had virtually identical T_ms. Taken together, this strongly suggests that arabinosyladenine cannot stably base-pair within a duplex and preferentially occupies the bulged position in the UGCUGAC · GUAGUA context.

Substitution of arabinosyladenine at position 154 resulted in a dramatic reduction in both steps of splicing (Fig. 2A, lanes 19–27). The lag time before E1 and IVS–E2 appearance was extended by >20 min, after which the rate of accumulation of these RNAs was decreased ~10-fold relative to all-ribose RNA (Table 1). Essentially no second step products could be detected. In arabinose-containing IVS–E2 [generated from E1–IVS–E2 prepared with [α-^{32}P]CTP and unlabeled arabinosyladenine in RNA(146–156)], no branches utilizing the 2′-OH of arabinose were detectable (data not shown). This was not unexpected because of the altered stereochemistry of the 2′-OH of arabinose compared with ribose. More interestingly, neither was the 2′-OH of guanosine 153 the primary branch site, as digestion of IVS–E2 RNA labeled with [α-^{32}P]GTP generated predominately ‡pG

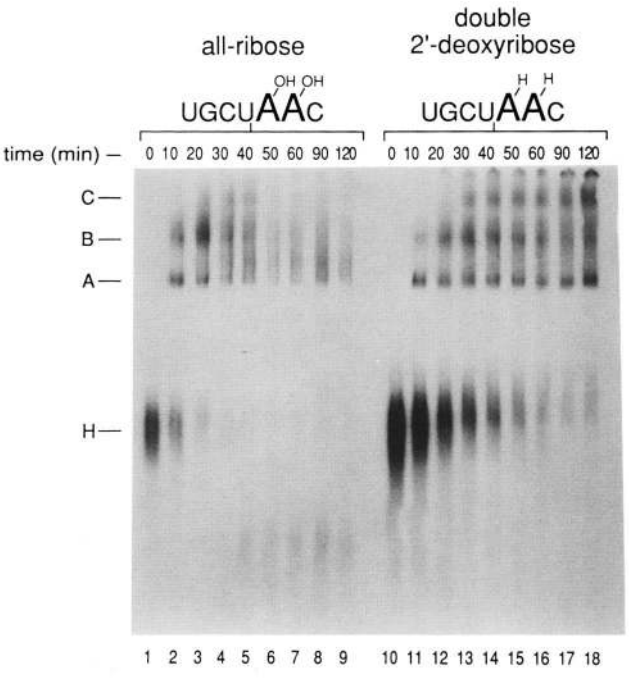

Figure 4. Time course of splicing complex assembly for all-ribose and double-2′-deoxyadenosine-containing E1–IVS–E2 RNAs. Uniformly labeled E1–IVS–E2 RNAs containing the branch sequence UGCUAAC with all-ribose (lanes *1–9*) or two 2′-deoxyadenosines (lanes *10–18*) were incubated under splicing conditions for the times indicated, adjusted to 0.5 mg/ml of heparin, and loaded onto a native 4% polyacrylamide gel.

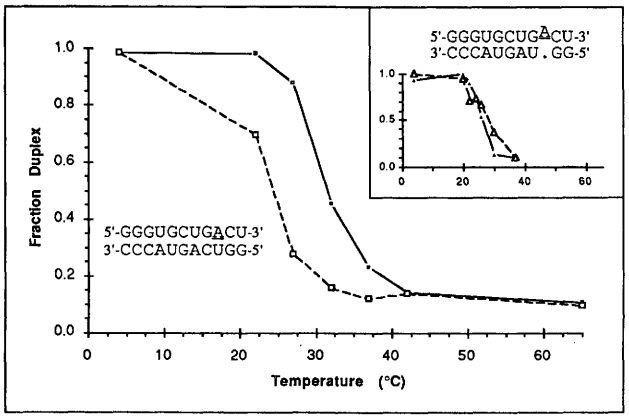

Figure 5. Melting temperatures of RNA · RNA duplexes containing a single ribo-adenosine or a single arabinosyladenine. Linear duplex (completely complementary). (*Inset*) bulged duplex. Duplexes were separated from monomer RNAs by electrophoresis on a 15% polyacrylamide (19:1) gel and quantitated using a Molecular Dynamics PhosphorImager. (Broken line) single arabinosyladenine RNA; (solid line) all–ribose RNA.

(Fig. 3, lane 21) and only a small amount (<5%) of pG^{p}G^{HO}pA (where the leftward-pointing HO- on adenosine indicates arabinose).

Several other labeling schemes and ribonuclease digestions revealed that the primary branch sites in the arabinosyladenine-containing RNA were guanosine(157) and guanosine(158) (data not shown). The sequence around positions 157 and 158 shows partial complementarity to U2 snRNA and is typical of other cryptic branch site sequences (Padgett et al. 1985; Rautmann and Breathnach 1985; Ruskin et al. 1985; Hornig et al. 1986). Thus, the arabinose substitution suppressed splicing within the preferred consensus sequence and only a low level of branches at a nearby cryptic site were detected. Most significantly, the arabinose substitution at position 154 did not allow efficient utilization of the adjacent guanosine(153) residue as the branch.

The most active cryptic branch sites in all-ribose RNA are positions 157 and 158

Because the arabinosyladenine modification predominantly caused utilization of a nearby cryptic branch site, we compared its effects with those caused by mutations. To reveal the most active cryptic branch site in PIP85.B RNA, the consensus branch sequence UGCUGAC was changed to GUCGUAC, by exchanging all of the guanosine and uridine residues. This mutant sequence effectively disrupts any complementarity to U2 snRNA, but it retains the single adenosine residue at position 154. The first step rate of splicing of the GUCGUAC mutant E1–IVS–E2 RNA was only reduced twofold relative to that of the consensus UGCUGAC RNA (Table 1). Primer extension analysis coupled with single [^{32}P]phosphate labeling and RNase digestions similar to those above showed that guanosine(157) and guanosine(158) were the primary branch sites used in the mutant RNA (data not shown). Thus, mutation of the consensus branch sequence resulted in efficient activation of the same nearby cryptic branch sites as those observed at low levels in the arabinose- and double 2′-deoxy-containing RNAs.

For subsequent experiments (see below) it was also important to ascertain the effects of all four possible purine–purine combinations (AA, AG, GA, GG) at position 153 and 154 within the context of an otherwise consensus branch site sequence (UGCURRC). This comparison confirmed the marked preference for usage of positions 154, and, when guanosine and adenosine are compared for relative activity at a given position, the adenosine is always preferred (Table 2). For example, positions 153 and 154 are utilized in a 1:24 ratio in the AA set as compared with a 1:1 ratio in the AG set. The bias for an adenosine in the reactive site is also apparent at position 153 where usage is 1:1 in the AG set compared with 1:3 for the GG set.

Splicing of E1–IVS–E2 RNAs containing modified sugars at position 153

Because the 2′-OH at position 153 was used as the branch site nucleophile when position 154 contained 2′-H, the effects of modifications at this position were also tested. To incorporate 2′-H or 2′-arabinose nucleotides at position 153 specifically, RNA(146–156) with the branch sequence UGCUAGC was used. The 2′-deoxyadenosine(153) substitution in this sequence had little effect on the first step compared with the all-ribose RNA of the same sequence (Fig. 2C, cf. lanes 10–18 and 1–9). But, as would be expected for branch formation at guanosine(154), the rate of the second step was reduced. Subsequent branch site analysis similar to those described above confirmed that >95% of the branches formed on the 2′-deoxyadenosine(153) RNA were to guanosine(154).

Substitution of arabinosyladenine at position 153, however, produced quite different effects. The total accumulation of intermediates relative to all-ribose RNA was decreased (down twofold; Fig. 2C, cf. lanes 19–27 and 1–9) and all splicing was arrested prior to step 2. The IVS–E2 that did accumulate contained branches predominantly at the cryptic sites 157 and 158, with the use of guanosine(154) being only 4% (Table 2). Therefore, whereas substitution of 2′-deoxyadenosine at position 153 resulted in exclusive use of the preferred site (154), arabinosyladenine at position 153 did not allow efficient usage of position 154. Coupled with the previous results from substitutions at position 154, these observations demonstrate that the presence of an arabinose suppresses utilization of the adjacent nucleotide in branch formation.

Discussion

Previous genetic studies in yeast and mammalian systems have shown that recognition of the branch site re-

gion by U2 snRNA via an RNA · RNA duplex is essential for pre-mRNA splicing (Parker et al. 1987; Wu and Manley 1989; Zhuang and Weiner 1989). The present biochemical analyses of branch site utilization provide evidence that commitment to a branch region does not require the 2′-OH nucleophile and that either of two adjacent bases in the branch site · U2 snRNA duplex may shift into a bulged position (i.e., unpaired within the duplex) and be activated to form the branch. As discussed below, this suggests that the branch site · U2 snRNA duplex is an active component in the first transesterification reaction.

Commitment to a branch region

Branch site sequences in mammalian introns are not highly conserved, and mutation of the most frequently used site commonly results in utilization of a nearby cryptic site (Padgett et al. 1985; Rautmann and Breathnach 1985; Ruskin et al. 1985; Hornig et al. 1986). This suggests a dynamic process in which several potential branch sites compete for spliceosome components in formation of an initial stable complex (Fig. 6). Following this initial competition, the results of the present study suggest that the spliceosome strongly commits to a particular site for branch formation. The primary data for this conclusion are the total yields of lariat intermediate generated from substrates containing mutated versus chemically modified branch site sequences (Table 2). Extensive mutation of the consensus branch sequence resulted in an efficient reaction at cryptic branch sites nearby. Substitution of two 2′-deoxyadenosine residues in the consensus branch site sequence, however, resulted in a 20-fold overall reduction in branch formation, suppressing the alternative use of the cryptic site. We interpret this to reflect the stable binding of the spliceosome to the nonreactive consensus sequence, thereby precluding recognition of the cryptic site. Thus, recognition of an otherwise consensus, but nonreactive, branch sequence likely commits the spliceosome to this nonproductive site.

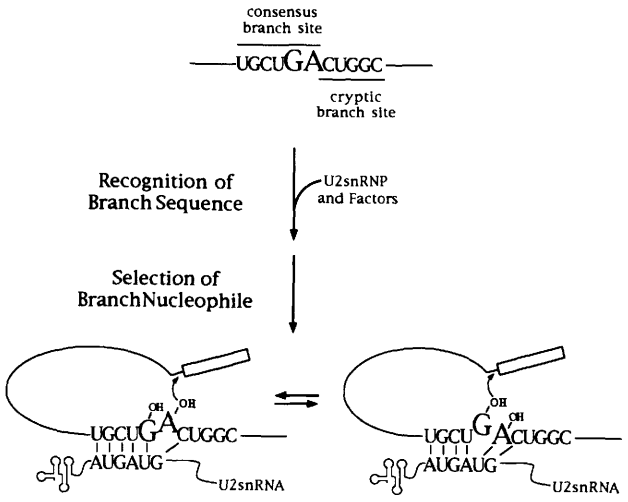

Figure 6. Bulged duplex model of branch site · U2 snRNA recognition and nucleophile selection for the first step of nuclear pre-mRNA splicing. A pre-mRNA substrate may contain several potential branch regions. U2 snRNP, when associating with the pre-mRNA substrate, may initially base-pair with a number of these regions, but the process eventually selects one sequence for stable base pairing. This commitment does not require the chemically active 2′-OH group. In pairing with U2 snRNA, either of two positions in the selected branch region can bulge from the duplex; although this is diagramed only for the consensus branch site above, the same alternative bulging occurs within complexes committed to the cryptic branch site. The 2′-OH of this bulged nucleotide in the duplex with U2 snRNA then becomes the active nucleophile for the first step of splicing.

Selection of the 2′-OH as nucleophile

In this study, two adjacent purines within a strong consensus sequence, UGCUR$_1$R$_2$C, were used alternatively for branch formation. Given the extensive complementarity between the branch site sequence and U2 snRNA, it is likely that this alternative reactivity of adjacent bases occurred from a single duplex formed with U2 snRNA (Fig. 6).

After base-pairing with U2 snRNA, what determines the particular nucleotide that becomes activated for branch formation? The reactive nucleotide must occupy a bulged position in the RNA duplex region [alternative pairing schemes denoted by interconversion (⇆) signs in Fig. 6]. The present evidence supporting this conclusion is severalfold. First, substitution of a 2′-deoxyadenosine into the preferred site for branch formation, the second position, resulted in efficient utilization of the adjacent guanosine in the first position. Second, and in contrast, substitution of an arabinosyladenine into the second position suppressed utilization of the adjacent guanosine, even though the arabinosyladenine was not used as a nucleophile itself. Based on the T_m effects observed with the different sugar substitutions, we propose that in the first case, 2′-deoxyadenosine can base-pair with the complementary uridine in U2 snRNA (GUAGUA) while the adjacent guanosine bulges from the RNA duplex. In the second case, arabinosyladenine cannot stably base-pair with the opposing uridine and remains bulged from the RNA duplex, suppressing the presentation of the adjacent guanosine. This interpretation is consistent with the further findings that substitution of an arabinosyladenine into the first position of the tandem purines strongly suppressed the use of the guanosine in the preferred second position. In comparison, substitution of 2′-deoxyadenosine in the first position allowed the guanosine in the second position to be used efficiently for branch formation. Thus, taken together, the effects of the 2′-deoxyribose and arabinose substitutions provide evidence for the bulged duplex model, that either of two positions may alternatively bulge from the branch site · U2 snRNA duplex thereby being selected as the branch site.

What is the status of the branch site · U2 snRNA in-

teraction at the time of the first chemical step? Formally, there are two possibilities. The first, and by far simplest, scenario is that the branch site · U2 snRNA pairing is maintained during the first step. In this case, the bulged duplex would function both to select the branch site nucleotide and to present the 2′-OH of that nucleotide directly into the catalytic site. The nucleophile would be thus specified by the unique structure of the sugar–phosphate backbone of the bulged duplex. In the second scenario, the branch region · U2 snRNA pairing would be disrupted prior to the chemical step, thereby uncoupling branch nucleotide selection from nucleophile activation. In this case, nucleophile presentation would occur from a structure other than the branch site · U2 snRNA bulged duplex. This would require that a different spliceosomal component bind the branch site in a manner that (1) is independent of base sequence, (2) is independent of 2′-OH, 2′-H, or arabinose sugars, (3) allows no slippage to adjacent nucleotides, and (4) can cycle back to the U2 snRNA pairing step to allow for alternative selection of the adjacent position. Taken together, these unusual characteristics make the existence of such a component highly unlikely. Thus, we conclude that the branch site region is base-paired with U2 snRNA at the time of the first transesterification reaction.

Position and base effects

Occupation of the bulged position in the RNA duplex is not the only criterion for reactivity in branch formation. An adenosine residue is more reactive than guanosine or other bases at this position (this study; Hornig et al. 1986). When all four purine–purine variants were tested for relative activity as a branch site within a consensus sequence, the comparison (Table 2) revealed a marked preference for usage of the second position, and at the second position, there was a distinct selection for adenosine. Thus, some aspect of the adenosine base contributes to proper presentation of the nucleophile into the active site for the first step. This could result from either direct recognition of some substituent on the base or the relative propensity of adenosine to stack into a helix when occupying a bulged position.

The bulged duplex model

The results presented here suggest the following model of branch nucleophile selection. A pre-mRNA may contain several potential branch regions (Fig. 6). U2 snRNP, when associating with the pre-mRNA, may initially base-pair with a number of these regions, but the process eventually selects one sequence, leading to formation of a stable complex. In pairing with U2 snRNA, either of two positions in the selected branch sequence can bulge from the duplex. There are preferences both for the second of the two positions to be bulged and for the bulged nucleotide to be an adenosine; if the second position is an adenosine, the effects are synergistic. The 2′-OH of this bulged nucleotide in the duplex with U2 snRNA then becomes the active nucleophile for the first step of splicing.

Materials and methods

Plasmids and RNA transcription

pPIP85.B is a splicing construct derived from pPIP85.A (Moore and Sharp 1992) to include a single adenosine in the branch region and a polypyrimidine tract modified for efficient use in RNA · RNA ligations. It was constructed by standard subcloning of reverse-transcribed cDNA of the three-part ligation product RNA(1–145)(146–156)(157–254) described in Figure 1A and contains the following 235-nucleotide sequence between the T7 promoter and *Hind* III site of pBS- (Stratagene): 5′-GGGCGAATTCGAGCTCACTCTCTTCCGCATCGCTG-TCTGCGAGGTACCCTACCAG↓GTGAGTATGGATCCCT-CTAAAAGCGGGCATGACTTCTAGAGTAGTCCAGGGTT-TCCGAGGGTTTCCGTCGACGATGTCAGCTCGTCTCGA-GGGTGCTGACTGGCTTCTTCTCTCTTTTTCCCTCAG↓-GTCCTACACAACATACTGCAGGACAAACTCTTCGCGG-TCTCTGCATGCAAGCTT-3′. Arrows indicate the 5′ and 3′ splice sites, and the underlined A indicates the preferred branch site. RNA(1–145) and RNA(1–234) were transcribed from plasmid pPIP85.B cleaved with *Xho*I or *Hind*III, respectively. A DNA template for transcription of RNA(157–234) was generated by PCR with a deoxyoligonucleotide primer that inserted a T7 promoter immediately 5′ to nucleotide 157, and M13 reverse primer, followed by cleavage with *Hind*III. These transcriptions were performed under standard conditions (see Moore and Sharp 1992, 1993).

Short RNAs containing the branch site region [RNAs(146–156), with sequences as described in the figure legends] were transcribed using T7 RNA polymerase and deoxyoligonucleotide templates (Milligan and Uhlenbeck 1989). Conditions generally were 40 mM Tris-HCl (pH 8.0), 1 mM spermidine, 20 mM $MgCl_2$, 0.01% Triton X-100, 50 μg/ml of BSA, 5 mM DTT, 100 μM template, 5 U/ml of T7 RNA polymerase (Stratagene), 8 mM GMP, and 2 mM each NTP, with incubations at 37°C for 2 hr, but were modified as described below for individual nucleotide incorporations. For inclusion of single 2′-deoxyadenosine or arabinosyladenine nucleotides, the ATP was replaced by 2 mM dATP (Pharmacia) or 2 mM arabinosyl–ATP (Sigma). For transcripts containing two 2′-deoxyribose residues, reactions additionally contained 3 mM $MnCl_2$ and were incubated at 30°C. For high specific activity RNAs, reactions included 40 μM NTP and 1 mCi of [α-^{32}P]NTP (New England Nuclear) or 200 μM dNTP and 1 mCi [α-^{32}P]dNTP. The 11-nucleotide transcription products were purified on 37% polyacrylamide (19:1) gels run in 1× TBE (89 mM Tris-borate, 2 mM EDTA) and visualized by autoradiography or UV shadowing.

Determination of melting temperatures

For the completely complementary duplexes, RNAs(146–156) containing all-ribose or a single arabinosyladenine were hybridized to the complementary ribo-oligonucleotide (Fig. 5, lower) at 65°C followed by slow cooling to 4°C in 100 mM KCl, 50 mM Tris-glycine, 1 mM EDTA. Bulged duplexes (Fig. 5, upper) were similarly hybridized, but in 250 mM KCl, 50 mM Tris-glycine, 10 mM $MgCl_2$. All RNAs were present at a concentration of 10 nM. Following hybridization, samples were adjusted to contain 20% glycerol, 0.05 mg/ml of *Escherichia coli* tRNA and 50 U/ml of RNasin, incubated at the temperatures indicated for 5 min, and directly loaded onto running 15% polyacrylamide (19:1) gels containing 50 mM Tris-glycine and 5% glycerol for the com-

pletely complementary duplexes, and additionally 50 mM KCl and 5 mM $MgCl_2$ for the bulged duplexes. Duplex and monomer species were quantitated using a Molecular Dynamics PhosphorImager.

Synthesis of E1–IVS–E2 RNA substrates

RNA(1–145), RNAs(146–156), and RNA(157–234) were hybridized to a complementary deoxyoligonucleotide [cDNA(169–136): 5′-GAGAGAAGAAGCCAGTCAGCACCCTCGAGACGAG-3′ or, for the poor U2-pairing sequence, 5′-GAGAGAAGAAGCCAGTACGACACCTCGAGACGAG-3′] by heating at 65°C for 5 min followed by cooling to ambient temperature over 5 min, and were joined using T4 DNA ligase (Moore and Sharp 1992, 1993). Ligation reactions typically included 50 mM Tris-HCl (pH 7.8), 10 mM $MgCl_2$, 20 mM DTT, 1 mM ATP, 1 U/ml of RNasin (Promega Biotech), 3% polyvinylpyrrolidone (PVP)-40, 2.5 U/ml of T4 DNA ligase (U.S. Biochemical), and 1 μM each RNA and cDNA, and were incubated at 30°C for 3–5 hr. Products were purified on 15% polyacrylamide (29:1), 8 M urea gels run in 1× TBE.

In vitro splicing assays and analyses of branch site positions

High specific activity branch region RNAs(146–156) joined to nonradioactive flanking RNA(1–145) and RNA(157–234) were incubated under splicing conditions as described previously (Grabowski et al. 1984), and the products were separated and purified on 15% polyacrylamide (29:1), 8 M urea gels run in 1× TBE. The pre-mRNA (E1–IVS–E2) and lariats (IVS–E2 and IVS) were digested for 30 min at 37°C followed by 15 min at 42°C, using 0.06 U/ml of RNase T2 (Calbiochem) in 30 mM NaOAc (pH 5.2) and 0.01 mg/ml of *E. coli* tRNA; using 0.06 U/ml of nuclease P1 (BMB) in 30 mM NaOAc (pH 5.2), 0.4 mM $ZnSO_4$, and 0.01 mg/ml of *E. coli* tRNA; or using 0.5 mg/ml of RNase A (BMB) and 1 U/ml of RNase T1 (Calbiochem) in 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, and 0.01 mg/ml of *E. coli* tRNA. One-half of each nuclease-digested sample was removed and adjusted to 0.05 U/ml of calf intestinal alkaline phosphatase (BMB) and further incubated at 37°C for 30 min. Products were separated by electrophoresis through a 25% (19:1) polyacrylamide, 8 M urea gel.

Acknowledgments

We thank B. Blencowe, J. Crispino, G. Huh, A. MacMillan, P. McCaw, R. Meyers, and J. Pomerantz for their critical review of this manuscript, M. Siafaca for her patience and ever-present assistance, and the members of the Sharp laboratory for their continued interest and support. C.C.Q. and M.J.M. are supported by Damon Runyon–Walter Winchell Cancer Research Fund and American Cancer Society postdoctoral fellowships, respectively. This work was supported by U.S. Public Health grant RO1-GM34277 from the National Institutes of Health to P.A.S. and partially by a Cancer Center Support (core) grant P30-CA14051 from the National Cancer Institute.

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

References

Blackburn, P. and S. Moore. 1982. Pancreatic ribonuclease. In *The enzymes* (ed. P.D. Boyer), pp. 317–433. Academic Press, New York.

Cech, T.R. 1986. The generality of self-splicing RNA: Relationship to nuclear mRNA splicing. *Cell* **44:** 207–210.

Domdey, H., B. Apostol, R.-J. Lin, A. Newman, E. Brody, and J. Abelson. 1984. Lariat structures are in vivo intermediates in yeast pre-mRNA splicing. *Cell* **39:** 611–621.

Freyer, G.A., J. Arenas, K.K. Perkins, H.M. Furneaux, L. Pick, B. Young, R.J. Roberts, and J. Hurwitz. 1987. *In vitro* formation of a lariat structure containing a G2′-5′G linkage. *J. Biol. Chem.* **262:** 4267–4273.

Gao, Y.-G., G.A. van der Marel, J.H. van Boom, and A.H.-J. Wang. 1991. Molecular structure of a DNA decamer containing an anticancer nucleoside arabinosylcytosine: Conformational perturbation by arabinosylcytosine in B-DNA. *Biochemistry* **30:** 9922–9931.

Grabowski, P.J., R.A. Padgett, and P.A. Sharp. 1984. Messenger RNA splicing in vitro: An excised intervening sequence and a potential intermediate. *Cell* **37:** 415–427.

Guthrie, C. 1991. Messenger RNA splicing in yeast: Clues to why the spliceosome is a ribonucleoprotein. *Science* **253:** 157–163.

Hornig, H., M. Aebi, and C. Weissmann. 1986. Effect of mutations at the lariat branch acceptor site on β-globin pre-mRNA splicing in vitro. *Nature* **324:** 589–591.

Jacquier, A. 1990. Self-splicing group II and nuclear pre-mRNA introns: How similar are they? *Trends Bio. Sci.* **15:** 351–354.

Koizumi, M., Y. Hayase, S. Iwai, H. Kamiya, H. Inoue, and E. Ohtsuka. 1989. Design of RNA enzymes distinguishing a single base mutation in RNA. *Nucleic Acids Res.* **17:** 7059–7070.

Konarska, M.M., P.J. Grabowski, R.A. Padgett, and P.A. Sharp. 1985. Characterization of the branch site in lariat RNAs produced by splicing of mRNA precursors. *Nature* **313:** 552–557.

Konforti, B.B., M.J. Koziolkiewicz, and M.M. Konarska. 1993. Disruption of base pairing between the 5′ splice site and the 5′ end of U1 snRNA is required for spliceosome assembly. *Cell* **75:** 863–873.

Lin, R.-J., A.J. Newman, S.-C. Cheng, and J. Abelson. 1985. Yeast mRNA splicing in vitro. *J. Biol. Chem.* **260:** 14780–14792.

Maschhoff, K.L. and R.A. Padgett. 1993. The stereochemical course of the first step of pre-mRNA splicing. *Nucleic Acids Res.* **21:** 5456–5462.

Milligan, J.F. and O.C. Uhlenbeck. 1989. Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol.* **180:** 51–62.

Moore, M.J. and P.A. Sharp. 1992. Site–specific modification of pre–mRNA: The 2′ hydroxyl groups at the splice sites. *Science* **256:** 992–997.

———. 1993. The stereochemistry of pre–mRNA splicing: Evidence for two active sites in the spliceosome. *Nature* **365:** 364–368.

Moore, M.J., C.C. Query, and P.A. Sharp. 1993. Splicing of precursors to mRNA by the spliceosome. In *The RNA world*, (ed. R. Gesteland and J. Atkins), pp. 303–357. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Noble, J.C.S., Z.-Q. Pan, C. Prives, and J.L. Manley. 1987. Splicing of SV40 early pre-mRNA to large T and small t mRNAs utilizes different patterns of lariat branch sites. *Cell* **50:** 227–236.

Noble, J.C.S., C. Prives, and J.L. Manley. 1988. Alternative splicing of SV40 pre-mRNA is determined by branch site selection. *Genes & Dev.* **2:** 1460–1475.

Padgett, R.A., M.M. Konarska, M. Aebi, H. Hornig, C. Weissmann, and P.A. Sharp. 1985. Nonconsensus branch-site sequences in the *in vitro* splicing of transcripts of mutant rab-

bit β-globin genes. *Proc. Natl. Acad. Sci.* **82:** 8349–8353.
Paolella, G., B.S. Sproat, and A.I. Lamond. 1992. Nuclease resistant ribozymes with high catalytic activity. *EMBO J.* **11:** 1913–1919.
Parker, R., P.G. Siliciano, and C. Guthrie. 1987. Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. *Cell* **49:** 229–239.
Perreault, J.-P., T. Wu, B. Cousineau, K.K. Ogilvie, and R. Cedergren. 1990. Mixed deoxyribo- and ribo-oligonucleotides with catalytic activity. *Nature* **344:** 565–567.
Rautmann, G. and R. Breathnach. 1985. A role for branchpoints in splicing in vivo. *Nature* **315:** 430–432.
Reed, R. and T. Maniatis. 1988. The role of the mammalian branchpoint sequence in pre-mRNA splicing. *Genes & Dev.* **2:** 1268–1276.
Reilly, D.J., J.C. Wallace, R.F. Melham, D.W. Kopp, and M. Edmonds. 1989. Isolation and characterization of branched oligonucleotides from RNA. *Methods Enzymol.* **180:** 177–191.
Rodriguez, J.R., C.W. Pikielny, and M. Rosbash. 1984. In vivo characterization of yeast mRNA processing intermediates. *Cell* **39:** 603–610.
Ruskin, B., A.R. Krainer, T. Maniatis, and M.R. Green. 1984. Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. *Cell* **38:** 317–331.
Ruskin, B., J.M. Green, and M.R. Green. 1985. Cryptic branch point activation allows accurate *in vitro* splicing of human β-globin mutants. *Cell* **41:** 833–844.
Sharp, P.A. 1985. On the origins of RNA splicing and introns. *Cell* **42:** 397–400.
Takahashi, K. and S. Moore. 1982. Ribonuclease T1. In *The enzymes* (ed. P.D. Boyer), pp. 435–468. Academic Press, New York.
Uchida, T. and F. Egami. 1971. Microbial ribonucleases with special reference to RNases T1, T2, N1, and U2. In *The enzymes* (ed. P.D. Boyer), pp. 205–250. Academic Press, New York.
Uhlenbeck, O.C. 1987. A small catalytic oligoribonucleotide. *Nature* **328:** 596–600.
Wu, J. and J. Manley. 1989. Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing. *Genes & Dev.* **3:** 1553–1561.
Wyatt, J.R., E.J. Sontheimer, and J.A. Steitz. 1992. Site-specific crosslinking of mammalian U5 snRNP to the 5′ splice site prior to the first step of premessenger RNA splicing. *Genes & Dev.* **6:** 2542–2553.
Zhuang, Y. and A.M. Weiner. 1989. A compensatory base change in human U2 snRNA can suppress a branch site mutation. *Genes & Dev.* **3:** 1545–1552.
Zhuang, Y., A.M. Goldstein, and A.M. Weiner. 1989. UACUAAC is the preferred branch site for mammalian mRNA splicing. *Proc. Natl. Acad. Sci.* **86:** 2752–2756.

Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model.

C C Query, M J Moore and P A Sharp

Genes Dev. 1994, **8:**
Access the most recent version at doi:10.1101/gad.8.5.587