Mutational analysis of a plant branchpoint and polypyrimidine tract required for constitutive splicing of a mini-exon

CRAIG G. SIMPSON, GRAHAM THOW, GILLIAN P. CLARK, S. NIKKI JENNINGS,
JENNY A. WATTERS, and JOHN W.S. BROWN

Unit of Gene Expression, Scottish Crop Research Institute, Invergowrie, Dundee, DD2 5DA, Scotland, United Kingdom

ABSTRACT

The branchpoint sequence and associated polypyrimidine tract are firmly established splicing signals in vertebrates. In plants, however, these signals have not been characterized in detail. The potato invertase mini-exon 2 (9 nt) requires a branchpoint sequence positioned around 50 nt upstream of the 5' splice site of the neighboring intron and a U₁₁ element found adjacent to the branchpoint in the upstream intron (Simpson et al., *RNA*, 2000, 6:422–433). Utilizing the sensitivity of this plant splicing system, these elements have been characterized by systematic mutation and analysis of the effect on inclusion of the mini-exon. Mutation of the branchpoint sequence in all possible positions demonstrated that branchpoints matching the consensus, CURAY, were most efficient at supporting splicing. Branchpoint sequences that differed from this consensus were still able to permit mini-exon inclusion but at greatly reduced levels. Mutation of the downstream U₁₁ element suggested that it functioned as a polypyrimidine tract rather than a UA-rich element, common to plant introns. The minimum sequence requirement of the polypyrimidine tract for efficient splicing was two closely positioned groups of uridines 3–4 nt long (<6 nt apart) that, within the context of the mini-exon system, required being close (<14 nt) to the branchpoint sequence. The functional characterization of the branchpoint sequence and polypyrimidine tract defines these sequences in plants for the first time, and firmly establishes polypyrimidine tracts as important signals in splicing of at least some plant introns.

Keywords: branchpoint; mini-exon; polypyrimidine tract; pre-mRNA splicing; splicing efficiency

INTRODUCTION

Commitment of a pair of intron splice sites to splicing requires the early recognition of pre-mRNA splicing signals. In metazoa, the splice sites at the 5' and 3' end of the intron and an internal branchpoint associated with a downstream polypyrimidine tract are essential. In early spliceosome assembly, U1snRNP is recruited to the 5' splice site where U1snRNA base pairs with the precursor messenger RNA (pre-mRNA). Branchpoint binding protein (BBP/SF1) and U2 auxiliary factor 65 (U2AF⁶⁵) bind cooperatively to the branchpoint and polypyrimidine tract and recruit U2snRNP, allowing base pairing of U2snRNA to the branchpoint sequence, and help define the 3' splice site. Interactions between the factors at the 5' splice site and branchpoint/polypyrimidine tract/3' splice site form the presplicing complex (Adams et al., 1996; Krämer, 1996; Reed, 2000; Hastings

& Krainer, 2001). Both intron and exon splicing enhancer and silencer elements have also been identified that modify the selection of splice sites through interactions with other protein factors. Many of these splice site selection processes in multiple intron pre-mRNA transcripts are coordinated through the activity of a number of splicing regulatory factors, including SR-rich proteins, that permit cooperative interaction between splice sites along the pre-mRNA transcript and lead to highly regulated splice site selection (Black, 1995; Wang & Manley, 1997; Blencow, 2000; Reed, 2000; Smith & Valcárcel, 2000; Hastings & Krainer, 2001).

Metazoan intron 5' and 3' splice sites consist of 9 nt (CAG:GUAAGU) and 5 nt (GCAG:G) consensus sequences respectively (highly conserved GU and AG are underlined; ":" denotes cleavage site; Krämer, 1996). Yeast introns contain similar sequences, but the 5' splice site is less variable in sequence. In vertebrate introns, the branchpoint sequence and downstream polypyrimidine tract are essential signals for efficient recognition of the 3' splice site and splicing. The vertebrate intron branchpoint region is variable with the consensus,

Reprint requests to: John W.S. Brown, Department of Gene Expression, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, Scotland, United Kingdom; e-mail: jbrown@scri.sari.ac.uk.

CURAY, in contrast to yeast where the branchpoint sequence, UACUAAC, is highly conserved.

Detailed characterization of branchpoint and polypyrimidine tract sequences have been obtained largely through mutational analysis and the availability of in vitro splicing extracts (Langford et al., 1984; Ruskin et al., 1985; Hornig et al., 1986; Zhuang et al., 1989; Roscigno et al., 1993; Query et al., 1994; Coolidge et al., 1997; Lund et al., 2000). In the absence of plant nuclear splicing extracts, it has not been possible to directly observe the formation of an intron lariat, which has made characterization of plant branchpoint consensus sequences difficult. However, plant U2snRNAs are highly conserved at the branchpoint recognition sequence (Vankan & Filipowicz, 1988) and mutational analysis of putative branchpoints has highlighted their importance in plant splicing (Simpson et al., 1996). In addition, branchpoints were mapped in two plant introns following mutation of G + 1 $\rightarrow$ A at their 5' splice sites, permitting only the first step of splicing to occur. The branchpoints mapped to adenines contained in a sequence similar to a vertebrate branchpoint consensus (Liu & Filipowicz, 1996). Computer analyses of plant intron sequences identified a branchpoint consensus that matched the vertebrate branchpoint consensus. The sequence and the position of these putative branchpoints, like their vertebrate counterparts, were highly variable (Brown, 1986; Brown et al., 1996; Simpson & Filipowicz, 1996; Simpson et al., 1996; Tolstrup et al., 1997). Finally, a temperature-sensitive splicing *Arabidopsis* mutant in *APETELA3* (*ap3-1*), containing a mutation in a 5' splice site, resulted in exon skipping. A suppressor mutant (*ap3-11*) was identified where mutation had generated a new putative branchpoint sequence in the intron upstream of the skipped exon, restoring splicing by compensating for the weakened 5' splice site (Sablowski & Myerowitz, 1998; Yi & Jack, 1998).

These various lines of evidence demonstrated the presence of a branchpoint sequence with similarity to that of vertebrate introns and the importance of such sequences in splicing and splice site selection in plants. However, to date, the importance of the sequence context surrounding plant branchpoint adenines remains to be determined. Plants do not contain pronounced polypyrimidine tracts as found in vertebrate introns, but tend to have U-rich sequences between putative branchpoints and the 3' splice site (Brown, 1986; Hanley & Schuler, 1988; Wiebauer et al., 1988; Goodall & Filipowicz, 1991; Baynton et al., 1996; Brown et al., 1996). However, U- or UA-richness is a general feature of plant introns, essential for efficient splicing (Goodall & Filipowicz, 1989). Mutations to U-rich regions in a number of plant introns (maize *Adh1* introns 1 and 3, the *Bz2* intron, pea *rbcS3A* intron 1, and potato invertase intron 1 and synthetic introns), by changing the position of the U-rich regions, by disrupting their length, or by

changing the sequence content, led to changes in splicing efficiency and activation of cryptic splice sites (Goodall & Filipowicz, 1989; Lou et al., 1993a, 1993b; McCullough et al., 1993; Luehrsen & Walbot, 1994a, 1994b; Baynton et al., 1996; Gniadkowski et al., 1996; Carle-Urioste et al., 1997; Merritt et al., 1997; Ko et al., 1998; Latijnhouwers et al., 1999; Simpson et al., 2000). More specifically, altering U-rich sequences upstream and downstream of putative branchpoint signals can affect the recognition of the intron and lead to the activation of cryptic splice sites (Luehrsen & Walbot, 1994a; Baynton et al., 1996; Merritt et al., 1997; Simpson et al., 2000). Because of the effects of U-rich sequences positioned throughout plant introns, it has not been possible to distinguish sequences that function as polypyrimidine tracts.

The potato invertase mini-exon is constitutively included in invertase transcripts and is dependent on strong constitutive splicing signals (Bournay et al., 1996; Simpson et al., 2000). The key elements that determine mini-exon inclusion are a branchpoint sequence, an adjacent downstream U-rich region, and the distance between these signals and the 5' splice site downstream of the mini-exon (Fig. 1). The requirement for these strong splicing signals makes this splicing arrangement sensitive to sequence changes in the splicing signals. It provides a system to investigate, in detail, the sequence and spatial requirements of these signals, for the first time in plants. This article addresses the current lack of detailed information of these important splicing signals through the systematic mutation of the invertase intron 1 branchpoint and polypyrimidine tract. The relative importance of nucleotides surrounding the branchpoint adenine for efficient splicing of the mini-exon and the contribution of these nucleotides towards the strength of plant branchpoint signals are determined. In addition, the U-rich sequence was analyzed, suggesting that it functions as a polypyrimidine tract. Further mutations define the key sequence components and the relative positions of branchpoint and polypyrimidine tract. The information from this study will be important in the analysis of alternative splicing regulation in plants.

RESULTS

Systematic mutational analysis of the invertase branchpoint sequence

The invertase intron 1 branchpoint sequence CUAU was systematically mutated in each position and the effect on splicing efficiency analyzed in vivo in tobacco protoplasts (Fig. 2, Table 1). Mutation of C⁻³ to U reduced splicing to 80% but mutation to a purine reduced splicing substantially (A⁻³, 39%; G⁻³, 18%). Thus a pyrimidine in this position with preference for C⁻³, is required for efficient splicing. Mutations to U⁻² se-

FIGURE 1. Genomic structure of invertase *GF* mini-exon. **A:** Schematic representation of proposed exon definition process that promotes splice site selection of the 9-nt invertase mini-exon (Simpson et al., 2000). Exons are indicated as boxes and introns by lines. Splice sites are shown as GU and AG. The branchpoint (BP1) and the associated U-rich element (U₁₁) are found over 50 nt away from the downstream 5' splice site in all dicot invertase genes isolated so far and are indicated by the double-headed arrow. **B:** RNA sequence of the 3' end of invertase *GF* intron 1. The branchpoint consensus (BP1), the mini-exon 2 (in capital letters and separated from the intron sequence by //), and the 5' splice site of intron 2 are shown. The branchpoint consensus (BP1) and the U₁₁ mutated in this article are indicated in capitals.

verely reduced splicing in all cases, giving less than 5% splicing, suggesting that U in this position is an essential nucleotide. Mutations of A⁻¹ to G reduced splicing only slightly (94%) whereas mutation to C or U had a greater effect (51% and 47%, respectively). Thus

a purine at position -1 is the preferred nucleotide but pyrimidines are tolerated. As expected, mutations of the branchpoint adenine (A⁰) severely inhibited splicing. A⁰ to G gave no splicing whereas A⁰ to U or C gave 19% and 7%, respectively. Finally, mutation of U⁺¹ to C did not affect splicing, whereas mutation to purines reduced splicing to 59% for A and 33% for G.

FIGURE 2. Splicing analysis of branchpoint mutations. GeneScan analysis of RT-PCR products from tobacco protoplasts transfected with constructs mutated at the invertase intron 1 branchpoint consensus. Seven examples showing the range of mini-exon inclusion obtained through changes to the branchpoint consensus are shown. For a full list of all the branchpoint mutations and their effect on mini-exon inclusion, see Table 1. M is a marker lane showing the position of labeled markers at 160 bp and 200 bp. The expected RT-PCR product size of 180 bp for mini-exon skipping is indicated (Sk) as is the 189-bp RT-PCR product that shows inclusion of the mini-exon (In). The nucleotide mutations (underlined) to the branchpoint sequence are shown on the left.

TABLE 1. Splicing efficiency of branchpoint mutations.

Clone	Branchpoint mutations					Splicing efficiency	Frequency ^a
	-3	-2	-1	0	+1		
inv1	C	U	A	A	U	99 ± 1	12
inv74	A	U	A	A	U	39 ± 0.3	
inv55	G	U	A	A	U	18 ± 0.4	
inv54	U	U	A	A	U	80 ± 3.2	5
inv56	C	A	A	A	U	3 ± 1.6	
inv75	C	C	A	A	U	4 ± 3.0	
inv57	C	G	A	A	U	2 ± 1.2	
inv73	C	U	C	A	U	51 ± 4.6	4
inv52	C	U	G	A	U	94 ± 2.4	5
inv53	C	U	U	A	U	47 ± 1.6	5
inv76	C	U	A	C	U	7 ± 1.5	
inv77	C	U	A	G	U	0.4 ± 0.4	
inv10	C	U	A	U	U	19 ± 3.8	
inv71	C	U	A	A	A	59 ± 0.2	5
inv51	C	U	A	A	C	99 ± 0.8	3
inv72	C	U	A	A	G	33 ± 1.2	
inv79	U	U	G	A	U	48 ± 3.4	5
inv80	U	U	U	A	U	11 ± 2.1	4
inv81	G	U	G	A	U	9 ± 5.4	
inv78	C	U	U	A	A	19 ± 0.3	5

^aFrequency of occurrence of putative branchpoint sequences is based on dicot introns (Simpson et al., 1996).

Bold nucleotides indicate changes from the branchpoint sequence CUAAU found in inv1.

Thus a pyrimidine is the preferred nucleotide at position +1. These results clearly show the importance of the U^{-2} and A^0 nucleotides and preferences for pyrimidines in positions -3 and +1, and purines at -1 (Fig. 2, Table 1).

In addition to single nucleotide mutations, double mutations were made in positions -3, -1, and +1 to investigate the effects of different combinations of pyrimidines and purines at these positions. Four double mutations were analyzed: UUUAU (inv80), UUGAU (inv79), GUGAU (inv81), and CUUAA (inv78), and in all cases the double mutation spliced with lower efficiency than each of the respective single mutations (Table 1). Mutation of C^{-3} to U had a small reduction in splicing to 80% and A^{-1} to U reduced splicing to 47%. The combination of these mutations in UUUAU drastically reduced inclusion of the mini-exon to only 11% (Table 1). Similarly, the C^{-3} to U and A^{-1} to G reduced splicing only slightly (80% and 94%, respectively) but the combination of the mutations reduced splicing efficiency to 48% of the wild type (Table 1). The C^{-3} -to-G mutation reduced splicing to 18%. This significant reduction was further affected in combination with the A^{-1} -to-G mutation to give a mini-exon splicing efficiency of 9% (Table 1). In addition, the series of mutations CUGAU, UUGAU, and GUGAU allows a direct comparison of different nucleotides at position -3 in the NUGAU sequence context. Splicing efficiencies were 94%, 48%, and 9% respectively confirming the preference for C over U at this position. The fourth double mutant, CUUAA, spliced with 19% efficiency, compared to the single mutations CUUAU (47%) and CUAAA (59%), again demonstrating the synergistic negative effect of the mutations when combined. The nucleotide preferences in each position of the invertase branchpoint sequence, determined experimentally here, are entirely consistent with the plant intron branchpoint consensus sequences: CURAY and YURAY (Brown, 1986; Brown et al., 1996; Simpson & Filipowicz, 1996; Simpson et al., 1996; Tolstrup et al., 1997).

Finally, the loose consensus branchpoint sequence of plant and vertebrate introns (YURAY) reflects considerable sequence variation. The frequency of occurrence of putative branchpoint sequences in dicot introns identified 12 different sequences with occurrence frequencies of 3–12% (the most frequently occurring putative branchpoint sequence was CUAAU, found in 12% of the introns analyzed; Table 1; Simpson et al., 1996). Of these sequences we have tested 10 (including the 8 most frequent) in the invertase mini-exon splicing system during the course of the analysis of single- and double-nucleotide mutations (Table 1). In the context of the mini-exon system, all but two of the branchpoint spliced with >45% efficiency. The poorly spliced constructs (<20% efficiency) were two of the double mutations with branchpoint sequences UUUAU and CUUAA, suggesting they are intrinsically weaker sequences.

The U_{11} sequence in potato invertase acts as a polypyrimidine tract

Dicot introns are UA rich, and U- or UA-rich elements are essential for efficient splicing. All of the dicot invertase gene sequences identified so far contain a distal branchpoint and an adjacent downstream U-rich sequence. These sequences vary in length and base composition, but in the wild-type potato *invGF* intron 1 sequence, the U-rich tract consists of 11 consecutive uridines (U_{11}). Reduction to U_{10} gave full inclusion of the mini-exon and this construct, *inv1*, forms the basis for further mutations to this important signal (Simpson et al., 2000; this article). An initial construct, UC_9U , gave a low level of splicing (10%) pointing to limited compensation of U by C in this sequence. A second mutation, introducing a U-rich sequence (U_{18}) upstream of the branchpoint in the UC_9U construct was unable to compensate for the loss of the U-rich region (Simpson et al., 2000). These two constructs suggested that the U-rich sequence downstream of the branchpoint in the invertase intron may function as a polypyrimidine tract. To examine whether the invertase U-rich sequence acted as a bona fide polypyrimidine tract or as an UA-rich element, two sets of mutants were constructed. In each set, the 5' most U_4 sequence was retained and As (inv38, 41, 58, and 60) or Cs (inv39, 42, 59, and 61) were introduced in increasing numbers (2–5) whereas the remaining Us decreased in number from U_4 to U_1 (Fig. 3, Table 2). In the first series, with increasing adenine nucleotides, there was rapid reduction in splicing efficiency (Fig. 3, Table 2). The reduced splicing efficiency of the $U_4A_3U_3$, $U_4A_4U_2$, and U_4A_5U constructs suggests that As compensate poorly for Us in the U_{10} sequence, and suggests that the U-rich sequence does not function as a UA-rich intronic element.

The second series of mutations was identical to the first except that Cs replaced the As. In this case, however, the reduction in splicing efficiency with increasing numbers of cytidines was considerably less. Both $U_4C_2U_4$ (inv39) and $U_4C_3U_3$ (inv42) gave complete inclusion of the mini-exon (Fig. 3, Table 2). Similarly, $U_4C_4U_2$ (inv59) and $U_5C_5U_1$ (inv61) both permitted greatly increased levels of splicing (83% and 60%, respectively) compared to their "A" counterpart (28% and 9%, respectively). Thus, cytidines compensate to a large extent for uridines in the U_{10} sequence whereas adenines do not, strongly suggesting that the U_{10} region acts as a polypyrimidine tract. In some animal intron systems where the branchpoint lies at an extended position upstream of its 3' splice site, the polypyrimidine tract is increased in size. In the potato invertase intron 1, 6 nt downstream of the U_{11} sequence is a second U-rich region ($UCUCUCUU$). To examine whether this sequence formed part of the polypyrimidine tract, the region was mutated to $ACACACUU$ (inv69; Table 2, Fig. 4). There was no reduction in splicing.

FIGURE 3. Splicing analysis of polypyrimidine tract. GeneScan analysis of RT-PCR products from tobacco protoplasts transfected with constructs mutated at the invertase intron 1 U_{11} element. The U_{11} element was altered to include increasing amounts of A (**A**) or C (**B**) nucleotides. All polypyrimidine tract mutations and changes in position are listed in Table 2. M is a marker lane showing the position of labeled markers at 160 bp and 200 bp. RT-PCR products of 180 bp and 189 bp for mini-exon skipping and inclusion respectively are labeled (Sk and In). The sequences indicating the mutational changes are shown on the left.

ing efficiency, suggesting that the U_{11} region was sufficient to support splicing.

Two groups of 3–4 uridines support splicing

The U-rich region of the potato *invGF* gene contains 11 consecutive Us whereas other dicot genes contain small groups of Us interspersed with other nucleotides. This, and the poor splicing of U_8 , U_6 , and U_4 mutations (30%, 3%, and 4%, respectively) suggests that a single group of consecutive Us is not sufficient to support splicing (Simpson et al., 2000). The series of constructs de-

TABLE 2. Splicing efficiency of polypyrimidine tract mutations.

Clone	Polypyrimidine tract mutations	Splicing efficiency (%)
inv1	U_{10}	99 ± 1
inv38	$U_4A_2U_4$	90
inv41	$U_4A_3U_3$	69 ± 1
inv58	$U_4A_4U_2$	28
inv60	$U_4A_5U_1$	9
inv39	$U_4C_2U_4$	98
inv42	$U_4C_3U_3$	99 ± 1
inv59	$U_4C_4U_2$	83
inv61	$U_4C_5U_1$	60
inv41	$U_4A_3U_3$	69 ± 1
inv62	$AU_3A_3U_3$	63
inv42	$U_4C_3U_3$	99 ± 1
inv63	$AU_3C_3U_3$	94 ± 6
inv43	$U_3AU_3AU_2$	75
inv42	$U_4C_3U_3$	99 ± 1
inv64	$U_4C_6U_3$	72 ± 4
inv65	$U_4C_9U_3$	41 ± 1
inv8	$U_1C_9U_1$	7 ± 2
inv69	$BPU_{10}UC_n$	99
inv44	BP-14nt- U_{10} -BP2	78 ± 2.5
inv45	BP-24nt-BP2- U_{10}	6 ± 1
inv66	BP-24nt-BP2- U_{10}	5 ± 3
inv48	BPU_{10} -BP2 U_{10}	38 ± 4
inv49	BP $U_4A_2U_4$ -BP2 U_{10}	8 ± 1
inv50	BP $U_4C_2U_4$ -BP2 U_{10}	40 ± 6
inv67	BPU_{10} ---- U_{10}	26 ± 3

Bold nucleotides indicate sequence change from U_{10} found in inv1. Strikethrough indicates mutational change of a putative splicing element.

scribed above, where U residues were replaced by As and Cs, can also be used to address the basic sequence requirements of the U-rich polypyrimidine tract (Fig. 3, Table 2). First, from the series where adenines replace uridines, $U_4A_2U_4$ permitted normal splicing in contrast to eight consecutive Us (U_8), which gave only 30% inclusion (Simpson et al., 2000). This suggests that the U-rich element may need to span at least 9–10 nt for efficient splicing. The low splicing levels of $U_4A_2U_2$ (28%) and $U_4A_5U_1$ (9%) agree with our previous results that a single U_4 group (4%) is not able to support efficient splicing (Simpson et al., 2000), and suggests that a second group of at least U_3 is needed. Second, when associated with cytidines, a U_4 group is able to support reasonable levels of splicing in that even the $U_4C_5U_1$ construct spliced with 60% efficiency. The relative splicing efficiencies of $U_4A_3U_3$ (inv41) and $U_4C_3U_3$ (inv42) were 69% and 99% inclusion, respectively. When the U_4 group was mutated to U_3 , splicing efficiency was reduced only slightly in both $U_3A_3U_3$ (inv62) to 63% and $U_3C_3U_3$ (inv63) to 94%. Thus two groups of U_3 were able to support splicing, but with slightly lower splicing efficiencies compared to the U_4/U_3

FIGURE 4. Constructs to examine positional relationships of branchpoint and polypyrimidine tract. The U_{10} element and U_{10} duplicates are boxed, branchpoints (BP1 and BP2) are underlined, and point mutational changes are indicated by arrows. The start of the mini-exon sequence is shown as a // and the efficiencies that these constructs included the mini-exon in tobacco protoplasts are shown to the right of the sequence (see also Table 2).

combination, and splicing remains strongest when associated with cytidines. Finally, the $U_4A_2U_4$ construct spliced at an efficiency of 90%. A different arrangement of the eight Us and two As, $U_3AU_3AU_2$ (inv43), reduced splicing to 75%, again suggesting that the U_4 groupings were more efficient in promoting splicing than U_3 (Table 2).

Increasing the spacing between U groups reduces splicing efficiency

On the basis that $U_4C_3U_3$ (inv42) gave complete splicing and given that Cs are able to compensate to some extent for Us, we examined the effect of increasing the spacing between the U_4 and U_3 groups by introducing further cytidines. Increasing the number of cytidines from C_3 (inv42) to C_6 (inv64) and C_9 (inv65) reduced splicing efficiency from 99% to 72% and 41%, respectively (Table 2). The small increase in splicing between

the $U_4C_5U_1$ (60%) and $U_4C_6U_3$ (72%) constructs suggests that even in association with cytidines, the presence of a U_3 group increases splicing. These results again suggest that a U_4 group is able to support splicing when associated with Cs, but also that the proximity of a second U_3 group is important to the function of the polypyrimidine tract.

Increasing the distance between the branchpoint and polypyrimidine tract decreases splicing efficiency

All of the dicot invertase genes contain U-rich tracts lying 2–5 nt downstream of the branchpoint sequence. To investigate whether the proximity of the polypyrimidine tract is required for efficient splicing, the distance between the branchpoint and U_{10} sequence in inv1 was increased from 2 nt to 14 nt (inv44) and 24 nt (inv45). The overall distance between the branchpoint and 5'

splice site was maintained in all constructs (Fig. 4). Splicing efficiency was reduced to 78% and 6% as the distance increased, suggesting that the proximity of these signals is important, presumably for interactions of protein factors (Table 2). However, the repositioning in *inv45* placed it 5 nt downstream of a second branchpoint-like sequence (Fig. 4). In the original construct, mutation of this sequence had no effect on splicing and was thought to be too close to the 5' splice site to allow exon bridging interactions to occur (Simpson et al., 2000). In its new position, however, the branchpoint lies 43 nt upstream of the 5' splice site and is now closely associated with the U_{10} element lying directly downstream. This raises the possibility that the branchpoint may operate within its new sequence context disrupting use of the original branchpoint 22 nt further upstream. Mutation of the putative branchpoint in *inv45* to create *inv66* showed no change in splicing (6% and 5%, respectively; Table 2). Thus it appears that the second branchpoint (BP2) is not used in splicing in *inv45* and the severe reduction in splicing in *inv45* is due to the increased distance between the authentic branchpoint and U_{10} sequence.

Competition between polypyrimidine tracts alters splicing of the mini-exon

The above results reinforce the importance of the branchpoint/polypyrimidine tract in mini-exon splicing. The extended distance between these signals and the 5' splice site of the mini-exon points to bridging interactions between factors at these sites. In an attempt to disrupt these interactions, we introduced a second U_{10} region downstream of BP2 (Fig. 4, *inv48*). This construct, therefore, contains two essentially competing branchpoint/polypyrimidine tract sequences. However, the downstream branchpoint lies only 33 nt from the 5' splice site. Although this branchpoint/polypyrimidine tract may be expected to associate with factors such as BBP and U2AF⁶⁵, it would not be expected to establish a functional splicing complex that permits inclusion of the mini-exon. This was confirmed by mutating the putative branchpoint, which resulted in little effect on mini-exon splicing efficiency (*inv67*; Table 2). In *inv48*, the mini-exon was included at 38% efficiency in the splicing assay compared to *inv1* (Fig. 4, Table 2). To demonstrate competition between the branchpoint/polypyrimidine tracts, the upstream polypyrimidine tract was weakened by mutation to $U_4A_2U_4$ (*inv49*). Splicing of the mini-exon was reduced to 8% and was fully restored when the two As were substituted by two Cs (*inv50*, 40% mini-exon splicing). Thus the majority of the observed splicing in these constructs required the upstream branchpoint/polypyrimidine tract, and the presence of the downstream signals interfered with the normal splicing process.

DISCUSSION

In this article, we have examined the sequence requirements for a plant branchpoint and an associated polypyrimidine tract in the context of splicing the invertase mini-exon *in vivo*. The sensitivity of this plant intron splicing system to mutational changes and the readily assayed splicing phenotype (inclusion or skipping of the 9-nt mini-exon) has permitted, for the first time, the detailed characterization of the sequence requirements of a plant intron branchpoint and polypyrimidine tract. The variability in plant branchpoint sequences and the potential requirement for polypyrimidine tracts in splicing of at least some plant introns draws further parallels between the plant and vertebrate splicing systems.

CURAY—an optimal plant intron branchpoint sequence

It is only comparatively recently that branchpoint sequences have become recognized signals in plant intron splicing (Liu & Filipowicz, 1996; Simpson et al., 1996, 2000; Yi & Jack, 1998). In yeast, the branchpoint sequence UACUAAC is highly conserved and base substitutions in most positions severely reduce or abolish splicing (Langford et al., 1984). In vertebrates, the branchpoint sequence is much more variable, but vertebrate intron splicing was more efficient when the branchpoint sequence matched the yeast branchpoint sequence (Zhuang et al., 1989). Furthermore, in an iterative splicing assay in HeLa nuclear extracts, optimal branchpoints were selected, and the yeast branchpoint sequence evolved rapidly in only seven rounds of selection (Lund et al., 2000). The invertase mini-exon branchpoint (UACUAAU) differs from the yeast branchpoint only in the last nucleotide. The presence of C or U in this position has little effect on splicing in yeast (Pascolo & Séraphin, 1997) and did not affect splicing in tobacco protoplasts (*inv51*; Table 1). Mutation of the invertase branchpoint CUAAU in virtually every other position led to some degree of reduction in splicing efficiency, suggesting that, as in yeast and vertebrate introns, this particular sequence represents a strong branchpoint signal. On the basis of relative splicing efficiencies of all of the mutations, the preferred or optimal plant branchpoint nucleotide composition is CURAY. This sequence is identical to the vertebrate branchpoint consensus. It is also identical to plant intron branchpoint consensus sequences derived from intron sequence comparisons (Brown, 1986; Brown et al., 1996; Simpson & Filipowicz, 1996; Simpson et al., 1996; Tolstrup et al., 1997), thus providing experimental evidence for the plant consensus, and defining the relative importance of individual nucleotides in the consensus sequence.

The key nucleotides were the branchpoint adenine, U at -2 , and a pyrimidine (C > U) at -3 . Of these, the

branchpoint adenine and U at -2 were the most important for mini-exon splicing. The importance of these two nucleotides is underlined by the corresponding nucleotides in yeast and vertebrate branchpoints being specifically recognized by mammalian and yeast BBP early in spliceosome assembly (Berglund et al., 1997, 1998). A putative plant BBP has been identified in the *Arabidopsis* database, but it remains to be determined whether this candidate gene can interact with plant branchpoint sequences. In the invertase mini-exon system, there was residual splicing in the branchpoint A-to-U or -C mutants (19% and 7%, respectively). In vertebrates, mutation to the branchpoint A and U at position -2 greatly reduced splicing and often led to use of cryptic branchpoints or nucleotides adjacent to the mutated branchpoint adenine (Ruskin et al., 1985; Hornig et al., 1986; Query et al., 1994). Selection of alternative branchpoint signals elsewhere within intron 1 is unlikely as strong branchpoint-like sequences lie 75–100 nt further upstream and lack polypyrimidine tracts. However, it is possible that other nucleotides in the branchpoint sequence are used, such as the adenine at -1 (Hornig et al., 1986; Query et al., 1994) or non-adenine branchpoint residues, as occur naturally in some animal introns (Adema et al., 1988; Hartmuth & Barta, 1988; Query et al., 1996).

Although CURAY sequences provided the most efficient splicing activity, branchpoints with sequences that vary from this consensus (e.g., the double mutants examined here) are still able to support splicing to some extent. In other naturally occurring introns, splicing efficiency is likely to be determined by a balance of the strengths of different splicing signals such that weaker branchpoint sequences (as defined in the context of the invertase intron) may still be utilized with high efficiency in other introns (Simpson et al., 1996; Brown & Simpson, 1998; Lorkovic et al., 2000).

The U₁₁ sequence acts as a polypyrimidine tract

Plant introns do not have prominent polypyrimidine tracts, but do show a preference for U-rich sequences at the 3' end of the intron (Brown, 1986; Wiebauer et al., 1988; Baynton et al., 1996; Brown et al., 1996). In addition, plants contain U2AF⁶⁵ orthologues, which have been shown to interact with mammalian polypyrimidine tracts and preferentially bind U-rich sequences (Domon et al., 1998). The general requirement for U- or UA-rich sequences throughout plant introns has made it difficult to identify functional polypyrimidine tracts experimentally. However, in the mini-exon system, increasing adenines at the expense of uridines in the U-rich region greatly reduced splicing, and cytidines were able to compensate for uridines to some extent. Taken with our previous result that a U₁₈ sequence placed directly upstream of the branchpoint was unable to support splic-

ing (Simpson et al., 2000), these results suggest that the U₁₁ element is a polypyrimidine tract rather than a general plant intronic U- or UA-rich element. Furthermore, our results are highly analogous to those obtained in the characterization of vertebrate polypyrimidine tracts: U₁₁ was found to be the strongest polypyrimidine tract, U₈ was the minimum number of consecutive Us for splicing, the requirement for Us can be partly supported by Cs, and U-rich regions containing purines can function but with much lower efficiency (Roscigno et al., 1993; Coolidge et al., 1997). That U₁₁ sequences are found in the potato invertase gene studied here, and in other plant invertase genes, suggests that a strong polypyrimidine tract is essential to invertase mini-exon splicing. This is further supported by the relative efficiency of splicing of the U₄A₂U₄ mutation in the normal and competition constructs where weakening the upstream polypyrimidine tract greatly reduced splicing (Fig. 4; Table 2). Therefore, the competition constructs accentuated the need for a strong polypyrimidine tract adjacent to the branchpoint in the invertase system, reflecting the particular arrangement of strong splicing signals for constitutive inclusion of the 9 nt exon.

The sequence composition and spatial arrangements that determine the functional strength of this polypyrimidine tract were also examined. The U-rich regions of many invertase genes from other plants are often more complex, containing two to four runs of Us, with at least one being U₄, interspersed with other nucleotides (Simpson et al., 2000). Mutational analysis showed that the minimum sequence requirements of a strong polypyrimidine tract for efficient mini-exon splicing are two closely positioned groups of 3–4 uridines, preferably separated by U or C residues. This sequence is reminiscent of the sequence (UC_{2–3}CU) identified from affinity selection with vertebrate U2AF⁶⁵ (Singh et al., 1995) and UCCU found in high frequency in polypyrimidine tracts produced in the iterative HeLa *in vitro* splicing selection assay (Lund et al., 2000). However, given the lower frequency of occurrence of pyrimidines at the 3' end of plant introns compared to vertebrate introns, there may be subtle differences in the binding specificity of plant U2AF⁶⁵, which, to date, is only known to have affinity for U-rich sequences (Domon et al., 1998).

In addition to the number of consecutive uridines and length of vertebrate intron polypyrimidine tracts, their position relative to the branchpoint and 3' splice site is also an important determinant of the strength of the signal (Coolidge et al., 1997). Separating a strong polypyrimidine tract from the branchpoint had little effect on splicing, but when the number of uridines was limiting, the polypyrimidine tract had to be located adjacent to the 3' splice site (Coolidge et al., 1997). In contrast, in the invertase mini-exon system, increasing the separation of the U₁₀ sequence from the branchpoint reduced splicing efficiency. Thus, for this intron system,

not only are a strong branchpoint and polypyrimidine tract required for efficient splicing, but also their proximity to one another is necessary. The requirement for this arrangement of signals is also reflected by the fact that the 22-nt sequence directly upstream of the 3' splice site is U-poor (see Fig. 1).

In vertebrate introns, the strength of the polypyrimidine tract can vary enormously. The utilization of weaker polypyrimidine tracts depends on the combination of other signals and factors, such as a strong branchpoint sequence (Mullen et al., 1991), exon splicing enhancers (Wang & Manley, 1997; Lopez, 1998) or association with U2AF³⁵ at the 3' splice site AG (Guth et al., 1999; Merendino et al., 1999; Wu et al., 1999; Reed, 2000). Clearly, naturally occurring plant introns also show great variation in putative polypyrimidine tract sequences. Although it is possible that some plant introns do not require a polypyrimidine tract for splicing, as was the case in a synthetic intron (Goodall & Filipowicz, 1989), from the parallels with the vertebrate systems, splicing of most plant introns are likely to require U2AF and thereby a polypyrimidine tract. The difference in sequence of polypyrimidine tracts among plant and animal introns may reflect subtle differences in sequence recognition by plant U2AF⁶⁵ due to the overall requirement for U- or UA-richness of plant introns. Finally, polypyrimidine tract strength can be an important determinant in alternative splicing systems. For example, in *Transformer* and *Sex-lethal* in *Drosophila*, U2AF⁶⁵ and other transacting proteins mediate 3' splice site choice, and in many systems, polypyrimidine tract-binding protein (PTB) regulates splice site selection (Adams et al., 1996; Lopez, 1998; Smith & Valcárcel, 2000). It remains to be seen to what extent plant polypyrimidine tracts regulate alternative splicing and gene expression in plants, but the characterization of both branchpoint and polypyrimidine tract sequences will greatly aid examination of alternatively spliced systems in plants.

MATERIALS AND METHODS

Construction of invertase intron mutants

All sequence changes made to the 3' end of intron 1 of the *GF* invertase gene (EMBL Accession Number AJ133765; Fig. 1b) and the insertion of a 5' splice site sequence in intron 2 were made based on *inv1* using the Stratagene Quick-change site-directed mutagenesis kit. *Inv1* is exactly as described previously (Simpson et al., 2000) and consists of 50 nt of exon 1, 219 nt of intron 1, 9 nt of exon 2, 108 nt of intron 2, and 70 nt of exon 3 (Fig. 1A) inserted into the unique *Bam*HI site of expression vector pDH515 (Simpson et al., 1996). A list of all the oligos used for mutagenesis are available on request. All clones were sequenced to confirm mutations and to confirm the remaining sequence was accurate. A list of the changes made are shown in Tables 1 and 2.

Splicing analysis by RT-PCR

Intron constructs were transfected into *Nicotiana tabacum* var. Xanthi and RNA isolated as described previously (Simpson et al., 1996). RT-PCR analysis was as described (Simpson et al., 2000) using the PCR primers O8, 5'-CCCAATTGTTCAACCCTAC labeled with the 5' fluorescent phosphoramidite 6-FAM and O9, 5'-GGTAAGATGCCTGTTGCGATTGC. O8 corresponds to the zein sequence 5' to the site of intron construct insertion and O9 complementary to the zein sequence 3' to the site of intron construct insertion in pDH515 (Simpson et al., 1996). Labeled RT-PCR products were separated on a 4% polyacrylamide denaturing gel on an ABI 377 DNA sequencing machine. Sizes of bands were calculated using GeneScan V.2.1 software (Perkin Elmer) by comparison with GeneScan-350 (TAMRA) size standards. Quantification of RT-PCR products was by measurement of the fluorescent peak areas of the detected fragments after 24 cycles. Previous quantification established PCR amplification was linear over a range of 15–24 cycles and Taq polymerase enzyme efficiency was the same for both mini-exon inclusion or exclusion products. Percentage inclusion or exclusion was calculated from the peak areas for each processed transcript. Unspliced products were not detected. Each construct was tested at least twice and standard errors were determined for constructs that were tested three or more times.

ACKNOWLEDGMENTS

This research was supported by the Scottish Executive Environment and Rural Affairs Department.

Received September 19, 2001; manuscript accepted without revision October 15, 2001

REFERENCES

- Adams MD, Rudner DZ, Rio DC. 1996. Biochemistry and regulation of pre-mRNA splicing. *Curr Opin Cell Biol* 8:331–339.
- Adema GJ, Bovenberg RAL, Jansz HS, Baas PD. 1988. Unusual branch point selection involved in splicing of the alternatively processed calcitonin CGRP-I pre-messenger RNA. *Nucleic Acids Res* 16:9513–9526.
- Baynton CE, Potthoff SJ, McCullough AJ, Schuler MA. 1996. U-rich tracts enhance 3' splice site recognition in plant nuclei. *Plant J* 10:703–711.
- Berglund JA, Abovitch N, Rosbash M. 1998. A cooperative interaction between U2AF⁶⁵ and mBBP/SF1 facilitates branchpoint region recognition. *Genes & Dev* 12:858–867.
- Berglund JA, Chua K, Abovitch N, Reed R, Rosbash M. 1997. The splicing factor BBP interacts specifically with the pre-mRNA branch-point sequence UACUAAC. *Cell* 89:781–787.
- Black DL. 1995. Finding splice sites within a wilderness of RNA. *RNA* 1:763–771.
- Blencow BJ. 2000. Exonic splicing enhancers: Mechanism of action, diversity and role in human genetic disease. *Trends Biochem Sci* 25:106–110.
- Bournay A-S, Hedley PE, Maddison A, Waugh R, Machray GC. 1996. Exon skipping induced by cold stress in a potato invertase gene transcript. *Nucleic Acid Res* 24:2347–2351.
- Brown JWS. 1986. A catalogue of splice junction and putative branch point sequences from plant introns. *Nucleic Acid Res* 14:9549–9559.
- Brown JWS, Simpson CG. 1998. Splice site selection in plant pre-mRNA splicing. *Annu Rev Plant Physiol Plant Mol Biol* 49:77–95.

- Brown JWS, Smith P, Simpson CG. 1996. *Arabidopsis* consensus intron sequences. *Plant Mol Biol* 32:531–535.
- Carle-Uriste JC, Brendel V, Walbot V. 1997. A combinatorial role for exon, intron and splice site sequences in splicing in maize. *Plant J* 11:1253–1263.
- Coolidge CJ, Seely RJ, Patton JG. 1997. Functional analysis of the polypyrimidine tract in pre-mRNA splicing. *Nucleic Acid Res* 25:888–896.
- Domon C, Lorkovic Z, Valcárcel J, Filipowicz W. 1998. Multiple forms of the U2 small nuclear ribonucleoprotein auxiliary factor U2AF subunits expressed in higher plants. *J Biol Chem* 273:34603–34610.
- Gniadkowski M, Hemmings-Mieszcak M, Klahre U, Liu H-X, Filipowicz W. 1996. Characterization of intronic uridine-rich sequence elements acting as possible targets for nuclear proteins during pre-mRNA splicing in *Nicotiana glauca*. *Nucleic Acid Res* 24:619–627.
- Goodall GJ, Filipowicz W. 1989. The AU-rich sequences present in the introns of plant nuclear pre-mRNAs are required for splicing. *Cell* 58:473–483.
- Goodall GJ, Filipowicz W. 1991. Different effects of intron nucleotide composition and secondary structure on pre-messenger-RNA splicing in monocot and dicot plants. *EMBO J* 10:2635–2644.
- Guth S, Concepción M, Gaur RK, Valcárcel J. 1999. Evidence for substrate-specific requirement of the splicing factor U2AF35 and for its function after polypyrimidine tract recognition by U2AF65. *Mol Cell Biol* 19:8263–8271.
- Hanley BA, Schuler MA. 1988. Plant intron sequences—Evidence for distinct groups of introns. *Nucleic Acid Res* 16:7159–7176.
- Hartmuth K, Barta A. 1988. Unusual branch point selection in processing of human growth-hormone pre-messenger RNA. *Mol Cell Biol* 8:2011–2020.
- Hastings ML, Krainer AR. 2001. Pre-mRNA splicing in the new millennium. *Curr Opin Cell Biol* 13:302–309.
- Hornig H, Aebi M, Weissmann C. 1986. Effect of mutations at the lariat branch acceptor site on beta-globin pre-mRNA splicing in vitro. *Nature* 324:589–591.
- Ko CH, Brendel V, Taylor RD, Walbot V. 1998. U-richness is a defining feature of plant introns and may function as an intron recognition signal in maize. *Plant Mol Biol* 36:573–583.
- Krämer A. 1996. The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu Rev Biochem* 65:367–409.
- Langford CJ, Klinz FJ, Donath C, Gallwitz D. 1984. Point mutations identify the conserved, intron-contained TACTAAC box as an essential splicing signal in yeast. *Cell* 36:645–653.
- Latijnhouwers MJ, Pairoba CF, Brendel V, Walbot V, Carle-Uriste JC. 1999. Test of the combinatorial model of intron recognition in a native maize gene. *Plant Mol Biol* 41:637–644.
- Liu H-X, Filipowicz W. 1996. Mapping of branchpoint nucleotides in mutant pre-mRNAs expressed in plant cells. *Plant J* 9:381–389.
- Lopez AJ. 1998. Alternative splicing of pre-mRNA: Developmental consequences and mechanisms of regulation. *Annu Rev Genet* 32:279–305.
- Lorkovic ZJ, Kirk DAW, Lambermon MHL, Filipowicz W. 2000. Pre-mRNA splicing in higher plants. *Trends Plant Sci* 5:160–167.
- Lou H, McCullough AJ, Schuler MA. 1993a. 3' splice site selection in dicot plant nuclei is position-dependent. *Mol Cell Biol* 13:4485–4493.
- Lou H, McCullough AJ, Schuler MA. 1993b. Expression of maize Adh1 intron mutants in tobacco nuclei. *Plant J* 3:393–403.
- Luehrsen KR, Walbot V. 1994a. Addition of A-rich and U-rich sequence increases the splicing efficiency of a deleted form of a maize intron. *Plant Mol Biol* 24:449–463.
- Luehrsen KR, Walbot V. 1994b. Intron creation and polyadenylation in maize are directed by AU-rich RNA. *Gene & Dev* 8:1117–1130.
- Lund M, Tange TØ, Dyhr-Mikkelsen H, Hansens J, Kjems J. 2000. Characterization of human RNA splice signals by iterative functional selection of splice sites. *RNA* 6:528–544.
- McCullough AJ, Lou H, Schuler MA. 1993. Factors affecting authentic 5' splice site selection in plant nuclei. *Mol Cell Biol* 13:1323–1331.
- Merendino L, Guth S, Bilbao D, Martínez C, Valcárcel J. 1999. Inhibition of *msl-2* splicing by Sex-lethal reveals interaction between U2AF³⁵ and the 3' splice site AG. *Nature* 402:838–841.
- Merritt H, McCullough AJ, Schuler MA. 1997. Internal AU-rich elements modulate activity of two competing 3' splice sites in plant nuclei. *Plant J* 12:937–943.
- Mullen MP, Smith CWJ, Patton JG, Nadal-Ginard B. 1991. Alpha-tropomyosin mutually exclusive exon selection: Competition between branchpoint/polypyrimidine tracts determines default exon choice. *Genes & Dev* 5:642–655.
- Pascolo E, Séraphin B. 1997. The branchpoint residue is recognized during commitment complex formation before being bulged out of the U2 snRNA-pre-mRNA duplex. *Mol Cell Biol* 17:3469–3476.
- Query CC, Moore MJ, Sharp PA. 1994. Branch nucleophile selection in pre-messenger RNA splicing—Evidence for the bulged duplex model. *Genes & Dev* 8:587–597.
- Query CC, Strobel SA, Sharp PA. 1996. Three recognition events at the branch-site adenine. *EMBO J* 15:1392–1402.
- Reed R. 2000. Mechanisms of fidelity in pre-mRNA splicing. *Curr Opin Cell Biol* 12:340–345.
- Roscigno RF, Weiner M, Garcia-Blanco MA. 1993. A mutational analysis of the polypyrimidine tract of introns. *J Biol Sci* 268:1122–11229.
- Ruskin B, Greene JM, Green MR. 1985. Cryptic branch point activation allows accurate in vitro splicing of human beta-globin intron mutants. *Cell* 41:833–844.
- Sablowski RWM, Meyerowitz EM. 1998. Temperature-sensitive splicing in the floral homeotic mutant *apetala3-1*. *Plant Cell* 10:1453–1463.
- Simpson CG, Clark GP, Davidson D, Smith P, Brown JWS. 1996. Mutation of putative branchpoint consensus sequences in plant introns reduces splicing efficiency. *Plant J* 9:369–380.
- Simpson GG, Filipowicz W. 1996. Splicing of precursors to messenger RNA in higher plants: Mechanism, regulation and sub-nuclear organization of the spliceosomal machinery. *Plant Mol Biol* 32:1–41.
- Simpson CG, Hedley PE, Watters JA, Clark GP, McQuade C, Machray GC, Brown JWS. 2000. Requirements for mini-exon inclusion in potato invertase mRNAs provides evidence for exon-scanning in plants. *RNA* 6:422–433.
- Singh R, Valcárcel J, Green MR. 1995. Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science* 268:1173–1176.
- Smith CWJ, Valcárcel J. 2000. Alternative pre-mRNA splicing: The logic of combinatorial control. *Trends Biochem Sci* 25:381–388.
- Tolstrup N, Rouzé P, Brunak S. 1997. A branchpoint consensus from *Arabidopsis* found by non-circular analysis allows for better prediction of acceptor sites. *Nucleic Acid Res* 25:3159–3163.
- Vankan P, Filipowicz W. 1988. Structure of U2 snRNA genes of *Arabidopsis thaliana* and their expression in electroporated plant protoplasts. *EMBO J* 7:791–799.
- Wiebauer K, Herrero JJ, Filipowicz W. 1988. Nuclear pre-messenger RNA processing in plants: Distinct modes of 3' splice site selection in plants and animals. *Mol Cell Biol* 8:2042–2051.
- Wang J, Manley JL. 1997. Regulation of pre-mRNA splicing in metazoa. *Curr Opin Gen Dev* 7:205–211.
- Wu SP, Romfo CM, Nilsen TW. 1999. Functional recognition of the 3' splice site AG by the splicing factor U2AF³⁵. *Nature* 402:832–835.
- Yi Y, Jack T. 1998. An intragenic suppressor of the *Arabidopsis* floral organ identity mutant *apetala3-1* functions by suppressing defects in splicing. *Plant Cell* 10:1465–1477.
- Zhuang Y, Goldstein AM, Weiner AM. 1989. UACUAAC is the preferred branchsite for mammalian mRNA splicing. *Proc Natl Acad Sci USA* 86:2752–2756.