Small sequence insertions within the branch point region dictate alternative sites of lariat formation in a yeast intron

Daniela Castanotto and John J. Rossi*

Department of Molecular Genetics, Beckman Research Institute of the City of Hope,
1450 East Duarte Road, Duarte, CA 91010, USA

Received August 21, 1992; Revised and Accepted November 11, 1992

ABSTRACT

The problem of intron recognition in *S.cerevisiae* appears to be in part solved by the strong conservation of intron encoded splicing signals, in particular the 5' GUAUGU and the branch point UACUAAC which interact via base pairing with the RNA components of U1 and U2 snRNPs respectively. Nevertheless, the mere presence of such signals is insufficient for splicing to occur. In the *S.cerevisiae* ACT1 intron, a silent UACUAAC-like sequence (UACUAAG) is located 7 nucleotides upstream of the canonical branch point signal. In order to investigate whether other factors, in addition to the U2-UACUAAC base-pair interactions, affect branch point selection in yeast, we created a cis-competition assay by converting the UACUAAG to a strong branch point signal (UACUAAC). If simply having a canonical UACUAAC sequence were sufficient for lariat formation, a 1:1 ratio in usage of the two signals should have been observed. In this double branch point intron, however, the downstream UACUAAC is utilized preferentially (4:1). Results obtained from the analyses of numerous sequence variants flanking the two UACUAAC sequences, demonstrate that non-conserved sequences in the branch point region are able to define lariat formation. Consequently, we conclude that U2 base-pairing is not the only requirement determining branch point selection in yeast, and local structure in the vicinity of the branch point could play a critical role in its recognition.

INTRODUCTION

From yeast to man, a fundamental step in gene expression is pre-messenger RNA splicing. This process involves a two step chemical reaction (1,2,3,4). In the first step, the 5' exon/intron phosphodiester bond is cleaved with the concomitant formation of a 5'-2' phosphodiester linkage between the first intron nucleotide and a branch point nucleotide (usually an adenosine) located upstream of the 3' splice site. In the second step, exon 1 is joined to exon 2 and the intron is released in the form of a lariat. The complete reaction is mediated by a multicomponent

RNA-protein complex termed the spliceosome (5,6) which contains five small nuclear RNAs (U1, U2, U4, U5 and U6) (7) and approximately 50 proteins (8). Specific spliceosomal components recognize key sequences within the intron, causing a series of conformational changes which ultimately fold the RNA into a substrate that can be spliced.

The mechanism by which splice sites are recognized is still poorly understood. While nuclear pre-mRNA splicing in yeast and metazoa seems to share the same molecular mechanisms, the contributions of the various intron encoded signals differ considerably. Introns in higher eukaryotes display only limited sequence requirements (9,10). In contrast, yeast intron consensus sequences are highly conserved, and mutations in these sequences severely inhibit the efficiency of splicing, sometimes completely abolishing the process (11,12). Nevertheless, it has been possible to obtain accurate splicing of a yeast intron in HeLa cell extracts; however, the position of the branch point varied from the canonical UACUAAC sequence, favoring a location closer to the 3' AG over the branch point sequence specificity (13). Determining the subtle differences of the splicing machinery that differentiate yeast from higher eukaryotes is essential in the attempt to understand the evolution of pre-mRNA splicing. Because the RNA-U2 snRNP complex is remarkably stable, it has been suggested that the formation of this complex, during the splicing reaction, commits that region of the RNA to the role of intron (14). RNase H oligonucleotide-directed nuclease sensitivity assays have shown that splicing factors associate with the branch point region at a very early stage of spliceosome assembly, and the UACUAAC sequence is necessary for protection at the 5' splice site (15,16). These, along with other results (17,18,20) indicate that recognition of the branch point occurs at least once, before the U2-pre-mRNA interaction. The factors mediating this recognition are still unknown.

In the present study we investigated whether other factors, in addition to the U2-UACUAAC base-pair interactions, affect branch point selection in yeast. To accomplish this, we have devised a cis-competition assay which allows quantitation of the relative usage of two competing branch points. Although both signals are the canonical UACUAAC, the downstream or endogenous signal is surprisingly favored by a ratio of 4:1 over

* To whom correspondence should be addressed

the upstream signal. To investigate the cause of this bias, a series of sequence changes both downstream and in-between the branch points were made. The results obtained add several new dimensions to our understanding of branch selection in yeast. These include 1) a dramatic influence of local sequence context, and perhaps structure, on branch point selection, 2) strong evidence against a unidirectional molecular scanning mechanism, and 3) lack of involvement of the conserved polyuridine tract in the first step of splicing.

MATERIALS AND METHODS

Materials and strains

Restriction endonuclease, T4 DNA ligase and T4 kinase were purchased from Bethesda Research Laboratories. Taq DNA polymerase was purchased from Perkin Elmer Cetus. ³²P-labeled nucleotides were purchased from NEN. Avian myeloblastosis virus (AMV) reverse transcriptase was obtained from Life Sciences. The oligonucleotides used for mutagenesis and for primer extension were either purchased from GENOSYS or were synthesized by the City of Hope DNA/peptide synthesis facility.

E. coli MC1061 (20) was used for routine manipulations. *E. coli* strain CJ236 (*dut*⁻, *ung*⁻, *thi*⁻, *relA*⁻) (Bio Rad) was used for mutagenesis. *S. cerevisiae* strain JM43 (*mat* α, *leu2*, *ura3*, *trp1*, *his3*) was kindly provided by J. McEwen. *E. coli* strains were transformed by the method of Kushner (21), and *S. cerevisiae* by the lithium acetate procedure (22).

DNA manipulations

A 510-bp *AluI-HindIII* fragment containing the *S. cerevisiae* *ACT1*-HIS4 fusion intron and regions of the flanking exons was excised from plasmid pYAHB-1, described previously (23), and cloned into the *SmaI-HindIII* site of pTZ18U (Bio Rad). The *Nru* site created during the construction of pYAHB-1 was converted back to the original *ClaI* site, resulting in the plasmid pTZ18-ACT. This construct was mutagenized to create all the sequence variants. Site-specific mutagenesis was performed as described by Kunkel (24) (Bio-Rad Mutagenesis Kit) or by using mutagenic oligonucleotides in a Polymerase Chain Reaction (PCR) (25). Mutations in plasmids pGCm2, pPolyUD2, and pPolyR2 were constructed by using the Bio Rad mutagenesis kit to generate mismatched heteroduplex molecules. The mutations contained in plasmids pD2, pSMBR1, pSCBR2, pPolyUBlock, pDRand, pPolyG, and pPolyC were introduced using pTZ18-ACT as a template for PCR mutagenesis. The mutagenic oligonucleotides all contain a *ClaI* restriction site at their 5' ends and incorporate various insertions between the two UACUAAC sequences. An oligonucleotide complementary to the T7 promoter sequence and a mutagenic oligonucleotide were used as primers in a PCR reaction to create a double stranded DNA fragment containing exon I and the entire intron upstream of the natural *ClaI* restriction site located just 3' of the UACUAAC sequence. These PCR products were individually digested with *EcoRI* and *ClaI* and ligated upstream of the *ACT1* 3' splice site into *EcoRI*-*ClaI*-cut pTZ18-ACT. pPolyUBr1 and pPolyA/PolyA were constructed by replacing the *ClaI-HindIII* fragment of pSMBR1 and pSCBR2 respectively, with the analogous fragment from pPolyUD2 (the *HindIII* site is located 80 nucleotides downstream of the *ACT1*-HIS4 fusion junction). Mutations were confirmed by dideoxy sequencing (26) of plasmid templates using the U.S.

Biochemical sequenase kit. The *EcoRI-HindIII* fragments from the resultant constructs were ligated into *EcoRI* and *HindIII* sites of Bluescript KS+, digested with *BamHI-SalI* and cloned into the same sites of pBM258 (27), which harbors the inducible GAL1-10 promoter used for inducible expression in *S. cerevisiae*.

RNA induction and analysis

Yeast transformants were grown in glucose minimal media (28) to an OD₆₀₀ of 0.8. The cells were washed once in T₁₀E₁ (10mM Tris-HCl pH 8.0, 1mM EDTA), and *ACT1*-HIS4 fusion RNA was induced by growing for 3 hours in 2% galactose minimal media. Preparation of RNA and primer extension analyses were performed according to Domdey (1). The oligonucleotides used in primer extension analyses are the 19mer ACT1 primer (5'-GGCTTTACACATACCAGAA-3'), which hybridizes 29 nucleotides downstream of the 3' splice site, and the 22mer 303-305 primer (5'-GCAGCAACGTGTAAACA-TATAA-3'), which is complementary to either side of the 3' splice site. The products of the primer extension reaction were electrophoresed in a 6% polyacrylamide 8 M urea sequencing gel. The radioactivity in the gel was quantitated with an AMBIS radioanalytic imaging system. The primer extensions and subsequent quantitations were repeated several times, demonstrating a variation of 5% or less in the observed ratios.

RESULTS

In the *S. cerevisiae* *ACT1* intron, a UACUAAG sequence is located 7 nucleotides upstream of the canonical branch point signal (UACUAAC). This defective sequence is never used as a site of lariat formation in the presence of the regular branch point (23,11), but can be activated by mutation (12) or deletion (23) of the canonical UACUAAC sequence. In order to identify intron sequences that might influence branch point utilization, we created a *cis*-competition assay by converting the UACUAAG

Figure 1. Construction of the double branch point intron. The plasmid pBTz258 containing the wild type *S. cerevisiae* *ACT1*-HIS4 gene fusion is shown. Distance from the native and cryptic branch points to the 3' splice site is given in nucleotides. Above the wild type sequence, are shown the nucleotide changes required to create the double branch point intron of pGCm2. pGCm2 contains three mutations; one converts the cryptic UACUAAG to UACUAAC and the others change the AG/AG of the 3' splice site to AC/AC. The mutated nucleotides are underlined.

sequence to UACUAAC and the 3' splice site AG to AC (plasmid pGCM2, Fig. 1). The AG to AC mutation does not alter the first step of splicing, but blocks the second step of the reaction resulting in the accumulation of intron lariats still covalently attached to exon II (29,12). This system allows a precise and quantitative determination of *in vivo* branch point utilization via primer extension analysis. Results obtained using total RNA isolated from *S.cerevisiae* transformants harboring pGCM2, indicate that the downstream UACUAAC is favored over the upstream site by a ratio of 4:1 (Fig. 2, Lane A). In our studies

the branch points have been mapped with two different primers, and each analysis has been repeated several times. Branches at both of these positions are also equally sensitive to debranching enzyme (data not shown). None of the intron variants used in this study resulted in detectable levels of precursor accumulation, indicating that these changes do not kinetically alter step 1 of splicing.

U2 Base pairing interactions cannot explain the preferential usage of the downstream branch point

Analyses of sequences surrounding both branch points suggested that the downstream UACUAAC might be favored because of the additional presence of a highly conserved adenine located two nucleotides downstream of the branch position (see Fig. 3A). This nucleotide is present in 70% of yeast introns (based on our recent survey of the Genbank database), and is believed to base pair with U2 (19,30). To determine whether or not the bias for the downstream branch point was a consequence of this potential extra base pair with U2, we converted a U to an A (plasmid pD2) at the analogous position in the upstream UACUAAC signal. When RNA from pD2 transformants was analyzed by primer extension, the same 4:1 ratio (downstream to upstream) in branch point utilization was observed. This result (Fig. 3B, lane A) shows that, if the conserved adenine forms a base pair with U2, this pairing does not exert a detectable influence on branch point selection, and suggests that factors other than U2 base-pairing play a role in branch selection.

Insertion of a poly U tract activates the upstream UACUAAC

A second hypothesis, which could explain the unequal branch point ratio seen with pGCM2, is the proximity of the downstream UACUAAC to the highly conserved polyuridine tract (Fig. 1). To examine the possibility that the polyuridine tract was influencing the first step of the splicing reaction, we introduced a replica of the downstream polyuridine tract between the two UACUAAC sequences (plasmid pSMBr1, Fig. 4A). As a control

Figure 2. Duplicated UACUAAC sites do not compete with equal efficiencies *in vivo* for 5'-2' lariat formation. An oligonucleotide complementary to exon II of the *S.cerevisiae* *ACT1* gene (19mer *ACT1*) was used in primer extension analysis of total RNA prepared from *S.cerevisiae* transformed with pGCM2 (lane A) and untransformed cells (lane B). The primer extension products are shown schematically beside the autoradiograph. The square and circle indicate a branch point to the upstream and downstream branch points respectively. The larger product seen in both lanes represents endogenous *ACT1* mRNA. Molecular weight markers in the left lane are 32 P-labeled *Hpa* II digested pBR322 restriction fragments.

Figure 3. Potential U2 base pairing with the highly conserved adenine following the second UACUAAC does not influence branch point utilization. A) A diagram is shown signifying how base pairing with U2 could facilitate usage of the downstream branch point in pGCM2 over the upstream one. The highly conserved adenine following the downstream UACUAAC can potentially base pair to U2 generating 7 base pairs at this branch point sequence, in contrast to the 6 base pairs in the upstream UACUAAC. B) The sequence of pD2, shown below the autoradiograph, contains a point mutation which converts the U immediately following the upstream UACUAAC to an A. Primer extension analysis was performed as described in Fig. 2.

Figure 4. The insertion of a poly U tract between the two UACUAAC sequences dramatically activates usage of the upstream branch point. **A)** The site of the insertions is shown by an arrow above the sequence of pGCm2, and the native poly U tract in the *S. cerevisiae* *ACT1* intron is underlined. A solid box below the sequence on either side of the insertion site shows nucleotides identical to those flanking the native poly U tract; these nucleotides are also signified with a solid box. Therefore, inserting the indicated seven nucleotides results in the creation of 12 nucleotides identical to the native poly U tract between both UACUAAC sites. Below are shown the poly U and poly A insertions used to create plasmids pSMBR1 and pSCBR2, respectively. The C residue found in the native poly U tract was included in these and all remaining nucleotide alterations for consistency. **B)** The 19mer *ACT1* primer was used in primer extension analysis of total RNA prepared from *S. cerevisiae* transformed with pSMBR1 (lane A), pSCBR2 (lane B), and untransformed cells (lane C). The solid box and circle represent lariats formed to the upstream and downstream branch points, respectively.

for spacing, a second plasmid, pSCBr2, was constructed in which a poly A tract was inserted at the analogous position (Fig. 4A). The results, presented in Fig. 4B (lanes A and B), demonstrate that the inserted polyuridine tract reverses the preference of lariat formation in favor of the upstream UACUAAC. Yeast transformants harboring plasmid pSMBR1 favor the upstream branch point (4:1), whereas the control transformants, harboring the polyadenine tract, still show a preferential utilization of the downstream site (3:2), although the bias is not as pronounced as in the parental construct (Fig. 2, lane A). These results suggested that, in the parental plasmid pGCm2, the downstream branch point was favored by its closer proximity to the naturally occurring polyuridine tract. Replacement of this tract should therefore result in an equal utilization of both UACUAAC sequences. We substituted the natural downstream poly U tract in pSMBR1, pD2 and pSCBr2 with adenine residues, creating plasmids pPolyUBr1, pPolyUD2, pPolyA/PolyA respectively (Fig. 5A) and carried out primer extension analyses. The results (Fig. 5B, lanes A, B, C) illustrate that this sequence change, in three different combinations, caused at most a 10% shift in favor of the upstream branch point (compare lane B of Fig. 5B to lane A of Fig. 2), demonstrating that the native yeast polyuridine tract does not determine the preferential choice of the downstream branch point.

Sequence composition within the branch point region dramatically affects branch point selection

Although altering the native polyuridine tract did not have a strong effect on branch utilization, insertion of a poly U tract between the two branch points completely reversed the ratio of their usage (pSMBR1 Fig. 4B, lane A). Several molecular characteristics of the pSMBR1 intron could account for this result. The dramatic activation of the upstream branch point could be due, for example, to factor(s) not necessarily involved in branch point selection, which bind the polyuridine tract and block lariat formation to the downstream branch. Alternatively, this new sequence could, by itself, 'mask' the site or interfere with U2/branch point base

pairing. To test these possibilities, an additional poly U tract was inserted in plasmid pSMBR1 just upstream (2 nts) of the first UACUAAC (pPolyUBlock, Fig. 5A). Primer extension analysis revealed branching still preferentially occurred at the upstream UACUAAC (Fig. 5B, lane E). This result demonstrates that the poly U tract of pSMBR1 does not activate the upstream branch point by merely blocking usage of the downstream one. An alternative explanation for the shift in ratios is that factors bound to the poly U tract directly enhance utilization of the upstream branch point. The insertion of a polyadenine tract between the two branch points in plasmid pSCBr2 (Fig. 4B, lane B) resulted in a small (20%) but reproducible shift in favor of the upstream UACUAAC. This polyadenine tract could itself be bound by a factor(s) resulting in a less pronounced shift in branch point utilization. To examine this possibility, a randomly generated sequence (CUCGUAU) was inserted between the two branch points (plasmid pDRan, Fig. 6A). This new sequence should not be bound by putative poly A or poly U recognition factor(s). Surprisingly, this construct showed an effect identical to that obtained with the poly U insertion at the same position (Fig. 6B, lane B). Substitution of the downstream, native poly U tract with the random sequence did not alter the efficiency of the downstream branch point selection (plasmid pPolyUR2, Fig. 6B, lane A). Because the randomly-chosen sequence insertion is pyrimidine rich (5 of the 7 bases are pyrimidines), we searched for a correlation between sequence composition and properties of pyrimidine versus purine tracts. We inserted a poly G or a poly C sequence between the two UACUAAC signals (Fig. 6A). The poly C insertion reproduced the same effect as poly U (compare Fig. 6B, lane B with Fig. 4B, lane A), while the poly G resulted in a less pronounced shift in favor of branching to the upstream signal (compare Fig. 6B, lane C with Fig. 4B, lane A). These results suggest that the shift seen in branch point utilization is not caused by sequence specific binding factor(s), but is due to the nucleotide composition within the branch point region; the base stacking potential of neighboring nucleotides could therefore affect branch point selection. In addition to the

Fig. 5. Substituting adenines for uridines in the native poly U tract has little or no effect on branch point selection. A) The native poly U tract shown at the top of the Fig. was substituted with adenines in three different backgrounds, as shown in A, B, and C. Plasmid names are underlined, followed by a letter which gives their respective lanes on the autoradiograph in panel B. pSCBr2 is repeated as a representative control because it is the parent plasmid of pPolyAPolyA. The parent plasmids of pPolyUBr1 and pPolyUD2 are pSMBR1 and pGCM2 shown in Fig.s 4 (lane A) and 2 (lane A) respectively. The last construct, pPolyUBlock, has a poly U tract, identical to the native one, inserted 2 nts. upstream of the first branch point. B) The results of primer extension analysis of total RNA prepared from *S. cerevisiae* transformants harboring the corresponding plasmids from panel A are shown. The primer used was an oligonucleotide complementary to the *ACT1* intron sequence (303–305 primer). The square and circle indicate lariat formation to the first and second UACUAAC sites, respectively.

expected products, the Poly C insertion resulted in an additional RT stop between the two branch points. This stop may represent a new site of lariat formation, although this needs to be further characterized. The percentages of branch point utilization for the various constructs are summarized in table 1.

DISCUSSION

We describe a system which allows the study of sequence context effects on the first chemical step of pre-mRNA splicing. Analysis of the parental construct, pGCM2, revealed that the downstream of two canonical branch point sequences is favored by a 4:1 ratio in branch point formation. The extra base-pairing potential of the downstream UACUAACA to the U2 snRNP does not determine this preference, suggesting that a critical step in branch point selection is occurring before the binding of U2. The bias between the two branch points is also not due to the presence of the naturally occurring polyuridine tract since replacement with a random or polyadenine tract has no significant effect on the ratio of branch utilization. In addition, the insertion of a 7 nt random pyrimidine rich, poly U or a poly C tract between the two UACUAAC sequences, all completely reversed branch point usage in favor of the upstream site, making it unlikely that Poly U is bound by sequence specific factor(s) involved in branch point selection. These results, along with those obtained with poly A and poly G insertions, show that the shifts seen in branch point utilization are determined by the nucleotide composition within the branch point region, conceivably even prior to U2 recognition, and dismiss the possibility of a unidirectional molecular scanning mechanism for lariat formation in yeast.

Role of the polyuridine tract in yeast introns

Recently, Patterson and Guthrie (31) showed that a polyuridine tract in yeast enhanced the usage of a 3' splice site (1991). However the mechanism by which the poly U tract increased usage of nearby AG dinucleotides as a 3' splice site, is not known.

The authors speculate that the polyuridine tract in yeast, as well as the polypyrimidine tract in mammals, is recognized by factor(s) influencing 3' splice site selection. The results of the present studies, however, demonstrate that, at least in the first step of splicing, the yeast polyuridine tracts serve a different role than that of the polypyrimidine tracts in mammalian introns (32,33), which are bound by factors essential for lariat formation (34,35,36). It is possible that because mammalian branch point sequences are highly variable, additional factors binding to the polypyrimidine tract are needed to ensure efficient binding of U2 snRNP to the branch site. It is noteworthy that, even in mammals, more UACUAAC-like sequences are preferred for branch point selection (37). The ideal base-pairing of U2 to the UACUAAC sequence in yeast may, therefore, provide an interaction with the branch point that is sufficiently stable to dismiss any requirement for additional factors. As a test of this hypothesis, we have created a condition more similar to that of metazoa by converting both UACUAAC sequences in our *ACT1* intron to weak branch point signals. However, even in this construct, the native polyuridine tract plays an insignificant role in branch selection (data not shown), indicating that the poly U tract does not bind a factor(s) which stabilizes U2 interaction with the branch point sequence.

We searched the Genbank database for all known yeast introns and found that, in general, polyuridine tracts are absent in introns where the 3' splice site is positioned close to the branch point. In contrast, in those introns where the branch point and the 3' splice site are well separated, the size and/or the number of the polyuridine tracts increases with increasing distance between the branch point and the 3' splice site. It is possible, thus, that the role of the yeast polyuridine tracts is to help to determine and/or maintain an order of structure necessary to produce critical and highly constrained interactions between snRNPs and specific sites within the intron. This would make them more similar to the AU-rich sequences of plant introns (38) than the polypyrimidine tracts of mammalian introns. In an evolutionary context, higher

Figure 6. A randomly chosen sequence insertion has the same effect as poly U insertions. In panel A are shown the sequences of pDPolyR (A) which has a randomly chosen sequence replacing the downstream native poly U tract, pDRan (B) which contains the same sequence between the two branch points, pDPolyG (C) and pDPolyC (D) which have respectively a poly G and a poly C tract inserted between the two branch points. In pDPolyG, an A was substituted with a T residue to avoid the creation of a new potential 3' splice site. Panel B shows primer extension analysis of total RNA prepared from *S. cerevisiae* transformants harboring the plasmids shown in panel A. The letters above each lane correspond to the plasmids indicated in panel A. The oligonucleotide used was the 303–305 primer.

Table I. Summary of branch point utilization and mutant plasmids used in this study

Plasmid Name	Relevant Mutations	Branch Point Utilization	
		BP1	BP2
pGCm2	Bp1.....Bp2....Poly U.....AC	20%	80%
pD2	Bp1A.....Bp2....Poly U.....AC	20%	80%
pSMBr1	Bp1.Poly U.Bp2....Poly U.....AC	80%	20%
pSCBr2	Bp1.Poly A.Bp2....Poly U.....AC	40%	60%
pPolyUD2	Bp1.....Bp2....Poly A.....AC	30%	70%
pPolyUBr1	Bp1.Poly U.Bp2....Poly A.....AC	80%	20%
pPolyAPolyA	Bp1.Poly A.Bp2....Poly A.....AC	45%	55%
pDPolyR	Bp1.....Bp2....Random.....AC	20%	80%
pPolyUBlock	Poly U.Bp1.Poly U.Bp2....Poly U.....AC	75%	25%
pDPolyC	Bp1.Poly C.Bp2....Poly U.....AC	90%	10%
pDPolyG	Bp1.Poly G.Bp2....Poly U.....AC	60%	40%
pDRand	Bp1.Random.Bp2....Poly U.....AC	80%	20%

The relevant mutations for the various plasmids are indicated with underlined bold print. Bp1 and Bp2 refer to the upstream and downstream branch points, respectively. The percentage of branch point utilization was determined as described in Materials and Methods.

eukaryotes could have developed a more complex splicing system, involving a larger number of interactions with sequences in the intron, in order to maintain splice site specificity with less sequence conservation. This increase in flexibility could induce the development of alternative splicing pathways, a fundamental mechanism for acquiring additional genetic diversity in metazoa.

Alternative branch point selection is determined by sequence context within the branch point region

The native polyuridine tract, if replaced, causes only a minor effect on branch formation. Thus, it cannot account for the 4:1 ratio observed with our original construct (pGCm2). Computer analyses of all mutant sequences revealed no link between the ratios of lariat formation and any potential for secondary structure in the branch point region. In our experiments, the poly C and the random insertions reproduced the same effect as poly U. The poly G insertion also resulted in favored usage of the upstream branch point, however this was less pronounced than the pyrimidine rich insertions. Because the various fragments inserted

are all of the same length, but activate the upstream branch point to different levels, we conclude that the choice of the branch is not simply dictated by a spatial constraint between the two UACUAAC sequences. The increase of spacing between the two branch point, indeed, may only accounts for 20% activation of the upstream branch point, since this is the minimum effect detected with a 7 nts insertion. In addition, some of the insertions display identical effects on branch selection, indicating that activation of the upstream UACUAAC is also not a simple sequence specific phenomenon. A combination of the above and other factors, such as base stacking of neighboring nucleotides, and the relative positions of the two branch points with respect to other critical spliceosomal factor binding sites, could all be determining parameters in the process of branch point selection. If this is the case, it would be extremely interesting to determine the contribution of these factors versus the U2 base-pairing, on the choice of the branch.

The branch point sequence is involved in the early events of splicing, and its recognition seems to take place prior to U2

association (17,18,19). Our results demonstrate that the mere presence of this sequence is insufficient to specify the sites of lariat formation. Additionally, our data dismiss the possibility of a simple unidirectional molecular scanning mechanism in branch selection. It is possible that, a specific interaction between 5' splice site and branch point recognition factors, defines the first requirement that initiates the cascade of events leading to spliceosome assembly. The insertions between the two branch points analyzed in this work could have altered the intron in a way that one of the two competing branch sites is alternatively positioned in a more favorable location for allowing this initial interaction. Distance, position, and inclination of the splice sites with respect to each other and to spliceosomal interaction sites, may be therefore the critical determinants in initiating intron recognition. Any alteration in these parameters could result in selection of new splice sites or even abort the splicing process. In this light it is interesting to note that synthetic introns containing all known consensus sequences are spliced at very low efficiency (39,40). In addition, many potentially 'good' splice sites are distributed randomly throughout genes, but are never selected by the splicing machinery. A recently described gene of *S.cerevisiae* (RAT1) encodes all of the canonical splice signals within the protein coding region, yet the RNA transcribed from this gene does not appear to be spliced (41).

It is widely believed that nuclear pre-mRNA splicing evolved from the self splicing mechanism of type II introns. The recent discovery of a small RNA that may function in trans to induce the formation of a group II-type structure required for splicing to occur (42), supports this theory. Yeast and mammalian pre-mRNA introns, however, seem to have lost the strict requirements for phylogenetically conserved structure present in type II introns. The assay described here demonstrates that local structural features could still be critical for intron recognition, perhaps a remnant of a type II progenitor.

ACKNOWLEDGMENTS

We would like to thank Michael Green for kindly providing the debranching enzyme. We are also grateful to Jim Deshler for helpful discussions and comments during this project. We thank Pete Gertson and Dave Elkins for critical reading of the manuscript. This work was supported by NIH grant AI29329 to JJR.

REFERENCES

- Domdey, H., Apostol, B., Lin, R.J., Newman, A., Brody, E. and Abelson, J. (1984) *Cell*, **39**, 611–621.
- Padgett, R.A., Konarska, M.M., Grabowski, P.J., Hardy, S.F. and Sharp, P.A. (1984) *Science*, **22**, 898–903.
- Ruskin, B., Krainer, A.R., Maniatis, T. and Green, M.R. (1984) *Cell*, **38**, 317–331.
- Zeitlin, S., and Efstratiadis, A. (1984) *Cell*, **39**, 589–602.
- Brody, E. and Abelson, J. (1985) *Science*, **228**, 963–967.
- Grabowski, P.J., Seiler, S.R. and Sharp, P.A. (1985) *Cell*, **42**, 345–353.
- Guthrie, C. and Patterson, B. (1988) *Annu. Rev. Genet.*, **22**, 387–419.
- Reed, R. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 8031–8035.
- Mount, S.M. (1982) *Nucl. Acids Res.*, **10**, 459–472.
- Shapiro, M.B., Senapathy, P. (1987) *Nucl. Acids Res.*, **15**, 7155–7174.
- Fouser, L.A., and Friesen, J.D. (1986) *Cell*, **45**, 81–93.
- Vijayraghavan, U., Parker, R., Tamm, J., Limura, Y., Rossi, J.J., Abelson, J. and Guthrie, C. (1986) *EMBO J.*, **5**, 1683–1695.
- Ruskin, B., Pikielny, C.W., Rosbash, M. and Green, M.R. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 2022–2026.
- Sharp, P.A. (1987) *Science*, **235**, 766–771.
- Rymond, B.C., and Rosbash, M. (1986) *EMBO J.*, **5**, 3517–3523.
- Pikielny, C.W., Teem, J.L. and Rosbash, M. (1983) *Cell*, **34**, 395–403.
- Ruby, S., and Abelson, J. (1988) *Science*, **242**, 1028–1035.
- Seraphin, B., and Rosbash, M. (1991) *EMBO J.*, **10**, 1209–1216.
- Guthrie, C. (1991) *Science*, **253**, 157–163.
- Casabadian, M.J. and Cohen, S. (1980) *J. Bacteriol.*, **143**, 971–908.
- Kushner, S.R. (1978) In J.W. Boyer and S. Nicosia (eds), *Proceedings of the 1 International Symposium on Genetic Engineering*. Elsevier/North Holland Biomedical Press. Amsterdam, pp. 17–23.
- Ito, H., Fukuda, Y., Murata, K. and Kimura, A. (1983) *J. Bacteriol.*, **153**, 163–168.
- Cellini, A., Parker, R., McMahon, J., Guthrie, C. and Rossi, J.J. (1986) *Mol. Cell. Biol.*, **6**, 1571–1578.
- Kunkel, T.A., Roberts, J.D. and Zakour, R.A. (1987) *Methods In Enzym.*, **154**, 367–382.
- Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) *Science*, **239**, 487–491.
- Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
- Johnston, M., and Davis, R. (1984) *Mol. Cell. Biol.*, **4**, 1440–1448.
- Sherman, F., Fink, G.R. and Hicks, J.B. (1986) *Methods in Yeast Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- Rymond, B.C., and Rosbash, M. (1985) *Nature*, **317**, 735–737.
- Guthrie, C., Riedel, N., Parker, R., Swerdlow, H. and Patterson, B. (1986) In *Yeast Cell Biology, UCLA Symposia on Molecular Biology, New Series*. Alan R. Liss, Inc., New York, Volume 33, pp 301–321.
- Patterson, B. and Guthrie, C. (1991) *Cell*, **64**, 181–187.
- Reed, R. (1989) *Genes & Dev.*, **3**, 2113–2123.
- Mullen, M.P., Smith, C.W.J., Patton, J.G., and Nadal-Ginard, B. (1991) *Genes & Dev.*, **5**, 642–655.
- Ruskin, B., Zamore, P.D., and Green, M.R. (1988) *Cell*, **52**, 207–219.
- Zamore, P.D. and Green, M.R. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 9243–9247.
- Zamore, P.D. and Green, M.R. (1991) *EMBO J.*, **10**, 207–214.
- Parker, R., Siliciano, P.G. and Guthrie, C. (1987) *Cell*, **49**, 229–239.
- Goodall, G.J., and Filipowicz, W. (1991) *EMBO J.*, **10**, 2635–2644.
- Nelson, K.K., and Green, M.R. (1988) *Genes & Dev.*, **2**, 319–329.
- Legrain, P., and Rosbash, M. (1989) *Cell*, **57**, 573–583.
- Amber, D.C., Goldstein, A.L., and Cole, C.N. (1992) *Genes & Dev.*, **6**, 1173–1189.
- Goldschmidt-Clermont, M., Choquet, Y., Girard-Bascou, J., Michel, F., Schirmer-Rahire, M. and Rochemaix, J.D. (1991) *Cell*, **65**, 135–143.