Insertion of non-intron sequence into maize introns interferes with splicing

Kenneth R. Luehrsen and Virginia Walbot

Department of Biological Sciences, Stanford University, Stanford, CA 94305, USA

Received April 16, 1992; Revised and Accepted August 20, 1992

ABSTRACT

Transposable element (TE) insertion into or near plant introns can cause intron skipping and alternative splicing events, resulting in reduced expression. To explore the impact of inserted sequences on splicing, we added non-intron sequence to two maize introns and tested these chimeric introns in a maize transient expression assay. Non-intron sequence inserted into *Adh1-S* intron 1 and actin intron 3 decreased expression from the luciferase reporter gene; the insertion sites tested were not in intron regions thought to be essential for splicing. Alternatively spliced mRNAs were not observed in transcripts derived from the insertion variants. In contrast, addition of an internal segment of an intron to *Adh1-S* intron 1 resulted in normal splice site selection and efficient processing. Because the normal intron sequence (including the conserved splice junctions) was retained in all constructs, we hypothesize that added non-intron sequence can interfere with intron recognition and/or splicing.

INTRODUCTION

Pre-mRNA introns are segments of non-coding RNA that are removed, or spliced, from a transcript prior to its translation. Splicing is a nuclear process and is mediated by spliceosomes, multisubunit particles composed of snRNPs, which are themselves composed of both proteins and RNA (1, 2). In yeast and mammalian cells, splicing is thought to initiate by base pairing of the 5' splice site to U1 snRNA, followed by binding of the U2 snRNP to the branch point (bp), 20–50 bases upstream of the 3' splice site. Two successive transesterifications result in 5' splice site cleavage, a 2'5' bond being made at the branch point, 3' splice site cleavage and ligation of the exons. Mutational analysis of yeast and animal introns with *in vivo* and *in vitro* splicing assays suggests that there are only four requirements for an intron: 1) the conserved 5' splice junction, 2) the conserved 3' splice junction, 3) the branch point sequence (and in mammals, a proximal polypyrimidine tract) and 4) a minimum length of ~70–80 nt (3).

The mechanism of splicing in plants is not as well understood. Plants contain snRNAs homologous to their yeast and animal counterparts and have the well-conserved GU and AG sequences at the 5' and 3' splice junctions, respectively. Plant introns have several distinctive properties when compared to introns of animals

and yeast. Attempts to identify a polypyrimidine tract or a branch point sequence have been unsuccessful (4). Although a few plant introns are several kb in length, the average is 249 bp, which is much shorter than introns of vertebrates that average 1127 bp but range from ~75 bp to >100 kb (5). Plant introns, with an average ~40% GC content, are distinctly AU-rich compared to the surrounding exon sequences which average ~60% GC (6). Maize introns are also AU-rich (59%), but show a bias for U (34%) (based on a survey of 32222 and 26394 bases of maize exon and intron sequence, respectively; K. Luehrsen, unpublished data).

As an *in vitro* splicing extract is not yet available for plants, most splicing studies have been done in protoplasts or transgenic plants. Goodall and Filipowicz (7), using artificial introns, have shown that dicot introns must be AU-rich for efficient splicing. Monocots do not share this absolute requirement as several maize introns are GC-rich; however, AU-richness can be functionally important for some introns (8). It has been shown that a minimum intron length of 65–70 nt is required for efficient processing in dicots (9). An alternative approach to understanding plant splicing has been to determine the similarities between plant and animal introns. Several studies (10, 11, 12) have led to the general conclusion that animal introns are not faithfully spliced in plants but that plant introns are generally correctly spliced in animal cells. These observations suggest that plant introns contain additional information apart from that at the splice junctions that is important for successful splicing (12).

Plant genes usually contain multiple introns, and each of these must be appropriately spliced for normal gene expression. The insertion of transposable elements (TE) can add RNA processing signals to a transcript, and the use of normal or cryptic transposon splicing signals is the defect in many mutant alleles of maize (13, 14). Surprisingly, many of the TE insertions in introns are not close to the conserved intron elements but nevertheless result in an observable phenotype. The autonomous maize TEs such as *Ac* and *Spm/En*, generally encode a transposase or accessory protein(s) and thus contain RNA processing signals for both splicing and polyadenylation. The non-autonomous elements *Ds*, *dSpm* and *Mu1* also contain RNA processing signals, even though these elements are not independently transcribed. Upon TE insertion, additional RNA processing signals present in the TE are added to the gene and can promote polyadenylation or activate alternative 5' or 3' splice sites or cryptic 3' splice sites in surrounding exon sequences (13). For example, a *Ds2* insertion in exon 13 of the *Waxy* gene causes alternative splicing between any of three *Ds2*-encoded cryptic 5' splice site (representing all

three reading frames), to a cryptic 3' splice site immediately following the *Ds2* element (13, 15). Similarly, a *dSpm* insertion downstream of the single intron of the *Bronze-1* gene causes skipping of the normal *Bz1* intron 3' splice site and utilization of cryptic 3'ss within the *dSpm* element (16). It has been hypothesized that some TEs have properties akin to a mobile intron (17) as they often create novel splicing events depending on where they insert. In some cases, TE insertions within or very near introns allow leaky gene expression because the TEs are spliced from the mRNA (18, 19).

In a previous study, we showed that the RNA processing phenotypes created by a *Mul* insertion in *Adh1-S* intron 1 could be reconstructed using plasmid facsimiles in a transient gene expression assay (14). Do TEs generally cause splicing failure when inserted in maize introns? In this study, we have constructed chimeric expression plasmids in which an intron was placed in a transcription unit containing the firefly luciferase gene as a reporter. In order to define the requirements for maize splicing and to assess the generality of impact of insertions on splicing, we generated intron variants containing foreign DNA and tested them in a transient assay using maize protoplasts. Luciferase expression and transcript structure were analyzed. Adding to intron length by the insertion of maize intron DNA was stimulatory while insertion of non-intron DNA decreased luciferase expression. We observed a polarity effect, with insertions near the splice junctions being more severe than those in the center of the intron. As no intron sequence was removed by these insertions and the conserved splice junctions were maintained, our results indicate that in maize, the intron is likely recognized *in toto* rather than as the simple assembly of splice junctions and a branch point.

MATERIALS AND METHODS

Plasmid constructions

All restriction and DNA modifying enzymes and molecular linkers were obtained from New England Biolabs or BRL and used according to the manufacturers' instructions. Plasmids pCaG, pAL57, pAL61, pAL73 and pAL74 and the RNase protection probes pKLL51, pJD350, pBSI₁/luc3 and pBSI₃/luc101 have been described in detail (20). The maps of all of the expression plasmids are shown in Figure 1. pAL61, containing *Adh1-S* intron 1, was the parent plasmid for the intron insertion mutants in Fig. 1a; pAL74, containing actin intron 3, was the parent plasmid for the intron insertion mutants in Fig. 1b.

The *Adh1* cDNA fragment chosen as an insertion element is a 894 bp *BalI* fragment spanning position 934 to 3035 (see the genomic sequence (21) for numbering) and contains no known promoter or polyadenylation signals; the sense and antisense orientations are referred to as 'a' and 'b', respectively. This *BalI* fragment was inserted in the *Adh1-S* intron 1 *StuI* site (113 bp upstream of the 3' splice site) of pAL61 in the 'a' and 'b' orientations resulting in pALAS42 and pALAS23, respectively. *BgIII* linkers (CAGATCTG) were added to the same *BalI* fragment and inserted into the *BgIII* site of pAL61 (285 bp and 245 bp down- and upstream of the 5' and 3' splice sites, respectively) resulting in pALAB2 and pALAB4 for the 'a' and 'b' orientations, respectively. The *BalI* fragment was also cut at an internal *HindIII* site (position 934 to 2602) and a *HindIII* linker (CAAGCTTG) was added to the remaining *BalI* blunt end. The resulting 649 bp fragment was inserted in the *HindIII* site of *Adh1-S* intron 1 (73 bp downstream of the 5' splice site) of

pAL61 in both orientations resulting in the plasmids pALAH32 and pALAH31 for the 'a' and 'b' orientations, respectively. The *Adh1* cDNA *BalI* fragment was also inserted in one or two copies into the maize actin intron 3 *StuI* site of pAL74 (142 bp upstream of the 3' splice site) to yield pAL74A111 ('a'), pAL74A11 ('aa'; two copies), pAL74A21 ('b') and pAL74A2 ('bb'; two copies).

The bacteriophage λ DNA was derived from a 1.4 kb *EcoRI/HindIII* fragment corresponding to position 26,104 to 27,479. Insertion at the *HindIII* site of *Adh1-S* intron 1 in pAL61 was accomplished by filling in the *EcoRI* cohesive end with Klenow and adding *HindIII* linkers (CAAGCTTG); the resulting plasmid was pALLH24. The λ *HindIII* fragment was excised from pALAH24, filled in with Klenow, and *BamHI* linkers (CGGATCCG) were added prior to insertion at the *Adh1-S* intron 1 *BgIII* site of pAL61, resulting in pALLB2. The λ fragment from pALLH24 was excised with *HindIII*, filled in with Klenow and inserted into the *Adh1-S* intron 1 *StuI* site of pAL61 to produce pALLS3.

A 729 bp fragment of actin intron 3 (14) was excised with *Fnu4HI* and *SspI*, the overhang was filled in with Klenow, and the fragment was inserted in the 'a' and 'b' orientations into the *StuI* site of pAL61 to result in pALI₃S2 (14) and pALI₃S1, respectively.

Transient assays

Transient assays were done as described previously (20) with the following modifications. Black Mexican Sweet (BMS) protoplasts were prepared using a cocktail of 0.3% cellulase (CELFL; Worthington), 1% cytolase (Genencor) and 0.02% pectolyase (Seishin Pharmaceuticals) for 3–5 hr with mild agitation. The protoplasts were subjected to a 10 min heat shock at 45°C and chilled on ice for >30 min prior to electroporation (X-Cell™ 450; Promega Corp). The electroporated protoplasts were incubated for 20–24 hr and were harvested for either enzyme assays or RNA recovery as previously described (20).

RNA analysis

The northern blot was prepared as described previously (20) using a ³²P-labeled RNA probe (pKLL51) representing the entire luciferase coding region. The blot was washed at 78°C in 0.1×SSC, 0.5% SDS. Exposure was for 72 h using two intensifying screens.

The RNase protection assays were done as described (20); the probes used are listed in the figure legend. Probes were synthesized using ³²P-labeled CTP (Amersham) as described (22). The gel exposure time was 24 hr using two screens. The ³²P content of the relevant bands was quantified with an LKB laser densitometer; the appropriate correction for labeled C content was made prior to calculating the relative transcript abundance and splicing efficiency.

RESULTS

Plasmid constructs and transient assay

All plasmid constructs have been described previously (14, 20) or appear in Figure 1. Briefly, the *Adh1-S* intron 1-containing plasmid pAL61 has the *Adh1* promoter, first exon, first intron and five bp of the second exon translationally fused with the firefly luciferase gene. The AUG in exon 1 is an efficient translation start (K.Luehrsen, unpublished data), precluding luciferase synthesis from unspliced transcripts. pAL57 is identical to pAL61 except that it lacks *Adh1-S* intron 1. pAL74 (20) is

Figure 1. Structure of the expression constructs used in this study. **a.** *Adh1-S* intron 1-containing constructs. Only the name and the length of the chimeric gene are shown; all expression cassettes are in pUC8. The thick arrow denotes the *Adh1* promoter. The ATG listed is the naturally occurring *Adh1* start codon and is the first present in the mRNA. The 534 bp *Adh1-S* intron 1 is labeled 1 in a blank box. Insertion sequences are drawn in different patterns and its orientation denoted by an arrow; > = 'a' or sense orientation and < = 'b' or antisense orientation. The luciferase cDNA fragment is shown as a striped box labeled L. N is the nos 3' polyadenylation region. **b.** Maize actin intron 3-containing constructs. The 882 bp actin intron 3 is labeled 3 and is drawn as a blank box.

similar to pAL61 except that *Adh1-S* intron 1 is replaced by actin intron 3 (along with 43 and 9 bp of actin up- and downstream exon sequence, respectively). pAL73 is identical to pAL74 except that it lacks actin intron 3.

All plasmid constructs were electroporated into BMS protoplasts and either total RNA was harvested or extracts containing reporter gene enzyme activity were prepared and assayed. In accord with previous studies (20, 23) the presence of either *Adh1-S* intron 1 (pAL61) or actin intron 3 (pAL74) in the transcription unit enhanced luciferase activity about three-fold (Tables 1 and 2), a phenomenon referred to as 'intron enhancement'.

Insertion of non-intron DNA into *Adh1-S* intron 1 reduces expression in a position-dependent manner

At 534 bp, *Adh1-S* intron 1 is relatively long for maize but with an AT-content of 57%, its base composition is typical of maize and other monocots. To assess the extent to which maize introns could accommodate the insertion of foreign DNA sequence and still splice, several insertion sequences were tested: 1) a 649 bp or 894 bp *Adh1* cDNA fragment (in the sense or antisense orientation) lacking RNA processing signals with 48% AT, 2) a 1.4 kb AT-rich (62%) fragment of bacteriophage λ DNA and 3) a 729 bp internal segment of actin intron 3 (78 and 76 bp removed from the 5' and 3' splice junctions, respectively, with 64% AT). Three convenient restriction sites (*Hind*III, *Bgl*II, *Sna*I) within *Adh1-S* intron 1 were chosen as insertion sites: the *Hind*III site is 73 bp downstream from the 5' splice junction, the *Bgl*II site is at the approximate middle of the intron (>245 bp from either splice junction) and the *Sna*I site is 113 bp upstream of the 3' splice junction. Using deletion analysis it was found that the three restriction sites and the inclusive 348 bp sequence could be removed without diminishing transcript levels or splicing efficiency (our unpublished observations); this indicates that the center of *Adh1-S* intron 1 is not essential for splicing. The expression results for the insertion constructs are shown in Table 1; all comparisons were made to the control plasmid pAL61.

Insertion of the actin intron 3 sequence ('a' or sense orientation; pALI₃S2) at the *Sna*I site of *Adh1-S* intron 1 resulted in a modest increase in expression, indicating that the sequences at the *Sna*I site could be interrupted without a detrimental effect and that a 1.3 kb intron could be efficiently spliced. Thus, large intron size does not preclude splicing: *Adh1-S* intron 1 can accommodate additional sequences. When inserted in the opposite ('b' or antisense; pALI₃S1) orientation, actin intron 3 sequence caused a modest decrease in luciferase expression. This indicates that insertion orientation can affect expression (the lower expression was accounted for by a decreased splicing efficiency; see below).

In contrast to the mild effects of actin intron 3 sequence additions, the insertion of *Adh1* cDNA fragments into *Adh1-S* intron 1 significantly reduced expression (Table 1). When the insertions were made in the 'a' orientation, expression was reduced to 12% or less (pALAH32, pALAB3 and pALAS42), indicating a severe mutation. When the same fragments were introduced at the same positions in the opposite, or 'b' orientation (pALAH31, pALAB4 and pALAS23), expression was again reduced but at levels ~5-fold above those found with the same inserts in the 'a' orientation. These results are shown graphically in Figure 2. We will show that the expression decreases are the result of a reduced steady state mRNA levels and/or a decreased splicing efficiency of the altered introns.

Artificial introns with high AU content are spliced more efficiently than GC-rich introns in both dicots and monocots (7, 8). We therefore inserted a fragment of bacteriophage λ DNA (62% AT) into *Adh1-S* intron 1 (57% AT). As shown in Table 1 (experiment 2), insertion of a 1.4 kb fragment of bacteriophage λ DNA into the *Hind*III, *Bgl*II or *Sna*I sites of *Adh1-S* intron 1 reduced expression to 25% to 55%. The qualitative position-dependent expression pattern was similar to the cDNA 'a' and 'b' insertion variants (Table 1 and Figure 2). Consequently, AT-rich fragments inserted into *Adh1-S* intron 1 can be detrimental. As we did not determine transcript structure for the λ insertion constructs, either fortuitous RNA processing signals or decreased splicing efficiency could be responsible for the expression decrease.

Table 1. Luciferase expression and RNA quantitation data for insertions into *Adh1-S* intron 1

Plasmid	Insertion	Position/ orientation	Expression Exp. 1			Expression Exp. 2			RNA Exp. 1		RNA Exp. 2	
			GUS units	lu/GUS	vs. pAL61 (%)	GUS units	lu/GUS	vs. pAL61 (%)	transcript	% splicing	transcript	% splicing
control			0	0	0	0	0	0	—		—	
pAL57			1005	13 ± 4	19	881	18 ± 7	36	+		++	
pAL61	none		1101	68 ± 10	100	974	50 ± 14	100	++++	70	++++	60
pAL ₃ S2	actin	<i>StuI</i> /'a'	764	181 ± 43	267		ND		++++	79	ND	
pAL ₃ S1	intron 3											
	actin	<i>StuI</i> /'B'	839	74 ± 4	110		ND		++++	45	ND	
pALAH32	<i>Adh1</i> cDNA	<i>HindIII</i> /'A'	984	3 ± 2	4	845	1 ± 0.3	3	+	58	+	42
pALAB3	<i>Adh1</i> cDNA	<i>BglII</i> /'a'		ND		911	6 ± 1	12	ND		++	40
pALAS42	<i>Adh1</i> cDNA	<i>StuI</i> /'A'	1045	3 ± 1	5	816	2 ± 1	5	++	31	++	18
pALAH31	<i>Adh1</i> cDNA	<i>HindIII</i> /'b'	1156	8 ± 3	12	762	8 ± 2	16	++	60	ND	
pALAB4	<i>Adh1</i> cDNA	<i>BglII</i> /'b'		ND		902	27 ± 4	55	ND		ND	
pALAS23	<i>Adh1</i> cDNA	<i>StuI</i> /'b'	932	29 ± 6	43	776	18 ± 4	35	+++++	43	ND	
pALLH24	λ	<i>HindIII</i>	802	14 ± 3	21	531	13 ± 2	25	ND		ND	
pALLB2	λ	<i>BglII</i>	620	63 ± 32	93	835	28 ± 3	55	ND		ND	
pALLS3	λ	<i>StuI</i>		ND		645	14 ± 2	27	ND		ND	

Exp. 1 and exp. 2 refer to independent experiments. The GUS units are expressed as pMol/min and luciferase activity is expressed as light units (lu; photons per 10 sec). The absolute luciferase values ranged from <100 lu for the no plasmid control to >50000 lu for pAL61. The column transcript is a measure of luciferase transcript present (detected by the luc probe in Figure 4) and is expressed as one to five + with the pAL61 control being arbitrarily set at four +. The splicing efficiency was calculated from the RNase mapping data shown in Figure 4. ND=not done.

Table 2. Luciferase expression and RNA quantitation data for the insertions into actin intron 3

Plasmid	Insertion	Position/ orientation	GUS units	Expression lu/GUS	vs. pAL74 %	RNA	
						transcript	% spliced
control			0	0	0	—	
pAL73			735	25 ± 2	30	ND	
pAL74	none		686	85 ± 0.2	100	++++	88
pAL74A111	<i>Adh1</i> cDNA	<i>StuI</i> /'a'	574	64 ± 20	76	++	85
pAL74A11	<i>Adh1</i> cDNA	<i>StuI</i> /'aa'	676	44 ± 10	52	++	84
pAL74A21	<i>Adh1</i> cDNA	<i>StuI</i> /'b'	622	107 ± 6	126	ND	
pAL74A2	<i>Adh1</i> cDNA	<i>StuI</i> /'bb'	635	64 ± 21	76	ND	

ND=not done. The GUS units are expressed as pMol/min and luciferase activity is expressed as light units (lu; photons per 10 sec). The absolute luciferase values ranged from <100 lu for the no plasmid control to >50000 lu for pAL74. The column transcript is a measure of luciferase transcript present detected by the luc probe in an RNase mapping experiment (not shown) and is expressed as one to five stars with the pAL74 control being arbitrarily set at four +. The splicing efficiency was also calculated from the RNase mapping data not shown.

The most severe mutations were the result of exon (an *Adh1* cDNA) insertion. When the *Adh1* cDNA was reversed (non-exon 'b' orientation), expression was nearer wild-type levels. Although quantitative differences in expression were observed for the *Adh1* cDNA and λ insertions, they shared a general pattern: insertions at the upstream *HindIII* site were always the most severe, insertions at the central *BglII* site were invariably the least severe and those at the downstream *StuI* site were intermediate. These data indicate that expression is sensitive to the placement as well as the sequence content of insertions into *Adh1-S* intron 1.

Splicing efficiency is reduced by inserts of foreign DNA

The above data indicate that insertion of non-intron DNA into the intron of a reporter construct reduces luciferase expression in a transient assay. To investigate the molecular mechanisms responsible for reduced expression, transcript RNAs were analyzed by northern blotting and RNase protection.

Conceivably, each of the insertion sequences could contain RNA processing signals. To test for this, a northern blot was used to estimate the size of each transcript synthesized, and the results are shown in Figure 3. For the pAL57 and pAL61 constructs the major product is ~2.1 kb, the length expected

Figure 2. Bar chart comparing the luciferase expression of several constructs. All luciferase values from Table 1 were normalized to pAL61 set at 100% expression. Note that pAL57 lacks *Adh1-S* intron 1.

Figure 3. Northern blot of total RNA isolated from protoplasts electroporated with several different expression constructs. Total RNA was prepared from maize protoplasts electroporated with the expression construct listed. The ^{32}P -labeled RNA probe used represented the entire luciferase cDNA. The thick arrow to the right points to the spliced luciferase mRNA band. The small arrow to the left of each lane indicate the positions of the unspliced transcripts from each expression construct. Sizes (in kb) of the RNA standards are given on the left margin. Although there are several lightly-hybridizing bands in the lanes representing the intron chimera constructs, these can also be seen in the control lane, indicating fortuitous hybridization of the luciferase probe to unidentified cellular transcripts.

for mature luciferase mRNA. In addition, the pAL61 lane has a band at ~ 2.6 kb representing unspliced RNA (2.1 kb plus the 0.5 kb *Adh1-S* intron 1); we have previously noted ~ 25 – 40% unspliced transcript from this construct (20). For each expression construct containing an insertion in *Adh1-S* intron 1, an unspliced RNA size class was also detected (larger than 2.6 kb because of the insert into intron 1). No other strongly hybridizing bands were detected, indicating that no additional RNA processing events were occurring to a significant degree. For the *Adh1* cDNA insertion constructs (except pALAS23), much less luciferase-containing transcript was observed (see Table 1) consistent with the reduced luciferase activity compared to the pAL61 standard.

RNAse mapping using a probe that spanned the intron/exon 2 border was employed to compare the ratio of spliced to unspliced RNAs. From this ratio, a splicing efficiency was calculated. While we have not directly compared the half-lives of the spliced and unspliced transcripts *in vivo*, we assume they are similar. The following evidence supports this assumption: 1. The splicing efficiencies do not vary greatly between experiments, 2. A study of *Bronze-2* transcripts in BMS protoplasts showed that the spliced and unspliced forms are polyadenylated (J.Nash and V.Walbot, manuscript submitted), and 3. The cytoplasmic half-lives of an intron-containing and a cDNA version of Bz2 transcripts were the same in BMS protoplasts (J.Nash and V.Walbot, manuscript submitted).

As shown in Figure 4 and numerically in Table 1, the splicing efficiency of the control plasmid pAL61 ranged from 60–70%, in accord with previous results (20). For only one construct, pALI₃S2, was the splicing efficiency increased, to 78%; this construct contains a 729 bp addition of actin intron 3 sequence

Figure 4. Representative RNase mapping gel. 15 μg of total RNA, derived from protoplasts electroporated with the labeled expression construct, was protected from RNase digestion with a probe (pBSI₁luc (20) linearized with *Stu*I) spanning the intron/exon border. Unspliced and spliced RNAs protect a 210 nt and a 97 nt region of the probe, respectively. This probe does not protect any of the insertion sequences. The splicing efficiency of each construct was calculated by comparing the ratio of spliced to unspliced RNA. The luc cds probe protects a 334 nt RNA in the 3' end of the luciferase coding region and the protected RNA includes both spliced and unspliced transcripts. The *Adh1-S* cds probe protects a 152 nt RNA from cellular *Adh1* expression and is used to correct for protoplast viability and RNA recovery.

in the sense orientation. In all other cases the splicing efficiency was adversely affected by the insertion sequence. In general, the splicing efficiencies for the insertions at the *Hind*III and *Bgl*II sites were ~ 10 – 20% reduced from the splicing efficiency of the pAL61 control. However, insertions at the *Stu*I site resulted in a $\sim 40\%$ decrease in the splicing efficiency. Thus, insertions 113 bp upstream of the 3' splice junction significantly reduced the splicing efficiency of *Adh1-S* intron 1, even though they are likely far removed from a branch point.

***Adh1* cDNA insertions in actin intron 3 also reduce expression**

Actin intron 3 has previously been shown to be adversely affected by the insertion of a *Mul*I element in parallel with observations on *Mul*I insertions in *Adh1-S* intron 1 (14). To determine whether the *Adh1* cDNA insertion effects were idiosyncratic to *Adh1-S* intron 1 or could be generalized to other introns, we inserted the 894 bp *Adh1* cDNA fragment into the *Stu*I site of actin intron 3. Actin intron 3 is longer than *Adh1-S* intron 1 (882 bp vs 534 bp) and is more AT-rich (65% vs 57%). In addition the *Stu*I site of actin intron 3 is 142 bp upstream of the 3' splice junction versus 113 bp for the *Stu*I site of *Adh1-S* intron 1; the *Stu*I site and surrounding sequences can be deleted without affecting intron function (our unpublished results). As with *Adh1-S* intron 1, the insertion of the cDNA fragment in the 'a' orientation, in one or two copies (pAL74A111 and pAL74A11), significantly decreased expression when compared to the pAL74 control.

Insertion in the 'b' orientation, again in one or two copies (pAL74A21 and pAL74A2), had a less severe effect on luciferase expression. Insertions at the *StuI* site of *Adh1-S* intron 1 significantly reduced splicing efficiency, but the same cDNA insertion at the *StuI* site of actin intron 3 did not result in reduced splicing efficiency; instead transcript abundance was ~2-fold reduced (Figure 3). These results show that expression from both actin intron 3 and *Adh1-S* intron 1 constructs is adversely affected by the insertion of foreign DNA.

DISCUSSION

Differential impact of intron and non-intron insertions

We wished to determine if the splicing defects caused by the insertion of TEs into introns were caused only by the introduction of RNA processing signals or whether the insertion of non-intron DNA could affect processing, intron recognition and/or transcript stability. In this study, we found that the insertion of fragments of *Adh1* cDNA and λ DNA into *Adh1-S* intron 1 or actin intron 3 adversely affected expression of the luciferase reporter gene. None of these insertions was at or near intron elements known to be important for splicing and neither insert contains normal splice sites. Indeed, the insertions were placed in intron regions shown by deletion analysis to be non-essential for splicing (our unpublished observations), indicating that the introduced sequences were not disrupting an important, as yet undefined functional element. Although the base content of maize introns averages 59% AU (34% U), both the GC-rich *Adh1* cDNA and the AT-rich λ sequences resulted in decreased expression. Thus, AU-richness alone is not sufficient to mimic intron sequence; there are likely functional sequence motifs that are present in genuine introns that are absent in the bacteriophage λ sequence. A computer search to find extended RNA secondary structure caused by the insertions was negative, making it unlikely that fortuitous stem-loop structures were responsible for the effects we observed.

In contrast to the *Adh1* cDNA and λ DNA insertions, the 729 bp internal fragment of actin intron 3 inserted in *Adh1-S* intron 1 stimulated expression. We conclude that the maize splicing machinery was able to accommodate an intron insertion at the *StuI* site and an increase in intron length from 534 bp to 1.3 kb. Transcript analysis demonstrated that the mRNA level was enhanced ~2.5-fold and that the splicing efficiency was slightly increased. Thus, the insertion of additional intron sequence stimulated *Adh1-S* intron 1 recognition and splicing. Insertion of the same fragment in the antisense or 'non-intron' orientation did not affect transcript length or abundance but did reduce splicing efficiency. These results confirm that important sequence motifs other than AU-richness *per se* are required for efficient splicing because AU content of the hybrid intron is the same for either orientation of the insertion. In the sense orientation the insert is 36% T and 28% A; this possibly indicates that the functional sequence motifs in maize pre-mRNAs involve U-rich sequence. For dicot introns, no functional differences could be attributed to A- or U-rich sequence (7).

Splicing failure from insertions in introns

In the cases for which mRNA structure was studied, we found that the *Adh1* cDNA insertions simply lengthened the transcript. This result is consistent with splicing failure. Northern blot analysis indicated that the insertions were not adding usable splice junctions to a detectable degree, even though the *Adh1* cDNA contained at least one usable cryptic 3' splice site in *Adh1* exon

4 (shown to be activated by a *Ds2* inserted into *Adh1* (24)). Thus, the *Adh1* cDNA insertions were like TEs in being mutagenic but unlike TEs in failing to promote alternative splicing. This lends credence to the hypothesis that maize TEs have evolved to contain RNA processing signals that promote alternative splicing; a kind of phenotypic rescue is therefore possible at the mRNA level. This strengthens the hypothesis that at least some maize TEs have evolved into mobile introns (17).

The splicing efficiency of *Adh1-S* intron 1 varied with the position of the *Adh1* cDNA insertion. When placed at the upstream *HindIII* or central *BglII* sites, the cDNA insertions caused only a modest decrease in splicing efficiency. However, insertion of the cDNA or actin intron 3 antisense fragment at the *StuI* site invariably resulted in a ~2- to 3-fold decrease in splicing efficiency, indicating that the splicing machinery could not accommodate non-intron sequence at this location. This observation is in contrast to *Adh1* cDNA insertions at the downstream *StuI* site in actin intron 3 which did not impair splicing. Splicing of individual introns is likely to behave idiosyncratically according to the sequence content of the insertion and its proximity to the splice junctions. This could be a consequence of the differential binding of hnRNP proteins (see Conclusions).

Intron enhancement

Intron enhancement of gene expression is well-documented in both mammals and plants and results in an increase of cytoplasmic mRNA (23, 25). *Adh1-S* intron 1 and actin intron 3 have been shown to enhance expression via this mechanism (20, 23). Transcript abundance showed a marked decline as the *Adh1* cDNA insertions in pAL61 were placed at the *HindIII* site near the 5' splice site of *Adh1-S* intron 1; in fact, transcript levels approached those seen for the intronless pAL57 construct. The diminished transcript levels might be a consequence of a decreased transcription rate caused by the proximity of the insertion to the promoter or by fortuitous polyadenylation caused by the cDNA insert. Alternatively, we suggest that intron enhancement is nullified by the insertion. Enhancement may be a consequence of the pre-mRNA aggregating with spliceosomes and proceeding through RNA processing and transport in an efficient pathway, while intronless mRNAs may enter a less efficient pathway of RNA processing (26); studies in mammals indicate that splicing leads to more efficient polyadenylation and hence increased mRNA stability (27). We propose that the 5' proximal insertions (at the *HindIII* site) interfered with the initial steps of intron recognition (possibly by inhibiting U1 snRNP binding), shunting the transcript through the intronless pathway. Distal insertions (at the *BglII* and *StuI* sites) would allow more of the intron (>245 bases and therefore larger than a typical intron) to be exposed before the non-intron sequence is transcribed, and increase the likelihood of intron recognition, entry into the splicing pathway and enhancement; this interpretation is consistent with the increased transcript levels found for the *Adh1* cDNA insertions at the *BglII* and *StuI* sites. Expression was most reduced using the *Adh1* cDNA insertion in the sense or 'a' orientation. This observation could mean that intron recognition is especially sensitive to the presence of genuine exon sequence in the intron. Maize exons might have common sequence motifs that bind particular hnRNP proteins and direct the splicing complexes to intron regions of the pre-mRNA. The antisense or 'b' orientation of the cDNA would not have such exon characteristics when transcribed into RNA, resulting in a more efficient recognition of the chimeric introns. Recalling that

the insertion sequences sometimes caused a reduction in splicing efficiency, our evidence suggests that intron features that promote recognition and commitment to the splicing pathway in maize may not be the same as those required for efficient splicing. The pALAS23 construct is illustrative in that transcript levels reflected intron enhancement but the splicing efficiency was reduced ~4-fold.

CONCLUSIONS

In summary, we found that non-intron sequence inserted in *Adh1* intron 1 or actin intron 3 decreased transcript abundance and/or splicing efficiency. Unlike TEs, the cDNA insertions did not add splicing signals to the pre-mRNA; rather, their mere presence reduced overall expression. This suggests that the maize splicing machinery did not recognize the chimeric introns and/or was unable to splice them efficiently. Our working hypothesis is that in plants, intron recognition depends on internal intron features; these are redundant as large internal intron deletions (~75% of the intron) are still spliced efficiently and large insertions of internal intron sequence do not depress splicing. It is likely that the role of AU-richness of maize introns is in part a signal to distinguish them from exons, which are GC-rich (7, 8). Introns might thus be defined by their AU- (or U-) richness, with the 5' border being chosen in part by homology with U1 snRNA and the 3' border by a YAG at the transition point between AU- and GC-rich RNA regions. This type of intron definition might be explained by differential binding of hnRNP proteins to the nascent pre-mRNA. For example, research on mammalian RNA processing has revealed that a subset of hnRNP proteins binds pre-mRNA to form the H complex (28, 29); the mixture of proteins bound is specific to each pre-mRNA tested (30). The H complex then acts as a scaffold for the addition of splicing components. Plant intron regions could complex with an hnRNP AU- or U-rich binding protein that either attracts the splicing machinery and/or directs it to splice at the appropriate sequences. Alternatively, there might be a GC-rich binding protein that tags exons, leaving introns exposed. hnRNP proteins with affinities for a particular nucleotide, including U, have been identified in mammals (31). The presence of non-intron inserts in introns might confuse the initial recognition of an intron region or impair the ability of a recognized intron to be efficiently assembled into spliceosomes. The insertion of exon sequence, which might attract a subset of hnRNP proteins different from those of introns, might be particularly detrimental to splicing. These possibilities are consistent with the data presented.

If differential hnRNP binding is responsible for the altered splicing patterns that we observed, it is apparent that the diverse TE insertions in introns and neighboring exons provide a rich source of genetic variation. TEs might reduce expression by interfering with intron recognition and consequently intron enhancement, especially in transcripts having only one intron. As the TE becomes more AU-rich and intron-like, more opportunities for alternative splicing are created. Future work will concentrate on defining the intron motifs necessary to insure accurate intron recognition and efficient splicing.

ACKNOWLEDGMENTS

We thank Hanya Chrispeels, Maria-Ines Benito and Jane Hersberger for helpful comments. We also thank two anonymous reviewers for helpful comments. This research was supported by a grant from the NIH (GM 32422).

REFERENCES

1. Sharp, P.A. (1987) *Science*, **235**, 766–771.
2. Ruby, S.W. and Abelson, J. (1991) *Trends in Genet.*, **7**, 79–85.
3. Wieringa, B., Hofer, E. and Weissmann, C. (1984) *Cell*, **37**, 915–925.
4. Hanley, B.A. and Schuler, M.A. (1988) *Nucl. Acids Res.*, **16**, 7159–7176.
5. Hawkins, J.D. (1988) *Nucl. Acids Res.*, **16**, 9893–9908.
6. Csank, C., Taylor, F.M. and Martindale, D.W. (1990) *Nucl. Acids Res.*, **18**, 5133–5141.
7. Goodall, G.J. and Filipowicz, W. (1989) *Cell*, **58**, 473–483.
8. Goodall, G.J. and Filipowicz, W. (1991) *EMBO J.*, **10**, 2635–2644.
9. Goodall, G.J. and Filipowicz, W. (1990) *Plant Mol. Biol.*, **14**, 727–733.
10. Pautot, V., Brzezinski, R. and Tepfer, M. (1989) *Gene*, **77**, 133–140.
11. van Santen, S.V. and Spritz, R.A. (1987) *Gene*, **56**, 253–265.
12. Wiebauer, K., Herrero, J.-J. and Filipowicz, W. (1988) *Mol. Cell. Biol.*, **8**, 2042–2051.
13. Weil, C.F. and Wessler, S.R. (1990) *Annu. Rev. Plant Physiol.*, **41**, 527–552.
14. Luehrsen, K.R. and Walbot, V. (1990) *Plant Cell*, **2**, 1225–1238.
15. Wessler, S.R., Baran, G. and Varagona, M. (1987) *Science*, **237**, 916–918.
16. Kim, H.-Y., Schiefelbein, J.W., Raboy, V., Furtek, D.B. and Nelson, O.E., Jr. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 5863–5867.
17. Wessler, S.R. (1991) *Mol. Cell. Biol.*, **11**, 6192–6196.
18. Wessler, S.R. and Varagona, M. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 4177–4181.
19. Varagona, M.J., Purugganan, M. and Wessler, S.R. (1992) *Plant Cell*, **4**, 811–820.
20. Luehrsen, K.R. and Walbot, V. (1991) *Mol. Gen. Genet.*, **225**, 81–93.
21. Dennis, E.S., Gerlach, W.L., Pryor, A.J., Bennetzen, J.L., Inglis, A., Llewellyn, D., Sachs, M.M., Ferl, R.J. and Peacock, W.J. (1984) *Nucl. Acids Res.*, **12**, 3983–4000.
22. Krieg, P.A. and Melton, D.A. (1987) *Meth. Enzymol.*, **155**, 397–415.
23. Callis, J., Fromm, M. and Walbot, V. (1987) *Genes Devel.*, **1**, 1183–1200.
24. Simon, R. and Starlinger, P. (1987) *Mol. Gen. Genet.*, **209**, 198–199.
25. Buchman, A.R. and Berg, P. (1988) *Mol. Cell. Biol.*, **8**, 4395–4405.
26. Yu, X., Gelembiuk, G.W., Wang, C.-Y., Ryu, W.-S. and Mertz, J.E. (1991) *Nucl. Acids Res.*, **19**, 7231–7234.
27. Huang, M.T.F. and Gorman, C.M. (1990) *Nucl. Acids Res.*, **18**, 937–947.
28. Michaud, S. and Reed, R. (1991) *Genes Dev.*, **5**, 2534–2546.
29. Reed, R. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 8031–8035.
30. Bennett, M., Pinol-Roma, S., Staknis, D., Dreyfuss, G. and Reed, R. (1992) *Mol. Cell. Biol.*, **12**, 3165–3175.
31. Swanson, M.S. and Dreyfuss, G. (1988) *Mol. Cell. Biol.*, **8**, 2237–2241.